Libckpt: Transparent Checkpointing under Unix

James S. Plank
Micah Beck
Gerry Kingsley

Department of Computer Science
University of Tennessee
Knoxville, TN 37996
[plank,beck,kingsley]@cs.utk.edu

Kai Li

Department of Computer Science
Princeton University
Princeton, NJ 08544
li@cs.princeton.edu

Appearing in:
USENIX Winter 1995 Technical Conference
New Orleans, Louisiana
January 16-20, 1995

**Error Correction:**

In the published Usenix paper, two columns of Table 2 were swapped. These were the columns marked “Avg Ckpt Time” and “Avg Ckp Size.” This error has been corrected in this copy of the paper.
Libckpt: Transparent Checkpointing under Unix

James S. Plank, Micah Beck, Gerry Kingsley
University of Tennessee

Kai Li
Princeton University

Abstract

Checkpointing is a simple technique for rollback recovery: the state of an executing program is periodically saved to a disk file from which it can be recovered after a failure. While recent research has developed a collection of powerful techniques for minimizing the overhead of writing checkpoint files, checkpointing remains unavailable to most application developers. In this paper we describe libckpt, a portable checkpointing tool for Unix that implements all applicable performance optimizations which are reported in the literature. While libckpt can be used in a mode which is almost totally transparent to the programmer, it also supports the incorporation of user directives into the creation of checkpoints. This user-directed checkpointing is an innovation which is unique to our work.

1 Introduction

Consider a programmer who has developed an application which will take a long time to execute, say five days. Two days into the computation, the processor on which the application is running fails. If the programmer has not planned for this event, his only choice is to restart the program and lose two days of work. Upon restarting the program, he still needs five days of continuous failure-free processor time to complete the job.

Libckpt is a checkpointing library designed for such a programmer. To use libckpt, all he must do is change one line of his source code and recompile with the library libckpt.a. No other modifications need to be made. Upon execution, the program will periodically save its execution state to disk (at the default interval: every 10 minutes). Upon a processor failure, the programmer need only restart the program with the command line flag =recover, and the program will roll back to the most recently checkpointed state. In the example above, at most ten minutes of work will be lost, and three more days of non-continuous failure-free processor time will be needed to complete the job.

Libckpt is a tool for transparent checkpointing on uniprocessors running Unix. It implements incremental and copy-on-write checkpointing, two optimizations well-known in the literature [2, 3, 9, 10]. Libckpt is a user-level library and uses only facilities which are commonly available under Unix. Libckpt has been ported to and tested on a variety of architectures and operating systems with no kernel modifications. Source code for libckpt can be obtained at no cost by anonymous FTP from cs.utk.edu:/pub/plank/libckpt.

In this paper, we show the performance gains available in libckpt through transparent incremental and copy-on-write checkpointing. In addition, we introduce a new optimization technique, implemented in libckpt, called user-directed checkpointing. User-directed checkpointing works under the assumption that a little information from the user can yield large improvements in the performance of checkpointing. We demonstrate that this assumption is often valid.

2 Transparent Checkpointing

The goal of checkpointing is to establish a recovery point in the execution of a program, and to save enough state to restore the program to this recovery point in the event of a failure. The most straightforward method for establishing a recovery point under Unix is to suspend execution of the application while the entire contents of a process's memory and registers are written to a file. This is called sequential checkpointing because disk trans-
ners are not interleaved with program execution. Recovery is effected by reloading the executable from its original file, and then reconstructing the memory and register state from the checkpoint file. This is akin to creating a core file, from which a user may recover using the undump utility and execute().

We say that checkpointing is transparent when no changes need to be made to the application program. While transparency is easy to obtain at the kernel level, it is harder to achieve in a user level checkpointing library. All current implementations of checkpointing share this limitation: They operate transparently and correctly so long as the application is well-behaved in a sense we will define in Section 4. Libckpt and all other user-level checkpoints can cause a correct but ill-behaved application to fail or to produce incorrect output upon recovery.

Checkpointing with libckpt is not completely transparent. The name of the initial procedure in C must be changed from main() to ckpt_target(). This enables libckpt to gain control of the program as it starts, check the command line for the recover flag, read a file called .ckptrc to set checkpointing parameters, and begin checkpointing. In FORTRAN, libckpt is enabled by changing the main program module to SUBROUTINE ckpt_target(). No other program modifications are needed.

