Proceedings
of the
4th South African Computer Symposium

Holiday Inn, Pretoria
1 - 3 July 1987

edited by

Pieter Kritsinger
Computer Science Department
University of Cape Town
PREFACE

Computer science is an emerging discipline which is having difficulty in being recognised as a worthy member of the sciences. I will paraphrase John Hopcroft, co-winner of the 1986 Turing Award, when, during a recent interview, he said that the primary reason for the lack of recognition, is the age of our researchers. Probably not one of the researchers who presented their work at this symposium is older than 45. I know of no computer scientist in South Africa who is in a position where (s)he can affect funding priorities. As far as I know we have no representation on any of the committees of the Foundation for Research Development and for our Afrikaans speaking fraternity, none who is a member of the Akademie vir Wetenskap en Kuns. It will take time and conscious effort to establish our presence. The same is true of course for our universities. Again, with one exception, I know of no dean of a science faculty, vice-principal or principal who is a computer scientist. We consequently spend an enormous amount of time trying to explain the needs of computer science and its difficulties. I believe this symposium is a further step towards accreditation by our peers and superiors from the other sciences.

The total number of papers submitted to the Programme Committee for consideration was 34. Each paper was reviewed by three persons knowledgeable in the field it represents. Of those submitted, 23 were finally selected for inclusion in the symposium. As a result the overall quality of the papers is high and as a computer science community in Africa we can be justly proud of the final programme.

This is the fourth in the series of South African computer symposia. This year the symposium is sponsored by the Computer Society of South Africa (CSSA), the South African Institute for Computer Scientists and the local IFIP Committee. The executive director of the CSSA and his staff deserve warm thanks for handling the organisation as well as they have, while the Organising Committee provided Derrick and I with very valuable advice.

Finally I would like to express my sincere appreciation to the authors, to the members of the Programme Committee and particularly the reviewers. Without the kind cooperation of everyone, this symposium would not have taken place.

Pieter Kritzinger
SYMPOSIUM CHAIRMAN: PS Kritzinger, University of Cape Town.
SYMPOSIUM CO-CHAIRMAN: D Kourie, University of Pretoria.
MEMBERS OF THE PROGRAMME COMMITTEE

Judy Bishop, Witwatersrand University
Chris Bornman, UNISA
Hannes de Beer, Potchefstroom University
Gideon de Kock, Port Elizabeth University
Jaap Kies, Western Cape University
Derrick Kourie, Pretoria University
Pieter Kritzinger, Cape Town University
Tony Krzesinski, Stellenbosch University
Michael Laidlaw, Durban Westville University
Peter Lay, Cape Town University
Ken MacGregor, Cape Town University
Theo McDonald, Orange Free State University
Jan Oosthuizen, University of the North
Dennis Riordan, Rhodes University
Alan Sartori-Angus, Natal University
John Shochot, Witwatersrand University
Theuns Smith, Rand Afrikaans University
Trevor Turton, ISM (Pty) Ltd
Gerrit Wiechers, Infoplan.
LIST OF REVIEWERS

BERMAN Sonia
BISHOP Judy
BORNMAN Chris
CAREY Chris
CHERENACK Paul
DE BEER Hannes
DE VILLIERS Pieter
GORRINGE Pen
KIES Jaap
KOURIE Derrick
KRITZINGER Pieter
KRZESINSKI Tony
LAIDLAW Michael
LAY Peter

MacGREGOR Ken
MATTISON Keith
McDONALD Theo
RENNHACKKAMP Martin
RIORDAN Denis
SATORI-ANGUS Alan
SCHOCHOT John
SMITH Theuns
TURTON Trevor
VAN DEN HEEVER Roelf
VAN ROOYEN Hester
VON SOLMS Basie
VOS Koos
# TABLE OF CONTENTS

**Keynote Address**

"An Extensible System and Programming Tool for Workstation Computers." ........................................... 1  
Niklaus Wirth, ETH, Zurich

**Invited Lectures**

"The Relationship of Natural and Artificial Intelligence." ........... not included in Proceedings.  
G Lasker, University of Windsor, Ontario.

"Software Engineering: What Can We Expect in the Future?" ........... not included in Proceedings.  
D Teichrow, University of Michigan, U.S.A.

**Computer Languages I**

"SPS-Algol: Semantic Constructs for a Persistent Programming Language." ......................................... 13  
S Berman, University of Cape Town.

"Petri Net Topologies for a Specification Language." .... 25  
R Watson, University of the Witwatersrand.

