A STATIC ANALYSIS OF CSP PROGRAMS

Krzysztof R. APT

LITP, Université Paris 7
2, Place Jussieu
75251 PARIS
France

Abstract A static analysis is proposed as a method of reducing complexity of the correctness proofs of CSP programs. This analysis is based on considering all possible sequences of communications which can arise in computations during which the boolean guards are not interpreted. Several examples are provided which clarify its various aspects.

1. INTRODUCTION

Correctness proofs of concurrent and distributed programs are complicated because in general they are of the length proportional to the product of the lengths of the component programs. We claim in this paper that in the case of the CSP programs the length and the complexity of these proofs can be substantially reduced by carrying out first a preliminary static analysis of the programs. This analysis allows to reduce the number of cases which have to be considered at the level of interaction between the proofs of the component programs.

The analysis is quite straightforward and contains hardly any new ideas. It is based on considering all possible sequences of communications which can arise in computations during which the boolean guards are not interpreted. In this respect it bears a strong resemblance to the trace model for a version of CSP given in [H 1].

We apply this analysis to three types of problems. The first one consists of determining which pairs of input-output commands (i/o commands) may be synchronized during properly terminating computations. The second one consists of determining all possible configurations in which deadlock occurs. Finally we provide a sufficient condition for safety of a decomposition of CSP programs into communication-closed layers, a method of decomposition which has been recently proposed by Elrad and Francez [EF].

A similar analysis can be carried out for other programming languages which use rendez-vous as a sole means for communication and synchronization. In fact while writing this paper we encountered in the last issue of the Communications of ACM a paper by Taylor [T] in which such an analysis is carried out for ADA programs.

The only difference is in the presentation of this approach. R.N. Taylor presents an algorithm which computes all rendez-vous which may take place during execution of a program and all configurations in which deadlock may arise. His algorithm is also capable of determining which actions may occur in parallel. We on the other he present the analysis in a formal language theory framework providing the rigorous definitions which can be used in the case of concrete examples. We also link this analysis with a subsequent stage being the task of proving correctness of the programs.

The paper is organized as follows. In the next section we introduce the basic definitions. In section 3 we provide three applications already mentioned above. Section 4 is devoted to a more refined analysis which takes into account the problem of termination of the repetitive commands. Finally in section 5 a numbe: of conclusions is presented.

2. BASIC DEFINITIONS

We assume that the reader is familiar with the original article of Hoare [H] Throughout the paper we consider programs written in a subset of CSP. We disallow nested parallel composition, assume that all variables are of the same type and consequently omit all the declarations. Additionally we allow output commands to be used as guards. For the reasons which will become clear later we label each occurence of an input or output command by a unique label.

By a parallel program we mean a program of the form $P_1 \| \ldots \| P_n$ where each P_i is a process. For simplicity we drop the process labels. So according to the notation of [H] each process is identified with the command representing its body. The name of the process can be uniquely determined from the position of the command within the parallel composition.

The analysis carried out here can be straightforwardly extended to the full CSP.

Throughout the paper we denote by S,T arbitrary (sequential) commands, by g guards, by b,c boolean expressions, by t expressions and by α,β i/o commands. Labels of the i/o commands are denoted by the letters k,l,m. Finally, we write $[\ \square_{i=1}^{m} g_i \to S_i\ ]$ instead of $[g_1 \to S_1 \square \ldots \square g_m \to S_m]$.

Consider now a parallel program $P_1 \| \ldots \| P_n$. We proceed in two stages.

1°) With each process P_i we associate a regular language $L(P_i)$ defined by structural induction. We put

$L(x:=t) = L(skip) = \{\epsilon\}$,

$L(l:P_j!t) = \{l : \langle i,j \rangle\}$,

$L(l{:}P_j?x) = \{l : \langle j,i\rangle\}$,

$L(S_1;S_2) = L(S_1)L(S_2)$,

$L(g \to S) = L(g)L(S)$,

$L(b) = \{\epsilon\}$, $L(b;l{:}\alpha) = L(b)L(l{:}\alpha)$ $(=L(l{:}\alpha))$,

$L([\,\overset{m}{\underset{i=1}{\Box}}\, g_i \to S_i\,]) = \overset{m}{\underset{i=1}{\cup}} L(g_i \to S_i)$,

$L(*[\,\overset{m}{\underset{i=1}{\Box}}\, g_i \to S_i\,]) = L([\,\overset{m}{\underset{i=1}{\Box}}\, g_i \to S_i\,])^*$

Note that $L(P_i)$ is the set of all a priori possible communication sequences of P_i when the boolean guards are not interpreted. Each communication sequence consists of elements of the form $l{:}\langle i,j\rangle$ or $l{:}\langle j,i\rangle$ where l is a label of an i/o command uniquely identified and $\langle i,j\rangle$ ($\langle j,i\rangle$) records that fact that this i/o command stands for a communication from $P_i(P_j)$ to $P_j(P_i)$.

It is important that we associate with assignment and skip statements the set $\{\epsilon\}$ and not the empty language $\emptyset$. Otherwise not all communication sequences would be recorded in $L(P_i)$. The following example clarifies this issue.

Example 1

Let

$$P_1 \equiv [b_1 \to \text{skip} \;\Box\; b_2 \to k{:}P_2!x] \;;$$
$$*[l{:}P_2?y \to \ldots \;;\; m{:}P_2!y]$$

where ... stands for a "private part" of P_1, i.e. a command not involving any i/o commands. Then

$$L(P_1) = \{(l{:}\langle 2,1\rangle)(m{:}\langle 1,2\rangle)\}^*$$
$$\cup\; \{k{:}\langle 1,2\rangle\}\{(l{:}\langle 2,1\rangle)(m{:}\langle 1,2\rangle)\}^*$$

If we associated with skip the empty language then the first part of $L(P_1)$ would not be present even though it represents possible communication sequences.