By default, once libckpt gets control of a program, it generates a timer interrupt every ten minutes, and takes a sequential checkpoint at each interrupt. This and other defaults can be changed by placing appropriate lines in the .ckptrc file. In this section, we describe all options, where appropriate, as they would appear in the .ckptrc file.

Placing the line “checkpointing <on|off>” in the .ckptrc file turns checkpointing on or off. If off, libckpt will take no checkpoints and will not affect the execution of the application. The default is on.

dir <directory> specifies the directory in which checkpoint files are created and found. The default is the current directory.

maxtime <seconds> defines the interval between checkpoints. At the beginning of the program, and after each checkpoint, libckpt calls alarm(seconds) and takes a checkpoint upon catching each ALRM signal. Setting the timer interval to zero turns off all timer-based checkpointing. The default value of maxtime is 600 (10 minutes).

Many optimizations to simple sequential checkpointing have been described in the literature. Libckpt implements all published optimizations that are applicable to general-purpose uniprocessor checkpointing, as well as the new user-directed optimization. In the remainder of this section, we consider each of them in turn.

2.1 Incremental Checkpointing

When a checkpoint is taken, only the portion of the checkpoint that has changed since the previous checkpoint need to be saved. The unchanged portion can be restored from previous checkpoints. Incremental checkpointing [2, 3, 18] uses page protection hardware to identify the unchanged portion of the checkpoint. Saving only the changed portion reduces the size of each checkpoint, and thus the overhead of checkpointing.

incremental <on|off> turns incremental checkpointing on or off. The default is off.

In general, the size of a non-incremental checkpoint grows very slowly over time if at all. Moreover, only the most recent checkpoint file needs to be retained for recovery — older ones may be deleted. In contrast, old checkpoint files cannot be deleted when incremental checkpointing is employed, because the program’s data state is spread out over many checkpoint files. The cumulative size of incremental checkpoint files will increase at a steady rate over time, since many updated values may be saved for the same page. In order to place an upper bound on the cumulative size of incremental checkpoint files, it is necessary to coalesce all old checkpoint files into one new file, and then discard the old files. For this purpose, libckpt includes a utility program ckpt_coa, which coalesces a collection of incremental checkpoint files into a single checkpoint file.

maxfiles <n> sets the maximum number of incremental checkpoint files to n. After n checkpoint files have been created, libckpt invokes ckpt_coa to coalesce them into one file. If n = 1, then no incremental checkpointing can occur. Values of n greater than one allow the user to strike a balance between the time and space overhead of incremental checkpointing. The default is n = 1.

Libckpt uses page protection to identify which pages should be included in incremental checkpoints. Specifically, after initialization and after each checkpoint, the mprotect() system call is invoked to set the protection of all pages in the data
space to read-only. When a write occurs to a memory location in a protected page, the SEGV signal is caught by a handler in libckpt. The faulting page has its access protection set to read-write, and the page is marked as dirty. When libckpt takes the next checkpoint, only the dirty pages are included.

### 2.2 Forked Checkpointing

Incremental checkpointing as described in the previous section is still sequential: Execution of the application program is suspended while the checkpoint file is written out. An alternative is to make a copy of the program’s data space and to use an asynchronously executing thread of control to write the checkpoint file. This is called “main-memory checkpointing” [10], and improves checkpoint overhead if there is enough physical memory to hold the checkpoint, as the saving of the checkpoint to disk is overlapped with the execution of the application.

The Unix fork() primitive provides exactly the mechanism needed to implement main-memory checkpointing [7, 12]. When forked checkpointing is specified, libckpt forks a child process, which creates and writes the checkpoint file while the parent process returns to executing the application. The fork() system call provides the child with a fixed snapshot of the parent’s data space and a separate thread of control.

```bash
fork <on|off> in the .ckptrc file turns main forked checkpointing on or off. The default is off.
```

An important improvement to main-memory checkpointing is copy-on-write checkpointing [2, 9, 10]. Here the copy of main memory is taken using copy-on-write [4, 16]. Many implementations of fork() use a copy-on-write mechanism to optimize the copying of the parent’s address space [15]. Thus, forked checkpointing corresponds to either main-memory checkpointing or copy-on-write checkpointing, depending on the operating system’s implementation of fork().

### 2.3 Checkpoint Compression

With checkpoint compression, a standard compression algorithm like LZW [17] is used to shrink the size of the checkpoint [8, 13]. While this may be successful at reducing checkpoint size, it only improves the overhead of checkpointing if the speed of compression is faster than the speed of disk writes, and if the checkpoint is significantly compressed. For uniprocessor checkpointing this is not the case. Compression has only been shown to be effective in parallel systems with disk contention [13]. For this reason, checkpoint compression is not implemented in libckpt.

