LPS ALGORITHMS

Andy Lowry, Stephen Taylor
& Salvatore J. Stolfo

CUCS-203-84

第五世代コンピュータ国際会議

PROCEEDINGS OF THE

INTERNATIONAL CONFERENCE ON FIFTH GENERATION COMPUTER SYSTEMS 1984

LPS Algorithms[1]

Andy Lowry, Stephen Taylor, Salvatore J. Stolfo
Columbia University, Department of Computer Science
New York, New York 10027, USA

Abstract

LPS is a Logic Programming System currently under development and specifically targeted for implementation on massively parallel architectures. We present a detailed explanation of algorithms under development for parallel execution of LPS programs. The explanation is significantly more detailed than those published previously. An abstract proof procedure is developed which encompasses these algorithms and several variants, as well as the standard sequential Prolog algorithm. This abstract procedure provides a conceptual basis for our discussion and for a critical analysis of various execution strategies.

The algorithms have been successfully implemented and demonstrated in simulation on a number of small programs. Work is currently underway to transfer this implementation to a working prototype machine based on the DADO parallel architecture.

1 Introduction

Logic programming has attracted a great deal of attention as a medium for the development of software for parallel execution. Two major factors contributing to this perception are the demonstrated suitability of logic programming for the expression of a wide variety of software tasks, and the identification of several sources of parallelism inherent in the logic formalism itself. Thus logic programming languages appear to offer a framework in which programs naturally lend themselves to efficient parallel execution, but in which the programmer need not be overly cognisant of this goal.

With this view in mind we have developed methods for the execution of logic programs written in a language we call LPS, under a particular parallel execution model (12: Taylor et al. 1984; 14: Taylor et al. 1984). Our methods are not well characterised by any of the sources of parallelism identified in (Conery 1983), although they bear some resemblance to OR and AND parallelism. We unify a conjunction of goals simultaneously throughout a network of what may be considered intelligent memory devices. Each of these devices receives the entire goal list and attempts unification of each goal with every literal in its own local store. Upon completion of this activity, a series of network queries and combining operations results in the construction of a single relation representing all potential solutions of the original conjunction. The cycle repeats by selecting one member of that relation and producing from it a new conjunction to be solved.

We may view our proof search as a perusal through a tree of goal lists, where each node gives rise to children that can be obtained via resolution of one or more of its goals with clauses in the program. The structure of this tree depends on which goals are chosen for resolution in each node. In particular, we note that the standard sequential Prolog algorithm[2] chooses exactly one goal in each node, whereas the *current* LPS algorithms[3] always resolve every goal in the goal list. Both algorithms pursue a depth first search, although the LPS search tree, in comparison to the Prolog search tree, is characterised by:

- Shorter paths to leaves
- Earlier termination of unproductive paths
- Earlier consideration of most goals, causing earlier branching but not necessarily higher branching factors
- A substantially reorganised leaf structure, resulting in a different order to the construction of solutions

Although the LPS algorithms may appear to exhibit something of a breadth first nature due to the simultaneous construction of all children for whichever node is under consideration, that view is misleading. Although the children are constructed in unison, one child's subtree is searched before any other child is considered, so that the search pattern itself is purely depth first. The process may be viewed as a hill-climbing strategy in which all branches are equally favored.

In this paper we begin by presenting an abstract proof procedure that encompasses both the LPS and the Prolog algorithms, as well as many variations. We proceed with a specific example of the algorithm at work, followed by detailed explanation of the current LPS implementation in terms of the abstract algorithm. Finally, several alternative execution strategies are developed and analysed in the context of the abstract proof procedure.

We include discussion of the trade-offs among various execution strategies in terms of performance, storage requirements, and appropriateness to various types of logic programs. Much of the analysis presented here is intuitive in nature, due to a lack of observed performance measurements. Meaningful measurements are difficult to obtain because:

- Our current implementation is in the form of a simulation on a sequential machine, so that sample execution of any but the tiniest programs is prohibitively expensive. Implementations on a functioning parallel machine are

[1] This research is supported cooperatively by: Defense Advanced Research Projects Agency under contract N00039-82-C-0427, New York State Science and Technology Foundation, Intel Corporation, Digital Equipment Corporation, Valid Logic Systems Inc., Hewlett-Packard, AT&T Bell Laboratories and International Business Machines Corporation.

[2] See (Warren 1977). We will henceforth refer to this algorithm as simply the "Prolog algorithm"

[3] We note that the algorithms are under ongoing development

currently underway.

- The algorithms do not as yet provide for extensions to the Horn clause formalism such as negated condition elements, evaluable predicates, and goals with side-effects. These features are generally required by logic programs that attempt to do anything substantial and useful, so most existing programs cannot be executed in our current framework.

It is hoped that future work will remove these obstacles and allow for statistical analyses providing greater insight into the effects of the various strategies. This should in turn suggest opportunities for a more general mathematical analysis.

For an introduction to logic programming methods the reader is referred to (Robinson 1965; Robinson 1979; Kowalski 1979). A very brief description of the Prolog language, on which much of LPS has been modeled, may be found in (Shapiro 1982); for complete details see (Bowen et al. 1982). A description of the computing model for which our algorithms are targeted may be found in (12: Taylor et al. 1984). The DADO architecture, for which a specific implementation is underway, is described in (Stolfo and Shaw 1982; Stolfo et al. 1983). The reconciliation operation which we use may have been independently discovered by Pollard (Pollard 1981), although we have encountered significant difficulty in obtaining this reference. Related algorithms are described in (Khabaza 1984).

2 An Abstract Proof Procedure

2.1 Proofs

We define a *proof* for a given directive to be sequence of goal lists beginning with an instance of the directive and terminating in the empty goal list. Each goal list is composed of contributions from the individual goals in the preceding goal list, where each goal contributes any one of the following:

- Itself, as a singleton goal list. In this case we say the goal has been *retained*.
- The empty goal list, if the goal is satisfied via some fact. In this case we say the goal has been *removed*.
- The instance, under some substitution, of a rule body whose rule head, under the same substitution, is identical to the goal. Here we say the goal has been *expanded*.

Our proof procedure can then be viewed as the search for such a sequence. In addition, if a proof is found, the minimal substitution that transforms the directive into the first goal list in the sequence is displayed. We call this substitution a *solution* for the directive.

Since there may be more than one way to satisfy any given goal, one goal list may give rise to more than one successor goal list, any or all of which may lead to a successful proof. Thus there may be several proofs for a single directive. In general we will want our proof procedure to be capable of pursuing all possible proofs in a systematic fashion.

The difference stated in the Introduction between the search trees traversed by the Prolog and LPS algorithms may now be restated as follows: The Prolog algorithm pursues proofs in which each proof step consists of either removing or expanding the first goal in a goal list and retaining all other goals. In the current LPS algorithms no goal is ever retained in a goal step; rather, each goal is either removed or expanded.

2.2 The Procedure

Our description of what constitutes a proof allows us to quite readily verify proofs that are handed to us, but it is substantially more difficult to discover correct proofs when they exist. Two processes allow us to identify the substitutions that give rise to proofs: *unification* and *reconciliation*.

Unification (Robinson 1965) provides a method for determining whether a substitution exists that will transform two terms into identical terms. Such a substitution is called a *unifier*, although in the sequel we shall use this term to refer specifically to the *most general unifier*. By "most general" we mean that if U is the most general unifier of terms T_1 and T_2, and S is any other unifying substitution, then $S(T_1)$ is an instance of $U(T_1)$.

