VAX/VMS操作系统

进程、调度、管理、控制及通讯

国防科技大学研究所六〇二教研室

一九八五年六月

前　言

我们系统地研究和剖析了VAX/VMS操作系统。在此基础上，将陆续地创出一批资料，介绍ＶＭＳ操作系统。

在这部分资料中，主要介绍ＶＭＳ操作系统中，有关进程管理，进程低调，进程控制及进程通讯等方面的内容。主要从操作系统内部出发，阐述基本概念，数据结构，及实现的方法和策略。因为这一部分是ＶＭＳ操作系统最核心的内容之一，所以对许多问题，我们作了较详细的介绍和分析，核心的程序模块也给出了实现的框图或流图。

文章中错误的地方，请批评指正。

第一章　概　　述

VAS/VMS操作系统是VAX—11/780机的核心资源。它是一些控制程序和例行程序的集合。ＶＭＳ提供了范围很广的系统服务。系统服务只是一些过程。作为操作系统的一部分。以控制系统资源的合理使用。提供进程间的通讯和控制。实现操作系统的基本功能。

§１・１　系统的分层设计

VAX/VMS操作系统的结构是按层次设计法设计的。它的层次结构见图１—１。进程的调度和控制在层次结构的核心层。VAX/ＶＭＳ系统中的调度程序是使各进程得以合理运行的根本。调度主要负责各进程合理的占用ＣＰＵ。控制进程在各状态间的转换等等。

ＶＭＳ的核心层由三个相互独立的子系统组成。在各子系统内部是以传统的模块化方法设计的。三个子系统中。虚贮管理子系统和Ｉ／Ｏ子系统均有相应的介绍材料。本部分仅介绍：进程控制子系统的主要内容。

大部分的系统服务（也称广义指令）运行在ＶＭＳ的核方式（Ｋ）。ＶＭＳ中的系统服务是做为用户进程与操作系统间的界面。外部程序只能通过调用系统服务。才可能得到操作系统的支持。系统服务在ＶＭＳ中的分派控制在中断系统中介绍。详细的说明见相应的介绍材料。记录管理服务（ＲＭＳ）在执行方式下运行（Ｓ）。文件系统的主要内容见“ＲＭＳ服务介绍”相应的分派控制也在中断系统中介绍。操作系统的最外层是命令解释程序。命令语言（主

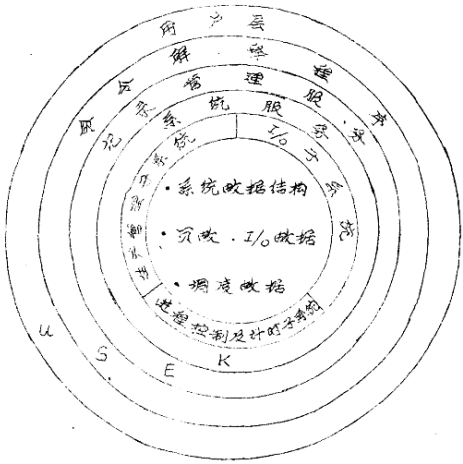


图 1 — 1　VAX/VMS 操作系统的分层设计

要是 D C L 语言）是终端用户的系统间的界面。V M S 的命令语言是功能很强的语言，有关的细节见相应的资料。在用户层主要运行着用户进程和大部分的系统实用程序，以及库程序。

在介绍这一部分之前，我们详细阅读了大量的资料，并用“反汇编器”，反出并阅读了若干核心的程序模块，基本搞清了V M S 的主要控制结构。文章中的框图给的稍粗了些，主要是为了阅读方便。文章中肯定有不少不成熟的东西，甚至有许多错误，望批评指正。