### 3 User-Directed Checkpointing

All the optimizations presented so far maintain the transparency of checkpointing through techniques that are not visible to the typical application program: signal handlers, page protection, and the creation of child processes. In this section, we consider a different approach that can improve on the performance of these transparent techniques and can also substitute for them when automatic mechanisms are not available. We call this approach “user-directed checkpointing.” We consider two ways in which user-supplied directives can improve the performance of checkpointing: memory exclusion and synchronous checkpointing.

#### 3.1 Memory Exclusion

There are two situations where the values of memory locations can be excluded from a checkpoint file: when the locations are dead and when they are clean. In the case of dead locations, the values in memory will never be read or written, and thus do not need to be saved. In the case of clean locations, the values in memory exist in a previous checkpoint and have not been changed. Thus they need not be saved in the current checkpoint. While the identification of excludable areas of memory can sometimes be automated (as in incremental checkpointing), libckpt also allows the programmer to declare them explicitly.

For example, suppose the user allocates a large temporary array T to make a calculation. When the lifetime of the data in array T is over, it will never be referenced again — the next use of array T will overwrite the old values. If a checkpoint is taken outside of the lifetime of array T, then it can be safely excluded from the checkpoint. Any computation proceeding from this point will not need to use the current values stored in array T.

The stack is a run-time mechanism that helps the checkpoint to determine the lifetime of local variables. This is one form of memory exclusion: Only the live portion of the stack is saved. Unfortunately, this does not work for heap variables or for variables which reside in the statically allocated data segment.
The basis of incremental checkpointing is that clean data need not be repeatedly written to disk. In order to implement automatic incremental checkpointing, libckpt monitors page modifications using the mprotect() system call and a handler for the SEGV signal. This approach has a few weaknesses: It can only operate at the page granularity; system calls can fail rather than generating a SEGV signal when asked to write to a protected page; and on some systems mprotect() is not reliable.

In those cases where automatic mechanisms cannot determine all possible memory exclusions, the performance of checkpointing can suffer. For this reason, libckpt allows the programmer to manage memory exclusion explicitly through two procedure calls:

```
exclude_bytes(char *addr, int size, int usage)
include_bytes(char *addr, int size)
```

exclude_bytes() tells libckpt to exclude the region of memory specified from subsequent checkpoints. It may be called when the user knows that these bytes are not necessary for the correct recovery from the program. Usage is an argument which currently may have one of two values: CKPT_READONLY or CKPT_DEAD. If the former, then exclude_bytes() has been called because the specified memory will not be written to until the user calls include_bytes() on it. Consequently, libckpt includes this memory in the next checkpoint, but excludes it from subsequent checkpoints until the memory is included with include_bytes(). If CKPT_DEAD is specified, then the memory is dead — it will not be read before it is next written. Thus, libckpt excludes this from the next and subsequent checkpoints, until it is explicitly included with include_bytes().

include_bytes() tells libckpt to include the specified region of memory in the next and subsequent checkpoints. Thus, include_bytes() cancels the effect of calls to exclude_bytes(), although calls to include_bytes() do not have to match calls to exclude_bytes(). By default, libckpt includes all bytes in a process’s active stack and data segments that have not been explicitly excluded.

User-directed memory exclusion can dramatically reduce the size of sequential and incremental checkpoint files, but it must be used very carefully. If a live region of memory is mistakenly excluded from a checkpoint, then a subsequent failure and recovery can cause an otherwise correct application to fail or to generate incorrect results.

### 3.2 Synchronous Checkpointing

In the previous section we discuss a mechanism for optimizing asynchronous checkpointing by excluding certain areas of memory. This allows the checkpointer to make use of data lifetime information which would not otherwise be available to it. However, the amount of data which can be excluded from the checkpoint is determined by the program’s state when the checkpoint is taken. If the stack is large, or the size of excluded memory is small, then memory exclusion will have little effect.

Synchronous checkpointing is a user directive that allows the programmer to specify points in the program where it is most advantageous for checkpointing to occur. These are called “synchronous” checkpoints because they are not initiated by timer interrupts. Synchronous checkpoints should be inserted by the programmer at points where memory exclusion can have the greatest effect.

checkpoint_here() is a procedure call specifying where a synchronous checkpoint can be taken.