Reconciliation (Pollard 1981; Khabaza 1984) is a procedure for determining whether two substitutions are compatible, and if so, producing the "most general" substitution that subsumes both. By this we mean that if R is the reconciliation of substitutions S_1 and S_2, then for any term T, R(T) is an instance of both $S_1(T)$ and $S_2(T)$. As with unification, by "most general" we mean that any other substitution with this property, when applied to any term T, gives rise to an instance of R(T).

Given the mechanisms of unification and reconciliation, the construction of a solution for a directive can be accomplished as shown in Figure 2-1. Starting with the directive itself as a goal list, the algorithm produces successive goal lists until either an empty goal list is constructed or a failure condition is encountered. Upon successful termination, Substitution_List contains a sequence of substitutions whose composition is a solution for the directive.

Construction of a new goal list from its predecessor proceeds as follows:

1. Each goal is analysed individually to produce: its *contribution* to the new goal list; a substitution (which we call an *instantiator*) that will be applied to the contribution before its addition to the new goal list; and another substitution comprising *constraints* on the overall solution.
2. The constraining substitutions are combined via reconciliation to produce a substitution supporting this goal step as a whole. This substitution is saved as a component of the solution that we seek.
3. All instantiators are updated through composition with the above reconciliation.
4. Each contribution is passed through its corresponding instantiator, and the results are collected into a single goal list.

2.2.1 Contributions

Contributions (in their pre-instantiated form) are determined as follows:

- A RETAINED GOAL contributes itself, verbatim.[4]
- A REMOVED GOAL contributes nothing.
- An EXPANDED GOAL contributes the body

[4] Keep in mind that we are presenting an abstract proof procedure which encompasses several practical strategies. Thus although we have stated that the LPS algorithms never retain a goal, we include goal retention in our abstract procedure in order to accomodate both the Prolog algorithm and several variants on the LPS algorithms

```
Goal_List := Directive;
Substitution_List := NIL;

WHILE Not Empty(Goal_List) DO

  Constraint_Set := NIL;

  FOREACH goal G in Goal_List DO
    Decide whether G is to be retained, removed, or
      expanded;
    IF retaining G THEN
      Contribution(G) := G;
      Instantiator(G) := NIL;
    ELSE IF removing G THEN
      Find a fact unifying with G, call the unifier U;
        IF none can be found, FAIL;
      Contribution(G) := NIL;
      Instantiator(G) := NIL;
      Restrict U to bindings for variables in G, add
        the result to Constraint_Set;
    ELSE IF expanding G THEN
      Find a rule R whose head unifies with G, call the
        unifier U; IF none can be found, FAIL;
      Contribution(G) := rule body of unifying rule;
      Instantiator(G) := U restricted to variables in R;
      Insert bindings to new created variables into
          Instantiator(G) for all variables from R not bound
          by U;
      Restrict U to bindings for variables in G, add
        the result to Constraint_Set;
    FI;
  OD;

  Compute reconciliation of all substitutions in
    Constraint_Set, call the result Rec; IF reconciliation
    fails, FAIL;
  Add Rec to Substitution_List;

  New_Goal_List := NIL;
  FOREACH goal G in Goal_List DO
    Instantiator(G) := Instantiator(G) composed with R;
    Instantiate Contribution(G) using Instantiator(G),
      and add the result to New_Goal_List;
  OD;

  Goal_List := New_Goal_List;
OD;
```

Figure 2-1: Abstract Proof Procedure

---

of the rule with whose head it unifies, verbatim.

2.2.2 Instantiators

Non-empty instantiators are only produced for expanded goals. It would be pointless to compute an instantiator for a removed goal since its contribution is always empty; in the case of a retained goal, all instantiation information comes from the constraints imposed by unification of non-retained goals, so an empty instantiator is set in place awaiting composition with the reconciliation of those constraints.

The instantiator for an expanded goal is simply the unifier that resulted from unification of the goal with a rule head. We only include bindings for variables that are contained in the rule (*rule variables*), since other bindings cannot contribute to instantiation of the rule body. We also insure that every rule variable is represented in the instantiator by binding any unbound rule variables to new created variables. Such a binding adds no information; the objective is to insure that the instantiated rule body will contain none of the original rule variables.

2.2.3 Constraints

Constraints are produced by unification of removed goals with facts and expanded goals with rule heads. Each unifier is added to a constraint set, after restricting it to variables that occurred in the goal (*goal variables*). The constraint set is used to produce a consistent substitution for the preceding goal list which supports its transformation into the succeeding goal list. Thus the only bindings of interest are those for goal variables, which is why the unifiers are pruned before adding them to the constraint set. Indeed, if the same fact or rule head is used to unify with more than one goal, inconsistent bindings for non-goal variables might result, but these must not prevent the proof from progressing. For example, consider the following program:[5]

```
Rule 1: tasty(X) :- sweet(X).
Fact 1: sweet(cookies).
Fact 2: sweet(cake).

Directive: tasty(cookies), tasty(cake).
```

We suppose that (as would be the case with LPS) our algorithm chooses to expand both of the original goals in its first step, using Rule 1. Unification of tasty(cookies) with tasty(X) produces the unifier [X/cookies], while unification of tasty(cake) with tasty(X) produces [X/cake]. Reconciliation of these two unifiers cannot succeed since variable X cannot be bound to both cookies and cake simultaneously. Clearly, though, the directive is provable. This problem of unwanted binding interaction does not occur if we discard bindings for X prior to reconciliation. Note that these bindings remain in instantiators so that they may be used for instantiation of rule bodies.

Similar reasoning shows why it is necessary to include "dummy bindings" for non-unified rule variables in the instantiators for expanded goals. If this were not done, those rule variables might end up occuring in two or more goals at some point during the proof. This would cause unwanted interactions since the algorithm would insure that only mutually compatible bindings were produced for all occurrences of those variables, while the separate occurrences should in fact be treated independently.

The purpose of composing each instantiator with the constraint set reconciliation is to insure that each goal list is cast in terms of the current state of knowledge of the solution under construction. That solution is constructed as a sequence of component substitutions, where each proof step produces one component. If goal lists are not kept up to date in this fashion, the same variable may end up bound by two or more different components. During later composition of the components, all but the first of these bindings would be completely lost. For example, the composition of [X/cookies] with [X/cake] is simply [X/cookies]. In general, it will be the case that no goal list will ever contain a variable for which a binding exists anywhere in the component substitutions produced thus far in the proof procedure.

---

[5] For our examples we adopt the Prolog convention that symbols beginning with a capital letter are considered variables, while all others are considered predicate and function symbols

2.3 Some Observations

Due to the "most general" nature of unification and reconciliation, our algorithm computes the most general solution that will support the constructed proof. This translates into conciseness in the solution set reported for a directive, although it does not guarantee that no solution will be an instance of another. This may arise if there are multiple proof paths for some particular solution.

Upon failure of a particular proof path, both the LPS and Prolog algorithms backtrack to the most recent choice point and pursue an alternate path. In the LPS algorithms we find that all of these alternate paths have already been started by the simultaneous construction of all possible successor goal lists from the choice point. The Prolog algorithms do not benefit from such a head start. As mentioned in the Introduction, this feature may easily mislead one to suspect that the LPS search strategy includes some breadth first component rather than being strictly depth first.

Finally, it will be seen that in LPS the composition of the component substitutions is performed incrementally as each component is produced, rather than computing the entire composition at the end of the proof.