§1·2 VMS提供的访问方式

从图1—1可以看到，VAX/VMS共分了四层。VAX/VMS分别给各层赋予不同的访问特权。离核心越近，所具有的特权越高。四种方问方式为：

内核方式（K）：供操作系统核心使用（包括页面管理，调度和I／O子系统），以及大部分系统服务。

执行方式（E）：部分系统服务调用和记录管理系统。

管理方式（S）：用于命令解释之类的服务。

用户方式（U）：用于用户级程序代码，实用程序，编译程序，调试程序等。

每种访问方式在每进程的进程控制区内均有它自己的堆栈。因此，每个进程可有四个堆栈。每种访问方式一个堆栈，而所以四种访问方式可存在于同一个虚拟地址空间内。

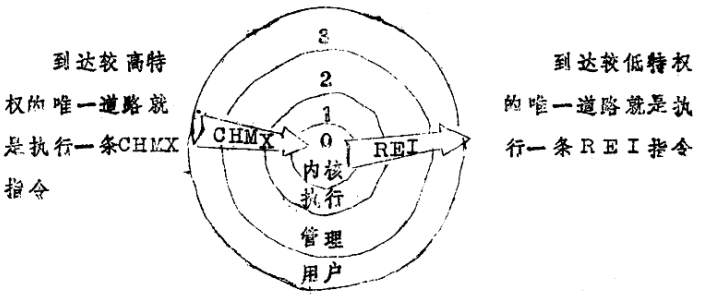


图1—2 改变访问方式的方法

由于不同的访问方式决定了不同的特权，所以，外层为了得到内层的服务必须改变其访问方式，为此，VAX/VMS规定了改变访问方式的方法，即利用CHMX和REI指令实现各种方式间的转换。

§1·3 进程及其有关概念

① 进程及进程关联：

· 进程：

所谓进程就是一个具有独立功能的程序关于某个数据集的一次运行活动，在VAX/VMS系统中，一个进程是由系统软件调度的可执行的基本实体，它提供了映象执行的内容，所以进程也是完成用户作业工作的基本实体，当用户向系统注册时，系统就为程序映象的执行建立一个进程。

· 进程关联

一个进程是由虚地址空间及硬件和软件的关联所组成，见图1－2。

进程的软件关联是由系统对进程进行调度和控制所需的所有数据组成，它主要包括进程的软件优先级、当前的调度状态、进程特权、限额和限制及其它一些信息，如进程名、进程ID等等。

软件关联包括：进程控制块PCB，作业信息块JIB以及进程标题PHD，见附录。

硬件关联的信息主要保存在硬PCB中，而且硬PCB结构本身又是嵌套在进程标题结构中，硬PCB中的域见附录。

在过程关联中，软PCB空间是在建立进程时，动态地从系统

空间中申请，并常驻内存。进程标题位于系统空间的平衡槽中，可被换出内存。

② 关联转换：

当前运行的进程退出运行时，保留进程现场的工作和可执行进程调度执行时恢复现场的工作，称之为关联转换。关联转换主要是涉及硬PCB中的各种寄存器内容。由SVPCTX、LDPCTX 指令完成。

③ 映象：

在进程关联中执行的程序称为映象。一个映象由若干过程和数据组成，这些数据和过程已由连接程序在一起，即它们已解决了模块间的符号引用，以及分配地址空间。

④ 工作集：

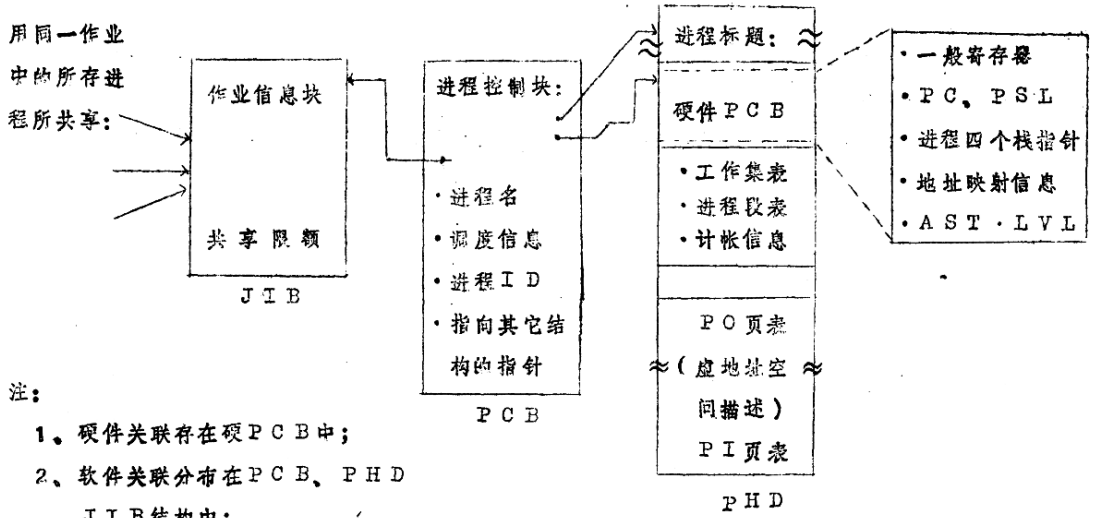
一个进程执行时，该进程虚地址空间页面的一个子集驻留在物理存贮器中，这个子集就称为进程的工作集。

