



## UvA-DARE (Digital Academic Repository)

### Thread extraction for polyadic instruction sequences

Bergstra, J.A.; Middelburg, C.A.

**Publication date**

2009

**Document Version**

Submitted manuscript

[Link to publication](#)

**Citation for published version (APA):**

Bergstra, J. A., & Middelburg, C. A. (2009). *Thread extraction for polyadic instruction sequences*. ArXiv. <http://arxiv.org/abs/0802.1578>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Thread Extraction for Polyadic Instruction Sequences<sup>\*</sup>

J.A. Bergstra and C.A. Middelburg

Programming Research Group, University of Amsterdam,  
Kruislaan 403, 1098 SJ Amsterdam, the Netherlands  
J.A.Bergstra@uva.nl, C.A.Middelburg@uva.nl

**Abstract.** Instruction sequences are often fragmented. An important reason for instruction sequence fragmentation is that the execution architecture at hand to execute instruction sequences sets bounds to the size of instruction sequences. In this paper, we study instruction sequences that have been split into fragments. The purpose is to develop a theoretical understanding of this matter. The possible joint behaviours exhibited by a collection of fragments on execution are explained in terms of threads as considered in basic thread algebra. In this way, a setting is provided in which the slow-down results of instruction sequence fragmentation can be analysed.

*Keywords:* instruction sequence fragmentation, polyadic instruction sequence, thread extraction, basic thread algebra.

*1998 ACM Computing Classification:* D.3.1, D.3.3, F.1.1, F.3.2, F.3.3.

## 1 Introduction

In this paper, we consider fragmentation of sequential programs that take the form of instruction sequences. With that we carry on the line of research with which a start was made in [3]. This line of research concerns the development of a theoretical understanding of possible forms of sequential programs and associated ways of programming. It is a fact of life that instruction sequences are often fragmented. An important reason for instruction sequence fragmentation is that the execution architecture at hand to execute instruction sequences sets bounds to the size of instruction sequences. However, there may also be other reasons for instruction sequence fragmentation, for instance business economical reasons. In this paper, we look for a direct approach to the formalization of instruction sequences that have been split into fragments and we investigate their relationship with sequential programs.

The question is how a joint behaviour of the fragments in a collection of fragments is achieved. The view of this matter is that there can only be a single

---

<sup>\*</sup> This research has been partly carried out in the framework of the Jacquard-project Symbiosis, which is funded by the Netherlands Organisation for Scientific Research (NWO).

fragment being executed at any stage, but the fragment in question may make any fragment in the collection the one being executed by means of a special instruction for switching over execution to another fragment. This does not fit in very well with the conception that the collection of fragments constitutes a sequential program. To our knowledge, a theoretical understanding of instruction sequences that have been split into fragments has not been developed. This has motivated us to take up this topic.

As the view is taken in the line of research carried on in this paper that the behaviour exhibited by a sequential program on execution is a thread as considered in basic thread algebra [3],<sup>1</sup> so the view is taken here that the possible joint behaviours exhibited by the fragments on execution are threads as considered in basic thread algebra. In execution architectures, a fragment must be loaded in order to become the one being executed. Hence, making a fragment the one being executed can be looked upon as loading that fragment for execution.

The instruction sequences taken for fragments are called polyadic instruction sequences in this paper. We introduce polyadic instruction sequences in the setting of program algebra [3], an algebra of programs in which programs are looked upon as instruction sequences. In [3], a hierarchy of program notations rooted in program algebra is presented as well. Included in this hierarchy are very simple program notations which are close to existing assembly languages up to and including simple program notations that support structured programming by offering a rendering of conditional and loop constructs. All of these program notations are referred to in this paper, but only one of them is actually used. That program notation is introduced under the name PGLD in [3].

This paper is organized as follows. First, we review basic thread algebra, program algebra, and the program notation PGLD (Sections 2, 3, and 4). Next, we introduce polyadic instruction sequences in the setting of program algebra, explain the possible joint behaviours of a collection of polyadic instruction sequences using basic thread algebra, and give an example of the use of polyadic instruction sequences (Sections 5 and 6). Following this, we extend basic thread algebra to allow for threads to make use of services and give a description of instruction register file services (Sections 7 and 8). After that, we show that, for each possible joint behaviour of a collection of polyadic instruction sequences, a sequential program can be synthesized from the collection of polyadic instruction sequences that exhibits on execution essentially the behaviour in question by making use of an instruction register file service (Section 9). Finally, we make some concluding remarks (Section 10).

In this paper, we do not give models of BTA and its extensions. For the preferred models, called projective limit models, the reader is referred to [10].

---

<sup>1</sup> In [3], basic thread algebra is introduced under the name basic polarized process algebra. Prompted by the development of thread algebra [6], which is a design on top of it, basic polarized process algebra has been renamed to basic thread algebra.

**Table 1.** Axiom of BTA

$$\underline{x \triangleleft \iota \triangleright y = x \triangleleft \iota \triangleright x} \quad \mathbf{T1}$$

## 2 Basic Thread Algebra

In this section, we review BTA (Basic Thread Algebra). BTA is a form of process algebra which is concerned with the behaviour that sequential programs exhibit on execution. Those behaviours are called *threads*.

In BTA, it is assumed that fixed but arbitrary finite sets  $\mathcal{A}$  and  $\mathcal{I}$  with  $\mathcal{A} \cap \mathcal{I} = \emptyset$  and  $\mathbf{tau} \in \mathcal{I}$  have been given. The members of  $\mathcal{A}$  are called *basic actions* and the members of  $\mathcal{I}$  are called *internal actions*. The members of  $\mathcal{A} \cup \mathcal{I}$  are referred to as *actions*. In previous work, we take in essence the singleton set  $\{\mathbf{tau}\}$  for  $\mathcal{I}$ . The generalization made here permits internal actions with differences relevant for analysis to be distinguished.

A thread performs actions in a sequential fashion. Upon each action performed, a reply from the execution environment of the thread determines how it proceeds. The possible replies are **T** and **F**. Performing an internal action, always leads to the reply **T**.

Although BTA is one-sorted, we make this sort explicit. The reason for this is that we will extend BTA with an additional sort in Section 7.

BTA has one sort: the sort **T** of *threads*. To build terms of sort **T**, BTA has the following constants and operators:

- the *deadlock* constant  $\mathbf{D} : \mathbf{T}$ ;
- the *termination* constant  $\mathbf{S} : \mathbf{T}$ ;
- for each  $a \in \mathcal{A} \cup \mathcal{I}$ , the binary *postconditional composition* operator  $\_ \triangleleft a \triangleright \_ : \mathbf{T} \times \mathbf{T} \rightarrow \mathbf{T}$ .

We assume that there are infinitely many variables of sort **T**, including  $x, y, z$ . Terms of sort **T** are built as usual (see e.g. [11, 12]). We use infix notation for the postconditional composition operator. We introduce *basic action prefixing* as an abbreviation:  $a \circ p$  abbreviates  $p \triangleleft a \triangleright p$ .

The thread denoted by a closed term of the form  $p \triangleleft a \triangleright q$  will first perform  $a$ , and then proceed as the thread denoted by  $p$  if the reply from the execution environment is **T** and proceed as the thread denoted by  $q$  if the reply from the execution environment is **F**. The threads denoted by  $\mathbf{D}$  and  $\mathbf{S}$  will become inactive and terminate, respectively.

BTA has only one axiom. This axiom is given in Table 1. In this table,  $\iota$  stands for an arbitrary member of  $\mathcal{I}$ .

Notice that each closed BTA term denotes a thread that will become inactive or terminate after it has performed finitely many actions. Infinite threads can be described by guarded recursion.