3 A Proof Example

Consider the following program:

```
Rule 1: can_eat(X) :- food_store(S), open(S,now),
                      has_money(X).
Rule 2: has_money(X) :- friend(Y,X), has_money(Y).
Fact 1: food_store(mama_joys).
Fact 2: food_store(take_home).
Fact 3: friend(chris,andy).
Fact 4: friend(tori,chris).
```

Suppose the author is interested in whether or not he is currently able to eat. First, from general knowledge of neighborhood food stores, and by subtly questioning his friends, he arrives at the following additional facts:

```
Fact 5: open(mama_joys,now).
Fact 6: has_money(tori).
```

Next he invokes the proof algorithm with the directive **can_eat(andy)** and observes the following execution:

1. The initial goal list is {**can_eat(andy)**}. We choose to expand the single goal via Rule 1. Unification with the rule head produces the substitution [X/andy].

Our goal's pre-instantiated contribution is the rule-body, {**food_store(S), open(S,now), has_money(X)**}. The instantiator is [X/andy, S/_1], where _1 is a created variable to which S is bound since it was not bound during unification. This expansion contributes nothing to the constraint set since no goal variables were bound during unification (indeed, there were no goal variables to be bound!).

Reconciliation of our (empty) constraint set produces an empty substitution, so our instantiator is not affected, and the next goal list is {**food_store(_1), open(_1,now), has_money(andy)**}.

2. Current goal list: {**food_store(_1), open(_1,now). has_money(andy)**}

Retain goal **food_store(_1)**:

Contribution: **food_store(_1)**
Instantiator: **NIL**
Constraint: **NIL**

Remove goal **open(_1,now)** via Fact 5:

Contribution: **NIL**
Instantiator: **NIL**
Constraint: [_1/mama_joys]

Expand goal **has_money(andy)** via Rule 2:

Contribution: {**friend(Y,X), has_money(Y)**}
Instantiator: [X/andy,Y/_2]
Constraint: **NIL**

The overall-constraint set is {[_1/mama_joys]}, whose reconciliation is just [_1/mama_joys]. The only instantiator that is affected by this reconciliation is the first, which becomes [_1/mama_joys]. Instantiating all of the contributions with their instantiators then produces the new goal list: {**food_store(mama_joys), friend(_2,andy), has_money(_2)**}.

3. Current goal list: {**food_store(mama_joys), friend(_2,andy), has_money(_2)**}

Remove goal **food_store(mama_joys)** via Fact 1:

Contribution: **NIL**
Instantiator: **NIL**
Constraint: **NIL**

Remove goal **friend(_2,andy)** via Fact 3:

Contribution: **NIL**
Instantiator: **NIL**
Constraint: [_2/chris]

Expand goal **has_money(_2)** via Rule 2:

Contribution: {**friend(Y,X), has_money(Y)**}
Instantiator: [X/_3, Y/_4]
Constraint: [_2/_3]

The overall constraint set is {[_2/chris], [_2/_3]}, whose reconciliation is [_2/chris, _3/chris]. This affects the instantiator for the third goal, which becomes [X/chris,Y/_4]. Instantiating all of the contributions with their instantiators yields the new goal list: {**friend(_4,chris), has_money(_4)**}.

4. Current goal list: {**friend(_4,chris), has_money(_4)**}

Remove goal **friend(_4,chris)** via Fact 4:

Contribution: **NIL**
Instantiator: **NIL**
Constraint: [_4/tori]

Remove goal **has_money(_4)** via fact 6:

Contribution: **NIL**
Instantiator: **NIL**
Constraint: [_4/tori]

The overall constraint set is {[_4/tori], [_4/tori]},[6] whose

[6] Of course, this constraint set is not really a set since it contains duplicate entries. However, the terminology is useful in a loose sense, and the current LPS implementation will in fact go through the work of reconciling two identical constraints rather than removing the duplicity

reconciliation is [_4/tori]. All contributions are nil, so the new goal list is empty.

5. Current goal list: {}

The algorithm terminates successfully upon encountering an empty goal list.

The sequence of reconciliations that was generated by the algorithm is:

```
[]
[_1/mama_joys]
[_2/chris, _3/chris]
[_4/tori]
```

The composition of these components yields the overall substitution: [_1/mama_joys, _2/chris, _3/chris, _4/tori]. The sequence of generated goal lists is:

```
{can_eat(andy)}
{food_store(_1), open(_1,now), has_money(andy)}
{food_store(mama_joys), friend(_2,andy),
        has_money(chris)}
{friend(_4,chris), has_money(_4)}
NIL
```

If we apply the overall substitution to this sequence of goal lists, we arrive at our final proof:

```
{can_eat(andy)}
{food_store(mama_joys), open(mama_joys,now),
        has_money(andy)}
{food_store(mama_joys), friend(chris,andy),
        has_money(chris)}
{friend(tori,chris), has_money(tori)}
NIL
```

4 The Current LPS Implementation

The LPS algorithms that we have formulated can most easily be understood as comprising three computational phases: *unification*, *join*, and *substitution*. In this section we will discuss an actual LPS implementation in terms of these components, relating each functionally to the abstract algorithm outlined above.

The implementation is based on the computing model described in (12: Taylor et al. 1984). Very briefly, we envision a network of independent *processing elements* (PE's) each equipped with a moderate local storage capacity. The network is controlled by a *control processor* (CP) which coordinates global communication and invokes individual instructions as well as local procedures in unison throughout the PE network. Global communication consists of *broadcast* messages from the CP to the network, and *reports* solicited by the CP from individual PE's.

4.1 The Binding Set Representation

A binding set represents the result of applying a single step of our proof procedure to a goal list. It contains the following information:

- The reconciliation of the constraint set produced by unification of goals with facts and rule heads.
- A list of rule body keys by means of which rule bodies may be obtained at the CP for instantiation and inclusion in a new goal list. Note that a single rule body key may appear more than once. This will be the case if the same rule head was used to expand more than one goal in the goal list.

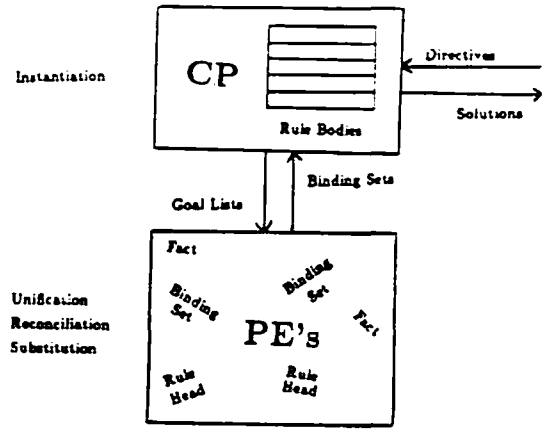

Figure 4-1: Flow of Data in LPS Execution

- An instantiator for each rule body key contained in the binding set. If a key appears more than once, each is associated with its own instantiator.

Recall that the current LPS algorithms never retain goals from one goal list to the next. Thus the above set of information includes everything required to construct the successor goal list as well as the solution component produced by this goal step. The overall data structure may be viewed as comprising several "layers," each identified with a layer "marker." Each layer contains a substitution of some sort – either the single reconciliation carried by the binding set or one of the possibly many instantiators. In the former case, the layer is called the *common layer* owing to its nature as a substitution that encompasses all the constraint set components contributed by the unifications. The layer marker for the common layer is the atom, COMMON. A layer containing an instantiator is called a *rule layer*, since a non-empty instantiator is produced only for a goal that is expanded by unification with some rule head. The marker for a rule layer is a key identifying the rule that was used in the expansion.

A binding set with no rule layers is of special interest, and we call it a *simple binding set*. Other binding sets are symmetrically termed *complex binding sets*. A simple binding set is important because it is reported only at the completion of a successful proof.

4.2 Distribution of Data

As we shall see, all unification is performed in the individual PE's that form the processor network, whereas instantiation takes place in the CP. For this reason we store all facts and rule heads, (that is, all the positive literals of our program) in the PE network itself. Each literal resides in a single PE, although any PE may contain several literals. Rule bodies, on the other hand, are kept in the CP. Each rule head in the PE network is tagged with a key which can be used to identify the corresponding rule body in the table maintained by the CP.

