The correctness of the definite assignment analysis in C#
The correctness of the
definite assignment analysis in C#  

Nicu G. Fruja  
Computer Science Department, ETH Zürich, CH–8092 Zürich, Switzerland  
fruja@inf.ethz.ch

Abstract
The C# compiler must carry out a specific conservative flow analysis to ensure that every local variable is definitely assigned when any access to its value occurs. This prevents access to uninitialized memory and is a crucial ingredient for the type safety of C#. We formalize the rules of the definite assignment analysis of the C# compiler with data flow equations and we prove the correctness of the analysis, i.e. if the analysis will infer a local variable as definitely assigned at certain program point, then this variable will actually be assigned at that point during every execution of the program.

1 Introduction
Let us suppose that an attacker wants to fool the C# type system. His idea is expressed by the next block:

```csharp
{  
    int[] a;  
    try {a = (int[])(new object());} 
    catch(InvalidCastException)  
    {Console.WriteLine(a[7]);} 
}
```

A pure object is type casted into an array of integers. The attacker thinks the following will work: after the InvalidCastException which is thrown at runtime is caught, the object can be used in the handler of the catch clause, as an array to generate unpredictable behavior. Some people might think that a NullReferenceException is thrown at runtime when a[7] is accessed and thus the attacker will not succeed. Actually, his idea does not work since the block is rejected already at compile time due to the definite assignment analysis. Through this analysis, the C# compiler infers that a might not be assigned in one execution path to the access of a[7]. The analysis states that a is not definitely assigned at the beginning of the catch block since it is not definitely assigned at the beginning of the try statement.

A necessary condition for C# to be a type safe language is the following: whenever an expression is evaluated, the resulting value is of the type of the expression. If we suppose that a local variable is uninitialized when its value is required, the execution proceeds with the arbitrary value which was at the memory position of the uninitialized local variable. Since this value could be of any type, we would obviously violate the type safety of C# and we could easily produce unpredictable behavior.

Since local variables are not initialized with default values like for example static variables or instance variables of class instances, a C# compiler must carry out a specific conservative flow analysis to ensure that every local variable is definitely assigned when any access to its value occurs. This definite assignment analysis which is a static analysis (see [8, 5] for other static analyses) has to guarantee that there is an initialization to a local variable on every possible execution path before the variable is read. Since the problem is undecidable in general, the C Language Specification [1, §5.3] contains a definition of a decidable subclass. So far, the definite assignment analysis of the Java compiler has been formalized with data flow equations in the work of Stärk et al. [11] and related to the problem of generating verifiable bytecode from legal Java source code programs. A formalization of the analysis for Java which uses type systems is presented in [9]. Since in our case, the analysis involves a fixed point iteration, the presentation as type systems does not appear to be a feasible solution.

1
The formalization of the C♯ definite assignment analysis we provide, sheds some light in particular on the complications generated by the goto and break statements (incompletely specified in [1]) and by the method calls with ref/out parameters - these are crucial differences with respect to Java. We also use the idea of data flow equations (see [1]) but due to the goto statement, the formalization cannot be done like in Java. For a method body without goto, however, the equations that characterize the sets of definitely assigned variables can be solved in a single pass. If goto statements are present, then the equations defined in our formalization do not specify in a unique way the sets of variables that have to be considered definitely assigned. For this reason, a fixed point computation is performed and the greatest sets of variables that satisfy the equations of the formalization are computed. Another difference with respect to Java is the presence of structs. Regarding the correctness of the analysis, we prove that, these sets of variables represent exactly the sets of variables assigned on all possible execution paths and in particular they are a safe approximation.

The rest of the paper is organized as follows. Section 2 introduces the data flow equations which formalize the C♯ definite assignment analysis while Section 3 shows that there always exists a maximal fixed point solution for the equations. In order to define the execution paths in a method body, the control flow graph is introduced in Section 4. The paper concludes in Section 5 with the proof of the correctness of the analysis, Theorem 1. A number of bugs in the Mono C♯ compiler [7] (version 0.26) which were discovered during the attempts to build the formalization of the definite assignment are included in the Appendix.

2 The data flow equations

In this section, we formalize the rules of definite assignment analysis from the C♯ Specification [1, §5.3] by data flow equations. Since the definite assignment analysis is an intraprocedural analysis, we restrict our formalization only to a given method \textit{meth}. We use labels in order to identify the expressions and the statements. Labels are denoted by small Greek letters and are displayed as superscripts, for example, as in \textit{^o\_exp} or in \textit{^o\_stm}. We will often refer to expressions and statements using their labels.

In order to precisely specify all the cases of definite assignment, static functions \textit{before}, \textit{after}, \textit{true}, \textit{false} and \textit{vars} are computed at compile time. Note that \textit{true} and \textit{false} are only for boolean expressions. These functions assign sets of variables to each expression or statement \(\alpha\) and have the following meanings. 
\textit{before}(\(\alpha\)) contains the local variables definitely assigned before the evaluation of \(\alpha\) and \textit{after}(\(\alpha\)) the variables definitely assigned after the evaluation of \(\alpha\) when \(\alpha\) completes normally. \textit{true}(\(\alpha\)) and \textit{false}(\(\alpha\)) consist of the variables (in the scope of which \(\alpha\) is) definitely assigned after the evaluation of \(\alpha\) when \(\alpha\) evaluates to \textit{true} and \textit{false}, respectively. \textit{vars}(\(\alpha\)) contains the local variables in the scope of which \(\alpha\) is, i.e. the universal set with respect to \(\alpha\).

In order not to increase the number of equations, we skip those language constructs whose analysis is very similar to the one of the constructs dealt with explicitly in our framework; examples are alternative control structures (do, switch, foreach), pre- and postfix operators (++, --). On the other hand, we omit also the statements for and lock since they can be written in terms of constructs from our framework as observed by C♯ Specification in [§5.3.3.9] and [§8.12], respectively.

Before we state the data flow equations, we clarify some details concerning the definite assignment analysis for the struct type variables which represent a key difference with respect to Java. A struct type variable is considered definitely assigned if and only if each of its instance variables is considered definitely assigned [1, §5.3]. We formalize this idea of propagating the definitely assigned status as follows. If a local variable \textit{loc} of a struct type is assigned, then all its instance fields \textit{instfields}(\textit{loc}) get assigned and further, if there are struct type instance fields, then their instance fields get assigned as well and so on. In the left example from Figure 1 the instance field \textit{y} of \textit{p.x} is definitely assigned before it is printed since \textit{p} is definitely assigned.

The recursively defined set \textit{lookdown}(\textit{loc}) represents the local variables ‘derived’ from \textit{loc} which get assigned when \textit{loc} gets assigned.

\[
\text{lookdown}(\textit{loc}) = \begin{cases} 
\{\textit{loc}\}, & \text{if } \textit{loc} \text{ is not of a struct type} \\
\{\textit{loc}\} \cup \bigcup_{\text{field} \in \textit{instfields}(\textit{loc})} \text{lookdown}(\textit{field}), & \text{if } \textit{loc} \text{ is of a struct type}
\end{cases}
\]

At the same time, if a local variable \textit{loc}_1 which is an instance field of the struct type variable \textit{loc}_2

1. gets assigned, and
2. each instance field of \( \text{loc}_2 \) other than \( \text{loc}_1 \) either is definitely assigned, i.e. is in the set \( \text{after}(\alpha) \) or is going to be assigned, i.e. is in the set \( \text{next} \): \( \{ \text{loc} \} \) for a simple assignment to \( \text{loc} \) or \( \text{OutParams}(\text{arg}_1, \ldots, \text{arg}_k) \) for a method invocation with parameters \( \text{arg}_1, \ldots, \text{arg}_k \).

then the containing variable \( \text{loc}_2 \) gets assigned as well and the procedure is repeated for \( \text{loc}_3 \). In the right example from Figure 1 the struct type field \( \text{p}.y \) gets assigned when its instance field \( \text{v} \) gets assigned. Then the instance variable \( \text{p} \) gets assigned when its instance field \( \text{p}.x \) gets assigned. This is expressed by the computation of the set \( \text{lookup}(\text{loc}_1, \alpha, \text{next}) \).

\[
\text{lookup}(\text{loc}_1, \alpha, \text{next}) = \begin{cases} 
\{ \text{loc}_1 \} \cup \text{lookup}(\text{loc}_2, \alpha, \text{next}), & \text{if } \text{loc}_1 \text{ is an instance field of the} \\
\{ \text{loc}_1 \}, & \text{instfields}(\text{loc}_2) \setminus \{ \text{loc}_1 \} \subseteq \text{after}(\alpha) \cup \text{next} \\
\end{cases}
\]

Therefore the set \( \text{struct}(\text{loc}, \alpha, \text{next}) \) contains the local variables that get definitely assigned when \( \text{loc} \) gets assigned where \( \alpha \) and the set \( \text{next} \) depends on the context - whether \( \text{loc} \) gets assigned through a simple assignment or through a method call.

\[
\text{struct}(\text{loc}, \alpha, \text{next}) = \text{lookup}(\text{loc}, \alpha, \text{next}) \cup \text{lookdown}(\text{loc})
\]

We are now able to state all the data flow equations. A first equation is given by initial conditions: for the method body \( \text{mb} \) of \( \text{meth} \) we have \( \text{before}(\text{mb}) = \emptyset \). Actually we should consider the set of value and reference parameters of \( \text{meth} \) but there is no worry that an access to any of them could cause troubles since when \( \text{meth} \) is invoked they are supposed to be definitely assigned \([5.1]\).

For the other expressions and statements in \( \text{mb} \), instead of explaining how the functions are computed, we simply state the equations they have to satisfy. Table [1] contains the equations for boolean expressions. If \( \alpha \) is the constant \text{true}, then \( \text{false}(\alpha) = \text{vars}(\alpha) \) is a consequence of the definition of the \text{false} set and of the fact that \text{true} cannot evaluate to \text{false}. Similar arguments hold for \( \text{true}(\alpha) = \text{vars}(\alpha) \), if \( \alpha \) is the constant \text{false}. In addition, we have for all expressions in Table [1] the equation \( \text{after}(\alpha) = \text{true}(\alpha) \cap \text{false}(\alpha) \). For a boolean expression \( \alpha \) which is not an instance of one of the expressions in Table [1] we have \( \text{true}(\alpha) = \text{after}(\alpha) \) and \( \text{false}(\alpha) = \text{after}(\alpha) \).

Table [2] lists the equations specific to arbitrary expressions where \( \text{loc} \) stands for a local variable and \( \text{lit} \) for a literal. The conditional expression is not the same like in Table [1] since it is not necessarily boolean. A \text{ref} parameter is used for ‘by reference’ parameter passing, in which the parameter acts as an alias for a caller-provided argument. An \text{out} parameter is similar to a \text{ref} parameter, except that the initial value of the caller-provided argument is not important. Concerning the definite assignment analysis, the \text{ref} arguments must be definitely assigned before the method invocation while the \text{out} arguments should not necessarily be assigned before the method is invoked. On the other hand, the \text{out} arguments must be definitely assigned when the method returns. Note that following a method invocation, not only the \text{out} parameters \( \text{OutParams}(\text{arg}_1, \ldots, \text{arg}_k) \) become definitely assigned, but eventually also struct type variables or instance fields of struct type variables.
the data flow equations

true  \[\text{true}(\alpha) = \text{before}(\alpha), \text{false}(\alpha) = \text{vars}(\alpha)\]
false  \[\text{false}(\alpha) = \text{before}(\alpha), \text{true}(\alpha) = \text{vars}(\alpha)\]
(! \beta e)  \[\text{before}(\beta) = \text{before}(\alpha), \text{true}(\alpha) = \text{false}(\beta), \text{false}(\alpha) = \text{true}(\beta)\]
(\beta_0 ? \gamma_1 : \delta e_2)  \[\text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{true}(\beta), \text{before}(\delta) = \text{false}(\beta)\]
true(\alpha) = \text{true}(\gamma) \cap \text{true}(\delta), \text{false}(\alpha) = \text{false}(\gamma) \cap \text{false}(\delta)\]
(\beta_1 \&\& \gamma e_2)  \[\text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{true}(\beta), \text{true}(\alpha) = \text{true}(\gamma), \text{false}(\alpha) = \text{false}(\beta) \cap \text{false}(\gamma)\]
(\beta_1 || \gamma e_2)  \[\text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{false}(\beta), \text{false}(\alpha) = \text{false}(\gamma)\]
true(\alpha) = \text{true}(\beta) \cap \text{true}(\gamma)\]