Synchronous checkpoints may be placed in program locations that are reached often. Checkpointing too often, however, can lead to poor performance, and in order to avoid this libckpt allows a minimum interval between checkpoints to be specified.

```
mintime <seconds> specifies the minimum period of time that must pass between checkpoints. The default is zero. If mintime seconds have not passed since the previous checkpoint, then checkpoint_here() calls are ignored.
```

Synchronous and asynchronous checkpointing techniques can complement one another. If maxtime seconds have passed and no synchronous checkpoint has been taken since the last checkpoint, then an asynchronous checkpoint is still taken. However, the effect of memory exclusion is likely not to be as beneficial as in a synchronous checkpoint. If both the mintime and maxtime parameters are set, then the former specifies the minimum interval between synchronous checkpoints, and the latter specifies an interval after which an asynchronous checkpoint will be taken. Whenever
a checkpoint is taken, both the minimum and maximum interval timers are reset.

If \texttt{maxtime} is zero, then asynchronous checkpoints are disabled. In this case the specification of memory exclusion can be optimized for synchronous checkpoints, because there is no danger of asynchronous checkpoints being taken.

### 3.3 An Example

There are many examples where user-directed, synchronous checkpointing can yield large performance gains. Consider the program in Figure 1. This is a typical driver program for many kinds of programs that repeat calculations over numerous points in a data set. Figure 2 shows how one can checkpoint this program with synchronous, user-directed checkpointing in \texttt{libckpt}.

```c
main()
{
    struct data *D;
    FILE *fi, *fo;
    D = allocate_data_set();
    fi = fopen("input", "r");
    fo = fopen("output", "w");
    while(read_data(fi, D) != -1) {
        perform_calculation(D);
        output_results(fo, D);
    }
}
```

Figure 1: A typical scientific driver program, no checkpointing

```c
ckpt_target()
{
    struct data *D;
    FILE *fi, *fo;
    D = allocate_data_set();
    fi = fopen("input", "r");
    fo = fopen("output", "w");
    while(read_data(fi, D) != -1) {
        perform_calculation(D);
        output_results(fo, D);
        exclude_bytes(D, sizeof(struct data), CKPT_DEAD);
        checkpoint_here();
        include_bytes(D, sizeof(struct data));
    }
}
```

Figure 2: A typical scientific driver program with checkpointing

By specifying that the checkpoint must be taken at the \texttt{checkpoint here()} call, we are able to omit all of the variable \texttt{D} from the synchronous checkpoint. This is because \texttt{D} is initialized anew at each iteration of the program. If \texttt{D} is large, then user-directed checkpointing will be responsible for a significant savings in checkpoint overhead. Note that this will be a vast improvement over incremental checkpointing because the memory locations in \texttt{D} will be dirty at the time of the checkpoint.

Section 6 shows other successful examples of user-directed checkpointing.

### 4 The Mechanics of Checkpointing and Recovery

The motivation for checkpointing is to reconstruct the recovery point. We therefore begin with an overview of the recovery process before describing the details of checkpointing. Recovery has four parts: process creation, data state restoration, system state restoration, and processor state restoration.

1. Process creation is implemented by invoking the checkpointed program with a special command line argument for recovery. This automatically restores the text portion of the process's state and begins execution. \texttt{Libckpt} parses the command line, detects the argument for recovery, and calls the recovery routine.

2. The recovery routine performs the rest of the recovery. Data state restoration means reading the checkpoint file to recreate the contents of data memory: This consists of the process's stack and data segments.

3. System state restoration means restoring as much of the operating system state as possible to its state at the time of the checkpoint. Much of the operating system state, such as the process ID and parent process ID, is unrestorable. However, most applications that need checkpointing are what we call "well-behaved," and do not rely on such state. \texttt{Libckpt} determines the state of the open file table at each checkpoint, and saves it as part of each checkpoint. Upon recovery, \texttt{Libckpt} restores the system so that the state of open files is the same as it was at the time of the checkpoint. No other system state is either saved or restored by \texttt{Libckpt}. 
4. Processor state restoration requires that processor registers, including the program counter and stack pointers be restored to their values when the checkpoint was taken. In libckpt, we use \texttt{setjmp()} to store the processor state in memory. The processor state is restored using \texttt{longjmp()}. Thus the recovery routine never returns, and execution continues as an apparent “second return” from the \texttt{setjmp()} of the checkpointing routine.