2°) We associate with $P_1\|\ldots\|P_n$ a regular language $L(P_1\|\ldots\|P_n)$. Its letters are of the form $k,l{:}\langle i,j\rangle$ standing for an instance of a communication between the output command of P_i labeled by k and the input command of P_j labeled by l.

First we define a projection function $[.]_i$ $(1\le i\le n)$ from the alphabet of $L(P_1\|\ldots\|P_n)$ into the alphabet of $L(P_i)$. We put

$[k,l{:}\langle i,j\rangle]_i = k{:}\langle i,j\rangle$

$[k,l{:}\langle i,j\rangle]_j = l{:}\langle i,j\rangle$

$[k,l{:}\langle i,j\rangle]_h = \epsilon$ if $h \neq i,j$

and naturally extend it to a homomorphism from the set of words of $L(P_1\|\ldots\|P_n)$ into the set of words of $L(P_i)$.

We now define

$L(P_1 \| \ldots \| P_n) = \{h : [h]_i \in L(P_i),\ i = 1,\ldots,n\}$

Intuitively, $L(P_1 \| \ldots \| P_n)$ is the set of all possible communication sequences of $P_1 \| \ldots \| P_n$ which can arise in properly terminating computations during which the boolean expressions are not interpreted.

3. APPLICATIONS

1. Partial correctness

Given a parallel program $P_1 \| \ldots \| P_n$ we define

$STAT = \{(k:\alpha, l:\beta) : k:\alpha$ is from P_i, $l:\beta$ is from P_j,

$\&\ \exists h\ \exists a\ [h \in L(P_1 \| \ldots \| P_n)$, a is an element of h,

$L(k:\alpha) = \{[a]_i\}$ and $L(l:\beta) = \{[a]_j\}\}$.

Intuitively STAT (standing for static match) is the set of all pairs of i/o commands which can be synchronized during a properly terminating computation of $P_1 \| \ldots \| P_n$ which ignores the boolean guards.

The set STAT should be compared with two other sets of pairs of i/o commands :

$SYNT = \{(k:\alpha, l:\beta) : k:\alpha$ is from P_i, $l:\beta$ is from P_j and $k:\alpha$ and $l:\beta$ address each other (match)$\}$

$SEM = \{(k:\alpha, l:\beta) :$ in some "real" properly terminating computation of $P_1 \| \ldots \| P_n$ $k:\alpha$ and $l:\beta$ are synchronized$\}$.

In the proof systems of [AFR] and [LG] dealing with partial correctness of CSP programs the crucial proof rule is the one that deals with the parallel composition of the processes. First one introduces so called proof outlines for component processes. A proof outline of S is a special form of a proof of partial correctness of the program S in which each subprogram of S is preceded and succeded by an assertion. These assertions are supposed to hold at the moment when the control is at the point to which they are attached. As behaviour of each component process depends on the other processes we ensure the above property by comparing proof outlines of the component processes. Given a proof outline the only assertions which have to be justified using proof outlines ot other processes are those succeeding the i/o commands.

Thus one identifies all pairs of possibly matching i/o commands and checks that the assertions attached to them are indeed justified when the communication takes place. This part of verification of the proof outlines is called in [AFR] the cooperation test and in [LG] the satisfaction test.

If the proof outlines satisfy the test then one can pass to the conclusion stating partial correctness of the parallel program.

We now concentrate on the step consisting of identifying all pairs of possibly matching i/o commands. According to our definition this is the set SEM. But since SEM is in general not computable as a function of the program $P_1 \| \ldots \| P_n$, this set is replaced in [AFR] and [LG] by a larger set SYNT being obviously computable. We propose to replace in this analysis the set SEM by the set STAT.

Note that the following clearly holds.

Fact $SEM \subseteq STAT \subseteq SYNT$

Moreover, the set STAT is obviously computable. Using the set STAT instead of SYNT as an "approximation" for SEM is more economical as less checks in the cooperation (satisfaction) test phase are then needed. Also the proof outlines (and in the case of [AFR] - the global invariant) can be simplified.

As an illustration of the difference between the sets STAT and SYNT consider the following example :

Example 2

Let

$$P_1 \equiv k_1:P_2?x\ ;\ldots;k_2:P_2!z\ ;\ldots;$$
$$*[b_1 \rightarrow \ldots;\ k_3:P_2?x\ ;\ldots;\ k_4:P_2!z\ ;\ldots],$$

$$P_2 \equiv l_1:P_1!y\ ;\ldots;\ l_2:P_1?u\ ;\ldots;$$
$$*[b_2 \rightarrow \ldots;\ l_3:P_1!y\ ;\ldots;\ l_4:P_1?u\ ;\ldots]$$

Then

$$STAT = \{(k_i:\alpha_i,\ l_i:\beta_i) : 1 \leq i \leq 4\}$$

and

$$SYNT = \{(k_i:\alpha_i,\ l_j:\beta_j) :\ |i-j| \text{ is even}, 1 \leq i,j \leq 4\}.$$

Thus STAT has here 4 elements whereas SYNT has 8 elements.

The difference between STAT and SYNT becomes more evident for longer programs. For example if in the above programs both repetitive commands contained 2k instead of two alternating i/o commands in succession then STAT would contain $2(k+1)$ elements whereas SYNT would contain $2(k+1)^2$ elements.