A *guarded recursive specification* over BTA is a set of recursion equations  $E = \{X = t_X \mid X \in V\}$ , where  $V$  is a set of variables of sort **T** and each  $t_X$  is a BTA term of the form  $\mathbf{D}$ ,  $\mathbf{S}$  or  $t \triangleleft a \triangleright t'$  with  $t$  and  $t'$  that contain only variables

**Table 2.** Axioms for guarded recursion

$\langle X E \rangle = \langle t_X E \rangle$	if $X = t_X \in E$	RDP
$E \Rightarrow X = \langle X E \rangle$	if $X \in V(E)$	RSP

**Table 3.** Approximation induction principle

$\bigwedge_{n \geq 0} \pi_n(x) = \pi_n(y) \Rightarrow x = y$	AIP
$\pi_0(x) = D$	P0
$\pi_{n+1}(S) = S$	P1
$\pi_{n+1}(D) = D$	P2
$\pi_{n+1}(x \triangleleft a \triangleright y) = \pi_n(x) \triangleleft a \triangleright \pi_n(y)$	P3

from  $V$ . We write  $V(E)$  for the set of all variables that occur in  $E$ . We are only interested in models of BTA in which guarded recursive specifications have unique solutions, such as the projective limit model of BTA presented in [1]. A thread that is the solution of a finite guarded recursive specification over BTA is called a *finite-state* thread.

For each guarded recursive specification  $E$  and each  $X \in V(E)$ , we introduce a constant  $\langle X|E \rangle$  of sort  $\mathbf{T}$  standing for the unique solution of  $E$  for  $X$ . The axioms for these constants are given in Table 2. In this table, we write  $\langle t_X|E \rangle$  for  $t_X$  with, for all  $Y \in V(E)$ , all occurrences of  $Y$  in  $t_X$  replaced by  $\langle Y|E \rangle$ .  $X$ ,  $t_X$  and  $E$  stand for an arbitrary variable of sort  $\mathbf{T}$ , an arbitrary BTA term of sort  $\mathbf{T}$  and an arbitrary guarded recursive specification over BTA, respectively. Side conditions are added to restrict what  $X$ ,  $t_X$  and  $E$  stand for.

We will use the following abbreviation:  $a^\omega$ , where  $a \in \mathcal{A} \cup \mathcal{I}$ , abbreviates  $\langle X|\{X = a \circ X\} \rangle$ .

We will write BTA+REC for BTA extended with the constants for solutions of guarded recursive specifications and axioms RDP and RSP.

Closed terms of sort  $\mathbf{T}$  from the language of BTA+REC that denote the same infinite thread cannot always be proved equal by means of the axioms of BTA+REC. We introduce AIP (Approximation Induction Principle) to remedy this. AIP is based on the view that two threads are identical if their approximations up to any finite depth are identical. The approximation up to depth  $n$  of a thread is obtained by cutting it off after performing a sequence of actions of length  $n$ . In AIP, the approximation up to depth  $n$  is phrased in terms of the unary *projection* operator  $\pi_n : \mathbf{T} \rightarrow \mathbf{T}$ . AIP and the axioms for the projection operators are given in Table 3. In this table,  $a$  stands for an arbitrary member of  $\mathcal{A} \cup \mathcal{I}$ .

We will write BTA+REC+AIP for BTA+REC extended with the projection operators and the axioms from Table 3.

### 3 Program Algebra

In this section, we review PGA (ProGram Algebra). The perception of a program as a single-pass instruction sequence is the starting-point of PGA.

In PGA, it is assumed that a fixed but arbitrary set  $\mathfrak{A}$  of *basic instructions* has been given. PGA has the following *primitive instructions*:

- for each  $a \in \mathfrak{A}$ , a *plain basic instruction*  $a$ ;
- for each  $a \in \mathfrak{A}$ , a *positive test instruction*  $+a$ ;
- for each  $a \in \mathfrak{A}$ , a *negative test instruction*  $-a$ ;
- for each  $l \in \mathbb{N}$ , a *forward jump instruction*  $\#l$ ;
- a *termination instruction*  $!$ .

We write  $\mathfrak{J}$  for the set of all primitive instructions.

The intuition is that the execution of a basic instruction  $a$  produces either  $\mathsf{T}$  or  $\mathsf{F}$  at its completion. In the case of a positive test instruction  $+a$ ,  $a$  is executed and execution proceeds with the next primitive instruction if  $\mathsf{T}$  is produced. Otherwise, the next primitive instruction is skipped and execution proceeds with the primitive instruction following the skipped one. If there is no next instruction to be executed, deadlock occurs. In the case of a negative test instruction  $-a$ , the role of the value produced is reversed. In the case of a plain basic instruction  $a$ , execution always proceeds as if  $\mathsf{T}$  is produced. The effect of a forward jump instruction  $\#l$  is that execution proceeds with the  $l$ -th next instruction. If  $l$  equals 0 or the  $l$ -th next instruction does not exist, deadlock occurs. The effect of the termination instruction  $!$  is that execution terminates.

PGA has the following constants and operators:

- for each  $u \in \mathfrak{J}$ , an *instruction constant*  $u$ ;
- the binary *concatenation operator*  $- ; -$ ;
- the unary *repetition operator*  $-^\omega$ .

We assume that there are infinitely many variables, including  $x, y, z$ . Terms are built as usual. We use infix notation for the concatenation operator and postfix notation for the repetition operator.

A closed PGA term is considered to denote a non-empty, finite or periodic infinite sequence of primitive instructions.<sup>2</sup> Closed PGA terms are considered equal if they denote the same instruction sequence. The axioms for instruction sequence equivalence are given in Table 4. In this table,  $n$  stands for an arbitrary natural number greater than 0. For each PGA term  $P$ , the term  $P^n$  is defined by induction on  $n$  as follows:  $P^1 = P$  and  $P^{n+1} = P;P^n$ . The equation  $X^\omega = X;X^\omega$  is derivable. Each closed PGA term is derivably equal to one of the form  $P$  or  $P;Q^\omega$ , where  $P$  and  $Q$  are closed PGA terms in which the repetition operator does not occur.

Each closed PGA term  $P$  is considered to denote an instruction sequence of which the behaviour is a finite-state thread, called the *thread produced by*

---

<sup>2</sup> A periodic infinite sequence is an infinite sequence with only finitely many subsequences.

**Table 4.** Axioms of PGA

$(x ; y) ; z = x ; (y ; z)$	PGA1
$(x^n)^\omega = x^\omega$	PGA2
$x^\omega ; y = x^\omega$	PGA3
$(x ; y)^\omega = x ; (y ; x)^\omega$	PGA4

**Table 5.** Defining equations for thread extraction operation of PGA

$ a  = a \circ \mathsf{D}$	$ \#l  = \mathsf{D}$
$ a ; x  = a \circ  x $	$ \#0 ; x  = \mathsf{D}$
$ +a  = a \circ \mathsf{D}$	$ \#1 ; x  =  x $
$ +a ; x  =  x  \trianglelefteq a \triangleright  \#2 ; x $	$ \#l + 2 ; u  = \mathsf{D}$
$ -a  = a \circ \mathsf{D}$	$ \#l + 2 ; u ; x  =  \#l + 1 ; x $
$ -a ; x  =  \#2 ; x  \trianglelefteq a \triangleright  x $	$   = \mathsf{S}$
	$! ; x  = \mathsf{S}$

$P$ . The set  $\mathfrak{A}$  of basic instructions is taken for the set  $\mathcal{A}$  of basic actions. The *thread extraction* operation  $| \_ |$  determines, for each closed PGA term  $P$ , a closed BTA+REC term that denotes the thread produced by  $P$ . The thread extraction operation is defined by the equations given in Table 5 (for  $a \in \mathfrak{A}$ ,  $l \in \mathbb{N}$  and  $u \in \mathfrak{J}$ ) and the rule that  $|\#l ; X| = \mathsf{D}$  if  $\#l$  is the beginning of an infinite jump chain. This rule is formalized in e.g. [7].

The behaviour of each closed PGA term, is a thread that is definable by a finite guarded recursive specification over BTA. The other way round, each finite guarded recursive specification over BTA in which no internal actions occur can be translated into a closed PGA term of which the behaviour is the solution of the finite guarded recursive specification concerned.

Closed PGA terms are considered to denote programs and therefore they constitute an elementary program notation. Closed PGA terms are also called PGA programs.

