PerlDSM: A Distributed Shared Memory System for Perl

Norman Matloff
University of California, Davis
matloff@cs.ucdavis.edu

Proceedings of PDPTA '02
Updated July 19, 2002

Abstract

A suite of Perl utilities is presented for enabling Perl language programming in a distributed shared memory (DSM) setting. The suite allows clean, simple parallelization of Perl scripts running on multiple network nodes. A programming interface like that of the page-based approach to DSM is achieved via a novel use of Perl's tie() function.

Keywords and phrases: Perl; parallel processing; distributed shared memory.

1 Introduction

The Perl scripting language has achieved tremendous popularity, allowing rapid development of powerful software tools. It has evolved in reflection of general trends in the software world, one of these being concurrency, with recent versions of Perl featuring multithreading capabilities. Our interest here will be in enabling Perl to perform more general parallel processing functions, with the concurrency now being across multiple machines connected via a network. A typical setting is that of a subdividable, I/O-intensive task being performed by a network of workstations (NOWs).

For example, we may have a Web server written in Perl[1] which we would like to parallelize into a *server pool*. Here a "manager" server would operate a task farm of incoming requests, and "worker servers" at other network nodes would obtain work from the manager.

Or, we may have a Web search engine written in Perl, which we wish to parallelize over various network nodes, with each node searching a different portion of Web space.

Perl applications of this type have been written, but with the parallelization operations being handled on an *ad hoc basis* in each case, i.e. without drawing upon a general package of parallelization utilities.

One such package has been developed, PVM-Perl [7]. This enables the Perl programmer to do message-passing on a network of machines. However, many in the parallel processing community believe that the shared-memory paradigm allows for greater programming clarity [8]. Accordingly, a number of software packages have been developed which provide *software distributed shared memory* (DSM) [6], which provides the programmer with the illusion of shared memory even though the platform consists of entirely separate machines (and thus separate memories) connected via a network.

Due to the clarity and ease of programming of the shared-memory model, our focus here is on DSM systems. These have been designed mainly for C/C++, and before now, DSM had not been developed for Perl. This paper presents such a system, PerlDSM. This set of Perl utilities enables the Perl programmer to do shared-memory paradigm programming on a collection of networked machines.

DSM itself is divided into two main approaches, the *page-based* and *object-based* varieties. The page-

[1] Or possibly with the server consisting of a Perl driver which runs C/C++ modules.

based approach is generally considered clearer and easier to program in. However, its implementation for C/C++, e.g. in Treadmarks [1], requires direct communication with the operating system. This kind of solution does not apply to interpreted languages like Perl. At first glance, therefore, it would appear that implementation of a page-based DSM in Perl would necessitate modification of Perl's internal mechanisms.[2]

Yet we have been able to design PerlDSM in a paged-based manner (i.e. giving the programmer an interface like those of paged-based systems) without resorting to modification of Perl internals. This is due to a novel use of Perl's tie() construct.

The details of the PerlDSM system and its implementation are presented in Sections 3 and 4, followed by an example in Section 5. In Section 6, we discuss other issues, such as planned extensions.

2 Page- vs. Object-Based DSM Systems

In a page-based DSM system, C code which operates on shared variables, say x and y, might look something like this:

```
shared int x,y;
...
x = 21;
y = x + 5;
```

First x and y must be set up as shared variables.[3] Then x and y are accessed in the normal C fashion; all the operations for sharing with other network nodes — e.g. fetching the latest copy of x from other nodes — are done behind the scenes, transparent to the programmer.

By contrast, in object-based systems, the code

```
x = 21;
y = x + 5;
```

might look like this:

```
x = 21;
put_shared("x",x);
x = get_shared("x");
y = x + 5;
put_shared("y",y);
```

Here the programmer must bear the burden of explicitly invoking the sharing operations via calls to the DSM system.

There are advantages and disadvantages to each of the two approaches to DSM [1]. The object-based DSM approach has certain performance advantages, but the page-based approach is generally considered clearer and more natural, as can be seen in the comparison above. Thus our goal was to develop some kind of page-based DSM for Perl.[4]

3 Implementation of a Page-Like Programming Interface in PerlDSM

The page-based approach relies on UNIX system calls which allow modification (actually replacement) of the page-fault handler in the machine's virtual memory system.[5] Only the pages currently owned by the given network node are marked as resident, and when the program reaches a page it does not currently own,

[2] Perl actually is designed to facilitate working at the internals level, but still this would not be a simple project.