"Towards a Programming Environment Standard in LISP." .... 45  
R Mori, University of Cape Town.

"ADA for Multiprocessors: Some Problems and Solutions." .. 63  
J Bishop, University of the Witwatersrand.

**Computer Graphics**

"Polygon Shading on Vector Type Devices." ............... 75  
C F Scheepers, CSIR.

"Hidden Surface Elimination in Raster Graphics Using Visigrams." ................................................. 97  
P Gorringe, CSIR.

**Database Systems I**

"On Syntax and Semantics Related to Incomplete Information Databases." ................................. 109  
M E Orlowska, UNISA.

"Modelling Distributed Database Concurrency Control Overheads." .................................................. 131  
M H Rennhackkamp, University of Stellenbosch.

**Operating Systems**

"The Development of a Fault Tolerant System for a Real-time Environment." ............................... 149  
M Morris, CSIR.

"A New General-purpose Operating System." ............... 161  
B H Venter, CSIR.
Computer Languages II


"A Generalised Expression Structure." .............. 189 W van Biljon, CSIR.

Computer Networks and Protocols I


Computer Networks and Protocols II


Artificial Intelligence

"A Data Structure for Exchanging Geographic Information." ............................................... 267 A Cooper, CSIR.

"The Design and Use of a Prolog Trace Generator for CSP." ............................................ 279 D G Kourie, University of Pretoria.

Database Systems II

"An Approach to Direct End-user Usage of Multiple Databases." ........................................... 297 M J Phillips, CSIR.

"A Semantic Data Model Approach to Logical Data Independence." ......................................... 329 S Berman, University of Cape Town.

Information Systems

"The ELSIM Language: an FSM-based Language for the ELSIM SEE." ...................................... 343 L du Plessis and C Bornman, UNISA.

Computer Languages III

"Experience with a Pattern-matching Code Generator." ... 371
M A Mulders, D A Sewry and W R van Biljon, CSIR.

"Set-oriented Functional Style of Programming." .......... 385
C Mueller, University of the Witwatersrand.