It is important to note that the set STAT consists of pairs of i/o commands which can be synchronized during a properly terminating computation. The following two examples clarify this issue.

Example 3

Let

$$P_1 \equiv \ldots;\ k_1:P_1?x\ ;\ldots\ ,$$

$$P_2 \equiv \ldots;\ l_1:P_2!y\ ;\ldots;\ l_2:P_2!u\ ;\ldots$$

Then $L(P_1) = \{k_1:<2,1>\}$ and $L(P_2) = \{[l_1:<2,1>][l_2:<2,1>]\}$ so $L(P_1 \| P_2) = \emptyset$. Thus STAT = $\emptyset$ even though the i/o commands labelled by k_1 and l_1, respectively can be synchronized. On the other hand, since $L(P_1 \| P_2) = \emptyset$, there does not exist a properly terminating computation of $P_1 \| P_2$. Indeed, for any properly terminating computation the sequence consisting of its consecutive communications belongs to $L(P_1 \| P_2)$.

Example 4

Let

$$P_1 \equiv [b_1 \to \ldots \square\ b_2 \to \ldots\ ;\ k_1:P_2?x;\ldots;k_2:P_2!x],$$

$$P_2 \equiv [c_1 \to \ldots \square\ c_2 \to \ldots\ ;\ l_1:P_1!y;\ldots;l_2:P_1!y].$$

Then $L(P_1) = \{\epsilon,\ [k_1:<2,1>][k_2:<1,2>]\}$ and $L(P_2) = \{\epsilon,\ [l_1:<2,1>][l_2:<2,1>]\}$ so $L(P_1 \| P_2) = \{\epsilon\}$. Thus STAT = $\emptyset$. The i/o commands labeled by k_1 and l_1, respectively can be synchronized but not during a properly terminating computation.

The situation when $L(P_1 \| \ldots \| P_n) = \emptyset$ should be compared with the situation when $L(P_1 \| \ldots \| P_n) = \{\epsilon\}$. In the first case no properly terminating computation of $P_1 \| \ldots \| P_n$ exists. In the latter case the properly terminating computations of $P_1 \| \ldots \| P_n$ can exist but in none of them a communication will take place.

In both cases STAT = $\emptyset$ so no cooperation (resp. satisfaction) test will take place in the proof rule dealing with parallel composition. This is in accordance with the fact that partial correctness of programs refers to properly terminating computations only. In both cases above no communication will take place in any such computation.

Finally we consider the following example :

Example 5

Let

$$P_1 \equiv *[k_1:P_2?x \to \ldots]\ ;\ k_2:P_2!u,$$

$$P_2 \equiv *[l_1:P_1!y \to \ldots]\ ;\ l_2:P_1?z.$$

Then STAT = SYNT = $\{(k_i:\alpha_i,\ l_i:\beta_i)\ :\ i=1,2\}$. Note however that the communication between the i/o commands with labels k_2 and l_2, respectively cannot take place as none of the repetitive commands can terminate. In particular no computation of $P_1 \| P_2$ terminates. The tools used so far do not allow us to deduce these facts formally. We shall return to this problem later.

2. Proofs of deadlock freedom

In the proof systemsof [AFR] and [LG] one proves deadlock freedom of the parallel programs by identifying first the set of blocked configurations, i.e. the vectors of control points corresponding to a deadlock. Then for each blocked configuration one shows that the conjunction of the assertions attached to the corresponding control points (and the global invariant in the case of [AFR]) is inconsistent. Thus the length of the proof of deadlock freedom is proportional to the number of blocked configurations.

We now suggest a more restricted definition of a blocked configuration which is sufficient for proofs of deadlock freedom and results in shorter proofs. The control points which are of interest here are those when the control resides in front of an i/o command or at the end of a process. With each control point of the first type we associate a set of i/o commands which can be at this point executed. With the control point of the second type we associate the set $\{\text{end } P_i\}$ corresponding to the situation when the control is at the end of the process P_i.

We define

$$C(k{:}\alpha) = \{\{k{:}\alpha\}\}$$

where $k{:}\alpha$ occurs in the process as an atomic command,

$$C([b_1 \to S_1 \,\square \ldots \square\, b_m \to S_m \,\square\, k_1{:}\alpha_1 \to S_{m+1} \square \ldots \square\, k_n{:}\alpha_n \to S_{m+n} \,\square b_{m+1}; k_{n+1}{:}\alpha_{n+1} \to$$

$$S_{m+n+1} \square \ldots \square b_{m+p}; k_{n+p}{:}\alpha_{n+p} \to S_{m+n+p}]) = \{A : A = \{k_i{:}\alpha_i : i=1,\ldots,n\} \cup B$$

where $B \subseteq \{k_{n+i}{:}\alpha_{n+i} : i=1,\ldots,p\}$, where $m \geq 0$ and $n+p \geq 1$,

$$C(*[\,\square_{i=1}^{m} g_i \to S_i]) = C([\,\square_{i=1}^{m} g_i \to S_i]).$$

For other type of commands S C(S) is not defined. Note that a typical set A considered above consists of all i/o guards which occur without the boolean guards together with a subset of those i/o guards which occur with a boolean guard.

Given now a process P_i we define $C(P_i)$ to be the union of all sets $C(S)$ for S being a subprogram of P_i together with the element $\{\text{end } P_i\}$. Each element of $C(P_i)$ corresponds to a unique control point within P_i.