In [3], a hierarchy of program notations rooted in PGA is presented. In this hierarchy, the program notations PGLA, PGLB, PGLC, PGLD, PGLDg, PGLE, and PGLS appear. PGLA programs are translated into PGA programs by means of a function `pg1a2pga`, PGLB programs are translated into PGLA programs by means of a function `pg1b2pg1a`, PGLC programs are translated into PGLB programs by means of a function `pg1c2pg1b`, etcetera. These functions are called projections in [3]. Each of them translates each of the programs from a program notation higher in the hierarchy into a program from a program notation lower in the hierarchy that exhibits the same behaviour on execution. Moreover, PGA programs are translated into PGLA programs by means of a function `pga2pg1a`, PGLA programs are translated into PGLB programs by means of a function `pg1a2pg1b`, PGLB programs are translated into PGLC programs by means of a

function `pg1b2pg1c`, etcetera. These functions are called embeddings in [3]. Each of them translates each of the programs from a program notation lower in the hierarchy into a program from a program notation higher in the hierarchy that exhibits the same behaviour on execution. We remark that there does not exist a function by which each PGLE program is translated into a PGLS program that exhibits the same behaviour on execution: PGLS is strictly weaker than PGLE.

In Section 5, we will take special versions of the above-mentioned program notations for the ones that may be used for fragments. Moreover, we will use the function `pg1a2pga` and in addition the functions `pg1b2pga`, `pg1c2pga`, ... defined by `pg1b2pga(P) = pg1a2pga(pg1b2pg1a(P))` for all PGLB programs  $P$ , `pg1c2pga(P) = pg1b2pga(pg1c2pg1b(P))` for all PGLC programs  $P$ , etcetera. In Section 9, we will use the function `pg1c2pg1d` and in addition the function `pga2pg1c` defined by `pga2pg1c(P) = pg1b2pg1c(pg1a2pg1b(pga2pg1a(P)))` for all PGA programs  $P$ .

PGLC, PGLD and PGLS are the most interesting program notations of the ones mentioned above. PGLC and PGLD are close to existing assembly languages. The main difference between them is that PGLC has relative jump instructions and PGLD has absolute jump instructions. PGLS supports structured programming by offering a rendering of conditional and loop constructs instead of (unstructured) jump instructions.

## 4 The Program Notation PGLD

In this section, we review the program notation PGLD. This program notation is reviewed because it will be used later on in Sections 6 and 9.

In PGLD, like in PGA, it is assumed that there is a fixed but arbitrary finite set  $\mathfrak{A}$  of *basic instructions*. Again, the intuition is that the execution of a basic instruction  $a$  may modify a state and produces T or F at its completion.

PGLD has the following primitive instructions:

- for each  $a \in \mathfrak{A}$ , a *plain basic instruction*  $a$ ;
- for each  $a \in \mathfrak{A}$ , a *positive test instruction*  $+a$ ;
- for each  $a \in \mathfrak{A}$ , a *negative test instruction*  $-a$ ;
- for each  $l \in \mathbb{N}$ , an *absolute jump instruction*  $\#\#l$ .

PGLD programs have the form  $u_1; \dots; u_k$ , where  $u_1, \dots, u_k$  are primitive instructions of PGLD.

The plain basic instructions, the positive test instructions, and the negative test instructions are as in PGA. The effect of an absolute jump instruction  $\#\#l$  is that execution proceeds with the  $l$ -th instruction of the program concerned. If  $\#\#l$  is itself the  $l$ -th instruction, then deadlock occurs. If  $l$  equals 0 or  $l$  is greater than the length of the program, then termination occurs.

The function `pg1d2pga` from the set of all PGLD programs to the set of all PGA programs can be defined directly as follows:

$$\text{pg1d2pga}(u_1; \dots; u_k) = (\psi_1(u_1); \dots; \psi_k(u_k); !; !; \#0; \#0)^\omega,$$

where the auxiliary functions  $\psi_j$  from the set of all primitive instructions of PGLD to the set of all primitive instructions of PGA are defined as follows ( $1 \leq j \leq k$ ):

$$\begin{aligned}
\psi_j(\#\#l) &= \#l - j && \text{if } j \leq l \leq k , \\
\psi_j(\#\#l) &= \#k + 2 - (j - l) && \text{if } 0 < l < j , \\
\psi_j(\#\#l) &= ! && \text{if } l = 0 \vee l > k , \\
\psi_j(u) &= u && \text{if } u \text{ is not a jump instruction .}
\end{aligned}$$

## 5 Polyadic Instruction Sequences

In this section, we look for a direct approach to the formalization of instruction sequences that have been split into fragments. The fragments resulting from a splitting are considered instruction sequences with special instructions for switching over execution from one fragment to another. The instruction sequences concerned are called polyadic instruction sequences.

It is assumed that a special version of PGLA, PGLB, PGLC, PGLD, PGLDg, PGLE or PGLS is used for each polyadic instruction sequence. Moreover, it is assumed that a collection of polyadic instruction sequences between which execution can be switched takes the form of a sequence, called a polyadic instruction sequence vector, in which each polyadic instruction sequence is coupled with the program notation used for it.

Our general view on the way of achieving a joint behaviour of the polyadic instruction sequences in a polyadic instruction sequence vector is as follows:

- there can only be a single polyadic instruction sequence being executed at any stage;
- the polyadic instruction sequence in question may make any polyadic instruction sequence in the vector the one being executed;
- making another polyadic instruction sequence the one being executed is effected by executing a special instruction for switching over execution;
- any polyadic instruction sequence can be taken for the one being executed initially.

In addition to special instructions for switching over execution, polyadic instruction sequences may contain special instructions for putting instructions into instruction registers and special instructions which are actually instruction place-holders: each of them is replaced by the instruction contained in one of the instruction registers on making a polyadic instruction sequence the one being executed. The presence of special instructions of the latter kind turns a polyadic instruction sequence into a parameterized instruction sequence of which the parameters are filled in each time it is made the one being executed. This feature accounts for the use of the prefix polyadic. Its merit is primarily that it allows for execution to proceed in effect from different positions each time a polyadic instruction sequence is loaded for execution. An example of this is given in Section 6.

We take the line that different program notations can be used for different polyadic instruction sequences in a polyadic instruction sequence vector. On making a polyadic instruction sequence in the vector the one being executed, it is considered to be translated into a  $\text{PGA}_p$  program.

$\text{PGA}_p$  is a variant of PGA in which the above-mentioned special instructions are incorporated. In  $\text{PGA}_p$ , it is assumed that there is a fixed but arbitrary finite set  $\mathfrak{A}_c$  of *core basic instructions*. In  $\text{PGA}_p$ , a basic instruction is either a core basic instruction or a supplementary basic instructions.

$\text{PGA}_p$  has the following *core primitive instructions*:

- for each  $a \in \mathfrak{A}_c$ , a *plain basic instruction*  $a$ ;
- for each  $a \in \mathfrak{A}_c$ , a *positive test instruction*  $+a$ ;
- for each  $a \in \mathfrak{A}_c$ , a *negative test instruction*  $-a$ ;
- for each  $l \in \mathbb{N}$ , a *forward jump instruction*  $\#l$ ;
- a *termination instruction*  $!$ .

We write  $\mathfrak{I}_c$  for the set of all core primitive instructions. The core primitive instructions of  $\text{PGA}_p$  are the counterparts of the primitive instructions of PGA.

$\text{PGA}_p$  has the following *supplementary basic instructions*:

- for each  $i \in \mathbb{N}$ , a *switch-over instruction*  $\###i$ ;
- for each  $i \in \mathbb{N}$  and  $u \in \mathfrak{I}_c$ , a *put instruction*  $\$put:i:u$ ;
- for each  $i \in \mathbb{N}$ , a *get instruction*  $\$get:i$ .

We write  $\mathfrak{A}_s$  for the set of all supplementary basic instructions. In the presence of a polyadic instruction sequence vector, a switch-over instruction  $\###i$  is the instruction for switching over execution to the  $i$ -th polyadic instruction sequence in the vector. A put instruction  $\$put:i:u$  is the instruction for putting instruction  $u$  in the instruction register with number  $i$ . A get instruction  $\$get:i$  is the instruction place-holder of which each occurrence in a polyadic instruction sequence is replaced by the contents of the instruction register with number  $i$  on switching over execution to that polyadic instruction sequence.