Tutorial

The use of Modula-2 in Software Engineering." ............ 399
N Wirth, ETH, Zurich.
# DAY 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>07h30</td>
<td>Registration and Coffee.</td>
</tr>
<tr>
<td>08h45</td>
<td>Welcoming address, President of the South African Institute of Computer Scientists, Dr. G. Wiechers.</td>
</tr>
<tr>
<td>09h00</td>
<td>Invited Lecture. Professor D. Teichrow, University of Michigan.</td>
</tr>
<tr>
<td></td>
<td><em>Software Engineering, ... What Can We Expect in the Future.</em></td>
</tr>
<tr>
<td>10h00</td>
<td><strong>Computer Languages I.</strong> Chairman: G. Wiechers.</td>
</tr>
<tr>
<td>10h15</td>
<td>S. Berman, University of Cape Town. <em>SPS-Algo: Semantic Constructs for a Persistent Programming Language.</em></td>
</tr>
<tr>
<td>11h25</td>
<td>A. Mori, University of Cape Town. <em>Towards a Programming Environment Standard in USP.</em></td>
</tr>
<tr>
<td>11h50</td>
<td>J. Bishop, University of the Witwatersrand. <em>ADA for Multiprocessors: Some Problems and Solutions.</em></td>
</tr>
<tr>
<td>12h30</td>
<td>LUNCH</td>
</tr>
<tr>
<td>14h00</td>
<td><strong>Computer Graphics.</strong> Chairman: D. Kourie</td>
</tr>
<tr>
<td></td>
<td>C. F. Scheepers, CSIR. <em>Polygon Shading on Vector Type Devices.</em></td>
</tr>
<tr>
<td>14h35</td>
<td>P. Gorringe, CSIR. <em>Hidden Surface Elimination in Raster Graphics Using Visigraphics.</em></td>
</tr>
<tr>
<td>15h15</td>
<td><strong>Operating Systems.</strong> Chairman: K. MacGregor.</td>
</tr>
<tr>
<td></td>
<td>M. Morris, UNISA. <em>The Development of a Fault Tolerant System for a Real-time Environment.</em></td>
</tr>
<tr>
<td></td>
<td>B. H. Venter, CSIR. <em>A New General-purpose Operating System.</em></td>
</tr>
<tr>
<td>15h30</td>
<td><strong>Database Systems I.</strong> Chairman: B. von Solms.</td>
</tr>
<tr>
<td></td>
<td>M.E. Orlowska, UNISA. <em>On Syntax and Semantics Related to Incomplete Information Databases.</em></td>
</tr>
<tr>
<td>16h05</td>
<td>M.H. Rennhakkamp, Stellenbosch University. <em>Modelling Distributed Database Concurrency Control Overheads</em></td>
</tr>
<tr>
<td>18h00</td>
<td><strong>Computer Languages II.</strong> Chairman: J. Bishop.</td>
</tr>
<tr>
<td></td>
<td>W. van Biljon, CSIR. <em>A Generalised Expression Structure.</em></td>
</tr>
<tr>
<td>18h00</td>
<td>Cocktail Party in Cullinan Room A.</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>08h30</td>
<td>Keynote Address by Professor Niklaus Wirth, Swiss Federal Institute</td>
</tr>
<tr>
<td></td>
<td>for Technology, Zurich.</td>
</tr>
<tr>
<td></td>
<td>An Extensible System and a Programming Tool for Workstation Computers.</td>
</tr>
<tr>
<td></td>
<td><strong>Computer Networks and Protocols I.</strong> Chairman: P.S. Kritzinger.</td>
</tr>
<tr>
<td>09h30</td>
<td>A.E. Krzesinski, University of Stellenbosch.</td>
</tr>
<tr>
<td></td>
<td>An Approximate Solution Method for Multiclass Queueing Networks</td>
</tr>
<tr>
<td></td>
<td>with State Dependent Routing and Window Row Control.</td>
</tr>
<tr>
<td>10h05</td>
<td>J. Punt, University of Cape Town.</td>
</tr>
<tr>
<td></td>
<td>A Protocol Validation System.</td>
</tr>
<tr>
<td>10h30</td>
<td><strong>COFFEE</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Computer Networks and Protocols II.</strong> Chairman: R. van der Heever.</td>
</tr>
<tr>
<td>11h00</td>
<td>P.S. Kritzinger, University of Cape Town.</td>
</tr>
<tr>
<td>11h35</td>
<td>Invited Lecture by Professor G. Lasker, University of Windsor, Ontario.</td>
</tr>
<tr>
<td></td>
<td>The Relationship of Natural and Artificial Intelligence.</td>
</tr>
<tr>
<td>12h00</td>
<td><strong>LUNCH</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Artificial Intelligence.</strong></td>
</tr>
<tr>
<td></td>
<td>Chairman: G. Lasker.</td>
</tr>
<tr>
<td>14h00</td>
<td>A. Cooper, CSIR</td>
</tr>
<tr>
<td></td>
<td>A Data Structure for Exchanging Geographic Information.</td>
</tr>
<tr>
<td>14h35</td>
<td>A. I. Newcombe, University of Cape Town and A. Rado, National Library</td>
</tr>
<tr>
<td></td>
<td>of Medicine, Maryland.</td>
</tr>
<tr>
<td></td>
<td>Strategies for Automatic Indexing and Thesaurus Building.</td>
</tr>
<tr>
<td>15h15</td>
<td><strong>COFFEE</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Database Systems II.</strong></td>
</tr>
<tr>
<td></td>
<td>Chairman: C. Bornman.</td>
</tr>
<tr>
<td>15h30</td>
<td>M.J. Philips, CSIR</td>
</tr>
<tr>
<td></td>
<td>An Approach to Direct End-user Usage of Multiple Databases.</td>
</tr>
<tr>
<td>16h05</td>
<td>S. Berman, University of Cape Town.</td>
</tr>
<tr>
<td></td>
<td>A Semantic Data Model Approach to Logical Data Independence.</td>
</tr>
<tr>
<td>16h45</td>
<td><strong>Open Forum with professors G. Lasker, D. Teichrow and N. Wirth.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Symposium Banquet in Cullinan Room.</strong></td>
</tr>
<tr>
<td></td>
<td>Guest speaker, Dr. D. Jacobson., Group Executive: Technology, Allied</td>
</tr>
<tr>
<td></td>
<td>Technologies Limited.</td>
</tr>
</tbody>
</table>
08h00  Registration (Tutorial only).

08h30  Tutorial.
      The Tutorial will be given by Professor Niklaus Wirth, Division of Computer Science,
      Swiss Federal Institute of Technology, Zurich.

      The use of Modula-2 in Software Engineering.
      Topics to be covered include:

      What is Software Engineering?
      Data types and structures.
      Modularization and information hiding.
      Definition and implementation parts.
      Separate compilation with type checking.
      Facilities to express concurrency.
      Pompous programming style.
      What could be excluded?

12h15  Close of Symposium.