⑤ 平衡集：

在任何一个时刻驻留在物理存贮器中的所有进程工作集的集合称为平衡集。

⑥ 作业：

作业是完成一项特殊任务的进程族。在进程族中，具有父子关系的一组进程，协同完成一个用户提交的作业。



注：

1、硬件关联存在硬PCB中；

2、软件关联分布在PCB、PHD JIB结构中；

3、“虚地址空间的描述”存贮在PO和PI页表中；

4、用户程序及控制数据分别占据进程虚地址空间中的程序区和控制区。

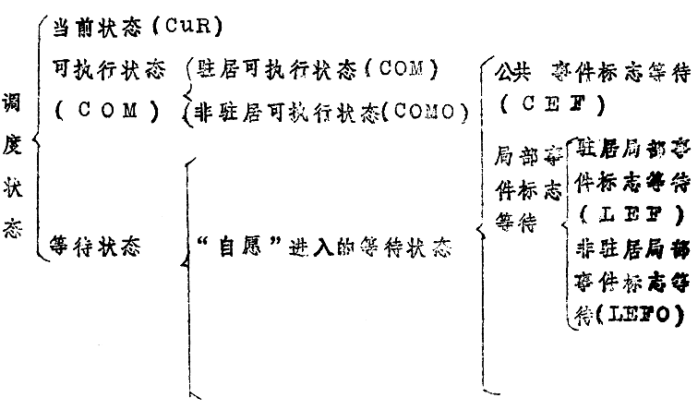
图1—2　进程关联

第二章　进程调度
状态及调度控制

所谓进程调度，指的是处理机调度，即“低调”。它主要完成各进程在调度状态间的转换，以及按照选定的策略和算法动态地变化进程的优先权并选择优先权最高的进程在CPU上运行。在多道进程程序的系统中调度策略的选择是操作系统的核心问题。VAX/VMS操作系统采用了一种较好的调度策略，充分体现了“分时”和“实时”的特征。本章主要讨论：进程的调度状态；驱动状态变化的系统事件；软件优先权；以及调度程序。

§2·1　调度状态及数据结构的表示

VMS操作系统将所有进入系统的进程分为如下的调度状态：



- 睡眠等待
  - 驻居睡眠等待(HIB)
  - 非驻居睡眠等待(HIBO)
- 挂起等待
  - 驻居挂起等待(SUSP)
  - 非驻居挂起等待(SUSPO)
- 存贮管理有关的等待状态
  - 页故障等待(PFW)
  - 自由页等待(FPG)
  - 冲突页等待(OLPG)
- 其它“杂类”的等待状态(MWAIT)

以下简单说明各种状态及数据结构表示：

下面所说的数据结构大都是定义在系统虚地址空间中的SDAT模块中。

1。当前状态(CUR Current state)

表示在某一特定时刻，正在CPU上运行的进程状态。因为单单处理机环境下，处于COR状态的进程最多只能有一个。故数据结构也较简单。SDAT模块中设有一指针：SCH $ GL—COR PCB指向当前正在运行进程的软PCB地址。

2。可执行状态(COM Computable state)

即所谓“就绪状态”，处于该状态的进程已具备了运行条件，只等时机成熟就获得运行。VMS是一个虚拟存贮的操作系统。所以把可执行状态又分为：存贮驻居的可执行状态(Computable

resident state）记为COM，和非存贮驻居的可执行状态（computable outswapped state）记为COMO。

VMS的调度算法是根据进程优先权的高低选择进程执行的。而系统中唯有处于这种状态的进程可能被调度运行。为了减少调度中选择最高优先权运行而带来的系统开销，对COM和COMO状态的进程设计了精巧的数据结构。

在SDAT模块中，根据VMS中的0～31个优先级划分，分别为COM和COMO状态建立了三十二个队列，和二个队列概览长字（QUEUe summary longword）。见图2－1。

图中表示的是32个队列头的结构，其中每一队列头由四倍字（quadword）组成，表示对应优先级队列的头、尾指针。处于该队列的每一软PCB用其自身的连接字，维持这条链。