Thus the mechanics of checkpointing are straightforward: When taking a checkpoint libckpt saves the processor state using \texttt{setjmp()} and records the state of the open file table. Then the data state, consisting of the program’s stack and data segments, is written to disk.

5 Experiments

In this section, we present the results of checkpointing five application programs using libckpt. The applications are long-running FORTRAN and C programs written by scientists to run under Unix. All are typical of programs that can benefit from checkpointing for fault-tolerance.

The experiments were performed on a dedicated Sparstation 2 running SunOS 4.1.3, and writing to a Hewlett Packard HP6000 disk via NFS. The specific instances of the applications are described in Table 1. We describe the applications below:

- **Matrix Multiplication** (MAT): This is a straightforward matrix multiplication. Two $615 \times 615$ matrices are read from disk and multiplied, and the product matrix is written to an output file.
- **Linear Equation Solver** (SOLVE): This is a testing program from LAPACK, a high-performance package of linear-algebra subroutines [1]. This program generates a system of 750 equations with 750 unknowns, uses LU decomposition to solve the system, and then writes the solution to disk. It repeats this process for seven separate systems of equations.
- **Cellular Automata** (CELL): This program executes a $2048 \times 2048$ grid of cellular automata for fifteen generations.
- **Shallow Water Model** (WATER): This is the program STSWM from the National Center for Atmospheric Research. The program is a shallow water model based on the spectral transform method [5]. The instance used here is “Zonal Flow over a Mountain” from their test suite, modeled at 15-minute intervals for six hours.
- **Multicommodity Flow** (MCNF): This program solves the multicommodity network flow problem using the simplex method [6]. The instance used here runs on a network of 100 vertices and 50 commodities.

Note that for the purposes of these experiments, input values have been chosen to give running times between thirteen and thirty minutes. Typically, the programs would be set up to run for much longer, thus making them ideal candidates for libckpt.

We present results pertaining to the three important metrics of checkpointing performance:

- **Checkpoint time**: This is the average duration of a checkpoint, from start to finish.
- **Checkpoint overhead**: This is amount of time added to the running time of the application as a result of checkpointing. Note that in sequential checkpointing, overhead is equal to the total checkpoint time. In main-memory and copy-on-write checkpointing, the overhead is smaller than the total checkpoint

<table>
<thead>
<tr>
<th>Application</th>
<th>Abbreviation</th>
<th>Language</th>
<th>Running Time (mm:ss)</th>
<th>Maximum Checkpoint Size (Mbytes)</th>
<th>Checkpoint Interval (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix Multiplication</td>
<td>MAT</td>
<td>C</td>
<td>15:20</td>
<td>4.6</td>
<td>2</td>
</tr>
<tr>
<td>Linear Equation Solver</td>
<td>SOLVE</td>
<td>FORTRAN</td>
<td>13:42</td>
<td>4.6</td>
<td>2</td>
</tr>
<tr>
<td>Cellular Automata</td>
<td>CELL</td>
<td>C</td>
<td>17:39</td>
<td>8.4</td>
<td>2</td>
</tr>
<tr>
<td>Shallow Water Model</td>
<td>WATER</td>
<td>FORTRAN</td>
<td>25:54</td>
<td>13.1</td>
<td>3</td>
</tr>
<tr>
<td>Multicommodity Flow</td>
<td>MCNF</td>
<td>FORTRAN</td>
<td>18:38</td>
<td>24.3</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Description of application instances
time because the disk writes are performed in parallel with the execution of the application.

- **Checkpoint size**: This is the average size of the checkpoint file.

The prime goal of checkpoint optimization is to minimize all three of these metrics, while still providing adequate fault-tolerance. Minimizing checkpoint overhead is the most important, because users would rather take the risk of failure than use a checkpointer that increases their applications’ running time significantly. Keeping the overhead of checkpointing under 10% of the program’s total running time is a reasonable goal [9, 13]. Minimizing checkpoint size is also important, as disk space rarely comes for free. Checkpoint time is the least important of the three metrics: When checkpointing for fault-tolerance, the only concern is that the current checkpoint complete before the user desires the next checkpoint to begin.

6 Results

All of the experimental results are contained in Table 2 in the appendix. All of the graphs and data in this section are drawn directly from Table 2.

6.1 Sequential Checkpointing

With no optimizations, checkpoint time and overhead should be the same, and should be directly proportional to the checkpoint size. Figure 3 confirms this prediction, showing checkpoint overhead and time vs. size for the sequential checkpointing runs described in Table 2.