The supplementary basic instructions of  $\text{PGA}_p$  can be viewed as built-in basic instructions. However, as laid down below, supplementary basic instructions do not occur in positive or negative test instructions. Thus, the core primitive instructions and supplementary basic instructions make up the primitive instructions of  $\text{PGA}_p$ .

$\text{PGA}_p$  has the following constants and operators:

- for each  $u \in \mathfrak{I}_c \cup \mathfrak{A}_s$ , an *instruction constant*  $u$ ;
- the binary *concatenation operator*  $- ; -$ ;
- the unary *repetition operator*  $-^\omega$ .

The axioms of  $\text{PGA}_p$  are the same as the axioms of PGA.

Suppose that in PGA the restriction is dropped that  $\mathfrak{A}$  must be a finite set. Then  $\text{PGA}_p$  can be viewed as the specialization of PGA obtained by taking the set  $\mathfrak{A}_c \cup \mathfrak{A}_s$  for  $\mathfrak{A}$  and excluding terms in which basic instructions from  $\mathfrak{A}_s$  occur in positive or negative test instructions. Henceforth, we actually drop the restriction

that  $\mathfrak{A}$  must be a finite set. This simplifies the definitions of the different program notations that can be used for polyadic instruction sequences and also enables the use of the functions `pgla2pga`, `pglb2pga`, etcetera for translating programs in those program notations into  $\text{PGA}_p$  programs.

The different program notations that can be used for polyadic instruction sequences are  $\text{PGLA}_p$ ,  $\text{PGLB}_p$ ,  $\text{PGLC}_p$ ,  $\text{PGLD}_p$ ,  $\text{PGLDg}_p$ ,  $\text{PGL}_p$ , and  $\text{PGLS}_p$ . The set of all  $\text{PGLA}_p$  programs is the subset of the set of all PGLA programs, taking the set  $\mathfrak{A}_c \cup \mathfrak{A}_s$  for  $\mathfrak{A}$ , in which the basic instructions from  $\mathfrak{A}_s$  do not occur in positive or negative test instructions. The other program notations are defined similarly. If the set  $\mathfrak{A}_c \cup \mathfrak{A}_s$  is taken for  $\mathfrak{A}$ , the function `pgla2pga` translates each  $\text{PGLA}_p$  program into a  $\text{PGA}_p$  program that exhibits the same behaviour on execution. Similar remarks apply to the other program notations.

A *polyadic instruction sequence* is either a  $\text{PGLA}_p$  program, a  $\text{PGLB}_p$  program, a  $\text{PGLC}_p$  program, a  $\text{PGLD}_p$  program, a  $\text{PGLDg}_p$  program, a  $\text{PGL}_p$  program or a  $\text{PGLS}_p$  program.

A *polyadic instruction sequence vector* is a sequence of pairs consisting of a polyadic instruction sequence and a member of the set  $\{A, B, C, D, Dg, E, S\}$  of *program notation indices*.

Let  $\alpha$  be a polyadic instruction sequence vector, let  $P_1, \dots, P_n$  and  $c_1, \dots, c_n$  be polyadic instruction sequences and program notation indices, respectively, such that  $\alpha = \langle (P_1, c_1) \rangle \circ \dots \circ \langle (P_n, c_n) \rangle$ ,<sup>3</sup> and let  $i \in [1, n]$ . Then we write  $pg(\alpha, i)$  and  $pgn(\alpha, i)$  for  $P_i$  and  $c_i$ , respectively.

Let  $\alpha$  be a polyadic instruction sequence vector of length  $n$ , and let  $i \in [1, n]$ . Then program notation index  $pgn(\alpha, i)$  indicates which program notation is used for polyadic instruction sequence  $pg(\alpha, i)$ .  $A$  stands for  $\text{PGLA}_p$ ,  $B$  stands for  $\text{PGLB}_p$ , etcetera. The program notation used is made explicit because it cannot always be determined uniquely from the polyadic instruction sequence concerned, whereas the behaviour that this polyadic instruction sequence exhibits on execution may be different for each of the program notations in question.

The set of instruction registers that contain an instruction and the contents of each of those registers matter when a polyadic instruction sequence is made the one being executed. That makes us introduce the notion of an instruction register file state and special notation relating to this notion.

An *instruction register file state* is a function  $\sigma : I \rightarrow \mathcal{I}_c$ , where  $I$  is a finite subset of  $\mathbb{N}$ .

Let  $p$  be a  $\text{PGA}_p$  program and  $\sigma$  be an instruction register file state. Then we write  $p[\sigma]$  for  $p$  with, for all  $i \in \text{dom}(\sigma)$ , all occurrences of `$get:i` in  $p$  replaced by  $\sigma(i)$ .

Let  $i, n \in \mathbb{N}$  be such that  $1 \leq i \leq n$ , let  $\alpha$  be a polyadic instruction sequence vector of length  $n$ , and let  $\sigma$  be an instruction register file state. Then we write

<sup>3</sup> We write  $D^*$  for the set of all finite sequences with elements from set  $D$ . We use the following notation for finite sequences:  $\langle \rangle$  for the empty sequence,  $\langle d \rangle$  for the sequence having  $d$  as sole element,  $\sigma \circ \sigma'$  for the concatenation of finite sequences  $\sigma$  and  $\sigma'$ , and  $\text{len}(\sigma)$  for the length of finite sequence  $\sigma$ .

$valid(\alpha, i, \sigma)$  to indicate that instructions of the form  $\$get:i$  do not occur in  $prj_{pgn(\alpha, i)}(pg(\alpha, i))[\sigma]$ .

We also use special notation relating to the program notation indices. Let  $c$  be a program notation index. Then we write  $prj_c$  for the projection  $pg1a2pga$  if  $c = A$ , the projection  $pg1b2pga$  if  $c = B$ , etcetera.

An obvious choice of the thread extraction operation of  $PGA_p$  is the thread extraction operation of  $PGA$ , taking the set  $\mathfrak{A}_c \cup \mathfrak{A}_s$  for  $\mathfrak{A}$ , restricted to the set of closed terms of  $PGA_p$ . This thread extraction operation is considered not to be the proper one, because it treats the supplementary basic instructions as arbitrary basic instructions and thus disregards the fixed effects that they produce on execution. Moreover, this thread extraction operation requires that in BTA the restriction is dropped that  $\mathcal{A}$  must be a finite set.

As regards the proper thread extraction for  $PGA_p$ , the idea is that it yields, for each  $PGA_p$  program  $P$ , a function that gives, for each polyadic instruction sequence vector  $\alpha$ , the thread that is the joint behaviour of  $P$  and the polyadic instruction sequences in  $\alpha$  if  $P$  is the polyadic instruction sequence being executed initially. Because this behaviour depends upon the set of instruction registers that contain an instruction and the contents of each of those registers, we need a thread extraction operation for each instruction register file state.

For each instruction register file state  $\sigma$ , we introduce the *thread extraction* operation  $|-|_\sigma$ . These thread extraction operations are defined by the equations given in Table 6 (for  $a \in \mathfrak{A}$ ,  $l, i \in \mathbb{N}$ ,  $u \in \mathfrak{I}_c \cup \mathfrak{A}_s$  and  $v \in \mathfrak{I}_c$ ) and the rule that  $|\#l; X|_\sigma(\alpha) = D$  if  $\#l$  is the beginning of an infinite jump chain. Here, it is assumed that  $gl \in \mathcal{I}$ . The internal action  $gl$  (generate and load) represents the internal activity involved in switching over execution. The internal actions  $\tau$  and  $gl$  are distinguished because  $\tau$  is considered to represent a negligible internal activity, whereas  $gl$  is considered to represent a substantial internal activity.