12h30  LUNCH
A NEW GENERAL-PURPOSE OPERATING SYSTEM

by

B.H. Venter

Computer Science Division, NRIMS, CSIR, Box 395, Pretoria

ABSTRACT

The current generation of widely-used, multi-user, general-purpose operating systems have evolved from versions that were designed when many of the issues that are important today were unimportant or not even thought of. This evolution has not been totally successful. In particular, the current generation is ill suited for implementation on loosely-coupled multi-processors.

A new operating system, designed with current requirements in mind, and flexible enough to adapt successfully to likely future requirements, has been developed as part of a project to build a loosely-coupled multi-processor system that should have the performance and functionality of a 'super main-frame' computer.

This paper concentrates on describing the fundamental mechanisms of the operating system: processes, inter-process communication, and servers. It also briefly outlines the support provided for data security, database applications, and real-time applications.

KEYWORDS AND PHRASES:

operating systems, system calls, inter-process communication, distributed systems, operating system security, operating system database support, real-time.
1. INTRODUCTION AND MOTIVATION

A modern general-purpose operating system can be expected to provide a high level of data security, to provide appropriate support for a comprehensive database management system, and to support real-time applications. Furthermore, such a system can be expected to isolate applications from the underlying hardware. That is, the operating system should be implementable on most current and future hardware systems, and an application written in a standard high level language should be able to run on any hardware system running under the control of such an operating system.

Moreover, a computer controlled by a modern operating system should be able to form part of a network of distributed computers, and provide users with efficient, transparent access to the resources available via the network. In particular, a modern operating system should be able to make effective use of the loosely-coupled multi-processor systems that are beginning to appear.

The current generation of widely-used, multi-user, general-purpose operating systems have gradually evolved from versions designed in the 1970's and 1960's. Then, many of the issues that are of considerable importance today, were unimportant or not even thought of. In particular, loosely-coupled multi-processor systems were unforeseen, and operating systems were designed with a single-processor mind set, making it difficult to port these systems to loosely-coupled multi-processors.

For example, current implementations of UNIX cannot even be ported to tightly-coupled multi-processors without major changes and preferably a complete rewrite [2]. The situation is even worse for loosely-coupled multi-processors: the latest 'standard', UNIX System V [4], allows arbitrary processes to share memory - which cannot be done efficiently on a loosely-coupled multi-processor. Furthermore, the message-passing mechanism - which is of critical importance to distributed applications - is far too complicated and cumbersome to be implemented efficiently. In fact, the designers of this message-passing mechanism have apparently assumed that the sender and receiver processes share access to a common physical memory. There are also other aspects of UNIX which, upon close examination, prove difficult or impossible to implement effectively on a loosely-coupled multi-processor.

The situation is not much better when one considers other widely-used current generation operating systems. Thus, if one aims to build a new computer system based on a loosely-coupled multi-processor, a substantial operating system development effort must be undertaken.

If compatibility with existing software is one's principal concern, one should aim to implement either a UNIX 'look-alike', or an MVS 'look-alike'. However, as pointed out above, UNIX is not well suited to the role. The same is true for MVS.
If, on the other hand, making do with limited resources is one's principal concern, then it makes sense to develop a new operating system, with full use being made of what has been learnt about operating systems since the 1970's.

The author is currently involved in the development of a loosely-coupled multi-microprocessor system. The aim is that this system should provide 'super main-frame' functionality and performance. The project is being undertaken with relatively limited resources, and building a working system with the available resources is considered more important than providing compatibility with some existing software base. Consequently, a new operating system has been designed for this computer, and a fairly complete 'quick-and-dirty' prototype has already been implemented. A full-scale implementation, using the prototype operating system as the development system, is currently underway.

The rest of this paper is a brief survey of the main features of the operating system design. The intention is not rigorously to justify design decisions, nor to point out what contribution the work makes, but rather to provide the reader with an overall description and enough detail to compare the new operating system to any existing operating system. A more detailed description can be found in [3].

2. PROCESSES

A new process can be created either by forking an existing process, in the style of UNIX, via the following two system calls: ('->' means 'returns')

```
fork_process (process_num, monitoring_io_port) -> process_num
replace_image (file_num, entry_point)
```

or by loading the process' starting image directly from a file:

```
load_new_process (file_num, monitoring_io_port) -> process_num
```

The latter method is more appropriate for a loosely-coupled multi-processor. It makes little sense to copy the current memory image of the parent process over the network to the processor that is to execute the new subprocess, only to replace it soon afterwards with a new image obtained from a file, as is usually the case. Forking is only provided to facilitate UNIX compatibility.