Table 1: Definite assignment for boolean expressions

<table>
<thead>
<tr>
<th>(\alpha \text{exp})</th>
<th>the data flow equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>loc</td>
<td>after(\alpha) = before(\alpha)</td>
</tr>
<tr>
<td>lit</td>
<td>after(\alpha) = before(\alpha)</td>
</tr>
<tr>
<td>(loc = \beta e)</td>
<td>before(\beta) = before(\alpha), after(\alpha) = after(\beta) \cup \text{struct}(loc, \beta, {\text{loc}})</td>
</tr>
<tr>
<td>(loc op = \beta e)</td>
<td>before(\beta) = before(\alpha), after(\alpha) = after(\beta)</td>
</tr>
<tr>
<td>(\beta_0 ? \gamma_1 : \delta e_2)</td>
<td>before(\beta) = before(\alpha), before(\gamma) = true(\beta), before(\delta) = false(\beta), after(\alpha) = after(\gamma) \cap after(\delta)</td>
</tr>
<tr>
<td>c.f</td>
<td>after(\alpha) = before(\alpha)</td>
</tr>
<tr>
<td>ref \beta \text{exp}</td>
<td>before(\beta) = before(\alpha), after(\alpha) = after(\beta)</td>
</tr>
<tr>
<td>out \beta \text{exp}</td>
<td>before(\beta) = before(\alpha), after(\alpha) = after(\beta)</td>
</tr>
<tr>
<td>c.m(\beta_1 \text{arg}_1, \ldots, \beta_n \text{arg}_k)</td>
<td>before(\beta_i) = before(\alpha), before(\beta_{i+1}) = after(\beta_i), i = 1, k - 1, after(\alpha) = after(\beta_k) \cup \bigcup_{\text{arg}_j \in \text{OutParams}(\text{arg}_1, \ldots, \text{arg}_k)} \text{struct}(\text{arg}_j, \beta_k, \text{OutParams}(\text{arg}_1, \ldots, \text{arg}_k))</td>
</tr>
</tbody>
</table>

Table 2: Definite assignment for arbitrary expressions

In cases not stated in Tables 1, 2 if \(\alpha \text{exp}\) is an expression with direct subexpressions \(\beta_1 e_1, \ldots, \beta_n e_n\), then the left-to-right evaluation scheme yields the general data flow equations:

\[\text{before}(\beta_1) = \text{before}(\alpha), \text{before}(\beta_{i+1}) = \text{after}(\beta_i), i = 1, n - 1 \text{ and after}(\alpha) = \text{after}(\beta_n)\]

The equations specific for statements can be found in Table 3. We assume that the try statements are either try-catch or try-finally statements (see 2 for a justification of this assumption). Special attention is paid to the labeled statement. The set of variables definitely assigned before executing a labeled statement consists of the variables definitely assigned both after the previous statement and before each corresponding goto statement or after any of the finally blocks of try-finally statements in which the goto is embedded (if any). More exactly, this can be formalized as follows. For two statements \(\alpha\) and \(\beta\), we consider \(\text{Fin}(\alpha, \beta)\) to be the list \([\gamma_1, \ldots, \gamma_n]\) of finally blocks of all try-finally statements in the innermost to outermost order from \(\alpha\) to \(\beta\). Then we define the set \(\text{JoinFin}(\alpha, \beta)\) of
the data flow equations

\[ \text{after}(\alpha) = \text{before}(\alpha) \]

\( (\beta \exp; ) \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{after}(\alpha) = \text{after}(\beta) \]

\( \{^\beta \text{stm}_1 \ldots ^\beta \text{stm}_n \} \)

\[ \text{before}(\beta_1) = \text{before}(\alpha), \text{after}(\alpha) = \text{after}(\beta_n) \cap \text{vars}(\alpha), \]

\[ \text{before}(\beta_{i+1}) = \text{after}(\beta_i) \cap \text{goto}(\beta_{i+1}), i = 1, n - 1 \]

\( \text{if} (\beta \exp) \gamma \text{stm}_1 \text{else} ^\delta \text{stm}_2 \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{true}(\beta), \]

\[ \text{before}(\delta) = \text{false}(\beta), \text{after}(\alpha) = \text{after}(\gamma) \cap \text{after}(\delta) \]

\( \text{while}(\beta \exp) \gamma \text{stm} \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{true}(\beta), \]

\[ \text{after}(\alpha) = \text{false}(\beta) \cap \text{break}(\alpha) \]

\( \text{goto L} \)

\[ \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{break} \)

\[ \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{return} \)

\[ \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{return } ^\beta \exp \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{throw} \)

\[ \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{throw } ^\beta \exp \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{after}(\alpha) = \text{vars}(\alpha) \]

\( \text{try } ^\beta \text{block} \)

\( \text{catch}(E_1 \, x_1) \gamma \text{block}_1 \)

\[ \text{before}(\beta) = \text{before}(\alpha), \]

\[ \text{after}(\alpha) = \text{after}(\beta) \cap \bigcap_{i=1}^{n} \text{after}(\gamma_i) \]

\( \vdots \)

\( \text{catch}(E_n \, x_n) \gamma \text{block}_n \)

\[ \text{before}(\gamma_i) = \text{before}(\alpha) \cup \{x_i\}, i = 1, n, \]

\[ \text{after}(\alpha) = \text{after}(\beta) \cap \bigcap_{i=1}^{n} \text{after}(\gamma_i) \]

\( \text{try } ^\beta \text{block}_1 \text{finally } ^\gamma \text{block}_2 \)

\[ \text{before}(\beta) = \text{before}(\alpha), \text{before}(\gamma) = \text{before}(\alpha), \]

\[ \text{after}(\alpha) = \text{after}(\beta) \cup \text{after}(\gamma) \]

\[ \text{JoinFin}(\alpha, \beta) = \bigcup_{\gamma \in \text{Fin}(\alpha, \beta)} \text{after}(\gamma) \]

Further, we define the set goto for a statement \( \beta \). For a labeled statement \( ^\beta \text{L} : \text{stm} \), the set goto(\( \beta \)) is defined as follows:

\[ \text{goto}(\beta) = \bigcap_{^\alpha \text{goto L};} (\text{before}(\alpha) \cup \text{JoinFin}(\alpha, \beta)) \]

where we take only the goto statements in the scope of \( \beta \). For all the other statements, as well for a labeled statement with no goto statements, goto(\( \beta \)) is the universal set \text{vars}(\( \beta \)). Now we are able to state the equation \( \text{before}(\beta_{i+1}) = \text{after}(\beta_i) \cap \text{goto}(\beta_{i+1}) \) from Table 3. In case of a labeled statement, the equation formalizes the above stated idea while for a non-labeled statement becomes \( \text{before}(\beta_{i+1}) = \text{after}(\beta_i) \).

The following example is a simplification of an example from the C\# Specification [1] §5.3.3.15:

```csharp
int i;
^\text{try} \{^\alpha \text{goto L;} \}
```

Table 3: Definite assignment for statements

definitely assigned variables after the execution of all these finally blocks:
finally γ{i = 3;}
β:L:Console.WriteLine(i);

The explanations that come with the example in the C# Specification state that i is definitely assigned before β, i.e., i ∈ before(β). Our equation before(β) = after(δ) ∩ goto(β) led us to the same conclusion since goto(β) = before(α) ∪ after(γ) and i ∈ after(γ) ⊆ after(δ) (see the equations for a try-finally in Table 3). Surprisingly, the example is rejected by the C compilers of .NET Framework 1.0 and Rotor SSCLI [10]: we get the error that i is unassigned. In the meantime, this was fixed in .NET Framework 1.1 [2] but still exists in Rotor.

The following explanation holds for the equation after(α) = after(β) ∩ vars(α) corresponding to a block of statements: the local variables which are definitely assigned after the normal execution of the block are the variables which are definitely assigned after the execution of the last statement of the block. However, the variables must still be in the scope of a declaration. Thus, let us consider the example:

```
{"int i; i = 1;\}  {int i; i = 2 * β i;\}
```

The variable i is not in after(α) since at the end of α, i is not in the scope of a declaration. Thus i ∈ before(β) and the block is rejected.

The idea for the equation which computes after(α) of a while statement α, is the same with that for a labeled statement. Similarly with the set goto, we define the set break(α) to be set of variables definitely assigned before all corresponding break statements (and possibly after appropriate finally blocks). This means that the set break(α) is defined by

\[ \text{break(α)} = \bigcap_\delta \left( \text{before(β)} \cup \text{JoinFin(β, α)} \right) \]

where we take only the break statements for which α is the nearest enclosing while. If the while statement does not have any break statements, then we define break(α) = vars(α). With this definition of break(α), we have the equation for after(α) as stated in Table 3.

There is one more technical detail to be decided. Suppose we want to state the equation for after of a jump statement. Let α be the following statement:

```
if(b) γ{i = 1;\} else δ: return;
```

It is clear that, the variables definitely assigned after α are the variables definitely assigned after the then branch and since our equation takes the intersection of after(γ) and after(δ), it is obvious that one has to require the set-intersection identity for after(δ). That is why we adopt the convention that after(α) is the universal set vars(α) for any jump statement α.

3 The maximal fixed point

The computation of the sets of definitely assigned variables from the data flow equations described in Section 2 is relatively straightforward. The key difference with respect to Java is the goto statement which brings more complexity to the analysis. Since the goto statement makes loops possible, the system of data flow equations does not have always a unique solution. Here is an example: if we consider a method which takes no parameters and has the following body

```
{"int i = 1;\} β:L:γ goto L;
```

then we have the following equations after(α) = \{i\}, before(β) = after(α) \cap before(γ) and before(γ) = before(β). After some simplification we find that before(β) = \{i\} \cap before(β) and therefore we get two solutions for before(β) (and also for before(γ)): \emptyset and \{i\}. This is the reason we perform a fixed point iteration - which is not the case in Java. The set of variables definitely assigned after α is \{i\} and since β does not ‘unassign’ i, i is obviously assigned when we enter β. Consideration of the example and the definition of definitely assigned show that the most informative solution is \{i\} and therefore the solution we require is the maximal fixed point MFP. For computing this solution, there exist various algorithms (see [3]).

Remark. Although the statements L: goto L; and while (true); look behaviorally similar, they are treated differently by the definite assignment analysis. Thus, if α denotes the labeled statement,
then the equation for \( \text{before}(\alpha) \) implies recursion (as noticed above). If \( \alpha \) is the \text{while} statement, then no equation corresponding to \( \alpha \) involves recursion. The set \( \text{after}(\alpha) \) of the above \text{while} statement can be computed according to the equations in a single step (i.e. with no fixed point iteration) as follows: 
\[
\text{after}(\alpha) = \text{false}(\alpha) \cap \text{break}(\alpha) = \text{vars}(\alpha) \cap \text{vars}(\alpha) = \text{vars}(\alpha).
\]
The set \( \text{before}(\alpha) \) is determined as for any regular statement using only the \( \text{after} \) set of the previous statement. Even if a \text{while} statement \( \alpha \) has an associated \text{continue} statement \( \gamma \), the equation for \( \text{before}(\alpha) \) does not involve the \text{continue} but only the previous statement \( \beta \). The reason is that, at the time the analysis is performed, the compiler is sure the \text{continue} is embedded in the \text{while} body and therefore the \( \text{set before}(\gamma) \) includes the variables in \( \text{after}(\beta) \) (if the \text{continue} is executed then necessarily \( \beta \) should have been executed). This is not always true for a labeled statement since the associated \text{gosub}s are not necessarily embedded in the labeled block (see the last example of Section 2) and that is why they are involved in the equation for \( \text{before}(\alpha) \).

In the rest of this section, we show that there always exists a maximal fixed point for our data flow equations. In order to prove the existence, one needs first to define the function \( F \) which encapsulates the equations. For the domain and codomain of this function, we need the set \( \text{Vars}(\text{meth}) \) of all local variables from the method body \( \text{mb} \). A simple inspection of the equations shows that they all have at the left side either a \text{before}, \text{after}, \text{true} or a \text{false} set and at the right side a combination of these kinds of sets and \text{vars} sets. We define the function \( F : D \rightarrow D \) with \( D = \mathcal{P}(\text{Vars}(\text{meth}))^T \) such that 
\[
F(X_1, \ldots, X_r) = (Y_1, \ldots, Y_r),
\]
where \( r \) is the number of equations and the sets \( Y_i \) are defined by the data flow equations. For example in the case of an \text{if-then-else} statement, if the equation for the \text{after} set of this statement is the \( i \)-th data flow equation, then the set of variables \( Y_i \) is defined by 
\[
Y_i = X_j \cap X_k
\]
where \( j \) and \( k \) are the indices of the equations for the \text{after} sets of the \text{then} and the \text{else} branch, respectively. Note that the sets \text{vars} are interpreted as constants.