We can couple nominal indices as labels with some of the polyadic instruction sequences in a polyadic instruction sequence vector. This would permit the use of alternative switch-over instructions with nominal indices instead of ordinal indices, like with the `goto` instructions from PGLDg. In the notational style of [2], the form of those alternative switch-over instructions would be  $\#\#\#[i]$ .

## 6 Example

In this section, we consider the splitting of a PGLD program  $P$  of 10000 instructions into two fragments.

We write  $\nu_1(l)$  for the number of absolute jump instructions  $\#\#l'$  with  $l' > 5000$  from position 1 up to position  $l$  and  $\nu_2(l)$  for the number of absolute jump instructions  $\#\#l'$  with  $l' \leq 5000$  from position 5001 up to position  $l$ .

The polyadic instruction sequence  $P'$  corresponding to the first half of  $P$  is obtained from the first half of  $P$  as follows:

- the instruction  $\$get:1$  is prefixed to it;

**Table 6.** Defining equations for thread extraction operations of  $\text{PGA}_p$

---

$ a _\sigma(\alpha) = a \circ \text{D}$	
$ a; x _\sigma(\alpha) = a \circ  x _\sigma(\alpha)$	
$ +a _\sigma(\alpha) = a \circ \text{D}$	
$ +a; x _\sigma(\alpha) =  x _\sigma(\alpha) \trianglelefteq a \triangleright  \#2; x _\sigma(\alpha)$	
$ -a _\sigma(\alpha) = a \circ \text{D}$	
$ -a; x _\sigma(\alpha) =  \#2; x _\sigma(\alpha) \trianglelefteq a \triangleright  x _\sigma(\alpha)$	
$ \#l _\sigma(\alpha) = \text{D}$	
$ \#0; x _\sigma(\alpha) = \text{D}$	
$ \#1; x _\sigma(\alpha) =  x _\sigma(\alpha)$	
$ \#l + 2; u _\sigma(\alpha) = \text{D}$	
$ \#l + 2; u; x _\sigma(\alpha) =  \#l + 1; x _\sigma(\alpha)$	
$ \! _\sigma(\alpha) = \text{S}$	
$ \! ; x _\sigma(\alpha) = \text{S}$	
$ \#\#\#i _\sigma(\alpha) = \text{gl} \circ  \text{prj}_{\text{pgn}(\alpha, i)}(\text{pg}(\alpha, i))[\sigma] _\sigma(\alpha)$	if $1 \leq i \leq n \wedge \text{valid}(\alpha, i, \sigma)$
$ \#\#\#i _\sigma(\alpha) = \text{D}$	if $1 \leq i \leq n \wedge \neg \text{valid}(\alpha, i, \sigma)$
$ \#\#\#i _\sigma(\alpha) = \text{S}$	if $i = 0 \vee i > n$
$ \#\#\#i; x _\sigma(\alpha) = \text{gl} \circ  \text{prj}_{\text{pgn}(\alpha, i)}(\text{pg}(\alpha, i))[\sigma] _\sigma(\alpha)$	if $1 \leq i \leq n \wedge \text{valid}(\alpha, i, \sigma)$
$ \#\#\#i; x _\sigma(\alpha) = \text{D}$	if $1 \leq i \leq n \wedge \neg \text{valid}(\alpha, i, \sigma)$
$ \#\#\#i; x _\sigma(\alpha) = \text{S}$	if $i = 0 \vee i > n$
$ \$\text{put}:i:v _\sigma(\alpha) = \text{tau} \circ \text{D}$	
$ \$\text{put}:i:v; x _\sigma(\alpha) = \text{tau} \circ  x _{\sigma \oplus [i \rightarrow v]}(\alpha)$	
$ \$\text{get}:i _\sigma(\alpha) = \text{D}$	
$ \$\text{get}:i; x _\sigma(\alpha) = \text{D}$	

---

- each absolute jump instructions  $\#\#l$  with  $l \leq 5000$  is replaced by the absolute jump instructions  $\#\#l'$ , where  $l' = l + \nu_1(l) + 1$ ;
- each absolute jump instructions  $\#\#l$  with  $l > 5000$  is replaced by the instruction sequence  $\$\text{put}:2:\#l'; \#\#\#2$ , where  $l' = (l - 5000) + \nu_2(l - 5000)$ ;

and the polyadic instruction sequence  $P''$  corresponding to the second half of  $P$  is obtained from the second half of  $P$  as follows:

- the instruction  $\$\text{get}:2$  is prefixed to it;
- each absolute jump instructions  $\#\#l$  with  $l > 5000$  is replaced by the absolute jump instructions  $\#\#l'$ , where  $l' = (l - 5000) + \nu_2(l - 5000) + 1$ ;
- each absolute jump instructions  $\#\#l$  with  $l \leq 5000$  is replaced by the instruction sequence  $\$\text{put}:1:\#l'; \#\#\#1$ , where  $l' = l + \nu_1(l)$ .

Notice that the positions occurring in jump instructions are adapted to the prefixing of a get instruction to each half of  $P$  and the replacement of each jump instructions that gives rise to a jump into the other half of  $P$  by two instructions.

For any instruction register file state  $\sigma$ , we have that  $|P|$  coincides with  $|\$put:1:\#1 ; \#\#\#1|_\sigma(\langle(P', D)\rangle \curvearrowright \langle(P'', D)\rangle)$  after the presence of the internal actions  $\tau$  and  $gl$  in the latter behaviour has been concealed. In Section 7, we will introduce operators to conceal the presence of internal actions.

In this section, we have illustrated by means of an example that splitting a program into fragments is relatively simple. In Section 9, we will show that synthesizing a program from a collection of fragments is fairly complicated.

## 7 Threads-Services Interaction and Abstraction

A thread may make use of services. That is, it may perform certain actions for the purpose of having itself affected by a service that takes them as commands to be processed. The processing of an action may involve a change of state of the service and at completion of the processing of the action the service returns a reply value to the thread. The reply value determines how the thread proceeds. In this section, we review the use operators, which are concerned with threads making such use of services.<sup>4</sup> We also introduce abstraction operators. They serve for concealment of the presence of internal actions, which arise among other things from the use operators. Both use operators and abstraction operators will be used later on in Section 9.

It is assumed that a fixed but arbitrary finite set  $\mathcal{F}$  of *foci* and a fixed but arbitrary finite set  $\mathcal{M}$  of *methods* have been given. Each focus plays the role of a name of a service provided by the execution environment that can be requested to process a command. Each method plays the role of a command proper. For the set  $\mathcal{A}$  of basic actions, we take the set  $\{f.m \mid f \in \mathcal{F}, m \in \mathcal{M}\}$ . Performing a basic action  $f.m$  is taken as making a request to the service named  $f$  to process command  $m$ .

A *service*  $H$  consists of

- a set  $S$  of *states*;
- an *effect* function  $eff : \mathcal{M} \times S \rightarrow S$ ;
- a *yield* function  $yld : \mathcal{M} \times S \rightarrow \{\mathbf{T}, \mathbf{F}, \mathbf{B}\}$ ;
- an *initial state*  $s_0 \in S$ ;

satisfying the following condition:

$$\forall m \in \mathcal{M}, s \in S \cdot (yld(m, s) = \mathbf{B} \Rightarrow \forall m' \in \mathcal{M} \cdot yld(m', eff(m, s)) = \mathbf{B}) .$$

The set  $S$  contains the states in which the service may be, and the functions  $eff$  and  $yld$  give, for each method  $m$  and state  $s$ , the state and reply, respectively,

<sup>4</sup> This version of the use mechanism was first introduced in [6]. In later papers, it is also called thread-service composition.