During execution of a logic program, goal lists are constructed in the CP, initially from the directive and subsequently from the goal list contributions carried in the binding sets. When a goal list is complete it is transmitted to the PE network where unification, reconciliation, and composition operations produce new binding sets. Of the possibly many binding sets produced, a single set is selected for transmission back to the CP, and the entire cycle is resumed while the other binding sets lie dormant in the PE network awaiting later selection. The operation is shown pictorially in figure 4-1.

4.3 The Unification Phase

The first phase of the LPS algorithm begins with the transmission of a goal list from the CP into the PE network. Residing in each PE is some (possibly empty) collection of facts and rule heads that were placed there when the program was initially loaded into the machine. Once the transmitted goal list has been captured, each PE unifies every goal with as many of its resident literals as possible, producing unifiers which are stored in the PE's local storage.

Unification with a fact produces a simple binding set whose common layer is the constraint set contribution specified by the abstract algorithm for a removed goal. That is, the unifier is stripped of all bindings for variables that were not present in the unified goal, and the resulting substitution becomes the common layer.

Unification with a rule head produces a complex binding set whose common layer is the unifier stripped of its non-goal variable bindings (same as the common layer for a removed goal). The rule layer is the instantiator for the expansion, as specified in the abstract algorithm. In other words, the unifier is stripped of all bindings for non-rule variables, and supplemented with bindings to new created variables for all unbound rule variables.[7] The marker for the rule layer is the key associated with the unifying rule head.

Each binding set produced during the unification phase is tagged with a *level number* which identifies, via its position within the transmitted goal list, the goal whose unification gave rise to the binding set. It will become clear during the discussion of the join phase why this tagging is required.

4.4 The Join Phase

We have named the second phase of our execution loop as the "join phase" due to a useful interpretation of the basic operation as an equi-join over a set of database relations. Indeed, if we recall that each goal in the transmitted goal set gave rise, during the unification phase, to a collection of binding sets with a common level number, we see that the level number provides us with a key to the "relation" defined by the corresponding goal. The database from which the relation was produced is the collection of literals (facts and rule heads) present in the PE network.

With this interpretation in mind, one sees that joining these several relations, using reconciliation as the basic pair-wise matching operation, computes reconciliations for all compatible combinations of unifiers for the goals in the transmitted goal list. At the completion of the join phase, every one of these binding sets will reside in the PE network and will be elegible for later selection and elaboration of the particular proof path it represents. Thus the transmitted goal list can be discarded at that point.

Any matching operation performed on two binding sets will require that the two bindings sets be accessible to the same processor. In general that will not be the case at the completion of the unification phase, since each binding set is stored in the PE containing the unifying literal. The join phase thus requires communication of binding sets around the network. This communication is coordinated by the CP.

[7] Note that variables created by two different PE's must be distinguishable. This is easily done if the PE's can be assigned unique identification tags, as those tags may then be incorporated into the created variable names. Such tags may be assigned at system startup using resolve and report operations. Alternatively, many existing and proposed machines fitting our model can generate unique ID's using various highly efficient methods.

The basic step in the join phase consists of selecting two relations out of the several to be joined and joining those two into a single relation, thus decreasing by one the number of relations to be joined. When only one relation remains, the join phase is complete.

In order to join two relations, one of the two is chosen to "feed into" the other. The CP loops over the feeder relation, extracting one member from the PE network during each iteration. As each element is obtained from the feeder it is broadcast to the entire PE network, and any PE that holds elements from the "consumer" relation attempts to reconcile the *common layer* of the feeder with each of its resident consumers (remember, the common layer is where the constraint set contributions were placed during the unification phase). Whenever reconciliation succeeds, a new binding set is created whose common layer contains the reconciliation. Any rule layer that appeared in either of the contributing binding sets is included in the new binding set, and the level number is set so as to identify the new binding set as belonging to the new joined relation under construction.

Each feeder binding set is discarded as soon as it has been matched against all possible consumers, and when the entire pair-wise join has been completed, the original consumer relation is discarded as well. Thus two relations have been discarded, and one has been produced, bringing us nearer to our goal of a single relation.

4.4.1 A Heuristic For Ordering The Join Phase

In our computing model communication should be held to a minimum since it must all be funneled through a single channel (the CP). Due to the commutative nature of the reconciliation operation, we may exercise a simple heuristic that should, under most circumstances, keep join phase communication close to minimal. Specifically, we always choose the smallest existing relation as the feeder, and the largest relation as the consumer. Cases can easily be constructed in which some other ordering turns out to be preferable, but the heuristic seems reasonable in the absense of methods for predicting the sizes of intermediate join results.

In the general case we choose to implement an approximation to the above heuristic since our computing model does not provide an efficient means of determining the size of a distributed relation.[8] We make use of a sequencing mechanism applied to the relation members. The idea is that within each relation the individual binding sets are assigned unique *sequence numbers* in the hope that the difference between the highest and lowest sequence numbers in a relation will generally be a useful estimate to the size of the relation.

In the current LPS implementation, sequence numbers are assigned during the unification phase according to the order in which the clauses were asserted during program loading. Thus any binding set that is produced by unification with the program's first clause is assigned a sequence number of one. Unification with the program's second clause yields sequence number two, and so on.

The assignment of sequence numbers to join results is analogous to the calculation of storage offsets to multi-dimensioned array elements. The first "dimension" is represented by the sequence number of the contributing binding set from the first relation (level number one), and so forth. The "offset" calculation can be performed efficiently by precomputing (in time linear in the number of relations) a "dope vector" similar to that used by

[8] Note, however, that many architectures fitting our model do in fact allow for fast network-wide sums, making the heuristic viable as presented. We hope to clarify the need for such a mechanism through statistical investigations.

many programming languages for array indexing. All sequence numbers are multiplied (again in linear time) by the dope vector elements corresponding to their level numbers prior to the commencement of the join operation. Then when two binding sets reconcile successfully, the sequence number for the new binding set is the sum of the two contributing sequence numbers.

In addition to their contribution to the join ordering heuristic, sequence numbers provide a method for ensuring a predictable perusal of the proof space by our implementation. Although from the point of view of pure theorem proving such predictability is inessential, under some circumstances such as I/O and recursion, it is crucial if the programming system is to be useful for a more general class of programs, as is the case with Prolog. Unfortunately, the sequence numbers as described here do not appear to provide an ordering that is easily comprehended or well suited for many programming tasks, so that alternatives must still be investigated.

4.4.2 Partition Of The Join Phase

For reasons that will become apparent in the upcoming discussion of variable purging, it may be desirable to impose a global constraint on the join phase ordering so that the relations arising from any single goal list contribution are fully joined among themselves prior to any attempt at combining results from different contributions. We adopt this strategy in the current LPS algorithms by conducting the join phase in two steps. First, a series of *partial joins* takes place in which each goal list contribution is reduced to a single relation in the PE network. When the partial joins have completed, a *final join* joins each of these relations into a single relation representing the successors to the goal list under consideration.

4.5 The Substitution Phase

The last task to be performed upon the discovery of a successful proof is the composition of the various substitutions that were generated along the way. As indicated in the abstract algorithm, these substitutions are the constraint set reconciliations computed to support the individual proof steps. Their composition is computed in the substitution phase of our algorithm.

As was briefly mentioned in the observations following the abstract proof procedure, we have chosen in our current implementation to compute this composition incrementally as the individual components are generated. Thus each time a new reconciliation is produced, we compute its composition with all prior reconciliations in its proof path. Once this has been computed, the individual reconciliation itself can be discarded.