The identification of all blocked configurations depends on the fact whether so called distributed termination convention (d.t.c) of the repetitive commands is taken taken into account. According to this convention a repetitive command can be exited when all processes addressed in the guards with boolean part true have terminated. This convention corresponds to the following definition of a guard being failed : a guard fails if either its boolean part evaluates to false or the process addressed in its i/o part has terminated. A repetitive command is exited when all its guards fail. If in the definition of a failure of a guard we drop the second alternative we obtain the usual termination convention of the repetitive commands. In [H] the

distributed termination convention is adopted.

Consider first the simpler case when the usual termination convention is used.

A triple $\langle A_1,\ldots,A_n\rangle$ from $C(P_1)x\ldots xC(P_n)$ is called blocked if

i) $\exists\, i \;\; A_i \neq \{\text{end } P_i\}$

(not all processes have terminated)

ii) $(\bigcup_{i\neq j} A_i \times A_j) \cap SYNT = \emptyset$

(no communication can take place)

Alternatively ii) can be stated as : no pairs of elements from A_i and A_j $(i\neq j)$ match. The notion of a blocked tuple is from [AFR].

Let Init(L) for a formal language L denote its left factor i.e. the set $\{u : \exists\, w(uw \in L)\}$. We now put

$$LP(P_1\|\ldots\|P_n) = \{h : [h]_i \in Init(L(P_i)),\ i=1,\ldots,n\}.$$

Intuitively, $LP(P_1\|\ldots\|P_n)$ is the set of all possible communication sequences of $P_1\|\ldots\|P_n$ which can arise in partial computations during which the boolean guards are not interpreted.

We now say that a tuple $\langle A_1,\ldots,A_n\rangle$ from $C(P_1)x\ldots xC(P_n)$ is statically blocked if

i) it is blocked

ii) $\exists\, h \in LP(P_1\|\ldots\|P_n) \;\; \forall\, i$

$$[A_i \neq \{\text{end } P_i\} \Rightarrow \forall\, \underline{d} \in A_i([h]_i a \in Init(L(P_i))) \text{ where } L(\underline{d}) = \{a\}$$

$$\wedge\; A_i = \{\text{end } P_i\} \Rightarrow [h]_i \in L(P_i)]$$

The second condition states that there exists a communication sequence which reaches the vector of the control points associated with $\langle A_1,\ldots,A_n\rangle$. Reachability is checked by considering the projections $[h]_i$ of the sequence h. If $A_i \neq \{\text{end } P_i\}$ then $[h]_i a$ for all $a \in \{L(\underline{d}) : \underline{d} \in A_i\}$ should be an initial part of a sequence from $L(P_i)$. If $A_i = \{\text{end } P_i\}$ then $[h]_i$ should be a sequence from $L(P_i)$.

If d.t.c. is used then we should add the following condition to the definition of a blocked triple

iii) For no $i_0,i_1,\ldots,i_k$ from $\{1,\ldots,n\}$, where $i_j \neq i_l$ for $j \neq l$: $A_{i_j} = \{\text{end } P_{i_j}\}$ and the processes addressed in the i/o commands of A_{i_0} are all among $\{P_{i_1},\ldots,P_{i_k}\}$.

This condition states that no exit can take place due to the distributed termination convention. Thus the set A_{i_0} should correspond to a repetitive command.

We denote the set of all statically blocked tuples by STATB and the set of all blocked tuples by SYNTB.

We now consider a couple of examples.

Example 6

Consider the processes P_1 and P_2 from the example 2. D.t.c. cannot be used here. It is easy to see that

$$\begin{aligned}SYNTB = &\{<\{k_i:\alpha_i\}, \{l_j:\beta_j\}> : |i-j| \text{ is odd}, 1 \le i,j \le 4\}\\ &\cup \{<\{k_i:\alpha_i\}, \{\text{end } P_2\}> : 1 \le i \le 4\}\\ &\cup \{<\{\text{end } P_1\}, \{l_j:\beta_j\}> : 1 \le j \le 4\}\end{aligned}$$

whereas

$$\begin{aligned}STATB = &\{<\{k_3:\alpha_3\}, \{\text{end } P_2\}>,\\ &<\{\text{end } P_1\}, \{l_3:\beta_3\}>\}\end{aligned}$$

Thus SYNTB has 16 elements whereas STATB has only two elements.

Example 7

Let

$$P_1 \equiv \ldots; k_1:P_2!x\ ;\ldots;\ k_2:P_2?z\ ;\ldots;\ *[b_1 \to k_3:P_2!x\ ;\ldots;\ k_4:P_2?z\ ;\ldots],$$

$$P_2 \equiv l_1:P_1?y\ ;\ldots;\ l_2:P_1u\ ;\ldots;\ *[l_3:P_1?y \to \ldots\ l_4:P_1!u\ ;\ldots].$$

This is a structure of the program partitioning a set studied in [D] and [AFR].

Consider first the case when the distributed termination convention is not used. Then SYNTB and STATB are the same as in the previous example.

Suppose now that d.t.c. is used. Then

$$\begin{aligned}SYNTB = &\{<\{k_i:\alpha_i\}, \{l_j:\beta_j\}> : |i-j| \text{ is odd}, 1 \le i,j \le 4\}\\ &\cup \{<\{k_i:\alpha_i\}, \{\text{end } P_2\}> : 1 \le i \le 4\}\\ &\cup \{<\{\text{end } P_1\}, \{l_j:\beta_j\}> : j = 1,2,4\}\end{aligned}$$

and

$$STATB = \{<\{k_3:\alpha_3\}, \{\text{end } P_2\}>\}.$$

Here SYNTB has 15 elements whereas STATB only one. Note that the only staticly blocked pair cannot arise in actual computations either. The only way P_2 can terminate is due to the termination of P_1. Thus if the control in P_2 is at its end then the same must hold for P_1. We note that our analysis is not precise enough in order to deal with this type of situations. The next example gives more evidence to this effect.