**Table 7.** Axioms for use operators

$S /_f H = S$		TSU1
$D /_f H = D$		TSU2
$(\iota \circ x) /_f H = \iota \circ (x /_f H)$		TSU3
$(x \trianglelefteq g.m \trianglerighteq y) /_f H = (x /_f H) \trianglelefteq g.m \trianglerighteq (y /_f H)$	if $f \neq g$	TSU4
$(x \trianglelefteq f.m \trianglerighteq y) /_f H = \mathbf{tau} \circ (x /_f \frac{\partial}{\partial m} H)$	if $H(\langle m \rangle) = \mathbf{T}$	TSU5
$(x \trianglelefteq f.m \trianglerighteq y) /_f H = \mathbf{tau} \circ (y /_f \frac{\partial}{\partial m} H)$	if $H(\langle m \rangle) = \mathbf{F}$	TSU6
$(x \trianglelefteq f.m \trianglerighteq y) /_f H = D$	if $H(\langle m \rangle) = \mathbf{B}$	TSU7

that result from processing  $m$  in state  $s$ . By the condition imposed on services, once the service has returned  $\mathbf{B}$  as reply, it keeps returning  $\mathbf{B}$  as reply.

Let  $H = (S, \text{eff}, \text{yld}, s_0)$  be a service and let  $m \in \mathcal{M}$ . Then the *derived service* of  $H$  after processing  $m$ , written  $\frac{\partial}{\partial m} H$ , is the service  $(S, \text{eff}, \text{yld}, \text{eff}(m, s_0))$ ; and the *reply* of  $H$  after processing  $m$ , written  $H(m)$ , is  $\text{yld}(m, s_0)$ .

When a thread makes a request to the service to process  $m$ :

- if  $H(m) \neq \mathbf{B}$ , then the request is accepted, the reply is  $H(m)$ , and the service proceeds as  $\frac{\partial}{\partial m} H$ ;
- if  $H(m) = \mathbf{B}$ , then the request is rejected and the service proceeds as a service that rejects any request.

We introduce the sort  $\mathbf{S}$  of *services* and, for each  $f \in \mathcal{F}$ , the binary *use* operator  $_{/f} : \mathbf{T} \times \mathbf{S} \rightarrow \mathbf{T}$ . The axioms for these operators are given in Table 7. In this table,  $\iota$  stands for an arbitrary member of  $\mathcal{I}$ ,  $f$  and  $g$  stand for arbitrary foci from  $\mathcal{F}$ , and  $m$  stands for an arbitrary method from  $\mathcal{M}$ .

Intuitively,  $p /_f H$  is the thread that results from processing all basic actions performed by thread  $p$  that are of the form  $f.m$  by service  $H$ . When a basic action of the form  $f.m$  performed by thread  $p$  is processed by service  $H$ , it is turned into the internal action  $\mathbf{tau}$  and postconditional composition is removed in favour of action prefixing on the basis of the reply value produced. This intuition is covered by axioms TSU5 and TSU6. Axioms TSU3 and TSU4 express that internal actions and basic actions of the form  $g.m$  with  $f \neq g$  are not processed. Axiom TSU7 expresses that deadlock takes place when a basic action to be processed is not accepted.

Let  $T$  stand for either BTA, BTA+REC or BTA+REC+AIP. Then we will write  $T$ +TSU for  $T$ , taking the set  $\{f.m \mid f \in \mathcal{F}, m \in \mathcal{M}\}$  for  $\mathcal{A}$ , extended with the use operators and the axioms from Table 7.

Both basic actions and internal actions are actions whose presence matters. For each  $\iota \in \mathcal{I}$ , we introduce the unary *abstraction* operator  $\tau_\iota : \mathbf{T} \rightarrow \mathbf{T}$  to conceal the presence of internal action  $\iota$  in the case where its presence does not matter after all.

The axioms for the abstraction operators are given in Table 8. In this table,  $a$  stands for an arbitrary action from  $\mathcal{A} \cup \mathcal{I}$ .

**Table 8.** Axioms for abstraction operators

$\tau_\iota(\mathbf{S}) = \mathbf{S}$		TT1
$\tau_\iota(\mathbf{D}) = \mathbf{D}$		TT2
$\tau_\iota(\iota \circ x) = \tau_\iota(x)$		TT3
$\tau_\iota(x \trianglelefteq a \triangleright y) = \tau_\iota(x) \trianglelefteq a \triangleright \tau_\iota(y)$	if $a \neq \iota$	TT4
$\bigwedge_{n \geq 0} \tau_\iota(\pi_n(x)) = \tau_\iota(\pi_n(y)) \Rightarrow \tau_\iota(x) = \tau_\iota(y)$		TT5

A main difference between the version of the use mechanism introduced here and the version of the use mechanism introduced in [9] is that the former version does not incorporate abstraction and the latter version incorporates abstraction.

Let  $T$  stand for either BTA, BTA+REC, BTA+REC+AIP, BTA+TSU, BTA+REC+TSU or BTA+REC+AIP+TSU. Then we will write  $T$ +ABSTR for  $T$  extended with the abstraction operators and the axioms from Table 8.

For each  $\iota \in \mathcal{I}$ , the equation  $\tau_\iota(\iota^\omega) = \mathbf{D}$  is derivable from the axioms of BTA+REC+AIP+ABSTR.

## 8 Instruction Register File Services

In this section, we describe services that make up register files with a finite set of registers that can contain instructions from a finite set of core primitive instructions. These services will be used in Section 9.

It is assumed that a fixed but arbitrary set  $I \subseteq \mathbb{N}$  such that  $I = [1, n]$  for some  $n \in \mathbb{N}$  and a fixed but arbitrary finite set  $U \subseteq \mathcal{I}_c$  have been given. The set  $I$  is considered to contain the positions of the registers in the instruction register file and the set  $U$  is considered to contain the instructions that can be put in those registers.

We write  $IRFS$  for the set  $\bigcup_{I' \subseteq I} I' \rightarrow U$ . The members of  $IRFS$  are considered to be the possible instruction register file states. It is assumed that a fixed but arbitrary bijection  $\theta : IRFS \rightarrow [1, \text{card}(IRFS)]$  has been given.

The instruction register file services accept the following methods:

- for each  $i \in I$  and  $u \in U$ , a *register put method*  $\text{put}:i:u$ ;
- for each  $j \in \text{rng}(\theta)$ , a *register file test method*  $\text{eq}:j$ .

We write  $\mathcal{M}_{\text{irf}}$  for the set  $\{\text{put}:i:u \mid i \in I \wedge u \in U\} \cup \{\text{eq}:j \mid j \in \text{rng}(\theta)\}$ .

It is assumed that  $\mathcal{M}_{\text{irf}} \subseteq \mathcal{M}$ .

The methods accepted by instruction register file services can be explained as follows:

- $\text{put}:i:u$ : the contents of register  $i$  becomes instruction  $u$  and the reply is  $\mathbf{T}$ ;
- $\text{eq}:j$ : if the state of the instruction register file equals  $\theta^{-1}(j)$ , then nothing changes and the reply is  $\mathbf{T}$ ; otherwise nothing changes and the reply is  $\mathbf{F}$ .

Let  $s \in IRFS \cup \{\uparrow\}$ , where  $\uparrow \notin IRFS$ . Then the *instruction register file service* with initial state  $s$ , written  $\mathcal{IRF}_s$ , is the service  $(IRFS, eff, yld, s)$ , where the functions  $eff$  and  $yld$  are defined as follows ( $\sigma \in IRFS \cup \{\uparrow\}$ ):<sup>5</sup>

$$\begin{aligned}
eff(\text{put}:i:u, \sigma) &= \sigma \oplus [i \mapsto u], & yld(\text{put}:i:n, \sigma) &= \top, \\
eff(\text{eq}:j, \sigma) &= \sigma, & yld(\text{eq}:j, \sigma) &= \top \quad \text{if } \theta(\sigma) = j, \\
& & yld(\text{eq}:j, \sigma) &= \text{F} \quad \text{if } \theta(\sigma) \neq j, \\
eff(m, \sigma) &= \uparrow \quad \text{if } m \notin \mathcal{M}_{\text{irf}}, & yld(m, \sigma) &= \text{B} \quad \text{if } m \notin \mathcal{M}_{\text{irf}}, \\
eff(m, \uparrow) &= \uparrow, & yld(m, \uparrow) &= \text{B}.
\end{aligned}$$

## 9 Program Synthesis

In order to establish a connection between collections of fragments and sequential programs, we show in this section that, for each possible joint behaviour of a collection of fragments, a sequential program can be synthesized from the collection of fragments that exhibits on execution essentially the behaviour in question by making use of an instruction register file service. More precisely, we show that, for each  $\text{PGA}_p$  program  $P$  and polyadic instruction sequence vector  $\alpha$ , a  $\text{PGA}$  program  $P'$  can be synthesized from  $P$  and  $\alpha$  such that, for all relevant instruction register file states  $\sigma$ ,  $|P'| \upharpoonright_{\text{irf}} \mathcal{IRF}_\sigma$  coincides with  $|P|_\sigma(\alpha)$  after the presence of the internal actions  $\text{tau}$  and  $\text{gl}$  has been concealed.