In order to achieve this strategy, we store in the common layer of a binding set, not the individual reconciliation that produced the binding set, but its composition with all prior reconciliations on its proof path. This is easily implemented because all of the binding sets produced by a join phase share a common proof history, and the cumulative substitution representing that history is exactly the substitution stored in the common layer of the complex binding set that gave rise to this proof step in the first place.

In our LPS implementation, then, the substitution phase is accomplished by transmitting the prior reconciliation history to the PE network following the join phase and computing in each PE the composition of that substitution with any new reconciliations.

Three possible benefits derive from our incremental substitution strategy. First, composition computations are performed in parallel in the PE network rather than individually for each reported solution by the CP. Second, debugging is easier because the progress represented by each binding set can be read directly in terms of the original directive variables rather than an obscure collection of created variables. Finally, we avoid a bookkeeping chore in the CP which, depending upon whether certain variants on the basic algorithms are chosen, may be extremely expensive in both time and space.

4.6 Managing Created Variables

In order to keep communication and processing costs to a minimum, it is desirable to discard bindings from our binding sets whenever they are no longer needed. In general the instantiator stored in a rule layer of a binding set will contain a binding for each variable appearing in the rule body, and no other bindings. Thus rule layers are not a problem in this respect. The common layer is more complicated.

In general there are two possible reasons for keeping a binding in the common layer of a binding set:

- The binding will be required in order to construct a solution, should the current proof path succeed.
- The binding might interact with other bindings to constrain the search space, so that discarding the binding could lead to incorrect proofs.

If at any point a particular binding can be determined not to fulfill either of these conditions, we may freely discard the binding and proceed with our proof.

When we report a solution, we limit the report to a display of a minimal substitution that will transform the directive into a satisfiable goal list. In particular, the intermediate goal lists are not displayed, in either their instantiated or uninstantiated form. Recall that our substitution phase is implemented incrementally, so that common layer substitutions always represent the total accumulated current knowledge of the solution being pursued. Thus we see that our first condition demands only that we not discard bindings for variables that appear in our original directive (*top-level variables*).

Other bindings are required for their constraining effects. However, we observe that once a binding has been produced for a variable, it is immediately used to remove all appearances of the variable from the binding set. Aside from this instantiation, the only way a binding can ever act to constrain the search space is through reconciliation with another binding for the same variable. But by the instantiation itself, we are guaranteed never to see the variable in a future goal list along the same proof path, so that no future bindings for it will ever be produced. Thus no further constraint by the variable is possible. We conclude that we need never maintain bindings for a variable (other than a top-level variable) once a binding for it has appeared at the end of a proof cycle.

We do not claim that the binding would not undergo further changes were it to be maintained throughout the remainder of the proof. For instance, if we produce the binding [_1/p(_2)] we may later produce the binding [_2/a]. The overall proof substitution would then include the binding [_1/p(a)]. However, the search constraints that are represented by this refinement are accomplished by the construction and reconciliation of bindings for _2; the refinement of _1's binding is a more or less passive side-effect. Since _1 is not a top-level variable, we have no interest in this side-effect, so there is really no point in producing it in the first place.

We see, then, that when a binding set is reported to the CP from the PE network its common layer should contain bindings only for top-level variables. However, more can be said about the other variables as well. In particular, we recall the join phase partitioning strategy discussed earlier, in which the join phase proceeds by a series of partial joins involving relations produced by common goal list contributions, followed by a final join of the partial join results. It turns out that many bindings can be pruned from the binding sets before the final join takes place, thus saving in communication costs during that join.

Recall that if a rule variable is not bound during unification the resulting instantiator is augmented by binding that variable to a new created variable. The created variable will thus appear in exactly one of the goal list contributions represented by the complete binding set, and hence in exactly one of the partial join result relations. Such a variable cannot constrain the final join, and since it is not a top-level variable, it will be discarded when the final join is complete. We can save communication costs in the final join if we discard the variable prior to the final join.

A list of such discardable variables may be computed easily by the CP during instantiation of a rule body by gathering together term sides of all variable/variable bindings in the instantiators. For example, if the binding [_34/_46] appears in an instantiator, we can safely discard all bindings for variable _46 prior to the ensuing final join.

We note here that if we are to discard bindings before the final join takes place, we must account for the possibility that some of our top-level variables are bound to terms that include discardable variables. Thus the composition operation that constitutes our substitution phase must in fact be performed prior to the final join. We may apply the operation simultaneously to all the relations that will take part in that join by waiting until all the partial joins have completed.

5 Alternative Unification Phase Strategies

We consider two strategies for the unification of goals in a goal list, which we call *asynchronous* and *synchronous unification*.

5.1 Asynchronous Unification

In the asynchronous case, a goal list is broadcast as a single unit to the PE network, and the PE's are instructed to go to work unifying the entire list of goals. The CP waits until all PE's have completed this task, at which point all possible unifications of the goals have taken place, and the resulting binding sets are resident in the PE network. This strategy allows overlapping of goal unification among the individual PE's. That is, each PE moves on to the next goal as soon as it has exhausted its own local supply of literals with which to attempt unification of the current goal, regardless of the state of progress in the other PE's.

As an example, consider the following somewhat idealised scenario:

Goals to be unified: a, b.
Literals resident in PE 1: a_1.
Literals resident in PE 2: b_1.

The following sequence of events results:

1. The CP broadcasts the goal list '(a, b)' to the PE network.
2. PE 1 begins unifying <a, a_1>.
3. At the same time PE 2 begins unifying <a, b_1>, fails quickly and progresses to unify <b, b_1>.
4. PE 1 completes unifying <a, a_1>, attempts to unify <b, a_1> and fails quickly.
5. PE 2 completes unifying <b, b_1>.
6. PE 1 and PE 2 have completed the unification phase.

As we see, unification of goal a in PE 1 is overlapped in time with unification of goal b in PE 2. Beneficial overlapping occurs for two reasons:

- A successful unification generally requires more time than an unsuccessful attempt. Failure is usually detected long before the two literals have been completely scanned (indeed, failure is immediate in the case of different predicate symbols), whereas success is not recognised until the scan is complete. Furthermore, additional work is required after a successful unification, for the construction of a binding set.
- One PE may need to attempt unification of a particular goal with more literals than another PE. If we assume a very small number of literals resident in each PE (due to a large PE network), we can expect that most PE's will be unable to unify most goals, so this will be a high source of overlap. Even without this assumption, various strategies for distributing the literals and indexing each PE's local literal pool by predicate symbol can increase the likelihood of this type of overlap.

Again assuming a very small number of literals resident in each PE, we see that the entire unification phase takes time that is linear in the size of the broadcast goal list. Furthermore, we expect the entire process to be exceptionally fast due to a small constant factor in our linear complexity. Contributing components are: (1) the time required to transmit the goal list, and (2) the time required for the PE's to individually scan the goal list and create binding sets for successful unifiers. The second component is linear because the basic unification algorithm is linear in the size of the terms being matched, and the sum of those sizes is no larger than the size of the entire goal list.

Generally, we would expect a single literal to unify with at most one goal in a goal list, so that the constant factor in our linear time complexity will be heavily dominated by the time for failure, rather than the time for successful unification. This accounts for the high performance we expect from asynchronous unification, since failure time is quite small.

5.2 Synchronous Unification

In the synchronous unification strategy the goals in a goal list are broadcast one at a time rather than as a single unit. Unification of each goal is performed in the PE's before the next goal is broadcast, so that none of the overlapping that we witness in the asynchronous strategy can occur.