We define now the relation \( \sqsubseteq \) on \( D \) to be the pointwise set inclusion relation: if \((X_1, \ldots, X_r) \in D \) and \((X'_1, \ldots, X'_r) \in D \), then we have \((X_1, \ldots, X_r) \sqsubseteq (X'_1, \ldots, X'_r)\) if \( X_i \subseteq X'_i \) for all \( i = 1, \ldots, r \). We are now able to prove the following result:

**Lemma 1** \((D, \sqsubseteq)\) is a finite lattice.

**Proof.** \( D \) is finite since for a given method body we have a finite number of equations and local variables and on the other hand, \( D \) is a lattice since it is a product of lattices: \((\mathcal{P}(\text{Vars}(\text{meth})), \sqsubseteq)\) is a poset since the set inclusion is a partial order and for every two sets \( X, Y \in \mathcal{P}(\text{Vars}(\text{meth})) \) there exists a lower bound \((X \cap Y)\) and an upper bound \((X \cup Y)\).

The following result will help us conclude the existence of the maximal fixed point.

**Lemma 2** The function \( F \) is monotonic on \((D, \sqsubseteq)\).

**Proof.** In order to prove the monotonicity of \( F = (F_1, \ldots, F_r) \), it suffices to remark that the components \( F_i \) are monotonic functions. This holds since they consist only of set intersections and unions which are monotonic (see the form of the equations).

The next result guarantees the existence of the maximal fixed point solution for our data flow equations:

**Lemma 3** The function \( F \) has a unique maximal fixed point \( \text{MFP} \in D \).

**Proof.** \((D, \sqsubseteq)\) is a finite lattice (Lemma 1) and therefore a complete lattice. But in a complete lattice, every monotonic function has a unique maximal fixed point (known also as the greatest fixed point). In our case, \( F \) is monotonic (Lemma 2) and the maximal fixed point \( \text{MFP} \) is given by \( \bigcap_k F^{(k)}(1_D) \). Here \( 1_D \) is the \( r \)-tuple \((\text{Vars}(\text{meth}), \ldots, \text{Vars}(\text{meth}))\), i.e. the top element of the lattice \( D \).

From now on, for an expression or statement \( \alpha \) we denote by \( \text{MFP}_{\text{b}}(\alpha) \), \( \text{MFP}_{\text{a}}(\alpha) \), \( \text{MFP}_{\text{t}}(\alpha) \) and \( \text{MFP}_{\text{f}}(\alpha) \) the components of \( \text{MFP} \) corresponding to \( \text{before}(\alpha) \), \( \text{after}(\alpha) \), \( \text{true}(\alpha) \) and \( \text{false}(\alpha) \), respectively.
The main result we want to prove is that, for an arbitrary expression or statement, the sets of local variables \( MFP_b \), \( MFP_a \) (and \( MFP_t \), \( MFP_f \) for boolean expressions) correspond indeed to sets of definitely assigned variables, i.e. variables which are assigned on every possible execution path to the appropriate point. The considered paths are based on the control flow graph \( CFG \). The nodes of the graph are actually points associated with every expression and statement. We suppose that every expression or statement \( \alpha \) is characterized by an \textit{entry} point \( B(\alpha) \) and an \textit{end} point \( A(\alpha) \). Beside these two points, a boolean expression \( \alpha \) has two more points: a \textit{true} point \( T(\alpha) \) (used when \( \alpha \) evaluates to \textit{true}) and a \textit{false} point \( F(\alpha) \) (used when \( \alpha \) evaluates to \textit{false}). The edges of the graph are given by the control transfer defined in the C\# Specification [1] §8. We show in Tables 4 and 5 the edges specific to each boolean and arbitrary expression, respectively. If the expression \( \alpha \) is not an instance of one expression in these tables

### Table 4: Control flow for boolean expressions

<table>
<thead>
<tr>
<th>( \alpha ) exp</th>
<th>edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>((B(\alpha), T(\alpha)))</td>
</tr>
<tr>
<td>false</td>
<td>((B(\alpha), F(\alpha)))</td>
</tr>
<tr>
<td>(! \beta e)</td>
<td>((B(\alpha), B(\beta)), (F(\beta), T(\alpha)), (T(\beta), F(\alpha)))</td>
</tr>
<tr>
<td>((\beta e_0 \ ? \ \gamma e_1 : \delta e_2))</td>
<td>((B(\alpha), B(\beta)), (T(\beta), B(\gamma)), (F(\beta), B(\delta)), (T(\gamma), T(\alpha)), (F(\gamma), F(\alpha))))</td>
</tr>
<tr>
<td>((\beta e_1 \ &amp; \ \gamma e_2))</td>
<td>((B(\alpha), B(\beta)), (T(\beta), B(\gamma)), (F(\beta), F(\alpha)), (T(\gamma), T(\alpha)), (F(\gamma), F(\alpha))))</td>
</tr>
<tr>
<td>((\beta e_1 \</td>
<td>\ \gamma e_2))</td>
</tr>
</tbody>
</table>

### Table 5: Control flow for arbitrary expressions

<table>
<thead>
<tr>
<th>( \alpha ) exp</th>
<th>edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>loc</td>
<td>((B(\alpha), A(\alpha)))</td>
</tr>
<tr>
<td>lit</td>
<td>((B(\alpha), A(\alpha)))</td>
</tr>
<tr>
<td>((\text{loc} = \beta e))</td>
<td>((B(\alpha), B(\beta)), (A(\beta), A(\alpha)))</td>
</tr>
<tr>
<td>((\text{loc} \ \text{op} = \beta e))</td>
<td>((B(\alpha), B(\beta)), (A(\beta), A(\alpha)))</td>
</tr>
<tr>
<td>((\beta e_0 \ ? \ \gamma e_1 : \delta e_2))</td>
<td>((B(\alpha), B(\beta)), (T(\beta), B(\gamma)), (F(\beta), B(\delta)), (A(\gamma), A(\alpha)), (A(\delta), A(\alpha)))</td>
</tr>
<tr>
<td>c.f</td>
<td>((B(\alpha), A(\alpha)))</td>
</tr>
<tr>
<td>\text{ref} \ \beta \ \text{exp}</td>
<td>((B(\alpha), B(\beta)), (A(\beta), A(\alpha)))</td>
</tr>
<tr>
<td>\text{out} \ \beta \ \text{exp}</td>
<td>((B(\alpha), B(\beta)), (A(\beta), A(\alpha)))</td>
</tr>
<tr>
<td>\text{c.m} ((\beta_1 \ \text{arg}_1, \ldots, \beta_k \ \text{arg}_k))</td>
<td>((B(\alpha), B(\beta_1)), (A(\beta_k), A(\alpha)), (A(\beta_i), B(\beta_{i+1})), i = 1, k - 1)</td>
</tr>
</tbody>
</table>

4 The control flow graph

The main result we want to prove is that, for an arbitrary expression or statement, the sets of local variables \( MFP_b \), \( MFP_a \) (and \( MFP_t \), \( MFP_f \) for boolean expressions) correspond indeed to sets of definitely assigned variables, i.e. variables which are assigned on every possible execution path to the appropriate point. The considered paths are based on the control flow graph \( CFG \). The nodes of the graph are actually points associated with every expression and statement. We suppose that every expression or statement \( \alpha \) is characterized by an \textit{entry} point \( B(\alpha) \) and an \textit{end} point \( A(\alpha) \). Beside these two points, a boolean expression \( \alpha \) has two more points: a \textit{true} point \( T(\alpha) \) (used when \( \alpha \) evaluates to \textit{true}) and a \textit{false} point \( F(\alpha) \) (used when \( \alpha \) evaluates to \textit{false}). The edges of the graph are given by the control transfer defined in the C\# Specification [1] §8. We show in Tables 4 and 5 the edges specific to each boolean and arbitrary expression, respectively. If the expression \( \alpha \) is not an instance of one expression in these tables
(e.g. \( \text{exp}_1 | \text{exp}_2 \)) and has the direct subexpressions \( \beta_1, \ldots, \beta_n \), then the left-to-right evaluation scheme adds to the flow graph also the following edges:

\[
(B(\alpha), B(\beta_i)), (A(\beta_i), B(\beta_{i+1})), i = 1, n - 1 \quad \text{and} \quad (A(\beta_n), A(\alpha))
\]

For every boolean expression \( \alpha \) in Table 4, we have the supplementary edges \((T(\alpha), A(\alpha))\) and \((F(\alpha), A(\alpha))\) which connect the boolean points of \( \alpha \) to the end point of \( \alpha \). These edges are necessary for the control transfer in cases when it does not matter whether \( \alpha \) evaluates to \text{true} or \text{false}. For example, if \( \beta \) is the method invocation \( \text{c.m} (\text{true}) \) and \( \alpha \) is the argument \text{true}, then the control is transferred from the end point of the last argument - that is \( A(\alpha) \) - to the end point of the method invocation - that is \( A(\beta) \). But since in Table 4 we have no edge leading to \( A(\alpha) \), we need to define also the supplementary edge \((T(\alpha), A(\alpha))\).

For a boolean expression \( \alpha \) which is not an instance of any expression from Table 4, we add to the graph the edges \((A(\alpha), T(\alpha))\), \((A(\alpha), F(\alpha))\). They are needed if control is transferred from a boolean expression \( \alpha \) to different points depending on whether \( \alpha \) evaluates to \text{true} or \text{false}. For example, if \( \alpha \) is of the form \( \text{exp}_1 | \text{exp}_2 \) and occurs in \( \beta (\text{true}) \), then the control is transferred from \( F(\alpha) \) to \( T(\beta) \) (if \( \alpha \) evaluates to \text{false}) or from \( T(\alpha) \) to \( F(\beta) \) (if \( \alpha \) evaluates to \text{true}). The necessity of the edges \((A(\alpha), T(\alpha))\), \((A(\alpha), F(\alpha))\) arises since, so far we have defined only \( \text{exp}_1 | \text{exp}_2 \) edges to \( A(\alpha) \).

Table 6 introduces the edges of the control flow graph for each statement. Note that we assume that the boolean constant expressions are replaced by \text{true} or \text{false} in the abstract syntax tree. For example, we consider that \text{true} | \text{b} is replaced by \text{true} in the following if statement:

```c
if \beta (\text{true} | \text{b}) \{ i = 1;
else \gamma \{ \text{int } j = i;\}
```

Although the new test (i.e. \text{true}) cannot evaluate to \text{false}, we still add to the graph the edge \((F(\beta), B(\gamma))\) since anyway the false point of \text{true} is not reachable (see Table 4).

In the presence of finally blocks, the jump statements \text{goto}, \text{break} and \text{continue} bring more complexity to the graph. Whenever such a jump statement exits one or more try blocks with associated finally blocks, the control is transferred first to the finally block (if any) of the innermost try statement. Further, if the control reaches the end point of the finally, then it is transferred to the next (with respect to the innermost to outermost order of the try statements) finally block and so on. If the control reaches the end point of the last finally block, then it is transferred to the target of the jump statement. For these control transfers we have special edges in our graph. But one needs to take care to some detail: these special edges cannot be used for paths other than those which connect the jump statement with its target. In other words, if a path uses such an edge, then necessarily the path contains the entry point of the jump statement. For this reason, we say that an edge \( e \) is conditioned by a point \( i \) with the meaning that \( e \) can be used only in paths that contain \( i \). If we do not make this restriction, then \([B(\text{mb})B(\alpha_1)B(\alpha_2)B(\alpha_3)B(\alpha_4)A(\alpha_5)B(\alpha_6)]\) would be a possible execution path to the labeled statement in the following method body

```c
a1:try  \{ 
  a2\{i = 1;\)
  goto L;
} finally a5\{\}
   L:Console.WriteLine(i);
```

in the theoretical case when the evaluation of \( \alpha_4 \) would throw an exception. But this does not match the control transfer described in the C# Specification.

The following sets introduce the above described edges. If \( \alpha \) and \( \beta \) are two statements and \( \text{Fin}(\alpha, \beta) \) is the list \( \gamma_1, \ldots, \gamma_n \), then the set \( \text{ThroughFin}_\alpha(\alpha, \beta) \) consists of the edges \((B(\alpha), B(\gamma_1)), (A(\gamma_n), B(\beta)), (A(\gamma_i), B(\gamma_{i+1})), i = 1, n - 1 \) all conditioned by \( B(\alpha) \) and the set \( \text{ThroughFin}_\alpha(\alpha, \beta) \) has the edges \((B(\alpha), A(\gamma_i)), (A(\gamma_n), A(\beta)), (A(\gamma_i), B(\gamma_{i+1})), i = 1, n - 1 \) all conditioned by \( B(\alpha) \). If \( \text{Fin}(\alpha, \beta) \) is empty, then the set \( \text{ThroughFin}_\alpha(\alpha, \beta) \) has only the edge \((B(\alpha), B(\beta))\) while \( \text{ThroughFin}_\alpha(\alpha, \beta) \) refers to the edge \((B(\alpha), A(\beta))\).