Example 8

Let for i=1,...,n

$P_i \equiv *[b_i \; ; \; P_{i-1}!x_i \rightarrow \ldots$

$\square \; c_i ; P_{i+1}!x_i \rightarrow \ldots$

$\square \; P_{i-1}?y_i \rightarrow \ldots$

$\square \; P_{i+1}?z_i \rightarrow \ldots$

where the addition and substraction is modulo n.

This is a structure of the distributed gcd program considered in [AFR]. The labels of i/o commands are omitted as they are not needed here.

We have in the case when d.t.c. is not used

$$\begin{aligned} SYNTB = \{<A_1,\ldots,A_n> : \exists\, i \; A_i \neq \{\text{end } P_i\} \\ \wedge \forall\, i[P_{i-1}!x_i \in A_i \rightarrow A_{i-1} = \{\text{end } P_{i-1}\} \\ \wedge \; P_{i+1}!x_i \in A_i \rightarrow A_{i+1} = \{\text{end } P_{i+1}\} \\ \wedge \; A_i \neq \{\text{end } P_i\} \rightarrow \{P_{i-1}?y_i,\; P_{i+1}?z_i\} \subseteq A_i]\} \end{aligned}$$

and

STATB = SYNTB.

Suppose now that d.t.c. is used. Then

$$\begin{aligned} SYNTB = \{<A_1,\ldots,A_n> : \exists\, i \;\; A_i \neq \{\text{end } P_i\} \\ \wedge \forall\, i \; [P_{i-1}!x_i \in A_i \rightarrow A_{i-1} = \{\text{end } P_{i-1}\} \\ \wedge \; P_{i+1}!x_i \in A_i \rightarrow A_{i+1} = \{\text{end } P_{i+1}\} \\ \wedge \; A_i \neq \{\text{end } P_i\} \rightarrow (\{P_{i-1}?y_i,\; P_{i+1}?z_i\} \subseteq A_i \\ \wedge \; (A_{i-1} \neq \{\text{end } P_{i-1}\} \vee A_{i+1} \neq \{\text{end } P_{i+1}\}))]\} \end{aligned}$$

and once again STATB = SYNTB.

We see that in this example all blocked tuples are statically possible. The reason for it is that recording sequences of communications does not suffice to distinguish between two control points : the beginning and the end of a repetitive command.

On the other hand a simple informal argument allows to reduce the number of blocked triples which can arise in actual computations to one. The argument runs as follows. Suppose that d.t.c. is not used. Then no process P_i can terminate. Assume now that this convention is used. If some process P_i has terminated then by d.t.c. his neighbours P_{i-1} and P_{i+1} must have terminated, as well. Thus no process can terminate as the first one. In other words no process can terminate.

Thus in both cases a blocked tuple $\langle A_1,\ldots,A_n\rangle$ with some $A_i = \{\text{end } P_i\}$ is not possible. This reduces the number of possible blocked tuples to one being $\langle A_1,\ldots,A_n\rangle$ where for $i = 1,\ldots,n$ $A_i = \{P_{i-1}?y_i, P_{i+1}?z_i\}$.

In the next section we propose a more refined analysis which leads to a more restricted notion of static match and staticly blocked configurations. These notions will allow to deal properly with the above examples.

3. Proofs of safety of a decomposition of programs into communication-closed layers

In a recent paper [EF] Elrad and Francez proposed a method of decomposition of CSP programs which simplifies their analysis and can be used for a systematic construction of CSP programs. It is defined as follows.

Suppose that we deal with a parallel program P of the form $P_1 \| \ldots \| P_n$ where for all $i=1,\ldots,n$ $P_i \equiv S_i^1 ;\ldots; S_i^k$. Some of the commands S_i^k can be empty. We call the parallel programs $T_j \equiv S_1^j \| \ldots \| S_n^j$ $(j=1,\ldots,k)$ the layers of P.

A layer T_j is called communication-closed if there does not exist a computation of P in which a communication takes place between two i/o commands from which one lies within T_j and the other outside T_j. A **decomposition** $T_1;\ldots;T_k$ of P is called safe iff all the layers T_j are communication-closed. In other words a decomposition $T_1;\ldots;T_k$ of P is safe if there does not exist a computation of P with a communication involving two i/o commands from different layers.

In [EF] also more general types of layers are considered whose boundaries may cross the repetitive commands. Our analysis does not extend to such decompositions. The interest in considering safe decompositions stems from the following observation.

Fact ([EF]) Suppose that $T_1;\ldots;T_k$ is a safe decomposition of the parallel program P. Then the programs $T_1;\ldots;T_k$ and P are input-output equivalent.

Proof (informal) Obviously every computation of $T_1;\ldots;T_k$ is also a computation of P. Consider now a properly terminating computation of P. Due to safety of the decomposition we can rearrange some steps of this computation so that it becomes a properly terminating computation of $T_1;\ldots;T_k$. Both computations terminate in the same final state.

Thus both programs generate the same pairs of input-output states. □

As an example of a safe decomposition consider the following program

$$P \equiv P_1?x \| P_2!y \| P_3?u \| P_4!z$$

Consider now the layers

$$T_1 \equiv P_1?x \| P_2y \| \lambda \| \lambda$$

and

$$T_2 \equiv \mathcal{A} \| \mathcal{A} \| P_3?u \| P_4!z$$

where $\mathcal{A}$ stands for the empty program.