Note that, unlike UNIX, a process can fork not only itself, but also one of its subprocesses. Furthermore, a parent process can be informed of all state changes in its subprocesses, and can exert complete control over them, with the ability to terminate, suspend, or restart a particular subprocess.

Furthermore, unlike UNIX, a new process is created in a suspended state, and execution must be started explicitly by its parent. This allows a process to load several subprocesses and to set up communication links between them while they are in a known state.

A parent process may also suspend an executing subprocess and then restart execution at a different instruction. The suspend/restart
mechanism is formulated such that the restarted subprocess can execute either an interrupt handler (eventually returning to the point of interruption), or an exception handler (never returning).

As is to be expected, a process can suspend or terminate itself and supply a reason code that will be received by its parent. A process can also dynamically obtain additional memory from the operating system, and release unused memory for use by other processes.

The operating system does not normally allow processes to share access to the same area of memory, since, in general, it is not possible or desirable to ensure that processes execute on physical processors that have access to a common physical memory bank. Memory sharing is therefore discouraged by limiting it to specially privileged 'server' processes. (Server processes are discussed in Section 4.)

It is, however, possible to simulate a form of memory sharing between different 'light-weight' processes executing on the same processor by incorporating a scheduler into the code of a single 'heavy-weight' operating system process. Such 'light-weight' schedulers can ignore most of the fairness, synchronization, and security issues that an operating system must address, and can thus provide a much cheaper form of single-processor concurrency than the operating system. A typical user of 'light-weight' processes would be a server process that must serve several clients concurrently.

3. INTER-PROCESS COMMUNICATION

As Hoare pointed out [1], transferring information from a process to a process does not differ from transferring information from a device to a process, or from a process to a device. In fact, in a modern operating system implementation, device drivers are likely to be implemented by processes.

Consequently, the operating system provides a generalized I/O call as the principal means whereby processes must interact with their environment. This system call corresponds to the classical I/O call of current generation operating systems in most respects, generalizing it only in so far as to allow I/O operations to be performed on processes as well as on files and devices, and by allowing a single operation to perform both output and input.

The I/O call is invoked as follows:

\[
\text{io (io_port, operation, address, out_len, out_buf, in_len, in_buf)}
\]

Note that all the parameters, except out_buf, may be updated by the call.

An I/O port corresponds to a UNIX file descriptor, but may represent another process, as well as a file or device. Furthermore, up to sixteen different I/O operations may be
carried out on an I/O port. For example, when an I/O port is linked to a file, the following operations are supported:

- \( 0 = \) read next string (length given in in_len)
- \( 1 = \) write next string (length given in out_len)
- \( 2 = \) read string at offset from file start (given in address)
- \( 3 = \) write string at offset from file start
- \( 4 = \) append string at end of file
- \( 5 = \) get current length of file (result in in_len)
- \( 6 = \) set current length of file to length in out_len
- \( 7 = \) flush updates to non-volatile storage

As can be seen, some operations are input operations, others perform output, and still others do neither. It is also possible to have operations that perform output as well as input. In general, the subset of operations that can be carried out on an I/O port and the effect of the operations depend on the kind of object to which the I/O port is linked. When an I/O port is used as an inter-process communication link, the programmer has full control over the subset of allowable operations and the interpretation given to them.

An I/O port is linked to a file, device, or server via:

```
open (process_num, io_port, object_num, desired_ops) -> available
```

Note that a process can open not only its own I/O ports, but also those of subprocesses; object_num identifies the file/device/server; desired_ops is a bitmap indicating the subset of the sixteen possible operations that the caller wishes to carry out on the I/O port. (For example, to open a file for sequential read-only access, desired_ops must have a value of 0000000000000001.)

An I/O port is linked to a process via:

```
connect (process_a, io_port_a, bitmaps_a, time_out_a, process_b, io_port_b, bitmaps_b, time_out_b)
```

A process may connect the I/O ports of its subprocesses to each other, or may connect its own I/O ports to the I/O ports of subprocesses. Thus a process can communicate with its subprocesses, siblings, parent, and files/devices/servers.

Note that each I/O port has a bitmap associated with it, indicating which I/O operations may be carried out on it, as well as a bitmap indicating which I/O operations perform output. Additionally, a time-out is associated with each port. These values are explicitly specified for ports opened with connect, but are determined by the object to which the port is linked for ports opened with open.