The synchronous strategy offers a potential benefit only in the case of the failure of a single goal throughout the entire PE network. In this case, the entire goal list can be thrown out immediately without attempting unification of the remaining goals.

Whether or not such opportunities arise with a frequency that merits adoption of a synchronous unification strategy is a question that will be investigated through statistical analyses of logic

programs (13: Taylor et al. 1984). We hope also to develop methods for identifying local characteristics of a search space that may indicate an increased likelihood for global failure of a single goal. If this can be done, a dynamic selection mechanism may be implemented that is capable of using asynchronous or synchronous unification depending on the proof history and current state.

A hybrid strategy may also be envisioned, in which the goal list is partitioned according to some suitable heuristic, and each portion is broadcast as a unit for asynchronous unification, while unification of the overall goal list is synchronous among the portions.

6 Alternative Join Phase Strategies

We note that our parallel execution of a pair-wise relational join results in a log(n) improvement in the time required for this operation by a sequential algorithm (14: Taylor et al. 1984). Our alternative join strategies investigate methods for minimising the number of pair-wise joins required in a single join phase, avoiding redundant computations, and controlling depletion of PE storage.

The reconciliation operation itself is performed as a series of refinements on a collection of bindings. The collection begins with the union of the bindings found in the two component substitutions. A refinement consists of identifying two bindings for the same variable and replacing one of them with the unifier of the two terms. When no such pair of bindings is left, reconciliation is complete.

Although pathological cases can be constructed, we believe that in practice the unification that takes place during reconciliation seldom produces new bindings for variables already bound, so that (again recalling that unification itself is linear in the size of the terms being combined) we will expect a performance that is roughly linear in the size of the component substitutions. We intend to investigate the validity of our assumption through statistical analysis.

6.1 Retaining Goals

It has been pointed out that our LPS implementation never retains a goal from one goal list to its successor. Instead, each goal is either removed or expanded. We expect this strategy to be quite beneficial in many applications, however there are at least two potential pitfalls, which we call the *big-small* problem and the *cartesian product* problem.

6.1.1 The Big - Small Problem

Consider the goal set {big(X), small(Y)} where, as its name suggests, big(X) is a goal that represents an extremely lengthy computation involving long chains of inference. Likewise, small(Y) is a goal that is very quickly satisfied, with multiple solutions. It turns out that a strict policy of non-retention of goals will perform substantially more work in locating multiple solutions of this directive than would a more flexible approach.

To see this, consider the very first step in the solution of our goal, and suppose that our database contains the two facts small(flea) and small(pebble). These two facts will unify immediately and produce simple bindings [Y/flea] and [Y/pebble], respectively. Meanwhile, big(X) unifies with a rule head somewhere in the database, producing a complex binding that represents a long computation in its infancy.

With these binding sets in place, our join phase will produce two complex bindings, both containing the status of the just-started big computation, and each containing one of the small

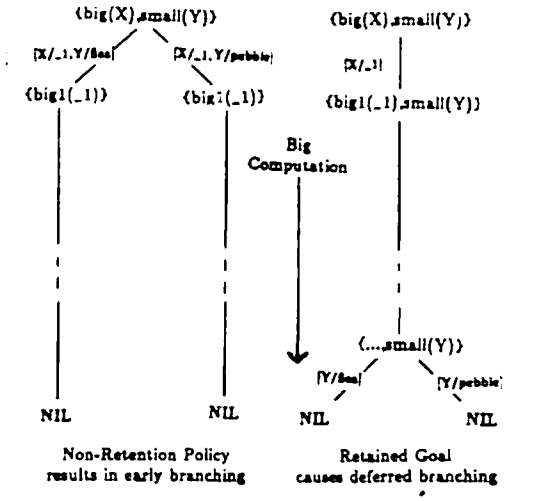

Figure 6-1: Goal retention postpones branching and may result in considerable savings of effort in the big - small problem

solutions. One of these complex bindings will be selected during the next cycle, and that selection will begin the long computation of big(X).

Meanwhile, the second complex binding will lie dormant in the PE network awaiting selection but not benefiting from the ongoing computation. When it is finally selected, the big computation will be repeated almost in its entirety. This is the duplication of effort that we would like to avoid.

To see a possible way out, consider the behavior of this goal in the standard sequential algorithm. Here the big(X) computation is carried out until a solution for it is achieved, all the while retaining the original small(Y) goal in the goal list. It is not until the big computation has terminated in a solution that the small goal is finally unified, resulting in its removal by the fact small(flea). Next the algorithm backs up its computation to its last choice point, which was its choice of a unifying fact for small(Y), and makes a new choice. This time the small goal is removed by the fact small(pebble) *without having to recompute the current solution for* big(X).[9]

Thus we see that by *retaining* the small goal we have avoided a large redundant computation.

It is instructive here to consider the tree of goal lists generated by our proof procedure, in which the decision to retain a goal at a certain point has the effect of postponing whatever branching might be caused by the choice of clauses with which the goal may unify. If each of the resulting branches gives rise to a deep subtree, the postponement may turn out to be quite beneficial. The case of our example is diagrammed in figure 6-1.

6.1.2 Cartesian Products

One other possible benefit of goal retention is containment of the potentially explosive growth in the number of binding sets resident in the PE network. As an example, suppose our goal list is {plentiful1(X), plentiful2(Y)}, where each goal unifies with a very large fact base (say M facts for plentiful1 and N facts for plentiful2). Thus in one proof step we achieve two

[9] Note that the Prolog algorithm would encounter the big-small problem if the goal list were reversed, as in {small(Y),big(X)}

large, independent relations, whose join is their complete cross product consisting of M x N binding sets. In such a situation it might be desirable (or even necessary) to limit the accumulation of binding sets by generating only one "slice" of the cartesian product at a time.

This might be accomplished by retaining plentiful2(Y) during the first cycle. The result will be M binding sets containing the solutions for plentiful1(X), and each containing our retained goal as well. These binding sets are selected one by one for further elaboration, and each one gives rise to N bindings sets that are reported and discarded in turn. The maximum number of binding sets resident in the network is thus M + N - 1, rather than M x N.

6.1.3 Implementation Of Goal Retention

Two problems need to be addressed if goal retention is to be accommodated in our algorithms:

1. The actual mechanisms for retaining goals must be worked into the implementation. This requires a slight modification of the binding set representation so that actual goals can be represented, as well as mechanisms that allow the CP to identify to the PE's which of the broadcast goals are to be retained.

2. The means by which goals are selected for retention must be decided. Possibilities include automatic selection based on static and/or dynamic program analysis; marking of procedures, rules, or even individual condition elements by the programmer; and combinations of these two strategies. We prefer a completely automatic mechanism, consistent with the philosophy that logic programming offers opportunities for parallelism without burdening the programmer with this goal.

6.1.4 Benefits Of A Non-Retention Policy

It should be noted here that a policy of non-retention of goals provides at least two potential benefits.

First, the total path length for any successful proof is minimized by such a policy, generally translating into reduced effort for a single proof. As the big-small problem illustrates, however, situations may easily arise in which much greater benefits due to commonality of proof paths are missed by this eager strategy.

