University of Groningen

Modelling of virtual memory computer systems with multiple classes of processes

Schoute, Albert Louis

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1978

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Schoute, A. L. (1978). *Modelling of virtual memory computer systems with multiple classes of processes*. s.n.

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 08-06-2022

Preface and Summary

The impact of many of the design decisions and scheduling rules in general purpose interactive computer system is often scarcely known owing to the complexity of the system itself and of the user demands.
Analysis and modelling of computer systems has therefore developed into an important subject in computer science.

Good interactive service to many users relies on a number of technical developments which have facilitated the preemption and reallocation of the main computer resources, such that service can be given at a rather equal pace. This has the obvious advantage that small demands are not delayed by large demands, so ensuring automatically that fair service is maintained.

Another consequence of the division of the computer resources amongst many users is that the utilisation of these resources is not so much dictated by the specific characteristics of the user demands but more by the allocation policy of the system itself and by the global characteristics of the workload presented by the users.
This may explain the relative success of the application of simple queueing network models as a representation of the dynamic execution of user programs which are simultaneously in a state of execution (multi-programming).

One of the technical developments which has improved interactive systems is the virtual memory technique (often implemented by means of paging), which makes the dynamic use of main memory considerably simpler. In respect of the use of main memory, it also seems desirable to divide it in a simple manner between user program which require service. However, the allocation of main memory appears to be a rather critical aspect, on the one hand because a sufficient allocation (according to the individual needs) is a prerequisite for satisfactory execution and on the other hand because reallocation implies unloading and reloading of parts of main memory and this causes processing overhead and delays.
The memory allocation may be considered as a flood-gate which plays an

important role with respect to a good interactive service and a good utilisation of the central processing capacity.

The optimal allocation of main memory is the most important objective of a classification of processes, according to their expected demands. As in the Edinburgh Multi Access System (EMAS) [Whitfield and Wight 1973, Shelness et al. 1974], which has to a large extent been used as the object of this study, the expectations could be based on the demands in the recent past.

In this thesis we consider modelling aspects of virtual memory systems with multiple classes of processes. The existence of these different classes could be considered as being in some way representative of the different types of demands generated by users.

Because of many practical aspects (such as limitations due to the finiteness of memory), queueing network models are either too complicated to be solved or form only a partial representation of the actual servicing. A great deal of mathematical analysis is therefore being done in relation to the performance analysis of computer systems. A generally used (and in many cases inevitable) way of solving queueing models for computer systems is by means of a hierarchical decomposition.
Courtois [1977] considers in his book on decomposability the justification of this type of solution method and the possible applications to computer system models. In this thesis decomposability aspects are treated by different methods.
We give in Chapter 1 a representation for the stationary solution of finite state Markov processes, which makes use of graphs. This representation, called the tree-form solution, is used to show the essential characteristics of systems which have a simple product-form expression as solution. In this framework questions of decomposability and state aggregation become more transparant and may lead to new ways of estimating the measure of decomposability (section 4.2).
Product-form queueing networks are in Chapter 2 related to special (balanced multi-dimensional birth and death processes.

Different levels of virtual memory system modelling are treated in Chapters

3 to 5. In Chapter 3 the basic operation of virtual memory systems with paging is sketched and basic assumptions concerning process behaviour are considered. It is shown that for adaptive memory allocation policies (e.g. working-set algorithms) the single-lifetime-function models are not adequate as description of the paging behaviour. Paging behaviour may be considered within the context of a classification system of processes. The effects of a selective allocation system, such as that implemented in EMAS, are analysed in sections 3.5 to 3.7.

In Chapter 4 we consider the use of queueing network models. The results of queueing network models with multiple classes of processes [Baskett et al 1975] are presented within the context of Chapter 2.
We investigate how more precise modelling with respect to priority service of the CPU and bulk service by the paging drum may be achieved.

At the global level the memory admission is an important factor, especially when different classes of processes are considered.
We give in Chapter 5 an account of the extent to which product-form models can be used in this area. Memory allocation requirements of different sizes make a realistic modelling of the admission process (by analytical means) extremely difficult. Either crude approximations must be made or numerical methods have to be used.

Some results from experiments on EMAS are given in the last Chapter and a number of modelling examples are worked out.