The general progression of an I/O call is as follows:

- send operation, address, out_len, in_len to the other process (append contents of out_buf if the operation calls for it)
- wait for a response from the other process (return with an error code if no response within time_out)
I/O state transition diagram

Process A

initial state

<table>
<thead>
<tr>
<th>calls I/O to perform output</th>
</tr>
</thead>
<tbody>
<tr>
<td>awaiting initiation</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>initiating I/O</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>I/O incomplete</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>response available</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>I/O complete</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>next I/O call</td>
</tr>
</tbody>
</table>

Process B

initial state

<table>
<thead>
<tr>
<th>calls I/O to perform input</th>
</tr>
</thead>
<tbody>
<tr>
<td>initiating I/O</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>I/O incomplete</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>response available</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>I/O complete</td>
</tr>
<tr>
<td>________________________________</td>
</tr>
<tr>
<td>next I/O call</td>
</tr>
</tbody>
</table>

A receives go-ahead
B transmits input request

A transmits output request
B receives response to input request

A receives response to I/O
B transmits request for next I/O

A transmits request for next I/O
B receives response

Figure 1
- return operation, address, out_len, in_len sent by the other process, and store rest of its response in in_buf (sender out_len = receiver in_len, and sender in_len = receiver out_len)

When an I/O port has just been connected to another, the first operation that outputs the contents of out_buf is delayed until the other process performs an input operation on its corresponding I/O port. The I/O call of the process that performed the input operation then completes, returning the parameters and buffer contents supplied by the process that performed the output operation (resulting in a data transfer taking place; see also figure 1). Meanwhile, the outputting I/O call is suspended while awaiting a response. This response is received when the process that performed the initial input operation performs its next I/O operation. The response consists of the parameters and possibly the contents of out_buf, supplied to the second I/O operation.

Thus, two communicating processes are always synchronized so that the initiation of an operation by one, causes the completion of an operation by the other, as well as a data transfer. The result is that communication takes on a 'hand-shaked' request-response nature, and that when one process sends output to another, the other has already set up an input buffer to receive the transferred data. This eliminates the need for a system buffer pool and takes care of flow-control and synchronization.

When a port is opened to a file/device/server, the operating system engages in a dialogue with a corresponding file-driver-process/device-driver-process/server-process, which results in an inter-process communication link being set up between the client process and a driver/server process. After the open, the driver/server is waiting to complete an I/O operation, and the client process has yet to perform its first operation. When the client performs its first I/O operation on the port, the incomplete I/O at the driver/server completes, resulting in the driver/server receiving the request made by the client. After serving the request, the driver/server responds by initiating a next I/O operation, which in turn causes the client's incomplete I/O operation to complete, returning the desired results.

Naturally, the I/O call allows a proceed option (indicated by setting a modifier bit in operation). A proceed I/O call sends the request to the other process, but does not wait to receive the response. The requesting process can later complete the call and receive the response by performing another I/O call on the port—setting a modifier bit to indicate whether or not to wait for the response, if this has not yet been received.

It is possible for a process to have a number of I/O ports on which proceed I/O operations were carried out. Therefore, it is possible to perform an I/O call that will complete any one of these incomplete calls. This facility will typically be used by a server process that serves many clients concurrently, and thus may be waiting for several requests (that is, have several incomplete I/O calls) at the same time.
The I/O call also supports broadcasting/multi-casting, as well as scatter/gather transfers.

4. SERVERS

Server processes are the principal means by which the operating system provides its services. Furthermore, servers can be used as the basic blocks for building distributed applications.

A server process has a globally visible name, drawn from the same name space used for files and devices. Thus, a prospective client establishes an inter-process communication link with a server by calling open. The operating system carries out a call to open by sending an unsolicited message to the target of the open call. Whether it is a file, device, or explicit server, the target of an open call is always a process; hence the term 'server' will from now on be understood to include files and devices.

When a server starts executing, it sets up a number of 'reconfigurable' I/O ports, using:

```c
make_reconfigurable
    (io_port, valid_op_bitmap, in_len, in_buf, time_out)
```

While reconfigurable, an I/O port does not represent a communication link to any particular process, but acts as a receiving port for unsolicited messages. Thus, when the operating system sends an unsolicited message to a server, one of the server's reconfigurable I/O ports is selected to hold the message, and the server will receive the message when it tries to complete an I/O call on that port (it will usually try to complete an I/O call on any port). Note that make_reconfigurable sets up the buffer to hold the unsolicited message.

The unsolicited message received by the server

a) can be trusted because it comes from the operating system;

b) fully identifies the prospective client, its privileges, the type of access required, and so on.