“队列概览长字”由32位组成，每一位对应一个队列，当某一位为0时，表示对应的队列为空。

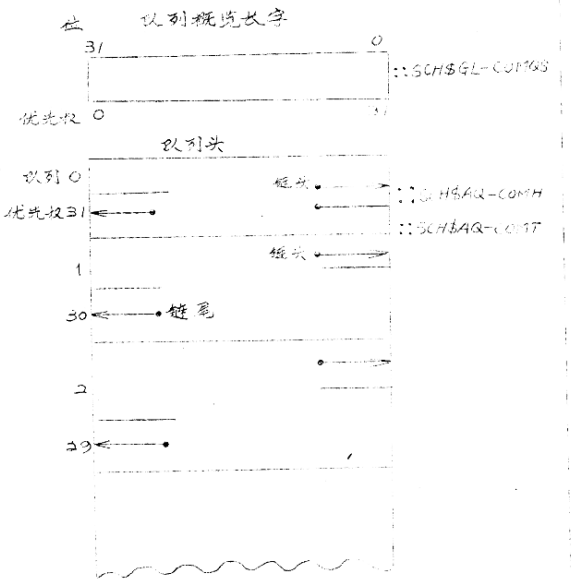
COMO状态的进程要由交换程序将其换入内存后，才有可能被调度运行。Com队列中的进程位是唯一能被调度执行的进程状态。

3、公共事件标志等待（CEF：common event Flag wait state）

等待一个或多个公共事件标志的进程到“公共事件块CEB”上去排队。公共事件块CEB是从系统空间的“非页动态存贮区”中分配的，队列的头指针SCH$GQ—CEBHD定义在SDAT模块中。有关这一状态的详细实施及数据结构在第四章中讨论。

以下各种等待状态，它们各自的状态队列具有相同的结构，每一队列头均由三个四倍字组成。见图2－2：

COM 状态队列：



COMO 状态队列：

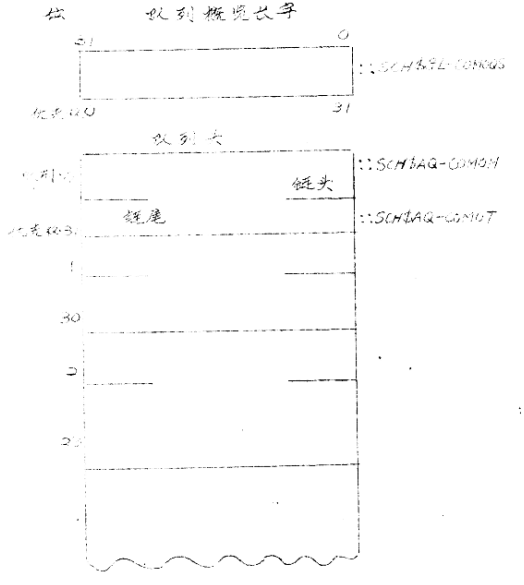


图 2 — 1　com　como　状态队列

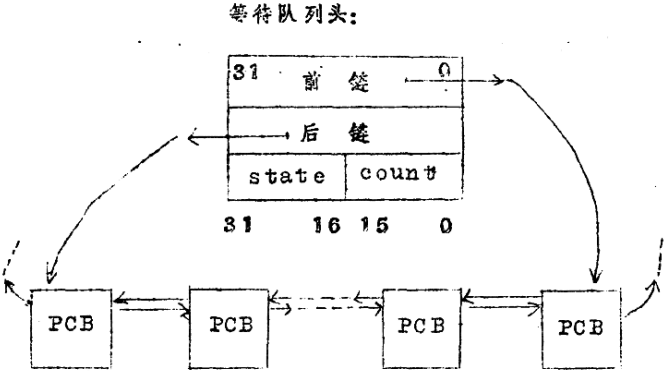


图2－2　除CEF状态外的所有等待状态的队列形成

4、志愿(Vluwntary)等待状态:

i) 局部事件等待LEF：根据是否驻居于内存分为LEF LEFO两个状态。用两条队列表示，形式如图2－2。从CUR状态变为LEF状态。一般是由系统服务：$WAITER 、$WF LOR $WFLAND等引起的；

ii) 睡眠状态(hibernate)：也分为驻居和非驻居两种：HIB 、HIBO 。状态队列如图2－2“睡眠”系统服务使进程从CUR状态转换成HIB状态。“$WAKE 系统服务、AST排队。删除进程“事件”，则能使HIB或HIBO转换成Com 或como 状态。

iii) 挂起状态（suspend）：也分为驻居和非驻居两：SUSP、SUSPO。状态队列如图2－2。同样。“挂起”系统服务能使进程从CUR状态变为CURO状态。而“$RESUME系统服务和进程删除”，则能使被挂起的进程变为com或Como状态。