Note that in Table 5 for \text{goto} and \text{continue}, the set of edges \( \text{ThroughFin}_\alpha \) is added to the graph, since after executing the finally blocks the control is transferred to the entry point of the labeled statement and \text{while} statement, respectively, while in case of \text{break} the set \( \text{ThroughFin}_\alpha \) is considered, since at the end, the control is transferred to the end point of the \text{while} statement.
Table 6: Control flow for statements

There are two more remarks concerning the try statement. Since in a try block can occur anytime a reason for abruption (e.g. an exception), we should have edges from every point in a try block to: every associate catch block, every catch of enclosing try statements (if the catch clause matches the type of the exception) and to every associate finally block (if none of the catch clauses matches the type of the exception). We do not consider all these edges, since from the point of view of the definite assignment analysis which is in particular an ‘over all paths’ analysis, it is equivalent to consider only one edge to the entry points of the catch and finally blocks - from the entry point of the try block (see Table 6).

The next remark is concerning the end point \(A(\alpha)\) of a try-finally statement \(\alpha\). The C\(_\text{\#}\) Specification states in [§8.10] that \(A(\alpha)\) is reachable only if both end points of the try block \(\beta\) and finally block \(\gamma\) are reachable. The only edge to \(A(\alpha)\) is \((A(\gamma), A(\alpha))\) and we know that the finally block can be reached either through a jump or through a normal completion of the try block. In case of a jump, if control

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(\beta)</td>
</tr>
</tbody>
</table>
reaches the end point \( \mathcal{A}(\gamma) \) of the finally, then it is transferred further to the target of statement which generated the jump and not to \( \mathcal{A}(\alpha) \). This means that all paths to \( \mathcal{A}(\alpha) \) contain also the end point \( \mathcal{A}(\beta) \) of the try block. That is why we require that the edge \((\mathcal{A}(\gamma), \mathcal{A}(\alpha))\) is conditioned by \( \mathcal{A}(\beta) \) (see Table 6) - otherwise in the following example, \( \mathcal{A}(\alpha) \) would be reachable in our graph (under the assumption that \( \mathcal{B}(\alpha) \) is reachable):

\[
\text{"try } \beta \{ \text{goto } L; \} \text{ finally } \gamma{\}}
\]

We define now the sets of valid paths to all points in the method body. We will not consider all the paths in the graph but only the valid paths - that is the paths \( p \) for which the following is true: if \( p \) uses a conditioned edge then it contains also the point which conditions the edge.

\[
\text{valid}([\alpha_1, \ldots, \alpha_n]) \equiv \text{ for every conditioned edge } (\alpha_i, \alpha_{i+1}) \exists j < i \text{ s.t. } \alpha_j \text{ conditions } (\alpha_i, \alpha_{i+1})
\]

If \( \alpha \) is an expression or a statement, then \( path_b(\alpha) \) and \( path_a(\alpha) \) are the sets of all valid paths from the entry point of the method body \( \mathcal{B}(mb) \) to the entry point \( \mathcal{B}(\alpha) \) and to the end point \( \mathcal{A}(\alpha) \) of \( \alpha \), respectively.

\[
\begin{align*}
\text{path}_b(\alpha) = & \{ [\alpha_1, \ldots, \alpha_n] \mid \alpha_1 = \mathcal{B}(mb), \alpha_n = \mathcal{B}(\alpha), (\alpha_i, \alpha_{i+1}) \in CFG, i = 1, n-1 \quad \\
& \quad \text{and valid}([\alpha_1, \ldots, \alpha_n]) \}\n\text{path}_a(\alpha) = & \{ [\alpha_1, \ldots, \alpha_n] \mid \alpha_1 = \mathcal{B}(mb), \alpha_n = \mathcal{A}(\alpha), (\alpha_i, \alpha_{i+1}) \in CFG, i = 1, n-1 \quad \\
& \quad \text{and valid}([\alpha_1, \ldots, \alpha_n]) \}\n\end{align*}
\]

Similarly, if \( \alpha \) is a boolean expression, then \( path_t(\alpha) \) and \( path_f(\alpha) \) are the sets of all valid paths from \( \mathcal{B}(mb) \) to the true point \( \mathcal{T}(\alpha) \) and to the false point \( \mathcal{F}(\alpha) \) of \( \alpha \), respectively.

In the proofs from the next section, we will use the following two notations. If \( p \) is a path, then \( p[i, j] \) is the subpath of \( p \) which connects the point \( i \) with the point \( j \). Also, over the set of all paths, we define the operation \( \ominus \) to be the concatenation of the lists of points which define the paths.

5 The correctness of the analysis

We prove that, when a C\(_{\#}\) compiler relies on the sets \( \text{MFP}_b, \text{MFP}_a, \text{MFP}_t \) and \( \text{MFP}_f \) derived from the maximal fixed point of the data flow equations in Section 2, the risk of accessing the value of a local variable which is not initialized does not exist. In other words, the correctness of the analysis means that, if a local variable is in one of the four sets - that is the analysis infers the variable as definitely assigned at a certain program point, then this variable will actually be assigned at that point during every execution of the program. A variable is assigned on a path if the path contains an initialization of the variable or a catch clause whose exception variable is our variable. Due to the special treatment of struct type variables, we define in a recursive way what we mean by initialization. We say that a path \( p \) contains an initialization of a local variable \( \text{loc} \) if at least one of the following is true:

1. \( p \) contains a simple assignment (not a compound assignment) to \( \text{loc} \), or
2. \( p \) contains a method invocation for which \( \text{loc} \) is an out parameter, or
3. \( \text{loc} \) is an instance field of a struct type variable \( x \) and \( p \) contains an initialization of \( x \), or
4. \( \text{loc} \) is of a struct type and \( p \) contains initializations of each instance field of \( \text{loc} \).

We prove actually more than the correctness. We show that the components of the maximal fixed point \( \text{MFP} \) are exactly (not only a safe approximation of) the sets of variables which are assigned on every possible path to the appropriate point. In order to formalize this, we define the following sets. If \( \alpha \) is an arbitrary expression or statement, then \( \text{AP}_b(\alpha) \) and \( \text{AP}_a(\alpha) \) denote the sets of local variables in \( \text{vars}(\alpha) \) (the variables in the scope of which \( \alpha \) is) which are assigned on every path in \( \text{path}_b(\alpha) \) and in \( \text{path}_a(\alpha) \), respectively.

\[
\text{AP}_b(\alpha) = \{ x \in \text{vars}(\alpha) \mid x \text{ is assigned on every path } p \in \text{path}_b(\alpha) \}\n\]
\[ \text{AP}_a(\alpha) = \{ x \in \text{vars}(\alpha) \mid x \text{ is assigned on every path } p \in \text{path}_a(\alpha) \} \]

For a boolean expression \( \alpha \), we have two more sets: \( \text{AP}_t(\alpha) \) and \( \text{AP}_f(\alpha) \) are defined similarly as above, but with respect to paths in \( \text{path}_t(\alpha) \) and \( \text{path}_f(\alpha) \), respectively.

The next result is used to prove Lemma 5.

**Lemma 4** For every expression or statement \( \alpha \), if \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \), then \( \text{MFP}_a(\alpha) \subseteq \text{vars}(\alpha) \). Moreover, if \( \alpha \) is a boolean expression, then we have also \( \text{MFP}_t(\alpha) \subseteq \text{vars}(\alpha) \) and \( \text{MFP}_f(\alpha) \subseteq \text{vars}(\alpha) \).

**Proof.** It is done for each expression and statement by induction over the abstract syntax tree, starting from the leaves. Thus, we will prove first the above stated implications for all possible leaves of the AST.

The expressions which are leaves in the AST are the empty-statement, \texttt{goto} L, \texttt{break}, \texttt{continue}, \texttt{return}, and \texttt{throw}. For the last five, from the equations we obviously have \( \text{MFP}_a(\alpha) \subseteq \text{vars}(\alpha) \). For the empty-statement, this is true as well since our hypothesis is \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \).

The statements considered leaves in an AST are the \texttt{empty-statement}, \texttt{goto} L, \texttt{break}, \texttt{continue}, \texttt{return}, and \texttt{throw}. For the last five, from the equations we obviously have \( \text{MFP}_a(\alpha) \subseteq \text{vars}(\alpha) \). For the empty-statement, this is true as well since our hypothesis is \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \).

In the next step, the so-called induction step, the implications for each expression and statement are proved, under the assumption that their children satisfy the implications.

The following lemma claims that, the MFP sets of an expression or statement \( \alpha \), consist of variables in the scope of which \( \alpha \) is.

**Lemma 5** For every expression or statement \( \alpha \) we have \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \) and \( \text{MFP}_a(\alpha) \subseteq \text{vars}(\alpha) \). Moreover, if \( \alpha \) is a boolean expression, then we have also \( \text{MFP}_t(\alpha) \subseteq \text{vars}(\alpha) \) and \( \text{MFP}_f(\alpha) \subseteq \text{vars}(\alpha) \).

**Proof.** We will show the above inclusions for all expressions and statements using an induction which starts from the root of the method body. So, under the assumption that a node of the abstract syntax tree AST satisfies the inclusions, we will prove that all its children satisfy the inclusions as well.

According to Lemma 4, it is enough to prove for all labels \( \alpha \): \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \).

For the method body, this is trivial: \( \text{MFP}_b(mb) \subseteq \text{vars}(mb) \) since \( \text{MFP}_b(mb) = \text{vars}(mb) = \emptyset \).

For the last step of the proof, which consists in showing for each expression or statement that, under the assumption \( \text{MFP}_b(\alpha) \subseteq \text{vars}(\alpha) \), each of its direct subexpressions or sub-statements \( \beta \) satisfies \( \text{MFP}_b(\beta) \subseteq \text{vars}(\beta) \).

The correctness of the definite assignment analysis in \( \mathcal{C}_2 \) is proved in the following theorem, which claims that the analysis is a safe approximation.

**Theorem 1** (safe approximation) For every expression or statement \( \alpha \), the following relations are true: \( \text{MFP}_b(\alpha) \subseteq \text{AP}_b(\alpha) \) and \( \text{MFP}_a(\alpha) \subseteq \text{AP}_a(\alpha) \). Moreover, if \( \alpha \) is a boolean expression, then we have \( \text{MFP}_t(\alpha) \subseteq \text{AP}_t(\alpha) \) and \( \text{MFP}_f(\alpha) \subseteq \text{AP}_f(\alpha) \).

**Proof.** We consider the following definitions. The set \( \text{AP}_b^n(\alpha) \) is defined in the same way as \( \text{AP}_b(\alpha) \), except that we consider only the paths of length less than \( n \). Similarly, we define also the sets \( \text{AP}_a^n(\alpha) \), \( \text{AP}_t^n(\alpha) \), \( \text{AP}_f^n(\alpha) \) (analogously, we have definitions for the sets of paths \( \text{path}^n \)). According to these definitions, the following set equalities hold for an arbitrary \( \alpha \): \( \text{AP}_b(\alpha) = \bigcap_n \text{AP}_b^n(\alpha) \), \( \text{AP}_a(\alpha) = \bigcap_n \text{AP}_a^n(\alpha) \) and if \( \alpha \) is a boolean expression, then \( \text{AP}_t(\alpha) = \bigcap_n \text{AP}_t^n(\alpha) \) and \( \text{AP}_f(\alpha) = \bigcap_n \text{AP}_f^n(\alpha) \). Therefore to complete the proof, it suffices to show for every \( n \): if \( \alpha \) is an expression or statement, then \( \text{MFP}_b(\alpha) \subseteq \text{AP}_b^n(\alpha) \) and \( \text{MFP}_a(\alpha) \subseteq \text{AP}_a^n(\alpha) \) and in addition, if \( \alpha \) is a boolean expression, \( \text{MFP}_t(\alpha) \subseteq \text{AP}_t^n(\alpha) \) and \( \text{MFP}_f(\alpha) \subseteq \text{AP}_f^n(\alpha) \). This is done by induction on \( n \).

**Basis of induction:** \( \mathcal{B}(mb) \) is the only path of length 1 (the entry point of the method body). Obviously, no local variable is assigned on this path and therefore we have \( \text{AP}_b^1(mb) = \emptyset \) which satisfies \( \text{MFP}_b(mb) \subseteq \text{AP}_b^1(mb) \) since from the equations \( \text{MFP}_b(mb) = \emptyset \). From the definition of \( \text{AP}_a^1 \), we get \( \text{AP}_a^1(mb) = \text{vars}(mb) = \emptyset \) and from the equations of a block, we derive also \( \text{MFP}_a(mb) \subseteq \text{vars}(mb) \) and implicitly \( \text{MFP}_a(mb) \subseteq \text{AP}_a^1(mb) \). If \( \alpha \neq mb \), then \( \text{AP}_t^1(\alpha) = \text{AP}_t^1(\alpha) = \text{vars}(\alpha) \) and if \( \alpha \) is a boolean expression \( \text{AP}_t^1(\alpha) = \text{AP}_f^1(\alpha) = \text{vars}(\alpha) \). If we apply Lemma 5, then the basis of the induction is complete.
\textit{Induction step:} The proof has the following pattern: we will show for all the expressions and statements from Tables 12 and 3 the relations for MFP\(_b\) and if it is case for MFP\(_t\) and MFP\(_f\) and for all their subexpressions and substatements, we will show the relations for MFP\(_a\). In this manner, all the relations are proved except that for MFP\(_b\)\((mb)\) which holds anyway since MFP\(_b\)\((mb)\) = \(\emptyset\).