After receiving such a message, the server must decide whether or not to accept the client and indicate this by outputting an appropriate message through the I/O port that received the request. If the client is accepted, the I/O call indicating the acceptance remains incomplete until the client performs its first I/O operation on the port it has just opened successfully. Thus, when a client is accepted via a reconfigurable I/O port, the I/O port is configured as a dedicated inter-process communication link between client and server. Note that, among other things, the server's acceptance message supplies information to be associated with the client's I/O port: the valid operations, the operations that output out_buf, and the time-out to be used when waiting for a response from the server.

A server process can be tied to a particular processor of a multi-processor system, in which case it is loaded when that processor bootstraps. Alternatively, a server can be 'untied', in which case it is loaded into any available processor when first
Designating a process as a server, thus giving it global visibility, is a privileged operation since a server is trusted to take part in the 'new client' protocol. Servers are also the only class of processes that can be allowed to perform certain privileged operations.

For instance, a server that is tied to a particular processor can be used as a device driver by allowing it the privilege to access I/O space and to install interrupt handlers. A tied server can also be allowed to share memory with other servers tied to the same processor.

It follows that an implementation of the operating system will itself largely consist of a set of server processes distributed among the various physical processors. Adding new servers to the collection making up the basic operating system will be straightforward, resulting in an 'open', extensible, adaptable operating system. Furthermore, structuring the operating system as a set of servers communicating via messages greatly facilitates the transparent integration of local resources (services) with resources available, via a network, from other systems.

The server mechanism also allows a single service (that is, a single object in the name space) to be implemented by several co-operating processes. One way to implement such a multi-process server is to designate several server processes as the members of a single server group, identified by a single 'group object' number. When a client performs an I/O operation on an I/O port linked to a group, the request message is broadcast to all members of the group. It is possible for the structure of the group to be invisible to the client, in which case the members must use a protocol to ensure that only one response will be generated. Alternatively, a client may be required to be aware of the group structure, in which case the client must set up additional I/O ports to receive the multiple responses (an I/O port can receive at most one response for each request).

It is also possible to implement a server group, without using broadcasting, by means of a co-ordinator process. In this case, client requests go to the co-ordinator, and the co-ordinator then 'subcontracts' them to the other processes co-operating to provide the service. A co-ordinator can subcontract all client interaction, in which case the 'request to become a client' message goes to the co-ordinator and is subcontracted, following which all further interaction is between client and subcontractor (unbeknown to the client). Alternatively, a co-ordinator can subcontract individual requests, in which case all requests first go to the co-ordinator and then to the subcontractors.

Note that subcontracting and broadcasting can be combined. It is thus possible to exploit multiple processors to achieve both speed and fault-tolerance, while providing clients with the illusion of dealing with a simple 'single' server.
5. THE FILE SYSTEM

The file system is intended to facilitate the storage and retrieval of arbitrary strings of bytes, such as text files, object files, and process images. The file system is not intended as an efficient or convenient store for records since it is more reasonable to use a database management system to store and retrieve such data.

The file system is implemented as a set of communicating servers, with each open file having a corresponding 'driver' process that actually receives and acts upon the I/O operations that client processes perform on I/O ports linked 'directly' to files. The file system servers are structured as a hierarchy and in such a way that the database management system can exist 'alongside' the file system, by making use of the lower level 'disk block' servers instead of the file system visible at system call level.

Files are identified by numbers drawn from a global name space that includes devices and servers. This allows user interfaces to use arbitrary symbolic name-to-file number mappings, further adding to the 'open', extensible, adaptable nature of the operating system.

Files are allocated in two steps:

\[
\text{create\_temporary\_file (logical\_vol\_num, size\_hint) -> file\_num}
\]

\[
\text{make\_file\_permanent (file\_num, future\_expansion\_hint)}
\]

Temporary files are automatically reclaimed when the process that allocated them terminates. By converting these temporary files into permanent files, rather than directly allocating permanent files, it is possible to follow a protocol that results in the allocation of a file and the recording of a symbolic name-to-file number mapping as an atomic operation.

The size hints that may be given when a file is allocated allow the operating system to pre-allocate space for a file so that sequential access is optimized. The file system need not heed the hints and will never refuse to allocate or grow a file because a suitable run of disk blocks is not available.

The design of the file system attempts to minimize the visibility of physical disk volumes to application programmers in order to promote program transportability: All permanently on-line storage media are consolidated into logical volume zero and files from all volumes are identified from a single name space.