6.2 Checkpointing with fork()

When checkpointing with fork(), the application writes its checkpoints to disk asynchronously. This enables it to run concurrently with the saving of the checkpoint, thereby reducing the overhead of checkpointing dramatically, as shown in Figure 4. This figure displays the percentage reduction in the overhead of checkpointing by using fork()\(^1\).

Because SunOS 4.1.3 implements fork() with copy-on-write, Figure 4 shows that copy-on-write improves the overhead of checkpointing by over 70 percent in almost all cases.

6.3 Incremental Checkpointing

Figure 5 is a graph showing the percentage reduction of checkpoint size and checkpoint overhead when using incremental checkpointing instead of simple sequential checkpointing. In three of the applications (MAT, WATER, and MCNF), only a fraction of the applications’ address spaces are modified between checkpoints, resulting in a significant reduction in the average checkpoint size. Correspondingly, the overhead of checkpointing is significantly reduced. In the other two programs, the entire address spaces of the programs are modified between checkpoints, yielding little to no reduction in the size of checkpoints. Therefore, in SOLVE and CELL, the overhead of checkpointing is increased due to the fact that the cost of handling page faults is not offset by a savings in the time to write the checkpoint to disk.

1 Ninety percent reduction in checkpoint overhead means that the overhead of checkpointing using fork() is ten percent of the overhead of sequential checkpointing.
6.4 User-Directed Checkpointing

In the previous three sections, the results corroborate published research concerning checkpointing optimizations [2, 3, 9, 10]. In this section, we evaluate the new technique: user-directed checkpointing. In three of the applications, we analyzed the application programs and inserted directives in the code. In each case, we were able to add under ten lines of code, making checkpoints synchronous, and excluding memory from these checkpoints. We describe the details of each application below.

**SOLVE**: Adding directives to the Linear Equation Solver was straightforward. At the end of each iteration, all of the program’s arrays are dead: The matrix of equations will be initialized anew for the next iteration, and the solution vector will be recalculated. Therefore at the end of each iteration, we insert `exclude_bytes()` calls for the equation matrix and solution vector, then a `checkpoint_here()` call, and finally `include_bytes()` calls to re-include the matrix and vector in case of an asynchronous checkpoint.

The results can be seen in Figure 6: The calls to `exclude_bytes()` and `checkpoint_here()` produce checkpoint files that are almost, reducing the checkpoint size and overhead by over 90 percent. This is significant, because it is an application where incremental checkpointing fails to improve the performance of checkpointing.

**CELL**: At the end of each generation of the cellular automaton application, the previous value of the automaton grid becomes dead — its values are not used for the calculation of the subsequent generations of the computation. Therefore we added user directives to checkpoint at the end of each generation, excluding the dead half of the grid from each checkpoint. In order to checkpoint at roughly the same interval as before, we also set `mintime` to 100, so that every second generation is checkpointed.

The results are in Figure 7. With our user directives, the checkpoint size is halved. Accordingly, the overhead of checkpointing is also halved. Thus, as in SOLVE, the calls to `exclude_bytes()` and `checkpoint_here()` succeed in improving the overhead of checkpointing in an application where incremental checkpointing fails.

**MAT**: In the matrix multiplication the two input matrices are read-only data. Moreover, once a product element is calculated it too is read-only. This is why incremental checkpointing works so well. In this application, we inserted `exclude_bytes()` calls (with `flag=CKPT_READONLY`) after reading the input ma-
7 Related Work

There has been much computer science research devoted to checkpointing. Checkpointing has been implemented on uniprocessors [8, 11], multiprocessors [9, 10], transputers [14], multicomputers [13], and ad distributed systems [2, 7]. Of these implementations, only two (Condor [11] and Fail-Safe PVM [7]) are publicly available code for Unix environments. Both implement sequential checkpointing with forking, and neither is designed for simple uniprocessor checkpointing; Condor is a system for batch programming using process migration, and Fail-Safe PVM requires the programmer to have access to the PVM infrastructure. Neither package implements any optimizations beyond calling fork().

User-directed checkpointing bears some similarity to checkpointers by Li and Fuchs [8], and Silva et al [14]. The former describes static checkpointing, which is similar to our synchronous checkpointing. The user places potential_checkpoint_here() calls into his program, and the compiler and/or runtime system decides which of those calls would be best for checkpointing. They call for no user assistance in determining the memory to exclude (they only exclude the stack and unallocated heap memory), and do not show the dramatic performance improvements gained by user-directed checkpointing in the SOLVE, CELL, and MAT applications.