Whenever we consider an \(x\) in a MFP set of a certain expression or statement \(\alpha\), from the induction hypothesis \(x\) is also in the corresponding AP\(_n\), and in particular in \(\text{vars}(\alpha)\).

We prove first the relations corresponding to boolean expressions from Table 1.

\textbf{Case 1.} \(^{\alpha}\text{exp} = \text{true}\)

We want to prove for the boolean constant \text{true} the following subset relation: MFP\(_t\)\((\alpha)\) \(\subseteq\) AP\(_{n+1}\)\((\alpha)\).

Let \(x\) \(\in\) MFP\(_t\)\((\alpha)\). Since MFP is a solution of the equations, we have MFP\(_t\)\((\alpha)\) = MFP\(_b\)\((\alpha)\) and implicitly \(x\) \(\in\) MFP\(_b\)\((\alpha)\). But according to the induction hypothesis, we have \(x\) \(\in\) AP\(_b\)\((\alpha)\). This means that \(x\) is assigned on every path in path\(_b\)\((\alpha)\). On the other hand, from the control flow graph CFG, we remark that the only way to reach the true point \(T\) ∈ path\(_b\)\((\alpha)\). In the same way can be done also the proof for MFP\(_f\)\((\alpha)\) \(\subseteq\) AP\(_{n+1}\)\((\alpha)\) since AP\(_f\)\(_{n+1}\)\((\alpha)\) = \(\text{vars}(\alpha)\) - there is no path to \(F\) - and from the equations we have MFP\(_f\)\((\alpha)\) \(\subseteq\) \(\text{vars}(\alpha)\).

\textbf{Case 2.} \(^{\alpha}\text{exp} = \text{false}\)

Similar to Case 1.

\textbf{Case 3.} \(^{\alpha}\text{exp} = (1 \ \beta \ e)\)

For the proof of MFP\(_b\)\((\beta)\) \(\subseteq\) AP\(_b\)\(_{n+1}\)\((\beta)\), we consider an \(x\) \(\in\) MFP\(_b\)\((\beta)\). Because MFP is a fixed point of \(F\), we have MFP\(_b\)\((\beta)\) = MFP\(_b\)\((\alpha)\) and consequently \(x\) \(\in\) MFP\(_b\)\((\alpha)\). From the induction hypothesis we get \(x\) \(\in\) AP\(_b\)\(_{n}\)\((\alpha)\). Since we have path\(_b\)\(_{n+1}\)\((\beta)\) = path\(_b\)\(_{n}\)\((\alpha)\) ∪ \(B\)\((\beta)\), we can conclude that \(x\) \(\in\) AP\(_b\)\(_{n+1}\)\((\beta)\). The proofs for the relations corresponding to MFP\(_t\)\((\alpha)\) and MFP\(_f\)\((\alpha)\) are similar.

\textbf{Case 4.} \(^{\alpha}\text{exp} = (\beta e_0 \ ? \ \gamma e_1 : \ \delta e_2)\)

The proofs for MFP\(_b\)\((\beta)\), MFP\(_b\)\((\gamma)\) and MFP\(_b\)\((\delta)\) are similar with the proofs from Case 3.

We are going to make now the proof corresponding to MFP\(_t\)\((\alpha)\), namely MFP\(_t\)\((\alpha)\) \(\subseteq\) AP\(_t\)\(_{n+1}\)\((\alpha)\). Let \(x\) be a local variable in MFP\(_t\)\((\alpha)\). From the equations, we get \(x\) \(\in\) MFP\(_t\)\((\gamma)\) ∩ MFP\(_t\)\((\delta)\) and further, from the induction hypothesis, we have \(x\) \(\in\) AP\(_t\)\(_n\)\((\gamma)\) and \(x\) \(\in\) AP\(_t\)\(_n\)\((\delta)\). A simple inspection of CFG shows that there are only two ways to reach to \(T\)\((\alpha)\): either through the edge \((T(\gamma), T(\alpha))\) or through the edge \((T(\delta), T(\alpha))\). This means that

\[
\text{path}_{t}^{n+1}(\alpha) = \text{path}_{t}^{n}(\gamma) \cup T(\alpha) \cup \text{path}_{t}^{n}(\delta) \cup T(\alpha)
\]

which implies \(x\) \(\in\) AP\(_t\)\(_{n+1}\)\((\alpha)\). In the same way can be done also the proof for MFP\(_f\)\((\alpha)\) \(\subseteq\) AP\(_f\)\(_{n+1}\)\((\alpha)\).

\textbf{Case 5.} \(^{\alpha}\text{exp} = (\beta e_1 \ \&\& \ \gamma e_2)\)

Similar to Case 4.

\textbf{Case 6.} \(^{\alpha}\text{exp} = (\beta e_1 \ || \ \gamma e_2)\)

Similar to Case 4.

We consider now for proving the relations corresponding to an arbitrary expression from Table 2.

\textbf{Case 7.} \(^{\alpha}\text{exp} = \text{loc}\)

Similar to the proof corresponding to MFP\(_b\)\((\beta)\) in Case 3.

\textbf{Case 8.} \(^{\alpha}\text{exp} = \text{lit}\)

Similar to the proof corresponding to MFP\(_b\)\((\beta)\) in Case 3.

\textbf{Case 9.} \(^{\alpha}\text{exp} = (\text{loc} = \beta e)\)

The proof corresponding to MFP\(_b\)\((\beta)\) is similar with the proof for MFP\(_b\)\((\beta)\) in Case 3.

We want to prove now MFP\(_a\)\((\alpha)\) \(\subseteq\) AP\(_a\)\(_{n+1}\)\((\alpha)\). If we consider an \(x\) \(\in\) MFP\(_a\)\((\alpha)\), from the equations we get \(x\) \(\in\) MFP\(_a\)\((\beta)\) \(\cup\) struct\((\text{loc}, \beta, \{\text{loc}\})\) where \(\beta\) is the expression which is assigned to \(\alpha\) (see Table 2). If \(x\) \(\in\) MFP\(_a\)\((\beta)\), from the induction hypothesis we get \(x\) \(\in\) AP\(_a\)\(_n\)\((\beta)\). The proof can be completed with the same arguments like before. We assume now that \(x\) is in the set struct\((\text{loc}, \beta, \{\text{loc}\})\) and then there two main cases to discuss:
• if \( x \in \text{lookdown}(\text{loc}) \), then either \( x \) is \( \text{loc} \) or \( x \in \bigcup_{\text{field} \in \text{instfields}(\text{loc})} \text{lookdown}(\text{field}) \) and \( \text{loc} \) is of a struct type. Further if \( x \) is \( \text{loc} \), then it is sure that \( x \) is assigned on any path to \( \mathcal{A}(\alpha) \), in particular also on paths of length at most \( n+1 \), i.e. \( x \in \text{AP}_n^{n+1}(\alpha) \). In the other case, we get that there exists an instance field \( y \) of \( \text{loc} \) such that \( x \in \text{lookdown}(y) \) and the procedure can be continued in this way. Since we obviously have to deal with a finite chain, \( x \) is one instance field in this chain. But since every path to \( \mathcal{A}(\alpha) \) contains an initialization of \( \text{loc} \), we say that it contains also an initialization of all the instance fields returned by the previous procedure (see the definition of initialization). Therefore we get \( x \in \text{AP}_n^{n+1}(\alpha) \).

• if \( x \in \text{lookup}(\text{loc}, \beta, \{\text{loc}\}) \), then either \( x \) is \( \text{loc} \) and the proof is the same like before or \( \text{loc} \) is an instance field of a struct variable \( y \) such that \( \text{instfields}(y) \subseteq \text{MFP}_n(\beta) \cup \{\text{loc}\} \) and \( x \in \text{lookup}(y, \beta, \{\text{loc}\}) \).

Let us suppose that the variable \( x \) is the struct variable \( y \). This implies that all the instance fields of \( x \) (except \( \text{loc} \)) are in the set \( \text{MFP}_n(\beta) \) and according to the induction hypothesis this means that the instance fields are assigned on every path to \( \mathcal{A}(\beta) \) of length at most \( n \). Since \( \text{loc} \) is assigned on every path to \( \mathcal{A}(\alpha) \) we obtain that \( x \) is in the set \( \text{AP}_n^{n+1}(\alpha) \).

If \( x \) is not the variable \( y \), then there exists a sequence \( y_1, \ldots, y_k \) of local variables such that \( x = y_k \), \( y = y_1 \), \( y_i \) is an instance field of the struct variable \( y_{i+1} \) for all \( i = 1, k - 1 \) and \( y_1 \) of \( y_i \), \( i = 1, k \) except \( \text{loc} \) of \( y_1 \) and \( y_i \), \( i = 1, k \) are in the set \( \text{MFP}_n(\beta) \) and implicitly in \( \text{AP}_n^{n+1}(\beta) \). With the same arguments as before, we can prove that \( y_1 \in \text{AP}_n^{n+1}(\alpha) \). In this manner, we can show by induction on \( i \) the desired result \( y_k \in \text{AP}_n^{n+1}(\alpha) \).

We have proved in both cases \( x \in \text{AP}_n^{n+1}(\alpha) \).

**Case 10.** \( a \exp = (\text{loc op} \Rightarrow \text{beta} e) \)

Similar to Case 9 with the difference that there is no case when \( x \) is in \( \text{struct}(\text{loc}, \beta, \{\text{loc}\}) \) since no initialization happens on the edge \((\mathcal{A}(\beta), \mathcal{A}(\alpha))\).

**Case 11.** \( a \exp = (\text{beta} e_0 \Rightarrow \text{beta} e_1 : \gamma e_2) \)

Similar to Case 4.

**Case 12.** \( a \exp = c.f \)

Similar to the proof for \( \text{MFP}_k(\beta) \) in Case 3.

**Case 13.** \( a \exp = \text{ref} \, \beta \exp \)

Similar to the proof for \( \text{MFP}_k(\beta) \) in Case 3.

**Case 14.** \( a \exp = \text{out} \, \beta \exp \)

Similar to the proof for \( \text{MFP}_k(\beta) \) in Case 3.

**Case 15.** \( a \exp = \text{c.m}((\beta_1 \arg_1), \ldots, (\beta_k \arg_k)) \)

The proofs corresponding to the arguments are similar with the proof for \( \text{MFP}_k(\beta) \) in Case 3.

We want show \( \text{MFP}_n(\alpha) \subseteq \text{AP}_n^{n+1}(\alpha) \). Let us consider a local variable \( x \) in the set \( \text{MFP}_n(\alpha) \). According to the equation satisfied by \( \text{MFP}_n(\alpha) \), we have also

\[ x \in \text{MFP}_n(\beta_k) \cup \bigcup_{\arg_i \in \text{OutParams}(\arg_1, \ldots, \arg_k)} \text{struct}(\arg_i, \beta_k, \text{OutParams}(\arg_1, \ldots, \arg_k)) \]

If \( x \in \text{MFP}_n(\beta_k) \), then the proof can be completed with the same arguments as in the corresponding proof for a simple assignment in Case 9. Let us assume now that there exists an \( \text{out} \) parameter \( \arg_i \) such that \( x \in \text{struct}(\arg_i, \beta_k, \text{OutParams}(\arg_1, \ldots, \arg_k)) \). The two possible cases \( x \in \text{lookdown}(\arg_i) \) and \( x \in \text{lookup}(\arg_i, \beta_k, \text{OutParams}(\arg_1, \ldots, \arg_k)) \) are treated similar with the corresponding cases of a simple assignment.

**Case 16.** The proofs corresponding to expressions which are not instances of expressions from Tables 1 and 2 make use of the general data flow equations and follow the same pattern like the proof corresponding to \( \text{MFP}_k(\beta) \) in Case 3.

We will consider now for proving the statements from Table 3
Case 17. $\alpha \text{stm} =$;
Similar to the first proof in Case 3.

Case 18. $\alpha \text{stm} = (^0 \text{exp;} )$
Similar to the first proof in Case 3.