[3] This is often accomplished by calls to a shared version of malloc().

[4] An alternate world view which shares some characteristics with the shared-memory paradigm is that of *tuple spaces*. This was originally proposed in the Linda language [4], and currently in use in JavaSpaces [5] and T-Spaces [9]. Perl versions of this have been proposed. However, these are very similar in terms of programmer interface to object-oriented DSM; the programmer must insert function calls to get tuples from, and put tuples into, the tuple space. We thus did not pursue the tuples approach.

[5] Hence the term "page" in "page-based DSM."

a page fault occurs. The DSM has set up the new handler to get the new page from whichever node currently owns the page, rather than from disk as usual.

This means that direct implementation of page-based DSM in Perl, however, would not be possible, as Perl is an interpreted language. True, the data for the Perl program is also data for the Perl interpreter, but reconciliation of all the correspondences would be very difficult.

At first glance, then, it would seem that implementation of a page-like DSM for Perl would require working at the level of Perl internals. Actually, Perl does expose quite a bit of this to the application programmer, but still such an implementation would be rather complex.

However, it turns out that an interesting feature of Perl, the tie() function, can be used to form an extremely elegant solution to this problem. This function associates a scalar variable (or array or hash) with an object.[6] We say that the scalar is *tied* to the object.

The object is required to have three methods, named TIESCALAR (TIEARRAY or TIEHASH in the case of array or hash variables), FETCH and STORE. In Perl, each time a scalar variable appears on the right-hand side of an assignment statement (or in some other context in which its value needs to be fetched), the FETCH method is called and its return value used in place of the scalar variable. Similarly, if the scalar appears on the left-hand side of an assignment, the value on the right-hand side is passed to STORE.

To see how PerlDSM uses tie(), consider our example with x and y above,

```
shared int x,y;
...
x = 21;
y = x + 5;
```

[6]Here the word "object" is used in the sense of "object-oriented programming," not in the sense of "object-based DSM."

In PerlDSM, this is written as[7]

```
tie $x,'DSMScalar','$x',$SvrSkt
tie $y,'DSMScalar','$y',$SvrSkt
...
$x = 21;
$y = $x + 5;
```

(In Perl, all scalar variable names begin with a dollar sign.)

The two **tie** statements do the association of scalars to objects as mentioned above. DSMScalar is the name of the PerlDSM built-in class for shared scalars, and $SvrSkt is a socket to the PerlDSM variable server.

Then, just as the actual executable code in the C/C++ version,

```
x = 21;
y = x + 5;
```

is written with no distracting and burdensome calls for network fetch and store operations, the same is true for PerlDSM. In other words, the interface afforded the programmer by PerlDSM is similar to those of page-based DSMs for C/C++.[8]

One of the network nodes runs the PerlDSM shared variable server, DSMSvr.pl.[9] The values of the shared variables are maintained by the server.

The call to tie() for, say $x, triggers a call to TIESCALAR. PerlDSM has TIESCALAR check in with the server, registering this shared variable. The server then adds an entry "$x" to a hash, @SharedVariables. In Perl, a hash is an associative array, indexed by character strings; @SharedVariables{"$x"} will contain the value of $x.

The statement

[7]The syntax has been changed slightly since this paper appeared.

[8]Even though there are no "pages" in PerlDSM.

[9]That same node could also be running the PerlDSM application program.

```
$x = 21;
```

triggers a call to STORE, with argument 21. STORE then sends a message

```
write $x 21
```

to the server, which places 21 into @SharedVariables{"$x"}.

Similarly, the appearance of $x on the right-hand side of

```
$y = $x + 5;
```

will trigger a call to FETCH, which will get the current value of $x from the server.[10] Finally, the appearance of $y on the left-hand side of the assignment triggers a call to STORE, etc.

Both FETCH and STORE in DSMScalar.pm perform the necessary socket communications with the server. But again, it must be emphasized that all of this is transparent to the programmer. The programmer does not see the calls to FETCH and STORE, and thus can concentrate on writing clear, natural code.