5、存贮管理的等待状态：

与存贮管理有关的三个等待状态，不论是否驻居内存，均合并在各自的状态队列中“某一进程被换出（或换入）”时，只是将该进程软PCB中PCB$L—STS域中的PCB$V—RES位清成0（或置成1），不需移动队列。

i) 页故障等待状态（PFW：page fault Wait state）进程所需的页面不在内存时，出现页故障。当“故障页面”需要从虚空间读入时，该进程被放于PFW状态。

ii) 自由页等待状态（FPG：Free Page Wait state）：当进程需要在自己的工作集中增加一页，但自由页表已没有可分配的自由页时，进程进入FPG状态。这实际上是一个资源等待。要等待记过“修改的页面写回”、“进程出交换”或“虚拟空间删除”等方法提供了自由页面后，才可能变为com或como状态。

iii) 冲突页等待状态（COLPG colided page wait state）：

出现这种等待一般是因为若干个进程同时出现了“共享页面”的页故障。最初故障的进程进入了PFW状态，而第二个或以后的进程则进入COLPG状态。另外，当某个进程需要一个局部（priUate）页面，而该页面又已经正在向内存读入时，该进程也进入COLP状态。

当故障页面读入后COLPG状态才可变为com或como状态。

6、其它杂类的等待状态（mwAIT misceLLanous wait state）：

mWAIT状态表示进程等待那些未包括在其它等待状态中的资源。这些所有的杂类等待，都链在一条队列中，如图2－2。这些等待的原因，也可分为两类：互斥信号（mutex）等待和资源等待。详细内容列在表2－3中（请见下页）。

i）资源等待：

资源等待状态是等待那些当时已经用完了的或锁住的某些资源。如非页式的动态存贮区已经用完，或公共信箱已经没有空余。等到那些资源重新可用时，处于MWAIT状态的进程才变成com或como状态。

对应资源等待的软PCB中的PCB$L—EFWm域存在的是：所等待的资源号（$RSNDEF宏定义的小整数）。

ii）互斥信号量等待：

表 2－3　　MWAIT 状态的类型

等待的原因	PCB$L—EFWm 域的内容（软 PCB 中）	
互斥信号等待	符　　号	数值（16进制）
系统逻辑名表	LOG$AL—muTEX	8000 24F4
同组逻辑名表		8000 24F8
I／O 数据基	IOC$GL—muTEX	8000 2620
公共事件块表	EXE$GL—CEBmTX	8000 2624
页式动态存贮区	EXE$GL—PGDYNmTX	8000 2628
全局段描述字表	EXE$GL—GSDmTX	8000 262C
共享存贮器全局段描述表	EXE$GL—SHMGSMTX	8000 2630
共享存贮器信箱	EXE$GL—SHMmBmTX	8000 2634
入队／出队表（未使用）	EXE$GL—ENQMTX	8000 2638
已知的文件项表	EXE$GL—KFIMTX	8000 263C
行打印机部件控制块	uCB$L—LP—MuTEX	注 1
资源等待	符　　号	数值（16进制）
AST等待（等待系统或特殊核AST	RSN$—ASTWAIT	00000001
信箱满	RSN$—MAILBOX	00000002
非页式的动态存贮区	RSN$—NPDYNmEm	00000003
页文件满	RSN$—PGFILE	00000004
页式动态存贮区	RSN$—PGDYNMEM	00000005
等待广播（broadcast）信息	RSN$—BRKTHRu	00000006
映象激活锁	RSN$—IACLOCK	00000007
作业共享限额（pooledquota）等待（未使用）	RSN$—JQuOTA	00000008

注：行打印机部件的互斥信号量无固定的地址，信号量的值依赖于该部件所分配的 uCB。

互斥信号量是用于保护某些数据结构访问的。若某个进程未获得所申请的信号量使用权，则系统例程便将其置为信号量等待状态，变为MuAIT。

对应互斥信号等待状态的软PCB中的PCB$LEFLOM域存放所需信号量的系统虚拟地址。

各进程调度状态间的转换关系见图2－4。

§2·2 与调度和调度状态有关的进程关联

31 软件进程控制块（软PCB）： 0

SQFL（前链长字）	
SQBL（后链长字）	
PRI（当前优先数）	
PHYPCB 硬PCB物理地址	
STS 软件状态标志	
PRIB 基优先数	STATE调度状态

详见附录。