Silva et al implement a checkpointing package for transputers in which the user specifies exactly what and where to checkpoint, but the process state is not included in checkpoints. Thus the user is responsible for rebuilding the call stack, although not the data, on recovery. Our approach differs because the checkpointer is responsible for the entire process state, and not just for the integrity of the data.
8 Conclusion

We have written a general-purpose checkpointing library, libckpt, that provides fault-tolerance for long-running programs under Unix. The strengths of this library are its ease of use and low overhead. **Libckpt** is currently available via anonymous FTP to cs.utk.edu in the directory pub/plank/libckpt.

Our experiments with libckpt show first and foremost that it is general-purpose and easy to use. We were able to checkpoint all five applications by changing one line of the applications' source code, and relinking with libckpt. Once enabled, these programs could save their state to disk periodically for fault-tolerance, using the fork() and incremental checkpointing optimizations if so desired. For all five applications, we were able to dramatically lower the overhead of checkpointing with copy-on-write, as implemented by libckpt's fork() optimization. Moreover, in three of the five applications, checkpoint size and overhead were reduced by over 60 percent using incremental checkpointing. Thus, libckpt is able to take efficient checkpoints using standard techniques from the checkpointing literature.

Libckpt also implements user-directed checkpointing, a new technique for improving the performance of checkpoint-based systems. It is assumed that a little user input to the checkpoint system can result in a large performance gain. Memory exclusion and synchronous checkpointing are the two ways in which a user can direct the checkpoint to checkpoint more efficiently. In our experiments, directives added to three of the applications yielded performance improvements in all three cases.

One avenue of future research is to employ compiler analysis to assist user-directed checkpointing. If the user places the checkpoint_here() calls, the compiler can use data dependence analysis to make calls to exclude_bytes() and include_bytes(). The benefits may be twofold. First, the compiler may discover dead variables to exclude that the user may omit. Second, the compiler can guarantee that its memory exclusion will yield correct checkpoints. In other words, whereas the user might err in excluding too much memory from a checkpoint, resulting in a faulty recovery state, the compiler can guarantee correctness.

It is the authors' opinion that checkpointing primitives such as those provided by libckpt should be implemented in the operating system. This will improve both the performance and the generality of checkpointing. Until such a time, users can make use of a tool such as libckpt to render their programs resilient to failure.

Acknowledgements

The authors thank Jian Xu, Jack Dongarra, Christian Halloy, and the National Center for Atmospheric Research for help in obtaining test programs. We also thank Nitin Vaidya, Mootaz Elnozahy, Heather Booth, and the referees for their valuable comments. James Plank is supported by NSF grant CCR-9409496. Kai Li is supported by ARPA and ONR under contracts N00014-91-J-4039, and Intel Supercomputer Systems Division.

References


Author Information

Jim Plank is an assistant professor in the Department of Computer Science at the University of Tennessee. He received his Ph.D. degree from Princeton University in 1993. His areas of interest are checkpointing, fault-tolerance, operating systems, and architecture.

Micah Beck is an assistant professor in the Department of Computer Science at the University of Tennessee. He received his Ph.D. degree from Cornell University in 1992. His areas of interest are program analysis and compilation for parallelism.

Gerry Kingsley is a Ph.D. student in the Department of Computer Science at the University of Tennessee. His research interests include compiler analysis techniques, user level checkpointing, and general fault tolerance.

Kai Li received his Ph.D. degree from Yale University in 1986 and is currently an associate professor of the Department of Computer Science, Princeton University. His research interests are in operating systems, computer architecture, fault tolerance, and parallel computing. He is an editor of the IEEE Transactions on Parallel and Distributed Systems, and a member of the editorial board of International Journal of Parallel Programming.

The first three authors’ address is: Department of Computer Science, University of Tennessee, 107 Ayres Hall, Knoxville, TN 37996. Their email addresses are [plank, beck, kingsley]@cs.utk.edu. Kai Li’s address is: Princeton University, 35 Olden Street, Princeton, NJ 08544-2087. He may be reached electronically at li@cs.princeton.edu.
## Appendix