4 Other PerlDSM Concurrency Constructs

PerlDSM also includes calls to lock and unlock lock variables, calls to wait for or set threads-like condition variables, and a two-phase barrier call.

Locks, condition variables and barriers are implemented internally as reads and writes. A PerlDSM application performs a lock operation as a read of the lock variable $LOCK[$I], and the unlock is done by a write, e.g.

[10] PerlDSM is also extensible, as discussed later, so that some optimization could be done here to prevent an extra trip to the server.

```
$Lock = $LOCK[$I];
... (critical section here)
$LOCK[$I] = 0;
```

Similarly, a statement in which a condition variable appears on the right-hand side of an assignment will cause the node to block (i.e. execute a "wait") until some other node executes an assignment statement in which that condition variable appears on the left-hand side (i.e. that other node executes a signal).

Barriers are invoked in the same manner, e.g.

```
$Barr = $BARR;
```

Locks, condition variables and the barrier variable must be tied by the application program. Note also that these operations are of course blocking, even though they appear syntactically as reads.

5 Example

Following is a sample complete PerlDSM application, which finds the average load average among all nodes, by running the UNIX **w** command at each node and then averaging over all nodes. It is very simple, for the sake of brevity, but illustrates all the PerlDSM constructs.

```
#!/usr/bin/perl

# example PerlDSM application; finds the
# average load average among all nodes

# "use" is like C's "#include"
use DSMScalar;
use DSMUtils;

package main;

# globals:
$SvrSkt;  # socket to server
```

```
$NumNodes;  # total number of nodes
$MyNode;  # number of this node
$SumAvgs;  # sum of all the load averages

# check in with server
($SvrSkt,$NumNodes,$MyNode) =
   DSMUtils::DSMCheckIn();
print "total of ", $NumNodes,
   " nodes, of which I am number ",
   $MyNode, "\n";

# tie shared variables
tie $BARR,'DSMScalar','$BARR',$SvrSkt;
tie $SumAvgs,'DSMScalar','$SumAvgs',$SvrSkt;
tie $LOCK[0],'DSMScalar','$LOCK[0]',$SvrSkt;

# initialize sum to 0 if I am node 1
if ($MyNode == 1)  {
   $SumAvgs = 0.0;
}

# barrier
$Barr = $BARR;

# get load average at this node, by running
# UNIX w command, and parsing the output
system 'w > tmpout';
open TMP,"tmpout";
$Line = <TMP>;
close TMP;
@Tokens = split(" ",$Line);
$MyAvg = $Tokens[9];

# add to total of all load averages
$Lock = $LOCK[0];  # lock
$SumAvgs = $SumAvgs + $MyAvg;
$LOCK[0] = 0;  # unlock

# wait for everyone to finish,
# then print answer if I am node 1
$Barr = $BARR;
if ($MyNode == 1)  {
   print $SumAvgs/$NumNodes, "\n";
}

DSMUtils::DSMCloseSocket($SvrSkt);
exit;
```

6 Efficiency

One question which arises is that of efficiency. It is somewhat less of an issue in the context of an interpreted scripting language like Perl than for C/C++. Parallelizable applications of Perl tend to be I/O-intensive rather than compute-bound, so computational efficiency is less of a concern. Nevertheless, the results of a small timing experiment are presented here.

We wrote a primitive "parallel make" program using PerlDSM, using it to compile the **bash** UNIX shell [2]. The section we timed consisted of compilation of 43 source files. Run time was measured according to the last PerlDSM node to finish.[11] The nodes shared a common file system. Results were as follows (quoted values are the medians from several runs):

nodes	time
1	60.9
2	34.9
3	24.6
4	19.3

Here "1 node" means that the compilation was done entirely sequentially, outside of PerlDSM.

7 Summary and Future Work

In developing PerlDSM, we have attained our goal of enabling Perl programming with a parallel DSM world view, with the simplicity and clarity of page-based C/C++ DSM systems. We have made PerlDSM available at `http://heather.cs.ucdavis.edu/~matloff/perldsm.html`.