Let  $P$  be a  $\text{PGA}_p$  program and  $\alpha$  be a polyadic instruction sequence vector. The general idea is that:

- each polyadic instruction sequence in  $\alpha$  is translated into a  $\text{PGA}_p$  program and an appropriate finite collection of instances of this  $\text{PGA}_p$  program in which occurrences of get instructions are replaced by core primitive instructions is generated;
- $P$  and all the generated programs are translated into  $\text{PGLC}_p$  programs and these  $\text{PGLC}_p$  programs are concatenated;
- the resulting  $\text{PGLC}_p$  program is translated into a  $\text{PGLD}_p$  program and this program is translated into a  $\text{PGLD}$  program by replacing all occurrences of the supplementary instructions by core primitive instructions as follows:
  - a switch-over instruction  $\#\#\#i$  is replaced by an absolute jump instruction whose effect is a jump to the beginning of an appended instruction sequence whose execution leads, after the state of the instruction register file has been found by a linear search, to a jump to the beginning of the right instance of the  $\text{PGA}_p$  program that corresponds to the  $i$ th polyadic instruction sequence in  $\alpha$ ;
  - a put instruction  $\$\text{put}:i:u$  is simply replaced by the plain basic instruction  $\text{irf.put}:i:u$ ;

<sup>5</sup> We use the following notation for functions:  $[\ ]$  for the empty function;  $[d \mapsto r]$  for the function  $f$  with  $\text{dom}(f) = \{d\}$  such that  $f(d) = r$ ; and  $f \oplus g$  for the function  $h$  with  $\text{dom}(h) = \text{dom}(f) \cup \text{dom}(g)$  such that, for all  $d \in \text{dom}(h)$ ,  $h(d) = f(d)$  if  $d \notin \text{dom}(g)$  and  $h(d) = g(d)$  otherwise.

- a get instruction  $\$get:i$  is simply replaced by the absolute jump instruction whose effect is a jump to the position of the instruction itself.

A collection of instances of the  $PGA_p$  program corresponding to a polyadic instruction sequence in  $\alpha$  is considered appropriate if it includes all instances that may become the one being executed.  $P$  and all the generated programs are translated into  $PGLC_p$  programs because  $PGLC_p$  programs are relocatable: they can be concatenated without disturbing the meaning of jump instructions. The  $PGLC_p$  program resulting from the concatenation is translated into a  $PGLD_p$  program before the supplementary instructions are replaced because the replacement of a switch-over instruction by an absolute jump instruction is simpler than its replacement by a relative jump instruction.

Following the general idea outlined above, we define a function  $pgap2pgld$  that yields, for each  $PGA_p$  program  $P$ , a function that gives, for each polyadic instruction sequence vector  $\alpha$ , a  $PGLD$  program  $P'$  such that, for each relevant instruction register file service state  $\sigma$ ,  $|pgld2pga(P')|_{/irf \mathcal{IRF}_\sigma}$  coincides with  $|P|_\sigma(\alpha)$  after the presence of the internal actions  $\tau$  and  $gl$  has been concealed.

The function  $pgap2pgld$  from the set of all  $PGA_p$  programs to the set all functions from the set of all polyadic instruction sequence vectors to the set of all  $PGLD$  programs is defined as follows:

$$\begin{aligned} pgap2pgld(x)(\alpha) = & \\ & translate(pg1c2pgld(expand(x)(\alpha))) ; \\ & +irf.eq:1 ; \#\#l_{1,1} ; \dots ; +irf.eq:n' ; \#\#l_{1,n'} ; \\ & \vdots \\ & +irf.eq:1 ; \#\#l_{n,1} ; \dots ; +irf.eq:n' ; \#\#l_{n,n'} , \end{aligned}$$

where  $n = \text{len}(\alpha)$ ,  $n' = \max(\text{rng}(\theta))$ , the function  $expand$  from the set of all  $PGA_p$  programs to the set all functions from the set of all polyadic instruction sequence vectors to the set of all  $PGLC_p$  programs is defined as follows:

$$\begin{aligned} expand(x)(\alpha) = & \\ & pga2pg1c(x) ; \\ & pga2pg1c(gen(\alpha, 1, \theta^{-1}(1))) ; \dots ; pga2pg1c(gen(\alpha, 1, \theta^{-1}(n'))) ; \\ & \vdots \\ & pga2pg1c(gen(\alpha, n, \theta^{-1}(1))) ; \dots ; pga2pg1c(gen(\alpha, n, \theta^{-1}(n'))) , \end{aligned}$$

where  $n = \text{len}(\alpha)$ ,  $n' = \max(\text{rng}(\theta))$ , and the function  $gen$  from the set of all polyadic instruction sequence vectors, the set of all natural numbers and the set of all instruction register file states to the set of all  $PGA_p$  programs is defined as follows:

$$\begin{aligned} gen(\alpha, i, \sigma) = prj_{pgn(\alpha, i)}(pg(\alpha, i))[\sigma] & \text{ if } 1 \leq i \leq \text{len}(\alpha) \wedge \text{valid}(\alpha, i, \sigma) , \\ gen(\alpha, i, \sigma) = \#0 & \text{ if } 1 \leq i \leq \text{len}(\alpha) \wedge \neg \text{valid}(\alpha, i, \sigma) , \\ gen(\alpha, i, \sigma) = ! & \text{ if } i = 0 \vee i > \text{len}(\alpha) , \end{aligned}$$

the function *translate* from the set of all PGLD<sub>p</sub> programs to the set of all PGLD programs is defined as follows:

$$\text{translate}(u_1 ; \dots ; u_k) = \psi_1(u_1) ; \dots ; \psi_1(u_k) ,$$

where the functions  $\psi_j$  from the set of all primitive instructions of PGLD<sub>p</sub> to the set of all primitive instructions of PGLD are defined as follows ( $1 \leq j \leq k$ ):

$$\begin{aligned} \psi_j(\#\#\#i) &= \#\#l_i && \text{if } 1 \leq i \leq \text{len}(\alpha) , \\ \psi_j(\#\#\#i) &= ! && \text{if } i = 0 \vee i > \text{len}(\alpha) , \\ \psi_j(\$put:i:u) &= \text{irf.put:i:u} , \\ \psi_j(\$get:i) &= \#\#j , \\ \psi_j(u) &= u && \text{if } u \text{ is not a supplementary basic instruction} , \end{aligned}$$

where for each  $i \in [1, \text{len}(\alpha)]$ :

$$\begin{aligned} l_i &= \text{len}(\text{pga2pg1c}(x)) \\ &+ \sum_{h \in [1, \text{len}(\alpha)], h' \in \text{rng}(\theta)} \text{len}(\text{pga2pg1c}(\text{prj}_{\text{pgn}(\alpha, h)}(\text{pg}(\alpha, h))[\theta^{-1}(h')])) \\ &+ 2 \cdot \max(\text{rng}(\theta)) \cdot (i - 1) , \end{aligned}$$

and for each  $i \in [1, \text{len}(\alpha)]$  and  $j \in \text{rng}(\theta)$ :

$$\begin{aligned} l_{i,j} &= \text{len}(\text{pga2pg1c}(x)) \\ &+ \sum_{h \in [1, i-1], h' \in \text{rng}(\theta)} \text{len}(\text{pga2pg1c}(\text{prj}_{\text{pgn}(\alpha, h)}(\text{pg}(\alpha, h))[\theta^{-1}(h')])) \\ &+ \sum_{h' \in [1, j-1]} \text{len}(\text{pga2pg1c}(\text{prj}_{\text{pgn}(\alpha, i)}(\text{pg}(\alpha, i))[\theta^{-1}(h')])) . \end{aligned}$$

The following theorem states rigorously that, for any PGA<sub>p</sub> program  $P$  and polyadic instruction sequence vector  $\alpha$ , for all relevant instruction register file states  $\sigma$ ,  $|\text{pg1d2pga}(\text{pgap2pg1d}(P)(\alpha))| \text{ /}_{\text{irf}} \text{IRF}_\sigma$  coincides with  $|P|_\sigma(\alpha)$  after the presence of the internal actions  $\tau_{\text{au}}$  and  $\text{gl}$  has been concealed.