Case 19. $\alpha \text{stm} = \{\beta_1 \text{stm}_1 \ldots \beta_n \text{stm}_n \}$
We want to prove now $\text{MFP}_b(\beta_{i+1}) \subseteq \text{AP}_n^{n+1}(\beta_{i+1})$ corresponding to an embedded statement. If we arbitrarily choose an $x$ in $\text{MFP}_b(\beta_{i+1})$, then we obtain $x \in \text{MFP}_a(\beta_i)$ and $x \in \text{goto}(\beta_{i+1})$ (from the flow equations). From the induction hypothesis, we get $x \in \text{AP}_n^a(\beta_i)$. Note that at this point, $\text{goto}(\beta_{i+1})$ depends only on MFP sets and on the control flow $\text{CFG}$.

- Let us consider first the case when $\beta_{i+1}$ is not a labeled statement. In this situation, we have $\text{goto}(\beta_{i+1}) = \text{vars}(\beta_{i+1})$.
  - If moreover $\beta_{i+1}$ is not a while statement, then there exists in $\text{CFG}$ only one edge to $\text{B}(\beta_{i+1})$, namely $(\text{A}(\beta_i), \text{B}(\beta_{i+1}))$. This means that

$$\text{path}_b^{n+1}(\beta_{i+1}) = \text{path}_a^n(\beta_i) \oplus \text{B}(\beta_{i+1})$$

Since from the induction hypothesis $x \in \text{AP}_n^a(\beta_i)$, we get $x \in \text{AP}_b^{n+1}(\beta_{i+1})$.

- If $\beta_{i+1}$ is a while statement, then there could be in $\text{CFG}$ more edges to $\text{B}(\beta_{i+1})$ (if there are continue statements pointing to this while).

  * If there are no continue statements corresponding to our while statement, then the proof is the same like before.

  * We want to show now that $x$ is assigned on every path to $\text{B}(\beta_{i+1})$ of length at most $n+1$ that pass through a continue statement and possibly through finally blocks of enclosing try-finally statements.

    If there exists such a path $p$ which contains a continue statement, then necessarily $p$ pass through $\text{A}(\beta_i)$ because the $\text{CFG}$ shows that it is not possible to jump into the while body where the continue is supposed to be embedded in. Since $x \in \text{MFP}_a(\beta_i)$, from the induction hypothesis we get $x \in \text{AP}_a^n(\beta_i)$. This means that $x$ is assigned on every path to $\text{A}(\beta_i)$ of length at most $n$. Now we are sure that $p$ assigns $x$.

Since we are sure that $x$ is assigned on each path to $\text{B}(\beta_{i+1})$ of length at most $n+1$, we can conclude $x \in \text{AP}_b^{n+1}(\beta_{i+1})$.

- Let us suppose now that $\beta_{i+1}$ is a labeled statement. Then like in the case of the while statement, there could be in $\text{CFG}$ more edges to $\text{B}(\beta_{i+1})$ (if there are goto statements pointing to our labeled statement).

  - If there are no corresponding goto statements, then the proof is the same like in the case of a while statement with no appropriate continue statements.

  - We want to show now that $x$ is assigned on every path $\text{B}(\beta_{i+1})$ of length at most $n+1$ that pass through a goto statement and possibly through finally blocks of enclosing try-finally statements.

    Let $p$ be such a path that pass through a goto statement $\gamma$. Since $x \in \text{goto}(\beta_{i+1})$, we get $x \in \text{MFP}_b(\gamma) \cup \text{JoinFin}(\gamma, \beta_{i+1})$.

    * If there are no finally blocks in $\text{Fin}(\gamma, \beta_{i+1})$, then $\text{JoinFin}(\gamma, \beta_{i+1}) = \emptyset$ and implicitly $x \in \text{MFP}_b(\gamma)$. Using the induction hypothesis, we obtain $x \in \text{AP}_b^n(\gamma)$. This means that $p$ which is of length at most $n+1$ and contains $\text{B}(\gamma)$ assigns $x$.

    * Let us suppose now that the list $\text{Fin}(\gamma, \beta_{i+1})$ is non-empty: $\text{Fin} = [\gamma_1, \ldots, \gamma_k]$. From the definition of $\text{JoinFin}(\gamma, \beta_{i+1})$, we get

$$x \in \text{MFP}_b(\gamma) \cup \bigcup_{j=1}^{k} \text{MFP}_a(\gamma_j)$$
If $x \in \text{MFP}_b(\gamma)$, then from the induction hypothesis we derive $x \in \text{AP}_b^n(\gamma)$ and we are sure that $p$ which pass through $\mathcal{B}(\gamma)$ assigns $x$.

If there is a finally block $\gamma_j$ such that $x \in \text{MFP}_a(\gamma_j)$, then the induction hypothesis implies $x \in \text{AP}_a^n(\gamma_j)$. And since necessarily $p$ pass through $\mathcal{A}(\gamma_j)$, we are sure that $p$ assigns $x$.

Thus, we analyzed every possible path to $\mathcal{B}(\beta_{n+1})$ of length at most $n+1$ and we showed that each such a path assigns $x \in \text{vars}(\beta_{n+1})$, i.e $x \in \text{AP}_b^{n+1}(\beta_{n+1})$.

Case 20. $\text{stm} = \text{if} \ (\beta_{\text{exp}} \gamma \text{stm}_1 \ \text{else} \ \delta \text{stm}_2$
Similar to Case 4.

Case 21. $\text{stm} = \text{while}(\beta_{\text{exp}} \gamma)$
The proofs corresponding to MFP$_b(\beta)$ and MFP$_b(\gamma)$ are similar with the proof for MFP$_b(\beta)$ in Case 3.

The proof for MFP$_a(\alpha)$ can be done in the same way as for the case of a labeled statement treated in Case 19.

Case 22. The proofs for the goto, break, continue, return, return exp, throw and throw exp statements (except those involving exp) are trivial: MFP$_a(\alpha) \subseteq \text{AP}_a^{n+1}(\alpha)$ holds since from the induction hypothesis MFP$_a(\alpha) \subseteq \text{AP}_a^n(\alpha)$ and on the other hand $\text{AP}_a^n(\alpha) = \text{AP}_a^{n+1}(\alpha) = \emptyset$. The proofs for the points involving exp follow the same pattern like the proof for MFP$_b(\beta)$ in Case 3.

Case 23. $\text{stm} = \text{try-catch}$.
The proofs for MFP$_b(\alpha)$ and MFP$_b(\gamma)$, $i = \overline{1,n}$ are similar with the proofs in Case 9. The proof for MFP$_a(\alpha)$ follows the same pattern like the proof for MFP$_a(\alpha)$ in Case 4.

Case 24. $\text{stm} = \text{try-finally}$.
The proof for MFP$_b(\beta)$ is similar with the proof for MFP$_b(\beta)$ in Case 3.

The next proof we are doing is of MFP$_b(\gamma) \subseteq \text{AP}_b^{n+1}(\gamma)$. Let us consider an $x \in \text{MFP}_b(\gamma)$.

Following the data flow equations, we get $x \in \text{MFP}_b(\gamma)$. The induction hypothesis implies $x \in \text{AP}_b^n(\gamma)$.

Now if we consider CFG, there could be more edges leading to $\mathcal{B}(\gamma)$: from the entry point of the corresponding try-finally statement $(\mathcal{B}(\delta), B(\gamma))$, from the end point of the corresponding try block $(\mathcal{A}(\beta), B(\gamma))$, from a goto, break or continue statement $(\mathcal{B}(\delta), B(\gamma))$ (within a conditioned path), from the end point of another finally block $(\mathcal{A}(\omega), B(\gamma))$ (within a conditioned path).

We claim now that, independent of the last edge of an arbitrary path $p$ to $\mathcal{B}(\gamma)$, $p$ pass through the beginning $\mathcal{B}(\alpha)$ of the try-finally statement.

- If the last edge of $p$ is $(\mathcal{B}(\alpha), B(\gamma))$, then it is nothing to prove.
- If the last edge is $(\mathcal{A}(\beta), B(\gamma))$, since the end point $\mathcal{A}(\beta)$ can be reached only through $\mathcal{B}(\alpha)$ (according to CFG, it is not possible to jump into the try block), then the claim holds.
- For the case when the last edge of $p$ is $(\mathcal{B}(\delta), B(\gamma))$, the claim can be argued in the same way like above, because the respective jump statements are supposed to be embedded in the try block.
- If the last edge of $p$ is a conditioned edge $(\mathcal{A}(\omega), B(\gamma))$, then necessarily the finally block $\omega$ (as well as the jump statement which triggered the conditioning) is embedded in our try block. It means that, in order to justify the claim, we can apply the same argument like above.

So all the paths to $\mathcal{B}(\gamma)$ should pass through $\mathcal{B}(\alpha)$ and since $x \in \text{AP}_a^n(\alpha)$ we can be sure that $x \in \text{vars}(\gamma)$ is assigned on every path to $\mathcal{B}(\gamma)$ of length at most $n+1$, i.e. $x \in \text{AP}_b^{n+1}(\gamma)$.

We want to prove now for a try-finally block: MFP$_a(\alpha) \subseteq \text{AP}_a^{n+1}(\alpha)$. For an arbitrary $x \in \text{MFP}_a(\alpha)$, we obtain from the data flow equations that $x \in \text{MFP}_a(\beta)$ or $x \in \text{MFP}_a(\gamma)$. Further from the induction hypothesis, we have $x \in \text{AP}_a^n(\beta)$ or $x \in \text{AP}_a^n(\gamma)$. As a consequence $x \in \text{vars}(\beta)$ or $x \in \text{vars}(\gamma)$, but because $\text{vars}(\alpha) = \text{vars}(\beta) = \text{vars}(\gamma)$, we have for sure $x \in \text{vars}(\alpha)$.

Note that if $\mathcal{A}(\beta)$ is not reachable in CFG, then every path to $\mathcal{A}(\alpha)$ is not valid. Then we would have $\text{AP}_a^{n+1}(\alpha) = \text{vars}(\alpha)$ and implicitly $x \in \text{AP}_a^{n+1}(\alpha)$.

If $\mathcal{A}(\beta)$ is reachable, then every valid path to $\mathcal{A}(\alpha)$ should have as last edge $(\mathcal{A}(\gamma), \mathcal{A}(\alpha))$ and should implicitly pass through $\mathcal{A}(\beta)$. That is why every valid path to $\mathcal{A}(\alpha)$ pass through both $\mathcal{A}(\beta)$ and $\mathcal{A}(\gamma)$. Considering that $x \in \text{AP}_a^n(\beta)$ or $x \in \text{AP}_a^n(\gamma)$ and the above remark, we can conclude $x \in \text{AP}_a^{n+1}(\alpha)$. □
We can prove actually more: the MFP solution is not only an approximation of \( \text{AP} \) but it is perfect (Theorem 2). For this, we need also the following theorem which states that the MFP solution contains the local variables which are initialized over all possible paths.

**Theorem 2** For every expression or statement \( \alpha \), the following relations are true: \( \text{AP}_b(\alpha) \subseteq \text{MFP}_b(\alpha) \) and \( \text{AP}_a(\alpha) \subseteq \text{MFP}_a(\alpha) \). Moreover, if \( \alpha \) is a boolean expression, then we have also \( \text{AP}_t(\alpha) \subseteq \text{MFP}_t(\alpha) \) and \( \text{AP}_f(\alpha) \subseteq \text{MFP}_f(\alpha) \).

**Proof.** Tarski’s fixed point theorem [13] states that MFP is the lowest upper bound (with respect to \( \subseteq \)) of the set \( \text{Ext}(F) = \{ X \in D \mid X \subseteq F(X) \} \). It suffices to show that the \( r \)-tuple consisting of the \( \text{AP} \) sets is an element of \( \text{Ext}(F) \) since MFP is in particular an upper bound of this set. Since \( \subseteq \) is the pointwise subset relation, the idea is to prove, for the data flow equations in Tables 1, 2 and 3 the left-to-right subset relations where instead of the sets before, after, true and false we have the sets \( \text{AP}_b, \text{AP}_a, \text{AP}_t \) and \( \text{AP}_f \), respectively.

We will make the proof first for the data flow equations of the boolean expressions from Table 1.

**Case 1.** \( \alpha \text{\ _exp} = \text{true} \)
We want to prove \( \text{AP}_t(\alpha) \subseteq \text{AP}_b(\alpha) \) and \( \text{AP}_t(\alpha) \subseteq \text{vars}(\alpha) \). If \( x \) is a local variable in \( \text{AP}_t(\alpha) \), then \( x \) is assigned on every path to \( T(\alpha) \) (see the definition of \( \text{AP}_t(\alpha) \)). Since in CFG all the paths to the true point \( T(\alpha) \) contains the entry point \( B(\alpha) \) and moreover there is no assignment to \( x \) in true, we deduce that \( x \) is assigned on every path to \( B(\alpha) \) and consequently \( x \in \text{AP}_b(\alpha) \), i.e. \( \text{AP}_t(\alpha) \subseteq \text{AP}_b(\alpha) \). The second subset relation holds because of the definition of \( \text{AP}_f(\alpha) \).