The decomposition $T_1;T_2$ of P is obviously safe. Note however that the program $T_1;T_2$ admits less computations than the original program P.

This property holds in general for safe decompositions of parallel programs with more than three components. Consequently the safe decomposition is in general easier to study than the original program.

A natural question now arises how to prove safety of a decomposition into layers of a given parallel program P. We propose a simple sufficient condition for safety of a decomposition. It has been suggested by H. Fauconnier.

We first slightly refine the definition of the set STAT. Let STAT' be defined in the same way as STAT but refering to $LP(P_1 \| \ldots \| P_n)$ instead of $L(P_1 \| \ldots \| P_n)$. Intuitively STAT' is the set of all pairs of i/o commands which can be synchronized during a computation of $P_1 \| \ldots \| P_n$ which ignores the boolean guards. Such a computation can be infinite or deadlocked. Clearly STAT $\subseteq$ STAT' but not necessarily conversely.

Let SEM' be defined by an analogous refinement of SEM.

Theorem Consider a decomposition $T_1;\ldots;T_k$ of a parallel program P. Suppose that there does not exist a pair $(k{:}\alpha,\ l{:}\beta)$ in STAT' (of P) such that $k{:}\alpha$ is from T_i and $l{:}\beta$ from T_j $(i \neq j)$. Then the decomposition $T_1;\ldots;T_k$ of P is safe.

Proof By definition the decomposition $T_1;\ldots;T_k$ of P is safe if for no pair $(k{:}\alpha,\ l{:}\beta)$ from SEM'(of P) $k{:}\alpha$ is from some T_i and $l{:}\beta$ from some T_j $(i \neq j)$. Since SEM' $\subseteq$ STAT', the result follows. □

As an illustration of the use of the above theorem consider its two simple applications. First, the above given decomposition $T_1;T_2$ of $P \equiv P_1?x \| P_2!y \| P_3?u \| P_4!z$ obviously satisfies the condition of the theorem thus it is indeed safe.

Secondly, consider the parallel program from the example 2.

Let $T_1 \equiv (k_1{:}P_2?x;\ldots;k_2{:}P_2!z;\ldots \| l_1{:}P_1!y;\ldots;l_2{:}P_1?u;\ldots)$

and

$$T_2 \equiv *[b_1 \rightarrow \ldots;k_3{:}P_2?x;\ldots;k_4{:}P_2!z;\ldots] \;\|\; *[b_2 \rightarrow \ldots;l_3{:}P_1!y;\ldots;l_4{:}P_1?u;\ldots]$$

Then the decomposition $T_1;T_2$ is safe because STAT'(of P) obviously satisfies the required condition of the theorem. Note that here STAT' = STAT.

We conclude by observing that whenever STAT' = SEM' (which is the case for many CSP programs suggested in the literature) then the condition from the theorem becomes equivalent to the safety of the decomposition $T_1;\ldots;T_k$ of P.

4. A MORE REFINED ANALYSIS

We now return to the problems signaled in section 3.2. We stated there that our analysis is not sufficiently precise to deal with some type of blocked configurations. We now refine some of the concepts in order to obtain an even more restricted notion of staticly blocked tuple. To this purpose we need three additional types of symbols. One is $\sqrt{_i}$ denoting a sucessful termination of the process P_i. This symbol is directly inspired by [H1]. The second new symbol is $\langle i\downarrow i_1,\ldots,i_k\rangle$ which marks termination of a repetitive command within the process P_i due to the termination of the processes $P_{i_1},\ldots,P_{i_k}$. The symbols $\sqrt{_i}$ and $\langle i\downarrow i_1,\ldots,i_k\rangle$ are used only when d.t.c. is assumed. Finally we adopt the symbol $\not\delta$ which is intended to indicate that a repetitive command cannot terminate. It will be used only when d.t.c. is not assumed.

Consider first the case when d.t.c. is not used. We now refine the definition of $L(S)$ (see section 2.1°) for a repetitive command as follows.

Let $S \equiv *\,S_1$ for an alternative command S_1. If all guards in S_1 contain a boolean part (distinct from <u>true</u>) then $L(S) = L(S_1)^*$, i.e. the former definition is retained. Otherwise $L(S) = L(S_1)^*\{\not\delta\}$.

Note that in the latter case S cannot terminate. An occurrence of $\not\delta$ in a communication sequence will mark the fact that an impossibility of termination of a repetitive command has been ignored. Consequently such sequences will not be admitted. A different possible approach to this problem is by admitting <u>infinite</u> communication sequences. We prefer to use the above approach since it is simpler. All other definitions including that of $L(P_1\|\ldots\|P_n)$, STAT and STATB are retained. Note that $\not\delta$ does not occur in any sequence of the form $[h]_i$ so no communication sequence from $L(P_1\|\ldots\|P_n)$ or $LP(P_1\|\ldots\|P_n)$ "violates" the impossibility of termination of a repetitive construct.

Let us now return to the examples 5, 7 and 8.

<u>ad Example 5</u>

With the new definitions we have $\forall\, w \in L(P_i)$ ($\not\delta$ is an element of w) for $i=1,2$. Thus $L(P_1\|P_2) = \emptyset$ and consequently STAT = $\emptyset$. This agrees with our informal definition of STAT.