As mentioned earlier, we anticipate that the main usage of PerlDSM will arise from a desire to use Perl in a parallel but convenient manner, in settings in

[11] We observed a considerable amount of load imbalance, but made no attempt at reducing it in this simple experiment.

which speed is an issue but not worth efforts at optimization. In our parallel make example presented earlier, for instance, we might have gotten a similar performance improvement over the standard make by using Perl's fork() call to set up processes on a single node. Yet arguably PerlDSM offered a simpler, more convenient way to accomplish the same thing. Threads are currently experimental in Perl, and again, we would argue that PerlDSM allows for easier coding in this particular case.

Moreover, there are a number of distributed applications whose most natural coding is that of the shared memory paradigm. PerlDSM now gives Perl programmers the ability to use Perl in such settings.

Nevertheless, there are certainly ways in which PerlDSM can be made to run faster for applications in which the additional speed is desired. For example, this first implementation uses straightforward Sequential Consistency. More efficient consistency protocols, say Release Consistency or Scope Consistency, could be implemented rather simply, again due to Perl's marvelous tie() function.

Another source of speedup would be to allow the PerlDSM application programmer to write his/her own operations. Consider for example an atomic increment operation, say on a variable $n:

```
tie $n,'DSMScalar','$n',$SvrSkt;
...
$Lock = $LOCK[0];
$m = $n;
$n = $m + 1;
$LOCK[0] = 0;
```

This requires sending a total of four requests to the server, i.e. eight network communications. Yet it would be easy to implement such an operation within the PerlDSM infrastructure, entailing only one request to the server and thus only two network communications.

A more complex performance enhancement would be to allow multiple servers. A number of other enhancements are being considered.

It is remarkable that Perl's tie() function allowed us to implement DSM in such an simple, elegant manner, enabling us to avoid working at the level of Perl internals. Indeed, Rice University's team, in developing a DSM package for another interpreted language, Java/DSM [10], needed to resort to modification of the Java Virtual Machine interpreter. This raises the question of whether our approach here could be extended to other languages.

In C++, for example, one could overload the operators for assignment, addition, and so on. For example, one could create a class named DSMInt. Code like

```
DSMInt x,y,z;
...
x = y + z;
```

would do the following, say in the context of a single server as we have in PerlDSM: Invocation of the '+' operator would result in a TCP/IP message being sent to the server, to fetch the latest values of the two variables 'y' and 'z', and add them. Invocation of the '=' operator would then result in another TCP/IP message, sending the sum to the server to store as the current value of the variable 'x'. Since operator overloading is allowed to be type-dependent, there would be no problem with accommmodating both the above statement and also

```
x = y + 5;
```

since '+' could have a version in which the second addend is a scalar. The author is currently working on this project.

The Python scripting language has a feature which appears to be analogous to most, though not all, aspects of Perl's tie(). Python's __getitem()__ and __setitem()__ allow the application programmer to intercept array and hash references, thus enabling the development of "PythonDSM" in the case of arrays and hashes (which arguably allow scalars as special, one-element-array cases).

References

[1] C. Amza *et al.* Treadmarks: Shared Memory Computing on Networks of Workstations, *IEEE Computer*, January 1995.

[2] Bash source code, `ftp://prep.ai.mit.edu/pub/gnu`.

[3] M. Bouchard. MetaRuby. `http://www.ruby-lang.org/en/raa-list.rhtml?name=MetaRuby`.

[4] Nicholas Carriero and David Gelernter. *How to Write Parallel Programs: a First Course*, MIT Press, 1990.

[5] Eric Freeman *et al. JavaSpaces(TM) Principles, Patterns and Practice*, Addison-Wesley, 1999.

[6] Peter Kelleher. Distributed Shared Memory Home Pages, `http://www.csm.ornl.gov/pvm/perl-pvm.html`.

[7] Edward Walker. PVM-Perl (software package), `http://www.csm.ornl.gov/pvm/perl-pvm.html`.

[8] Barry Wilkinson and Michael Allen. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*, Prentice Hall, 1999.

[9] P. Wyckoff. T Spaces. *IBM Systems Journal*, November 3, 1998.

[10] Weimin Yu and Alan Cox. Java/DSM: a Platform for Heterogeneous Computing, *Proceedings of the ACM 1997 Workshop on Java for Science and Engineering Computation*, June 1997.