**Case 2.** \( \alpha \text{\ _exp} = \text{false} \)
Similar to Case 1.

**Case 3.** \( \alpha \text{\ _exp} = (\neg e) \)
We show that the following holds: \( \text{AP}_b(e) \subseteq \text{AP}_b(\alpha) \). Let \( x \) be an arbitrary local variable in \( \text{AP}_b(e) \), i.e. \( x \) is assigned on every path to \( B(\alpha) \). We observe that all paths to \( B(\alpha) \) contain the point \( B(\alpha) \) but since no local variable is assigned on the edge \( (B(\alpha), B(\alpha)) \), we infer that \( x \) is assigned on every path to \( B(\alpha) \), i.e. \( x \in \text{AP}_b(\alpha) \).

Similar remarks hold also for the other two equations \( \text{AP}_t(\alpha) \subseteq \text{AP}_f(\beta) \) and \( \text{AP}_t(\beta) \subseteq \text{AP}_t(\delta) \): all the paths to \( T(\alpha) \) and to \( T(\delta) \) contain \( T(\beta) \) and \( T(\delta) \), respectively.

**Case 4.** \( \alpha \text{\ _exp} = (\beta \_e_0 \ ? \ \_e_1 : \_e_2) \)
The proofs of the first three equations are similar to Case 1. In order to prove \( \text{AP}_t(\alpha) \subseteq \text{AP}_t(\gamma) \cap \text{AP}_t(\delta) \), we notice that in CFG, the paths to \( T(\gamma) \) pass either through \( T(\gamma) \) or through \( T(\delta) \). If \( x \in \text{AP}_t(\alpha) \), then \( x \) is assigned on every path to \( T(\gamma) \) and on every path to \( T(\delta) \). Hence \( x \in \text{AP}_t(\gamma) \cap \text{AP}_t(\delta) \). The proof for \( \text{AP}_f(\alpha) \subseteq \text{AP}_f(\gamma) \cap \text{AP}_f(\delta) \) is similar.

**Case 5.** \( \alpha \text{\ _exp} = (\beta \_e_1 \ & \ \& \ \_e_2) \)
Similar to Case 4.

**Case 6.** \( \alpha \text{\ _exp} = (\beta \_e_1 \ || \ \_e_2) \)
Similar to Case 4.

We will prove now the relations corresponding to the arbitrary expressions in Table 2.

**Case 7.** \( \alpha \text{\ _exp} = \text{loc} \)
Similar to the first proof in Case 3.

**Case 8.** \( \alpha \text{\ _exp} = \text{lit} \)
Similar to the first proof in Case 3.

**Case 9.** \( \alpha \text{\ _exp} = (\text{loc} = \_e_1) \)
The proof of \( \text{AP}_b(e) \subseteq \text{AP}_a(\alpha) \) is similar to the first proof in Case 3.

We want to prove \( \text{AP}_a(\alpha) \subseteq \text{AP}_a(\beta) \cup \text{struct}(\text{loc}, \beta, \{ \text{loc} \}) \). Let \( x \) be an arbitrary local variable in \( \text{AP}_a(\alpha) \). Suppose that \( x \) is not in the set \( \text{AP}_a(\beta) \) otherwise the proof is complete. This means there exists a path \( p \) to \( A(\beta) \) on which \( x \) is not assigned. But \( x \) should be assigned on \( p \cup A(\alpha) \) since \( x \in \text{AP}_a(\alpha) \). Therefore, the assignment to \( \text{loc} \) to the edge \( (A(\beta), A(\alpha)) \) determines the initialization of \( x \). According to the definition of initialization, this happens only in the following cases:
• the simple assignment to loc on the edge \((A(\beta), A(\alpha))\) is actually an assignment to \(x\), i.e. \(x = loc\). So \(x \in \text{struct}(\text{loc}, \beta, \{\text{loc}\})\).

• \(x_i\) is an instance field of a struct type variable \(x_{i+1}, i = 1, n - 1\) where \(x_1 = x\) and \(x_n = loc\). In this case, we get \(x \in \text{lookup}(\text{loc}) \subseteq \text{struct}(\text{loc}, \beta, \{\text{loc}\})\).

• \(x_i\) is an instance field of a struct type variable \(x_{i+1}, i = 1, n - 1\) where \(x_1 = loc\) and \(x_n = x\). If all instance fields of \(x_{i+1}\) except \(x_i\) are in the set \(\text{AP}_a(\beta)\) for \(i = 1, n - 1\), then we get \(x \in \text{lookup}(\text{loc}, \beta, \{\text{loc}\}) \subseteq \text{struct}(\text{loc}, \beta, \{\text{loc}\})\). Let assume now there exists \(k\) such that an instance field \(y\) of \(x_k\) is not in \(\text{AP}_a(\beta)\); then there exists a path \(q\) (obviously other than \(p\)) on which \(y\) is not assigned. Clearly, \(y\) is not assigned also on \(q \oplus A(\alpha)\) and therefore also \(x\) is not assigned on \(q \oplus A(\alpha)\). But this contradicts \(x \in \text{AP}_a(\alpha)\).

**Case 10.** \(a \exp = (\text{loc} \ ops = \beta e)\)

Similar to the first proof in Case 3. This case is different from Case 9 since a compound assignment is not considered an initialization (because the variable is assumed to be already initialized). This is the reason there is no initialization on the edge \((B(\beta), B(\alpha))\).

**Case 11.** \(a \exp = (\beta e_0 ?^* e_1 : \gamma e_2)\)

Similar to Case 4.

**Case 12.** \(a \exp = c.f\)

Similar to the first proof in Case 3.

**Case 13.** \(a \exp = \text{ref } \beta \exp\)

Similar to the first proof in Case 3.

**Case 14.** \(a \exp = \text{out } \beta \exp\)

Similar to the first proof in Case 3.

**Case 15.** \(a \exp = c.m(\beta_1 \ arg_1, \ldots, \beta_k \ arg_k)\)

The proofs of \(\text{AP}_b(\beta_1) \subseteq \text{AP}_b(\alpha)\) and \(\text{AP}_b(\beta_{i+1}) \subseteq \text{AP}_a(\beta_i)\), \(i = 1, n - 1\) are similar with the first proof in Case 3.

We want show:

\[
\text{AP}_a(\alpha) = \text{AP}_a(\beta_k) \cup \bigcup_{\arg_i \in \text{OutParams}(\arg_1, \ldots, \arg_k)} \text{struct}(\arg_i, \beta_k, \text{OutParams}(\arg_1, \ldots, \arg_k))
\]

Let us consider a variable \(x\) in \(\text{AP}_a(\alpha)\) and let us suppose that \(x \notin \text{AP}_a(\beta_k)\). Then there exists a path \(p\) to \(A(\beta_k)\) on which \(x\) is not assigned. On the other hand, \(x\) is assigned on \(p \oplus A(\alpha)\) since \(x \in \text{AP}_a(\alpha)\). Therefore the initialization of \(x\) on the edge \((A(\beta_k), A(\alpha))\) is due to an initialization of an out parameter \(\arg_i\). Further the proof is similar with the proof in the case of a simple assignment to \(\arg_i\) (see Case 9).

**Case 16.** The proofs corresponding to the general data flow equations for the expressions which are not instances of expressions from Tables 1 and 2 are similar with first proof in Case 3.

We will consider now for proving the statements from Table 3

**Case 17.** \(a \text{stm } = \\)\n
Similar to the first proof in Case 3.

**Case 18.** \(a \text{stm } = (\beta \exp ;)\)

Similar to the first proof in Case 3.

**Case 19.** \(a \text{stm } = \{^\beta_1 \text{stm}_1 \ldots ^\beta_n \text{stm}_n\}\)

A special case of a block is our method body. In order to prove \(\text{AP}_b(mb) \subseteq \emptyset\), it is enough to notice that \(\text{AP}_b(mb) = \emptyset\). This is indeed so because the only path to \(B(mb)\) is \([B(mb)]\) on which there is no initialization of any local variable.

For an arbitrary block of statements, the proof of \(\text{AP}_b(\beta_i) \subseteq \text{AP}_b(\alpha)\) is similar to the first proof in Case 3. To show \(\text{AP}_b(\alpha) \subseteq \text{AP}_b(\beta_i) \cap \text{vars}(\alpha)\), it suffices to notice that we have \(\text{AP}_a(\alpha) \subseteq \text{AP}_a(\beta_i)\) (proved with similar arguments like in Case 3) and \(\text{AP}_a(\alpha) \subseteq \text{vars}(\alpha)\) (from the definition of \(\text{AP}_a(\alpha)\)).
We want to prove now the subset relation for an arbitrary index $i$: $\text{AP}_b(\beta_{i+1}) \subseteq \text{AP}_a(\beta_i) \cap \text{goto}(\beta_{i+1})$

Note that here, $\text{goto}(\beta_{i+1})$ depends only on the AP sets and on the control flow graph $\text{CFG}$. We have two possible cases:

- If the statement $\beta_{i+1}$ is not a labeled statement, then according to the definition of $\text{goto}(\beta_{i+1})$, we have $\text{goto}(\beta_{i+1}) = \text{vars}(\beta_{i+1}) = \text{vars}(\beta_i)$ and since $\text{AP}_a(\beta_i) \subseteq \text{vars}(\beta_i)$, we need to prove only $\text{AP}_b(\beta_{i+1}) \subseteq \text{AP}_a(\beta_i)$ which is obvious since, all the paths to $\text{A}(\beta_i)$ are, except the edge $(\text{A}(\beta_i), \text{B}(\beta_{i+1}))$, also paths to $\beta_{i+1}$ and no variable is assigned on this edge. Note that it does not matter for the above proof whether $\text{A}(\beta_i)$ is reachable or not in $\text{CFG}$.

- If the statement $\beta_{i+1}$ is a labeled statement $L::stm$, then we could have more edges in $\text{CFG}$ leading to $\text{B}(\beta_{i+1})$. Let us consider a variable $x \in \text{AP}_b(\beta_{i+1})$, i.e. there exists an assignment to $x$ on every path leading to $\text{B}(\beta_{i+1})$. Exactly like in other case we can prove $\text{AP}_b(\beta_{i+1}) \subseteq \text{AP}_a(\beta_i)$. In order to show $x \in \text{goto}(\beta_{i+1})$, we need to prove that $x$ is in the set $\text{AP}_b(\gamma) \cup \text{JoinFin}(\gamma, \beta_{i+1})$ for every associated $\text{goto}$ statement $\gamma$.

Depending on whether there are $\text{goto}$ statements pointing to $\beta_{i+1}$ we have the following case distinction:

- If there is no associated $\text{goto}$ statement, then obviously we have $x \in \text{goto}(\beta_{i+1})$ since in this case $\text{goto}(\beta_{i+1}) = \text{vars}(\beta_{i+1})$.