<u>ad Example 7</u>

According to the new definition $\forall\, w \in L(P_2)$ ($\not\delta$ is an element of w). Thus by the definition $\langle\{k_3:\alpha_3\},\ \{\text{end } P_2\}\rangle \notin$ STATB. We have here

$$STATB = \{<\{end\ P_1\},\ \{l_3:\beta_3\}>\}.$$

Note that here $L(P_1 \| P_2) = \emptyset$ for the same reasons as above, so STAT = $\emptyset$.

ad Example 8

We have $\forall\, w \in L(P_i)$ ($\in$ is an element of w) for $i=1,\ldots,n$. Thus by the definition of STATB if $<A_1,\ldots,A_n>$ is a staticly blocked triple then for all i $A_i \neq \{end\ P_i\}$. We thus have

$$STATB = \{<A_1,\ldots,A_n> : A_i = \{P_{i-1}?y_i,\ P_{i+1}?z_i\},\ i=1,\ldots,n\}.$$

We now pass to the case when d.c.t. is used.

Let S be a repetitive command within the process P_i. Let $\{i_1,\ldots,i_k\}$ be the set of indices of the processes addressed in the guards of S which do not have a boolean part. Note that if S terminates in a computation of $P_1 \| \ldots \| P_n$ then at least the processes $P_{i_1},\ldots,P_{i_k}$ must have terminated at this moment.

We now refine the definition of $L(S)$ by identifying $<i\downarrow>$ with ϵ and putting $L(S) = L(S_1)^*\{<i\downarrow i_1,\ldots,i_k>\}$ where S_1 is the alternative command such that $S = *S_1$. All other clauses for sequential commands remain the same. We now put $L'(P_i) = L(P_i)\ \{V_i\}$ and define $L'(P_1 \| \ldots \| P_n)$ and $L'P(P_1 \| \ldots \| P_n)$ by defining first

$$[<i\downarrow i_1,\ldots,i_k>]_i = <i\downarrow i_1,\ldots,i_k>$$

$$[<i\downarrow i_1,\ldots,i_k>]_j = \epsilon \quad \text{if } i \neq j$$

$$[V_i]_i = V_i$$

$$[V_i]_j = \epsilon \quad \text{if } i \neq j$$

and putting

$$L'(P_1 \| \ldots \| P_n) = \{\, h : [h]_i \in L'(P_i),\ i=1,\ldots,n \wedge \underline{A}\}$$

$$L'P(P_1 \| \ldots \| P_n) = \{h : [h]_i \in Init(L'(P_i)),\ i = 1,\ldots,n \wedge \underline{A}\}$$

where the condition $\underline{A}$ is defined as follows

$$\underline{A} \equiv \forall\, p,j\ (h = a_1 \ldots a_p \wedge a_j = <i\downarrow i_1,\ldots,i_k> \rightarrow \forall\, l \in \{1,\ldots,k\}\ \exists\, m < j\ \ a_m = V_{i_l}).$$

The condition $\underline{A}$ states that if in a communication sequence h an exit from a repetitive command in P_i has been recorded then necessarily all the processes on which termination this loop exit depends have terminated before this exit took place.

The set STAT is defined as before but with reference to $L'(P_1 \| \ldots \| P_n)$ instead of $L(P_1 \| \ldots \| P_n)$. Similarly the set STATB is defined in the same way as before but with reference to $Init(L'(P_i))$ and $L'(P_1 \| \ldots \| P_n)$ instead of

Init(L (P_i)) and $L(P_1 \| \ldots \| P_n)$, respectively. Also the condition iii) in the definition of a blocked tuple is adopted.

We now return to the examples 7 and 8.

ad Example 7

It is easy to see that the new definition of STATB is more restricted than the previous one considered in section 3, so

STATB $\subseteq$ $\{\langle\{k_3:\alpha_3\}, \{\text{end } P_2\}\rangle\}$. We now show that $\langle\{k_3:\alpha_3\},\{\text{end } P_2\}\rangle \notin$ STATB, i.e. that STATB is empty.

We have

$L'(P_1) = \{k_1:\langle 1,2\rangle\}\{k_2:\langle 2,1\rangle\}(\{k_3:\langle 1,2\rangle\}\{k_4:\langle 2,1\rangle\})^* \{V_1\}$

and

$L'(P_2) = \{l_1:\langle 1,2\rangle\}\{l_2:\langle 2,1\rangle\}(\{l_3:\langle 1,2\rangle\}\{l_4:\langle 2,1\rangle\})^*\{\langle 2\downarrow 1\rangle V_2\}$.

Suppose that $\langle\{k_3:\alpha_3\}, \{\text{end } P_2\}\rangle \in$ STATB.

Then there exists $h \in L'P(P_1 \| \ldots \| P_n)$ such that $[h]_1(k_3:\langle 1,2\rangle) \in \text{Init}(L'(P_1))$ and $[h]_2 \in L'(P_2)$.

By the form of $L'(P_2)$ $[h]_2$ so a fortiori h contains the element $\langle 2\downarrow 1\rangle$. Since $h \in L'P(P_1 \| \ldots \| P_n)$, by the condition A h contains the element V_1. But this is impossible because by the above $[h]_1$ does not contain V_1. Contradiction. Thus STATB is indeed empty.

Note that

$L'(P_1 \| \ldots \| P_n) = \{\{k_1,l_1:\langle 1,2\rangle\}\{l_2,k_2:\langle 2,1\rangle\}$

$\{k_3,l_3:\langle 1,2\rangle\}\{l_4,k_4:\langle 2,1\rangle\}\}^*\{V_1\langle 2\downarrow 1\rangle V_2\}$

so STAT = $\{\{k_i:\alpha_i,l_i:\beta_i\}, i = 1,\ldots,4\}$ in contrast to the case when d.t.c. was not used.

ad Example 8

Suppose that $h \in L'P(P_1 \| \ldots \| P_n)$. We prove that h does not contain any element of the form $\langle i\downarrow i-1, i+1\rangle$. Suppose otherwise. Let $\langle i\downarrow i-1,i+1\rangle$ be the first element of this type in h. By the condition A some earlier element of h must be of the form V_{i-1}. We have $[h]_{i-1} \in \text{Init}(L'(P_i))$ and by the form of $L'(P_i)$ V_{i-1} must be preceded in $[h]_{i-1}$, so also in h, by $\langle i-1\downarrow i-2,i\rangle$. Contradiction.