However, it is necessary to introduce the concept of distinct volumes to cope with removable disk packs. Clearly, the operating system cannot just incorporate removable disk packs into a global pool and allow arbitrary files to be allocated on removable disk packs. Consequently, removable disk packs are grouped into one or more logical volumes, any of which can be off-line. File allocations on these volumes must be indicated explicitly and can only be performed by suitably authorized users.
6. Security

Users gain access to the operating system by interacting with a user authentication server. It is possible to associate an arbitrary authentication server with a given user access device, and to restrict a given server to admitting only a subset of users. Thus, it is possible to make it reasonably easy to gain entry as a casual user, while making it arbitrarily difficult to gain entry as a privileged user.

When a user gains entry into the system, the corresponding user access device is assigned to an appropriate user interface process, which is 'executing on behalf of' the user. The 'executing on behalf of' property of a process is inherited by all subprocesses, and can only be changed if a process is a specially privileged server process (for example, an authentication server). Thus, all actions initiated by a user are carried out by processes authentically executing on behalf of the user.

Files, devices, and servers all have owners, and are accessible only to processes executing on behalf of the owner or a user explicitly mentioned in an access list - a file listing all the users that may access the protected object, with each entry indicating an individual set of privileges for the user. (Note that a server may reject a client even if the client is listed in its access list.)

Additionally, files, devices, servers, and processes are associated with 'information categories'. An arbitrary number of information categories can be established, and the operating system enforces a set of rules that ensure that information cannot 'flow' from one category to another, unless specifically permitted. This is basically achieved by limiting a process to accessing only objects associated with the same information category, while also allowing a process to have read-only access to objects associated with categories that have flow paths to the category of the process. A process may only change its category if its user is permitted to operate in the new category and if the change cannot result in an illegal flow of information.

The operating system also includes mechanisms for limiting software piracy, denial of resources, Trojan Horse attacks, and the use of covert channels to subvert information flow controls.

7. DATABASE SUPPORT

Database managers are principally supported by the server mechanism. Furthermore, the file system is structured in such a way that a database manager can access disk blocks directly.

The only additional support provided by the operating system is in the form of a transaction co-ordinating server. This server is accessed by client processes via the system calls:

\[
\begin{align*}
\text{start\_transaction} \\
\text{commit\_transaction} & \to \text{success\_indicator} \\
\text{abandon\_transaction}
\end{align*}
\]
These calls keep the transaction co-ordinator up to date on the transaction status of processes, and it, in turn, keeps 'transaction supporting' servers up to date on the transaction status of their clients.

The transaction co-ordinator co-ordinates a two-phase commit, provides a 'centralized' deadlock detection service, and allows transaction supporting servers to use a wide range of synchronization strategies, including optimistic strategies.

While it is logically a centralized service, the transaction co-ordinator can be implemented as a distributed set of co-operating processes, and thus does not preclude the implementation of a distributed database manager.

8. REAL-TIME SUPPORT

The operating system allows the minimum rate of execution of a process to be specified directly rather than be influenced indirectly by means of priorities. Furthermore, processes are classed as 'real-time' or 'ordinary', with real-time processes pre-empting ordinary processes. All operating system server processes execute as ordinary processes, thus giving real-time application programmers complete control over processor allocation.

Note also that server processes can be granted the privilege of handling interrupts directly, as well as to share memory with similar servers on the same processor. Furthermore, the operating system can be configured with a separate 'real-time' file system that serves only real-time processes and thus isolates real-time processes from interference by non-real-time processes in this aspect as well.

9. CONCLUSION

This paper has briefly outlined the design of an operating system that is hardware independent, provides a high level of data security, and supports distributed applications, database applications, as well as real-time applications.

This has not been achieved by introducing radically new concepts, but by carefully reformulating, generalizing, and integrating the basic concepts that any operating system must support.

Firstly, the process concept has been stripped of restrictions and assumptions, leaving a straightforward mechanism that can be utilized efficiently for a wide range of applications. Secondly, the input/output mechanism has been generalized to provide a straightforward, efficient form of inter-process communication. Thirdly, the concept of a device has been generalized into the concept of a server, on which the operating system itself is largely based. Servers make the operating system flexible and extendible, and provide a suitable mechanism for distributed implementations.
REFERENCES

Communicating Sequential Processes
Communications of the ACM, 21(8), August 1978, pp. 666-677

Adapting UNIX for a multiprocessor environment
Communications of the ACM, 29(9), September 1986, pp. 895-901

[3] Venter, B.H.
The Design of a new general-purpose operating system
Ph.D. Thesis, University of Port Elizabeth, April 1987

Perkin-Elmer Corporation, New Jersey, 1984