Second, a retained goal does not constrain the search space under consideration. One benefit of the reconciliation model over the depth-first search strategy of the Prolog algorithms is that a much larger range of interactions are possible among the goals in a single goal list. In the Prolog strategy, the effects of computations on a goal may only propagate forward in the goal list, whereas if several goals are unified in one step, constraining interactions are carried in both directions. The program presented below is an example where the Prolog algorithms will never terminate, whereas a non-retention strategy terminates quite quickly. Here the second goal in the goal list constrains the first goal so as to avoid the infinite search that the first goal produces on its own. Since this "backward" constraint is not possible in the Prolog algorithm, we find the unconstrained first goal generating an infinite sequence of results, all but the first of which are disallowed by the second goal.

```
Rule 1: append(cons(A,T1),L,cons(A,T2)) :-
        append(T1,L,T2).
Fact 1: append(NIL,L,L).

Directive: append(X1,X2,Y1), append(Y1,Y2,NIL).
```

6.2 Single Feed Joins

The cartesian product problem mentioned in the last section is just one particularly severe case of the general problem that our parallel execution model may tend to accumulate binding sets that are waiting for selection. Goal retention was seen as one strategy for alleviating this problem by expanding the search space one "slice" at a time.

Another strategy is to perform our join operations in small steps by broadcasting only one feeder relation member to the consumer relation at a time. The binding sets produced by that single feeder are processed one by one until they run out, at which point the next feeder from the suspended join is broadcast.

This single-feed strategy offers a second possible benefit aside from containment of the binding set population. In many cases, a query will be presented with the intention of producing only a single solution, rather than pursuing all possible solutions. In this case, much of the effort that goes into our join operations will be wasted since if a solution is encountered early in the search space, a large percentage of the binding sets generated from joins will be discarded. The single-feed strategy defers this effort until it is required in order to continue the search.

It is expected that the implementation of a single feed strategy will require considerably more complicated control mechanisms than are needed for the eager join strategy. At this point in time no such implementation has been attempted, nor has careful thought been given as to the exact control mechanisms that would be required.

6.3 Redistribution of Binding Sets

One final approach to the problem of explosive growth in the binding set population takes a more local view. Specifically, what can be done about the case where binding sets begin accumulating at a few "hot spots" in the PE network?

In such a situation it would be beneficial to have a mechanism available whereby heavily loaded PE's could export some of their binding sets to other PE's. Such a mechanism is difficult to imagine in our computing model since all communication must be funneled through the CP. If, however, some direct PE-PE communication mechanism is provided[10] efficient redistribution might be realisable. We may even envision redistribution within the PE network overlapping computational tasks within the CP, such as the construction of a new goal list from a reported binding set.

6.4 Multiple Independent Joins

A particularly intriguing prospect for optimization of the join phase is the idea of performing two or more pair-wise joins in unison in the PE network. Our standard join algorithms may be adapted for this purpose by extending our model of computation to include a facility for temporarily partitioning the PE network into independently functioning subnetworks. One PE in each subnetwork would act as CP for the subnetwork.[11] With such a facility, our pair-wise join may be migrated to the subnetworks, so that several pairs of relations may be joined simultaneously. This strategy requires that the each subnetwork contain each relation to be joined, in its totality.

---

[10] Such a mechanism is available, for example, in the DADO binary tree architecture, in which tree neighbors may communicate without burdening the CP (Stolfo and Shaw 1982, Stolfo et al. 1983)

[11] The DADO architecture (Stolfo and Shaw 1982, Stolfo et al 1983), for example, allows for such a "multiple SIMD" execution mode

As an example of how a multiple join strategy might be realized, and to illustrate the potential savings, we consider a rather "brute force" approach. The PE network is divided into two subnetworks, and each fact and rule head is stored once in each subnetwork.

The unification phase will produce twice as many binding sets as in our standard model, each binding set appearing in both subnetworks. We note that the effort expended during the unification will be doubled in worst case, since the concentration of literals in the PE's has doubled so that each PE's scan of literals will take twice as long.

The join phase proceeds in two stages. During the first stage, one of the PE subnetworks joins half of the relations resulting from the unification phase while the other subnetwork simultaneously joins the other relations. The second phase is a single pair-wise join performed by the CP in the standard fashion, combining the results of the subnetwork joins. The total effort required by the join phase starting with n relations is that required for $n/2$ pair-wise joins, as compared with $n-1$ pair-wise joins if multiple independent joins are not utilised.

If we consider the overall savings realised by multiple joins in the above scenario we see that while the time for unification has been doubled in worst case, we have halved the time required in the join phase. In the case of communication costs, we see that there has been no increase during unification, whereas costs have been halved during the join phase. Further analysis may be able to identify more intelligent partitioning strategies, possibly based on data dependency analyses similar to those under investigation by Ishida (Ishida 1984) in his work on parallel execution of production systems.

6.5 Rule Layer Caching

We discuss one other join phase variant in which rule layers are stripped from binding sets whenever they pass through the CP and are replaced-by unique tags. The rule layers are stored in the CP and are retrievable via their tags. The advantage of this scheme is reduced communication of feeder binding sets during the pair-wise join operation. The strategy is justifiable on the basis that no use is made of rule layers except in the CP, so that they are little more than "excess baggage" in the binding sets during the join phase.

One major drawback of this scheme, however, is that it precludes the removal of common layer bindings from binding sets during the join. This is impossible because any common layer binding might be needed in order to update the instantiators in the binding set, and instantiators that are start out identical may in this way end up differing in their final form. In order to ensure correct updating of instantiators before instantiation, then, the common layer bindings must be fully maintained and reported to the CP along with the binding set.

In addition, such a scheme would probably require some method for determining when a rule layer may be discarded by the CP owing to all referencing binding sets having been reported and elaborated. Such a mechanism seems feasible given the current sequence number scheme.

The trade-offs involved have not yet been studied, although there seems to be reason to suspect that overall communication costs will not be greatly affected, the two effects largely cancelling each other.

7 Alternative Substitution Phase Strategies

The only major alternative under consideration for the substitution phase is the postponement of the composition of individual reconciliations in the proof path. Rather than keeping a completely updated reconciliation in each binding set, common layers would represent only the substitution required to complete the last proof step. The overall substitution would be computed by the CP whenever a solution was encountered.

The only substantial benefit that may be obtained from this strategy is that the entire substitution history, along with a history of goal sets that could also be maintained by the CP, would allow the reconstruction of the entire proof for reporting purposes. The drawbacks are several:

- The history mechanism required in the CP appears substantially more complicated than what is presently required. Binding sets would need additional tagging information to identify depth in the search space, and the CP's history mechanism would have to monitor this information in order to know whether to stack a new component, replace the top component, or pop the stack.
- The history mechanism would seem to prevent much flexibility in the order of selection of binding sets. A predictable order of traversal through the search space is potentially beneficial to programmers. The history mechanism would fit well into the ordering imposed by our current sequence number scheme, but as indicated earlier, it is questionable whether this ordering is useful. We hope to be able to identify a different ordering that fits well into the algorithms, but a history mechanism would severely constrain our options.
- We would no longer be able to remove common layer bindings prior to reporting the binding set to the CP.

7.1 A Previous Implementation

Earlier published work on LPS described a substitution phase that is substantially different from those currently under consideration. In fact, in early implementations the substitution phase was probably the most complex phase of the algorithm. The current approach is a direct result of investigations prompted by discontent with the earlier techniques. For historical completeness we briefly discuss this approach and relate it to current work.

The task of the substitution phase can be regarded as pushing forward a *frontier set* of bindings. Prior to a proof cycle, we are equipped with a collection of bindings that relate our top-level variables to variables in the rules about to be fired, as well as various created variables. We call these variables *middle-level* variables. As a result of unification and reconciliation, we are left with another collection of bindings, this time between middle-level variables and *bottom-level* variables, which are variables from the facts and rule heads with which our goals unified. The substitution phase must resolve these two collections into a new collection of bindings relating the top-level variables to the bottom-level variables. During the next proof step, of course, those bottom-level variables play the role of the middle-level variables, and the frontier set is advanced one more level.