- If there exists at least one associated $\text{goto}$ statement $\gamma$.
  - If $B(\gamma)$ is not reachable in $\text{CFG}$, then $\text{path}_b(\gamma) = \emptyset$ and consequently we get $x \in \text{AP}_b(\gamma) \cup \text{JoinFin}(\gamma, \beta_{i+1})$ because $\text{AP}_b(\gamma) = \text{vars}(\gamma) \supseteq \text{vars}(\beta_{i+1})$. The last subset relation holds because, in $\text{CFG}$ a $\text{goto}$ statement should be always in the scope of the corresponding labeled statement.
  - If $B(\gamma)$ is reachable in $\text{CFG}$, then let $p$ be an arbitrary path to $B(\gamma)$. We will make now another case distinction.
    - If there are no $\text{finally}$ blocks from $\gamma$ to $\beta_{i+1}$, i.e. $\text{Fin}(\gamma, \beta_{i+1}) = \emptyset$, then we get $\text{JoinFin}(\gamma, \beta_{i+1}) = \emptyset$ (since the union of an empty collection of sets is defined to be the empty set). This means that, in this case, it suffices to show $x \in \text{AP}_b(\gamma)$.
      In this case, we have in $\text{CFG}$ also the edge $(B(\gamma), B(\beta_{i+1}))$ that is defined by the set $\text{ThroughFin}_b(\gamma, \beta_{i+1})$. So, $p \oplus B(\beta_{i+1})$ is a path to $B(\beta_{i+1})$ and since $x \in \text{AP}_b(\beta_{i+1})$, we are sure that there is an assignment to $x$ also on $p$. Because $p$ was arbitrarily chosen, we get $x \in \text{AP}_b(\gamma)$.
    - Let us consider now the case when there are $\text{finally}$ blocks from $\gamma$ to $\beta_{i+1}$, i.e. $\text{Fin}(\gamma, \beta_{i+1}) = [\gamma_1, \ldots, \gamma_k]$. Accordingly, also the edges
      $$(B(\gamma), B(\gamma_1)), (A(\gamma_1), B(\beta_{i+1})), (A(\gamma_2), B(\gamma_2)), \ldots, (A(\gamma_k), B(\gamma_k))$$
    from the set $\text{ThroughFin}_b(\gamma, \beta_{i+1})$ are added to $\text{CFG}$. We will prove $x \in \text{AP}_b(\gamma) \cup \text{JoinFin}(\gamma, \beta_{i+1})$ by reduction to absurdum. Let us assume that $x \notin \text{AP}_b(\gamma) \cup \text{JoinFin}(\gamma, \beta_{i+1})$. This is equivalent with $x \notin \text{AP}_b(\gamma)$ and $x \notin \text{AP}_a(\gamma_j)$ for all $j = 1, \ldots, k$. This means that there exist the paths $p_j \in \text{path}_b(\gamma_j)$ for $j = 1, \ldots, k$ such that $x$ is not assigned on any of these paths. A simple inspection of $\text{CFG}$ shows that, necessarily the point $B(\gamma_j)$ occurs in the path $p_j$ for every $j = 1, \ldots, k$ since it is not possible to ‘jump’ into a $\text{finally}$ block. We want to prove now that the following
      $$p_0 \ominus p_1 \ominus p_2 \ominus \cdots \ominus p_k \ominus \text{B}(\gamma_k) \ominus \text{A}(\gamma_k) \oplus B(\beta_{i+1})$$
    is a valid path to $B(\beta_{i+1})$. The only problem that could arise is concerning the $\text{conditioned}$ edges. Remember that the edges $\text{conditioned}$ by a certain $\text{goto}$, $\text{break}$ or $\text{continue}$ statement can be used only in paths or subpaths that contain the entry point of the respective jump statement.
    The use of the edges $(B(\gamma), B(\gamma_1)), \ldots, (A(\gamma_k), B(\beta_{i+1}))$ is correct as long as our path contains $B(\gamma)$.  

19
Let us consider one of the subpaths \( p_j[B(\gamma_j), A(\gamma_j)] \) contained in the above defined path. If it contains a **conditioned edge**, then since the **conditioned** edges connect jump statements with **finally** blocks, we are sure that these embedded **finally** blocks are contained in our **finally** block \( \gamma_j \). Then the respective jump statement is embedded into the **try** blocks (corresponding to the **conditioned** connected **finally** blocks) which necessarily should be embedded in our **finally** block \( \gamma_i \) (this is an immediate consequence of the \( C_2 \) grammar). So the jump statement is embedded into our **finally** block. Considering that the subpath \( p_j[B(\gamma_j), A(\gamma_j)] \) contains **conditioned** edges, we get that also \( p_j \) uses the same **conditioned** edges and since we assumed that \( p_j \) is a valid path, necessarily \( p_j \) should contain the entry point of the respective jump statement which, as we proved above, is embedded in our **finally** block, and consequently appears in the subpath \( p_j[B(\gamma_j), A(\gamma_j)] \). This means that this subpath is valid. Obviously this is true for all the above considered subpaths.

The above defined path to \( B(\beta_{i+1}) \) does not assign \( x \). Obviously this contradicts \( x \in AP_b(\beta_{i+1}) \) and therefore our assumption is wrong. Hence, we obtain the desired \( x \in AP_b(\gamma) \cup JoinFin(\gamma, \beta_{i+1}) \).

There is one more special case, the **while** statements. In the same way like in case of a labeled statement, also for \( ^{i+1} \text{while} \) statement we may have several edges to \( B(\beta_{i+1}) \) if they are associated **continue** statements embedded into \( \beta_{i+1} \). But we do not care for the proof of this theorem about possible edges coming from **continue** statements: we need only to prove \( AP_b(\beta_{i+1}) \subseteq AP_b(\beta) \) which can be done exactly like in the case of a regular statement, considering only the paths via \( A(\beta) \).

This completes the proofs of the data flow equations for a block of statements.

**Case 20.** \( ^\alpha \text{stm} = \text{if } (^\beta \text{exp} \, \gamma \text{stm}_1 \ \text{else } ^\delta \text{stm}_2) \)

Similar to Case 4.

**Case 21.** \( ^\alpha \text{stm} = \text{while}(^\beta \text{exp}) \, \gamma \text{stm} \)

The proofs corresponding to the first two equations are similar with first proof in Case 3. We want now to prove the third equation:

\[
AP_a(\alpha) \subseteq AP_f(\beta) \cap break(\alpha)
\]

where the set \( break(\alpha) \) depends only on the \( AP \) sets and on the control flow \( CFG \). This proof is similar with corresponding proof for a labeled statement in Case 19 with the following observations: the role of \( A(\beta) \) is played now by \( F(\beta), B(\beta_{i+1}) \) by \( A(\alpha) \) and the goto statements are replaced by break statements.

**Case 22.** The proofs for the goto, break, continue, return, return exp, throw and throw exp statements (except those involving exp) are trivial: \( AP_a(\alpha) \subseteq vars(\alpha) \) is true because of the definition of \( AP_a(\alpha) \). The proofs for the points involving exp follow the same pattern like the first proof in Case 3.

**Case 23.** \( ^\alpha \text{stm} \) is a try-catch statement. The following subset relations

\[
AP_b(\beta) \subseteq AP_b(\alpha), AP_b(\gamma_i) \subseteq AP_b(\alpha) \cup \{x_i\}, i = 1, n
\]

\[
AP_a(\alpha) \subseteq AP_a(\beta), AP_a(\alpha) \subseteq AP_a(\gamma_i), i = 1, n
\]

are satisfied since in \( CFG \) we have defined the edges

\[
(B(\alpha), B(\beta)), (A(\beta), A(\alpha)), (B(\alpha), B(\gamma_i)), (A(\gamma_i), A(\alpha)), \ i = 1, n
\]

This completes the proofs for the try-catch statements.

**Case 24.** \( ^\alpha \text{stm} \) is a try-finally statement. The proofs for the first two relations are similar to Case 23.

We want to prove now \( AP_a(\alpha) \subseteq AP_a(\beta) \cup AP_a(\gamma) \). Note that there is no regular edge \( (A(\beta), A(\alpha)) \) or \( (A(\gamma), A(\alpha)) \) in \( CFG \). There is only the conditioned edge \( (A(\gamma), A(\alpha)) \).

We will make the proof by reduction to absurdum. Let us suppose there exists an \( x \in AP_a(\alpha) \) such that \( x \notin AP_a(\beta) \) and \( x \notin AP_a(\gamma) \). Then there exist the paths \( p_1 \in path_a(\beta) \) and \( p_2 \in path_a(\gamma) \) that do not
contain any assignments to \( x \). The construction of \( CFG \) shows that the point \( B(\gamma) \) occurs in any path to \( A(\gamma) \) (it is not possible to jump into the \text{finally} block), in particular in \( p_2 \).

The following
\[
p_1 \oplus p_2[B(\gamma), A(\gamma)] \oplus A(\alpha)
\]
is a valid path to \( A(\alpha) \) since the use of the conditioned edge is OK as long as the path contains \( A(\beta) \) and the subpath \( p_2[B(\gamma), A(\gamma)] \) is valid (with the same proof like in the case of a labeled statement).

Under our assumption, the above constructed path does not contain any assignments to \( x \) and this contradicts \( x \in AP_a(\alpha) \). It means that our assumption was wrong and we can conclude that \( AP_a(\alpha) \subseteq AP_a(\beta) \cup AP_s(\gamma) \).

The following result is then an obvious consequence of Theorem 1 and Theorem 2.

**Theorem 3** The maximal fixed point solution of the data flow equations in Tables 1, 2, 3 represents the sets of local variables which are assigned over all possible execution paths. More exactly, for every expression or statement \( \alpha \), the following are true: \( AP_b(\alpha) = MFP_b(\alpha) \) and \( AP_s(\alpha) = MFP_s(\alpha) \). Moreover, if \( \alpha \) is a boolean expression: \( AP_t(\alpha) = MFP_t(\alpha) \) and \( AP_f(\alpha) = MFP_f(\alpha) \).

6 Conclusion and future work

In this paper, we have formalized the definite assignment analysis of C\# by data flow equations. Since the equations do not always have a unique solution, we defined the outcome of the analysis as the solution of a fixed point iteration. We proved that there exists always a maximal fixed point solution \( MFP \). We showed the correctness of the analysis, i.e. \( MFP \) is a safe approximation of the sets of variables assigned over all possible paths. This is a key property for the type safety of C\# whose formalization is part of future work as well as proving the correctness of C\# compilers. This paper is part of a research project focusing on formalizing and verifying important aspects of C\#. So far, we have an ASM model for the operational semantics of C\# in [2]. During the attempts to build this model, there were discovered in [4] a few discrepancies between the C\# Specification and different implementations of C\#.

**Acknowledgment** We gratefully acknowledge partial support of the related work [2] by a Microsoft grant within the ROTOR project during the year 2002-2003.

**References**


7 Appendix

During the attempts to build the formalization of the definite assignment analysis in C♯ a few bugs were discovered in the Mono C♯ compiler [7] (version 0.26).

- **Test1**: Mono C♯ compiler claims that $i$ is unassigned.

```csharp
using System;

class Test1 {
    public static void Main() {
        bool b = false;
        int i;
        if (b) {
            i = 1;
            goto L;
        }
        return;
        L: Console.WriteLine(i);
    }
}
```

At the entry of the labeled statement, $i$ should be considered definitely assigned since it is assigned on the single path leading to the labeled statement - that is the path which contains the `goto` statement.

- **Test2**: Mono C♯ compiler accepts the program.

```csharp
using System;

class Test2 {
    public static void Main() {
        bool b = false;
        int i;
        if (b && (i = 1) >= 0)
            Console.WriteLine(i);
        else
            Console.WriteLine(i);
    }
}
```

If $b$ evaluates to `false`, then the test $(b && (i = 1) >= 0)$ is evaluated to `false` and the second operand of `&&`, i.e. $(i = 1) >= 0$ is not anymore evaluated. So in this case $i$ is not assigned. This means that $i$ is not definitely assigned at the entry point of the `else` branch. Although $i$ is not definitely assigned at that point, the Mono C♯ compiler does not reject the program.

- **Test3**: Mono C♯ compiler accepts the program.

```csharp
using System;

class Test3 {
    public static void Main() {
        bool b = true;
        int i;
        if (b || (i = 1) >= 0)
            Console.WriteLine(i);
        else
```
If $b$ evaluates to true, then the test $(b || (i = 1) >= 0)$ is evaluated to true and the second operand of 11, i.e. $(i = 1) >= 0$ is not anymore evaluated. So in this case $i$ is not assigned. This means that $i$ is not definitely assigned at the entry if the then branch. Although $i$ is not definitely assigned at that point, the Mono C$\#$ compiler does not reject the program.

• **Test4:** Mono C$\#$ compiler claims that $i$ is unassigned.

```csharp
class Test4 {
    public static void Main() {
        bool b = false;
        int i = 0;
        while (true) {
            if (b) {
                i = 1;
                break;
            }
            Console.WriteLine(i);
        }
    }
}
```

It is not clear at all from the C$\#$ Specification, what are the definitely assigned variables after the execution of the while statement in case there exists an appropriate break statement. Our equations answer precisely to this question: the variables which are definitely assigned both before the break statement and after the evaluation to false of the constant true (this happens for all local variables), in particular $i$. So the example should not be rejected by the Mono C$\#$ compiler.

• **Test5:** Mono C$\#$ compiler claims that $i$ is unassigned.

```csharp
class Test5 {
    public static void Main() {
        bool b = false;
        int i;
        while (b?true:true);
        Console.WriteLine(i);
    }
}
```

According to C$\#$ Specification [§5.3.3.7], the variables which are definitely assigned after the execution of the while statement are the variables which are assigned after the evaluation to false of the test expression $b?$true$:$true that variables which are definitely assigned after the evaluation to false of the constant true. This means that all the local variables in the scope of which while is should be considered as definitely assigned, in particular $i$. That is why, the Mono C$\#$ compiler should not reject the program. Note that the end point of while is considered to be reachable and therefore it is not the same situation like in the next example.

• **Test6:** Mono C$\#$ compiler claims that $i$ is unassigned.

```csharp
class Test6 {
    public static void Main() {
    }
}
```
According to C# Specification [1 §5.3.3.1], \(i\) is definitely assigned at the beginning of any unreachable statement. In our case, the \texttt{else} branch is unreachable since the test is the constant \texttt{true} [1 §8.7.1]. This means that \(i\) is definitely assigned at the beginning of the \texttt{else} branch and the program should not be rejected by the Mono C# compiler.