Suppose now that $\langle A_1,\ldots,A_n\rangle \in$ STATB. Let $h \in L'P(P_1 \| \ldots \| P_n)$ be a sequence certifying that $\langle A_1,\ldots,A_n\rangle$ is a blocked tuple. If for some i $A_i = \{\text{end } P_i\}$ then $[h]_i \in L'(P_i)$ i.e. by the definition of $L'(P_i)$ $[h]_i$ terminates with $\langle i\downarrow i-1,i+1\rangle V_i$. Thus h contains the element $\langle i\downarrow i-1,i+1\rangle$ which is impossible.

Thus for no i $A_i = \{\text{end } P_i\}$. We conclude that similarly as in the case when d.t.c. was not considered earlier in this section

STATB consists of exactly one element $\langle A_1,\ldots,A_n\rangle$ where for $i=1,\ldots,n$

$A_i = \{P_{i-1}?y_i, P_{i+1}?z\}$.

The same conclusions about STATB in the examples 7 and 8 above were reached in [AFR] by a formal reasoning within a proof system. The above proofs are more straightforward and moreover require only a limited knowledge about the programs under consideration.

As a final remark we observe that the use of the set STAT is not sufficient for the proofs of safety properties (in the sense of [OL])that are more general than partial correctness. In such cases more appropriate set to be used is STAT' in the definition of which one refers to arbitrary, possibly non-terminating or blocked computations.

Observe that in the case of example 8 (independently of the fact whether d.t.c. is used) STAT = Ø whereas

$$STAT' = \{(P_{i-1}?y_i,\ P_i!x_{i-1}) : i=1,\ldots,n\}$$
$$\cup\ \{(P_{i+1}?z_i, P_i!x_{i+1}) : i=1,\ldots,n\}$$

In [AFR] it is proved that a distributed gcd program whose structure is considered in the example 8 computes the g.c.d. of n numbers at the moment of reaching the only blocked configuration, the one discussed above. A proof of this fact within the proof system of [AFR] requires the use of the set STAT' and not STAT in the proof rule for parallel composition.

5. CONCLUSIONS

We have presented in this paper a method of analyzing the CSP programs which leads to simpler proofs of their correctness.

It can be easily automated and in fact such an algorithm for the case of ADA programs has already been described in [T].

It should be however noted that (as indicated in [T]) the algorithms computing the sets STAT, STAT' and STATB arising in this analysis are necessarily exponential. This can lead to inherent problems in the case of longer programs for which such an analysis is especially useful.

One could envisage a still more refined analysis in which one would take into account the boolean guards of the program under consideration. One could then infer for example that the second repetitive command in the process

$$P_i \equiv \ldots *[b \to S_1]\ ;\ *[b \to S_2]$$

cannot be entered so no i/o command from S_2 can be reached.

Such an analysis, however does not lead to any useful conclusions when applied to concrete examples. Any gain obtained by it is restricted to ill designed programs such as for example the above process P_i.

This leads us to an interesting question about the usefulness of the analysis presented in this paper. How accurate is it with respect to the semantical analysis ?

Consider first the sets STAT and STAT' which are used as "approximations" of the sets SEM and SEM', respectively. It is easy to design programs for which STAT (STAT') differs from SEM(SEM'). However, all such programs seem artificial. We observed that in all examples studied in [H] these sets do not differ and we conjecture that it is always the case for well-designed CSP programs. Of course such a conjecture is difficult to prove because no definition of a well-designed CSP program exists.

The situation changes when we compare the set STATB with the set of blocked configurations which can arise in actual computations. These sets may differ for simple and well-designed CSP programs. One can easily design a program of the form studied in the example 2 which is deadlock free whereas in this case STATB is not empty. An example of such a program is a slightly modified version of the program partitioning a set from [D].

REFERENCES

[AFR] K.R.APT, N. FRANCEZ & W.P. DE ROEVER, A proof system for communicating sequential processes, TOPLAS, vol. 2, N° 3, pp. 359-385, 1980.

[D] E.W. DIJKSTRA, A correctness proof for communicating processes : a small exercise, in : E.W. Dijkstra, Selected writings on computing : a personal perspective, Springer Verlag, New York, pp. 259-263, 1982.

[EF] T. ELRAD & N. FRANCEZ, Decomposition of distributed programs into communication-closed layers, to appear in SCP.

[H] C.A.R. HOARE, Communicating sequential processes, CACM, vol. 21, N° 8, pp. 666-677, 1978.

[H1] C.A.R. HOARE, A model for communicating sequential processes, in : R.M. McKeag, A.M. McNaughton, Eds., On the construction of programs, Cambridge University Press, pp. 229-243, 1980.

[LG] G. LEVIN & D. GRIES, A proof technique for communicating sequential processes, Acta Informatica, vol. 15, N° 3, pp. 281-302, 1981.

[OL] S. OWICKI & L. LAMPORT, Proving liveness properties of concurrent programs, TOPLAS, vol. 4, N° 3, pp. 455-495, 1982.

[T] R.N. TAYLOR, A general purpose algorithm for analyzing concurrent programs, CACM, vol. 26, N° 5, pp. 362-376, 1983.