<table>
<thead>
<tr>
<th>Application</th>
<th>User Directives</th>
<th>Incremental</th>
<th>Fork</th>
<th>Running Time (sec)</th>
<th>Overhead (sec)</th>
<th>% Overhead</th>
<th>Avg Ckpt Time (sec)</th>
<th>Avg Total Ckpt Size (Mbytes)</th>
<th>Num of Ckpts</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAT</td>
<td>No checkpoint</td>
<td></td>
<td></td>
<td>920.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>1160.0</td>
<td>239.3</td>
<td>26.0</td>
<td>21.2</td>
<td>4.59</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>961.3</td>
<td>40.7</td>
<td>4.4</td>
<td>3.6</td>
<td>0.57</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>969.5</td>
<td>48.8</td>
<td>5.3</td>
<td>22.4</td>
<td>4.59</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>931.0</td>
<td>10.3</td>
<td>1.1</td>
<td>3.4</td>
<td>0.57</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>963.5</td>
<td>42.8</td>
<td>4.7</td>
<td>3.6</td>
<td>0.59</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>970.5</td>
<td>49.8</td>
<td>5.4</td>
<td>3.5</td>
<td>0.56</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>930.5</td>
<td>9.8</td>
<td>1.1</td>
<td>5.0</td>
<td>0.59</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>931.0</td>
<td>10.3</td>
<td>1.1</td>
<td>3.3</td>
<td>0.56</td>
<td>11</td>
</tr>
<tr>
<td>SOLVE</td>
<td>No checkpoint</td>
<td></td>
<td></td>
<td>822.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>966.5</td>
<td>143.8</td>
<td>17.5</td>
<td>19.6</td>
<td>4.62</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>1000.7</td>
<td>178.0</td>
<td>21.6</td>
<td>24.7</td>
<td>4.55</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>857.6</td>
<td>34.9</td>
<td>4.2</td>
<td>19.6</td>
<td>4.62</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>864.8</td>
<td>42.1</td>
<td>5.1</td>
<td>24.7</td>
<td>4.55</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>836.3</td>
<td>13.6</td>
<td>1.7</td>
<td>1.3</td>
<td>0.11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>833.3</td>
<td>10.6</td>
<td>1.3</td>
<td>1.0</td>
<td>0.04</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>830.0</td>
<td>7.3</td>
<td>0.9</td>
<td>1.4</td>
<td>0.11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>829.3</td>
<td>6.6</td>
<td>0.8</td>
<td>1.1</td>
<td>0.04</td>
<td>7</td>
</tr>
<tr>
<td>CELL</td>
<td>No checkpoint</td>
<td></td>
<td></td>
<td>1059.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>1389.3</td>
<td>329.4</td>
<td>31.1</td>
<td>45.4</td>
<td>8.46</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>1455.4</td>
<td>395.5</td>
<td>37.3</td>
<td>53.0</td>
<td>8.44</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>1130.3</td>
<td>70.4</td>
<td>6.6</td>
<td>43.7</td>
<td>8.46</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>1143.9</td>
<td>84.0</td>
<td>7.9</td>
<td>50.0</td>
<td>8.44</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>1202.3</td>
<td>142.4</td>
<td>13.4</td>
<td>19.3</td>
<td>4.26</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>1210.3</td>
<td>150.4</td>
<td>14.2</td>
<td>20.3</td>
<td>3.63</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>1101.4</td>
<td>41.5</td>
<td>3.9</td>
<td>19.3</td>
<td>4.26</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>1111.1</td>
<td>51.2</td>
<td>4.8</td>
<td>20.4</td>
<td>3.63</td>
<td>7</td>
</tr>
<tr>
<td>WATER</td>
<td>No checkpoint</td>
<td></td>
<td></td>
<td>1553.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>2170.4</td>
<td>616.5</td>
<td>39.7</td>
<td>74.6</td>
<td>14.17</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>1800.3</td>
<td>246.4</td>
<td>15.9</td>
<td>29.9</td>
<td>5.23</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>1676.1</td>
<td>122.2</td>
<td>7.9</td>
<td>74.6</td>
<td>14.17</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>1612.3</td>
<td>58.4</td>
<td>3.8</td>
<td>28.5</td>
<td>5.15</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>1118.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MCNF</td>
<td>No checkpoint</td>
<td></td>
<td></td>
<td>1118.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>1681.8</td>
<td>563.6</td>
<td>50.4</td>
<td>139.3</td>
<td>24.31</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>1216.1</td>
<td>97.9</td>
<td>8.8</td>
<td>10.7</td>
<td>1.75</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>1175.2</td>
<td>57.0</td>
<td>5.1</td>
<td>131.3</td>
<td>24.31</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>1125.6</td>
<td>7.4</td>
<td>0.7</td>
<td>10.3</td>
<td>1.75</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2: Results of all checkpointing experiments