**Theorem 1.** *Let  $P$  be a PGA<sub>p</sub> program and  $\alpha$  be a polyadic instruction sequence vector, and let  $n$  be the highest number occurring in instructions of the form  $\$put:i:u$  or  $\$get:i$  in  $P$  or  $\alpha$ . Take the interval  $[1, n]$  for  $I$  and the set of all core primitive instructions occurring in instructions of the form  $\$put:i:u$  in  $P$  or  $\alpha$  for  $U$ , and let  $\sigma \in \text{IRFS}$ . Then  $\tau_{\text{au}}(|\text{pg1d2pga}(\text{pgap2pg1d}(P)(\alpha))| \text{ /}_{\text{irf}} \text{IRF}_\sigma) = \tau_{\text{au}}(\tau_{\text{gl}}(|P|_\sigma(\alpha)))$ .*

*Proof.* The proof of Theorem 1 follows the same line as the proof of Theorem 1 from [4] given in that paper. Here, we give only a brief outline of the proof of the current theorem.

The proof of this theorem proceeds as follows: (i) we give a set  $T$  of closed terms of sort  $\mathbf{T}$  with  $\tau_{\text{au}}(\tau_{\text{gl}}(|P|_\sigma(\alpha))) \in T$ , a set  $T'$  of closed terms of sort  $\mathbf{T}$  with

$\tau_{\text{tau}}(|\text{pgld2pga}(\text{pgap2pgld}(P)(\alpha))| /_{\text{irf}} \mathcal{IRF}_\sigma) \in T'$ , and a bijection  $\beta : T \rightarrow T'$ ;  
(ii) we show that there exists a set  $E$  consisting of one derivable equation  $p = p'$  for each  $p \in T$  such that, for all equations  $p = p'$  in  $E$ :

- $\beta(p) = p''$  is a derivable equation if  $p''$  is  $p'$  with, for all  $q \in T$ , all occurrences of  $q$  in  $p'$  replaced by  $\beta(q)$ ;
- $p' \in T$  only if  $p'$  can be rewritten to a  $q' \notin T$  using the equations in  $E$  from left to right.

This means that  $\tau_{\text{tau}}(\tau_{\text{gl}}(|P|_\sigma(\alpha)))$  and  $\tau_{\text{tau}}(|\text{pgld2pga}(\text{pgap2pgld}(P)(\alpha))| /_{\text{irf}} \mathcal{IRF}_\sigma)$  denote solutions of the same guarded recursive specification. Because guarded recursive specifications have unique solutions, it follows immediately that  $\tau_{\text{tau}}(|\text{pgld2pga}(\text{pgap2pgld}(P)(\alpha))| /_{\text{irf}} \mathcal{IRF}_\sigma) = \tau_{\text{tau}}(\tau_{\text{gl}}(|P|_\sigma(\alpha)))$ .  $\square$

In the proof outlined above, an apposite indexing of the closed terms in the sets  $T$  and  $T'$  facilitates the definition of the bijection  $\beta$ . Yet, this definition is much more complicated than the definition of the bijection needed in the proof from [4] referred to.

The definition of `pgap2pgld` shows that the synthesis of sequential programs from collections of fragments is fairly complicated.

## 10 Conclusions

We have given a formalization of instruction sequences that have been split into fragments. Thread extraction provides the possible joint behaviours of a collection of fragments. We have shown that, for each possible joint behaviour of a collection of fragments, a sequential program can be synthesized from the collection of fragments that exhibits on execution essentially the behaviour in question by making use of a service. This program synthesis is reminiscent of the service-based variant of projection semantics for program notations used in [5]. However, we consider the program synthesis too complicated to serve a semantical purpose.

In this paper, a fragment in a collection of fragments has two attributes: an ordinal index (position) and a program notation index. We have also mentioned that a nominal index (label) could be an optional attribute. Many other attributes that are relevant in practice can be imagined, e.g. modification date, author, tester, owner, user, and security level. In this paper, we have restricted ourselves to attributes that are indispensable for a theoretical understanding of instruction sequence fragmentation.

The question arises whether all aspects of instruction sequence fragmentation covered in this paper can be dealt with at the level of threads. This issue has been investigated in [8]. It happens that not all aspects can be dealt with at the level of threads. In particular, the ability to replace special instructions in an instruction sequence fragment by different ordinary instructions every time execution is switched over to that fragment cannot be dealt with at the level of threads. Threads turn out to be too abstract to deal with instruction sequence fragmentation in full.

## References

1. Bergstra, J.A., Bethke, I.: Polarized process algebra and program equivalence. In: J.C.M. Baeten, J.K. Lenstra, J. Parrow, G.J. Woeginger (eds.) Proceedings 30th ICALP, *Lecture Notes in Computer Science*, vol. 2719, pp. 1–21. Springer-Verlag (2003)
2. Bergstra, J.A., Bethke, I.: Predictable and reliable program code: Virtual machine based projection semantics. In: J.A. Bergstra, M. Burgess (eds.) Handbook of Network and Systems Administration, pp. 653–685. Elsevier, Amsterdam (2007)
3. Bergstra, J.A., Loots, M.E.: Program algebra for sequential code. *Journal of Logic and Algebraic Programming* **51**(2), 125–156 (2002)
4. Bergstra, J.A., Middelburg, C.A.: Instruction sequences with dynamically instantiated instructions. Electronic Report PRG0710, Programming Research Group, University of Amsterdam (2007). Available from <http://www.science.uva.nl/research/prog/publications.html> and <http://arxiv.org/ArXiv:0711.4217v2> [cs.PL]
5. Bergstra, J.A., Middelburg, C.A.: Instruction sequences with indirect jumps. *Scientific Annals of Computer Science* **17**, 19–46 (2007)
6. Bergstra, J.A., Middelburg, C.A.: Thread algebra for strategic interleaving. *Formal Aspects of Computing* **19**(4), 445–474 (2007)
7. Bergstra, J.A., Middelburg, C.A.: Program algebra with a jump-shift instruction. *Journal of Applied Logic* (2008). In press, doi: 10.1016/j.jal.2008.07.001
8. Bergstra, J.A., Middelburg, C.A.: Thread algebra for poly-threading. Electronic Report PRG0810, Programming Research Group, University of Amsterdam (2008). Available from <http://www.science.uva.nl/research/prog/publications.html> and <http://arxiv.org/ArXiv:0803.0378v2> [cs.L0]
9. Bergstra, J.A., Ponse, A.: Combining programs and state machines. *Journal of Logic and Algebraic Programming* **51**(2), 175–192 (2002)
10. Ponse, A., van der Zwaag, M.B.: An introduction to program and thread algebra. In: A. Beckmann, et al. (eds.) CiE 2006, *Lecture Notes in Computer Science*, vol. 3988, pp. 445–458. Springer-Verlag (2006)
11. Sannella, D., Tarlecki, A.: Algebraic preliminaries. In: E. Astesiano, H.J. Kreowski, B. Krieg-Brückner (eds.) Algebraic Foundations of Systems Specification, pp. 13–30. Springer-Verlag, Berlin (1999)
12. Wirsing, M.: Algebraic specification. In: J. van Leeuwen (ed.) Handbook of Theoretical Computer Science, vol. B, pp. 675–788. Elsevier, Amsterdam (1990)