The innovation that has allowed us to discard our old substitution algorithm is the filling out of instantiators with "dummy bindings" for unbound rule variables. As a result of this operation, our new frontier set and instantiator fall directly out of the composition procedure. Previously our approach was as follows:

1. Classify bindings into five different categories, as follows:

    - Upper level variable bound to lower level variable
    - Upper level variable bound to lower level ground term
    - Upper level variable bound to lower level non-ground term
    - Lower level variable bound to upper level ground term
    - Lower level variable bound to upper level non-ground term

2. List all possible combinations of a top-to-middle binding of one type and a middle-to-bottom binding of another type. The resulting set of twenty-five binding scenarios, along with the five cases where a top-to-middle binding is left by itself (unpaired with a middle-to-bottom binding) comprise all possible binding interactions.

3. Consider each binding interaction in turn and decide how it can be recognised and what contributions it can make to the resulting binding set.

4. Develop an algorithm to handle interactions according to the analysis just performed.

We do not intend to consider this approach further.

8 Conclusions and Future Work

It has not yet been established that the pilot algorithms presented in this paper can result in efficient interpreters for the execution of logic programs under the parallel computing model that we propose. A limited form of OR parallelism is achieved through simultaneous unification of individual goals with literals that are distributed over a large multiprocessor network, and a limited form of AND parallelism is achieved by satisfying an entire list of goals in a single algorithm cycle.

Our abstract proof procedure has provided a convenient basis for the specification and analysis of several alternative execution strategies. Although we have been able to identify some tradeoffs, it is apparent that no single choice of strategies will be optimal in all circumstances. Future research aims to further our understanding of these and other algorithms and to identify characteristics of logic programs that may be used as a criterion for strategy selection.

We are currently planning an implementation of a LPS interpreter on a prototype machine based on the DADO parallel architecture. One such prototype comprising fifteen PE's is currently functioning; a 1023-node prototype is under construction. Weisberg and Lerner are working on an implementation of a parallel version of Portable Standard Lisp for the DADO machine (Weisberg et al. 1984). As our simulation software was written in PSL, we expect that this effort will substantially simplify our implementation task by allowing a simple recompilation of large portions of the existing code for execution on the actual machine.

Taylor (13: Taylor et al. 1984) describes various methods currently under development for statistical analysis of logic programs. These include static, dynamic, and data-flow analyses intended to guide algorithmic decisions in the implementation of LPS. It is hoped that these analyses will quantify the potential for parallel execution, allow accurate performance estimates to be made, and isolate various qualities of logic programs which can be used in building intelligent compilers and interpreters.

Many features must be added to the LPS language in order to make it suitable for a wide range of applications. We intend to investigate such features as negated condition elements in rules, evaluable predicates, and condition elements with side effects. Khabaza's work (Khabaza 1984) appears promising as a basis for the implementation of negation as failure in the LPS framework. In addition, we will explore issues relating to control of program execution, including a more useful ordering of the solution set.

References

1. Bowen, D.L., Byrd, L., Pereira, F.C.N., Pereira, L.M., and Warren, D.H.D. *DECsystem-10 Prolog User's Manual.* University of Edinburgh, Dept of Artificial Intelligence, 1982.

2. Conery, J.S. *The AND/OR Process Model for Parallel Interpretation of Logic Programs.* Ph.D. Thesis, University of California Irvine, June 1983.

3. Ishida, T., Stolfo, S.J. Simultaneous Firing of Production Rules on Tree Structured Machines. Columbia University, New York, NY 10027, March, 1984.

4. Khabaza, T. Negation As Failure And Parallelism. 1984 International Symposium On Logic Programming, IEEE Computer Society, Technical Committee on Computer Languages, Atlantic City, February, 1984, pp. 70-75.

5. Kowalski, R. *Artificial Intelligence Series.* Volume 7: *Logic for Problem Solving.* North Holland, New York, 1979.

6. Pollard, G.H. *Parallel Execution of Horn Clause Programs.* Ph.D. Thesis, Department of Computing, Imperial College, 1981.

7. Robinson, J.A. "A Machine-Oriented Logic Based on the Resolution Principle." *Journal of the ACM Vol. 12*(1965), 23-44.

8. Robinson, J.A. *Logic: Form and Function.* Edinburgh University Press, 1979.

9. Shapiro, E.Y. *ACM Distinguished Dissertations Series. Algorithmic Program Debugging.* The MIT Press, Cambridge, MA, 1982.

10. Stolfo, S.J. and Shaw, D.E. "DADO: A Tree-Structured Machine Architecture for Production Systems." *Proceedings of the National Conference on Artificial Intelligence Vol. 1* (August 1982).

11. Stolfo, S.J., Miranker, D., Shaw, D.E. Architecture and Applications of DADO, A Large-Scale Parallel Computer for Artificial Intelligence. Proceedings of the Eighth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence, Inc., Karlsruhe, West Germany, August, 1983, pp. 850-854.

12. Taylor, S., Lowry, A., Maguire, G.Q.Jr., Stolfo, S.J. Logic Programming using Parallel Associative Operations. 1984 International Symposium on Logic Programming, Atlantic City, February, 1984, pp. 58-68.

13. Taylor, S., Lowry, A., Maguire, G.Q.Jr., Stolfo, S.J. Analyzing Prolog Programs. Columbia University, New York, NY 10027, March, 1984.

14. Taylor, S., Maio, C., Stolfo, S.J., Shaw, D.E. Prolog On The DADO Machine: A Parallel System for High-Speed Logic Programming. Third Annual Phoenix Conference On Computers And Communications, IEEE, March, 1984.

15. Warren, D.H.D. Implementing Prolog - Compiling Predicate Logic Programs. Tech. Rept. D.A.I 32/40, Department of Artificial Intelligence, Edinburgh University, May, 1977.

16. Weisberg, M.K., Lerner, M.D., Maguire, G.Q.Jr., Stolfo, S.J. ||PSL: A Parallel Lisp for the DADO Machine. Columbia University, New York, NY 10027, February, 1984.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION COLUMBIA UNIVERSITY 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION NAVELEX
6c. ADDRESS (City, State, and ZIP Code) 450 Computer Science Building Columbia University New York, NY 10027	7b. ADDRESS (City, State, and ZIP Code) 2511 Jefferson Davis Highway Arlington, VA 22202
8a. NAME OF FUNDING / SPONSORING ORGANIZATION DARPA 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) 1400 Wilson Boulevard Arlington, VA 22209	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. N00039-84-C-0165 / TASK NO. 2 / WORK UNIT ACCESSION NO
11. TITLE (Include Security Classification) LPS Algorithms	
12. PERSONAL AUTHOR(S) Lowry, A., S. Taylor and S. J. Stolfo	
13a. TYPE OF REPORT SPECIAL 13b. TIME COVERED FROM 6/84 TO 9/84	14. DATE OF REPORT (Year, Month, Day) 1984, October 15 15. PAGE COUNT 13
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB-GROUP	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) DADO, Production Systems, Parallel Computer, Fifth Generation, Logic Programming, AI, LISP.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

LPS is a Logic Programming System currently under development and specifically targeted for implementation on massively parallel architectures. We present a detailed explanation of algorithms under development for parallel execution of LPS programs. The explanation is significantly more detailed than those published previously. An abstract proof procedure is developed which encompasses these algorithms and several variants as well as the standard sequential Prolog algorithm. This abstract procedure provides a conceptual basis for our discussion and for a critical analysis of various execution strategies.

The algorithms have been successfully implemented and demonstrated in simulation on a number of small programs. Work is currently underway to transfer this implementation to a working prototype machine based on the DADO parallel architecture.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Salvatore J. Stolfo	22b. TELEPHONE (Include Area Code) (212) 280-8111 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE