

注：教师姓名后留有一个空格，后面填写教师职称。下面加下划线。

阅后删除此文本框。

**本科毕业设计**

**基于x86架构的微内核操作系统的实现**

|  |  |  |
| --- | --- | --- |
| **学生姓名** | **：** | 郑嘉滨 |
| **学号** | **：** | 171543346 |
| **学院** | **：** | 互联网金融与信息工程 |
| **专业** | **：** | 计算机科学与技术 |
| **指导教师** | **：** | 邓伟林 **职称：** |
| **提交日期** | **：** | 年 月 日 |

**本科毕业论文（设计）诚信声明**

本人郑重声明：所呈交的本科毕业论文（设计），是本人在指导老师的指导下，独立进行研究工作所取得的成果，成果不存在知识产权争议，除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

 学生签名：

时间： 年 月 日

**关于论文（设计）使用授权的说明**

本人完全了解广东金融学院关于收集、保存、使用学位论文的规定，即：

1.按照学校要求提交学位论文的印刷本和电子版本；

2.学校有权保存学位论文的印刷本和电子版本，并提供目录检索与阅览服务，在校园网上提供服务；

3.学校可以采用影印、缩印、数字化或其它复制手段保存论文。

本人同意上述规定。

学生签名：

时间： 年 月 日

摘 要

移动增值服务是在通信技术、计算机技术、互联网技术不断发展融合的基础上，在人们对以信息为基础的各种应用需求快速增长的激励下，在社会信息化水平日益提高的前提下，迅速发展的一种全新的服务方式。

随着中国联通CDMA 1X网络的建成，增值业务计费模式与以往话音计费有很大的不同，故对计费结算在增值业务支撑方面提出了新的要求。针对CDMA 1X增值业务计费系统的大并发量、高可靠性及可扩展性的要求，传统的集中式处理难以满足这些要求。而可以提供并行处理的分布式系统其巨大的灵活性可以应付增值业务爆炸式的增长。

通过对分布式系统的特点、设计中出现的问题以及软件可靠性等方面的分析，在系统故障表现形式以及故障处理等方面进行了研究，最终提出使用容错技术提高其可靠性。

在结合分布式增值业务计费系统进行可靠性设计时，分别从系统的硬件和软件两方面对系统进行设计。在硬件可靠性设计方面，采用双机容错冗余设计以及利用网络通信机制进行硬件状态查询。在软件方面，采用对各个处理线程的状态检查点监控，当线程出现故障时能够恢复或者重启，使得线程能够正常运行。

在可靠性设计的基础上，完成了分布式增值业务计费系统的可靠性实现。通过性能分析表明，研究成果和所描述的设计思路和实现方法是合理的、切实可行的，对于类似的分布式计费系统开发具有一定的参考价值。

**[关键词]**：计费；分布式系统；容错；软件可靠性

Abstract

Mobile value-added service is on the basis of communication technology, computer technology, Internet technology are being developed and merged constantly, with people‘s encouragement to various kinds of the application demand to increase fast based on information, the premise that the social information-based level improves day by day, a kind of brand-new method of service to develop rapidly.

Build up by network , CDMA 1X with CHINAUNICOM, value-added service charge mode and past speech have difference to charge, so settled account in charging and put forward the new demand in supporting in value-added service. To great erupting simultaneously amount, high reliability and request of the expansibility for CDMA 1X value-added service charge system, traditional centralized difficult to satisfy with these request. And the distributed system undertakes the parallel processing; its enormous flexibility should deal with the exploding growth of value-added service.

By These researches on pattern of trouble expression, analysis on the characteristic of the distributed system, problems in the design and software reliability, finally carry on improving its reliability by fault-tolerant technology.

With the reliability designs of the distributed value-added service charge system, carry on the reliability design in two respects of the hardware and software. In reliability design of the hardware, adopt one pair of machines to design fault-tolerantly and redundantly, the communication mechanism of the network to carry on the state inquiry of the hardware. In software, monitor each state check point which deal with the thread, so resume or restart when break down in thread, and make thread can normal running.

Based On the reliability design, the reliability of the distributed value-added service charge system is realized. Indicate through the analysis of performance, it is reasonable and practical to design ideas and methods showed in this thesis. And these are valuable for developing similar systems.

**[Key Words]:** Billing; Distributed system; Fault-tolerant; Software reliability

目 录

[摘 要 I](#_Toc65841343)

[Abstract II](#_Toc65841344)

[目 录 III](#_Toc65841345)

[1 绪论 1](#_Toc65841346)

[1.1 研究背景 1](#_Toc65841347)

[1.2 国内外研究现状 1](#_Toc65841348)

[1.3主要研究工作 1](#_Toc65841349)

[1.4 章节安排 1](#_Toc65841350)

[2 相关技术基础 1](#_Toc65841351)

[2.1微内核架构 2](#_Toc65841352)

[2.1.1 微内核架构简介 2](#_Toc65841353)

[2.1.2 微内核相对于宏内核的优点 2](#_Toc65841354)

[2.2 x86架构 2](#_Toc65841355)

[2.2.1 Intel 8086 2](#_Toc65841356)

[2.2.2 Intel 80386 2](#_Toc65841357)

[2.3 NASM 2](#_Toc65841358)

[2.4 GCC 2](#_Toc65841359)

[2.5 Bochs虚拟机 2](#_Toc65841360)

[3 x86架构中内核的总体设计和实现 1](#_Toc65841361)

[4 结论 1](#_Toc65841362)

[参考文献 2](#_Toc65841363)

[致 谢 3](#_Toc65841364)

[附录 内容名称 4](#_Toc65841365)

**基于x86架构的微内核操作系统的实现**

1 绪论

1.1 研究背景

操作系统（Operating system，OS）是一组主管并控制计算机操作，运用和运行硬件、软件资源和提供公共服务来组织用户交互的相互关联的系统软件程序，同时也是计算机系统的内核与基石。第一个操作系统（同时也是第一个批处理操作系统）是20世纪50年代中期由General Motors开发的，用在IBM 701上。**[1]**

在第一个简单批处理操作系统出现后，操作系统的发展经历了多道批处理系统、分时系统、实时系统、分布式操作系统等阶段，不同类型的操作系统有不同的应用场景。而在早期开发操作系统时，设计者通常只将关注点放在如何得到高效率上，对系统的设计缺乏首尾一致的思想**[2]**，导致随着系统的变大，整体呈现处混乱和安全性低的态势。不仅如此，随着代码的剧增，操作系统的可维护性大大降低，给维护人员增加负担。于是，操作系统结构化的思想应运而出。

操作系统的结构先后经历了无结构OS、模块化结构OS、分层式结构OS、微内核结构OS。现代的操作系统也可以大致分为宏内核OS和微内核OS。从当代主流操作系统的内核架构中，我们可以很清晰的看到，宏内核在当今的主流的操作系统架构中占据着绝对的优势。但是，如今操作系统的各种问题的出现，反过来暴露了宏内核的各种弊端。**[3]**宏内核一般包含了操作系统的大部分功能模块，代码高度集成，某个模块出现一个bug都有可能使整个系统崩溃。与之相对的，微内核架构的基本原则是机制跟策略相分离，把策略放置在操作系统较高层次，而机制置于微内核之中，也仅把系统最为核心、最为基础的模块放置于微内核中。同时，微内核操作系统给在外部空间的服务分配私有、独立的地址空间。这样，各服务间的隔离性得到保障，一个服务的崩溃不会对其他服务的正常工作产生影响。然而凡事都有两面性，微内核将部分服务放在用户空间必然导致进程之间的通信（Inter-process communication，IPC）频率的剧增，同时IPC带来的是频繁的上下文切换，即不断切换用户态和内核态。经过统计，执行微内核操作系统的绝大多数相关方法在用户态会产生一定的性能流失，约5%-10%左右。**[4]**但如今的计算机硬件性能与日剧增，考虑到微内核架构给整个系统带来的安全性，性能上的一点损失在一些条件下完全可以接受。

综上所述，相对于宏内核的高度集成和臃肿的设计架构，微内核优美的结构具有明显的优势，而这种优势随着当前嵌入式物联网技术的飞速发展愈发明显。微内核架构明显是未来的操作系统架构的趋势。

1.2 国内外研究现状

RC4000可能是第一个具备微内核架构思想的系统。这一系统可以用逐层添加操作系统的方法逐步扩充，以适应在进程调度与资源分配方面格式各样的不同需求。微内核术语早在1981年之前就出现了。**[3]**宏内核内高度集成的易变更的设备代码、协议、文件系统服务等使得宏内核随着规模的增大需要开发人员对这个内核代码进行复杂而又细致的管理，同时在某一项服务需要更升级更新时需要对整个内核的代码进行重新编译。为了克服这些问题，微内核思想由此产生。通过将一些需要频繁更新和修改的服务像其他用户服务一样，在用户空间实现这些服务，同时采用策略与机制分离的原则。机制放置于内核，策略放置于用户空间。

如今，经过数十年的发展，微内核已经发展到了第三代微内核。公认的第一代微内核操作系统是Mach，是由卡内基梅隆大学开发，为了用于操作系统研究，特别是分布式与并行运算上。但由于Mach初代内核提供了较多的服务，第一代Mach内核也被称为“胖微内核”。但Mach却有良好的跨平台性，Mach设计伊始将多处理器支持纳入考虑，是为千种处理器而设计的，移植到其他体系结构很容易。第二代微内核以L3和L4为代表，只提供最基本的操作系统服务并且注重性能优化。最显著的一点在于IPC中的优化，在同样的系统中Mach需要114毫秒来发送即便是最短的消息，而L3可以用少于10毫秒的时间来发送同样的消息。**[5]**L4内核是在L3的基础上发展起来的，对L3内核中具有的特性再次进行缩减， 就像L4/MIPS作者们所表述的：“一项特性当且仅当安全需要它在特权模式被实现时才应该在微内核里”。**[6]**由于它十分简单，L4很快被证明是具有高可移植性的。第三代微内核则主要重视系统的安全性问题等，其主要代表有：seL4、Fiasco、NOVA等。**[3]**

微内核不仅在学界取得了广泛的关注和影响，在业界也有逐渐颠覆宏内核的趋势。安卓(Android)操作系统是一个基于Linux内核与其他开源软件的开放源码的移动端操作系统。据操作系统行业数据分析显示：2019年手机端Android操作系统市场份额达68.63%。然而Google并没有止步于Android，在2016年之前Google就已经着手研发用于嵌入式系统的全新操作系统Google Fuchsia，这款OS不再是基于Linux这类宏内核，而是基于一个全新的名为Zircon的微内核；无独有偶，华为在2019年8月9日的开发者大会上，正式发布了鸿蒙操作系统（Harmony OS），一款全新面向全场景的分布式操作系统，而这款操作系统也是基于微内核架构的。

综上所述，由内核架构在学界和业界的发展历史可知，微内核架构的操作系统的已经是大势所趋，宏内核开始有被摒弃的迹象。尽管以当前市场上流行的操作系统来看，似乎印证了Linus Torvalds在那场和Andy Tanenbaum（MINIX之父）的口水仗所说的，微内核仅具有结构上的优美，而在性能上却完全不及宏内核。但日后，随着学界和业界对微内核的瓶颈问题——IPC的优化，同时凭借着微内核的良好的可移植性和可维护性，微内核还是非常有可能成为未来的操作系统的主流架构。

1.3主要研究工作

1.4 章节安排

2 相关技术基础

2.1微内核架构

2.1.1 微内核架构简介

现代的操作系统的内核通常分为三种：微内核（Micro kernel）、宏内核（Monolithic kernel）和混合内核（Hybrid kernel）。20世纪五十年代中期到后期，操作系统的开发者很少考虑结构问题，也不具有构造大型操作系统的经验。在一些单体系结构的操作系统中，任何一个过程都可以调用任何其它过程。因此，在早期操作系统的规模不大时，这种结构的缺点如缺乏扩展性、可移植性、灵活性等问题没有明显的暴露出来。然而随着操作系统代码量的增加和规模的增大，这种缺乏结构的方法就无法支撑了。

结构化的操作系统设计方法随之出现，首先出现的就是分层的操作系统，如图2.1所示：



图2.1 分层的内核

但是分层的结构也存在一些问题。由于相邻层之间存在许多交互，且每一层处理相当多的功能，使得一层中的大的变化会对相邻层（上一层或下一层）中的代码产生巨大的影响。其结果是，在基本的操作系统上很难通过实现增加或减少一些功能实现一个专用的版本。另外，相邻层之间的高频交互使得安全性很难的到保证。

第一代微内核结构出现于20世纪80年代，是由卡内基-梅隆大学（Carnegie-Mellon）的Avie Tevanian和Richard Rashid主导开发的Mach微内核。微内核的基本原理是，只有最基本的操作系统功能才能放在内核中。非基本的服务和应用程序在微内核之上构造，并在用户态下执行。尽管什么应该在微内核中、什么应该在微内核外，不同的设计有不同的分界线，但是共同的特点是许多传统上属于操作系统一部分的功能现在都是外部子系统，包括设备驱动程序、文件系统、虚存管理程序、窗口系统和安全服务。微内核结构采用水平分层的结构代替了传统的纵向分层结构，如图2-2所示：



图2.2微内核

微内核架构大多用在嵌入式操作系统中，包括：Mach，QNX，Sel4等标准微内核。其次之外，还有Windows NT，Mac OS X这类变种微内核。它们为了追求性能，将需要具备特权的服务组件放进核心空间，违反了微内核的基本设计原则，更接近宏内核的设计方式。这被称为混合内核，是介于微内核与宏内核之间的一种类型。

2.1.2 微内核相对于宏内核的优点

微内核设计的基本思想是简化内核功能，使在内核空间之外的用户空间尽可能多地实现系统服务，同时在系统服务之间加入安全保护措施。于此相对的，在宏内核架构中用户服务和内核服务在同一空间中实现，内核管理着进程调度，内存管理，文件管理和系统调用等模块的工作，由于用户服务在内核服务被实现在同一空间中，这样执行速度上要比微内核快。然而宏内核的劣势也显而易见，那就是内核中的某个服务崩溃，整个内核也崩溃了。另一点，想要在内核中添加新的功能就意味着内核的各个模块需要做相应的修改，因此其扩展性弱。

尽管微内核由于系统模块的分离导致性能上稍不如宏内核，但是微内核依然具有宏内核所不具备的巨大优势。微内核系统结构的核心优势有：（1）内核体积小，方便移植和进行维护；（2）进行拓展修改简易，这是因为微内核很多重要核心模块基本在核的外面完成；（3）系统使用人员很容易更新软件，是由于其仅是硬件的低级抽象，有很多其他模块的部分是到内核外面执行的，因此即使需要完成二次开发改造，也无需对内核再进行研发拓展；（4）可靠性比较高，这是因为每个核心模块都以离散部分模型存放在内核之外，当一个功能模块突然故障也不会干扰其他模块的正常运行。

但是，阻碍着微内核发展的恰巧也是微内核的优点所在之处。微内核架构本质上是一种客户-服务器模式（C/S Model），在微内核架构将众多系统服务放置到用户空间的同时，为系统服务间带来了地址空间边界，因此进程之间的通信（Inter-process communication，IPC）都需要经过内核再将消息发送给目标服务，这时就需要不断的切换用户态和内核态来实现系统。

总而言之，微内核架构所存在的一些问题可以归为以下几点：  
（1）进程或者服务间采用进程间通信的方式交换数据或者调用系统服务，而不是使用宏内核所使用的指针调用方式，这将造成额外的操作系统开销；  
（2）执行一些频繁使用的系统服务时，比如网络收发数据，造成频繁地上下文切换对操作系统来说也是一个负担；  
（3）由于系统服务高度模块化，系统服务之间存在大量的内存复制；  
（4）对互相之间存在的复杂调用关系的系统服务，难以合理地设计通信接口；  
（5）系统服务与内核在地址空间上分离，造成代码局部性差，降低了cache命中率。

2.2 x86架构

2.2.1 Intel 8086

8086是Intel公司第一款16位处理器，诞生于1978年，是整个Intel 32位架构处理器(IA-32)的开山鼻祖。8086体系架构由Stephen P. Morse设计，并在最后定案时得到Bruce Ravenel（8087体系结构设计者）的帮助。逻辑设计者是以Jim McKevitt与John Bayliss为首的硬件开发工程师团队。项目经理是William Pohlman。迄今为止8086的指令集仍然是PC机与服务器的基本指令集。

8086包含8个16位通用寄存器和4个16位段寄存器，通用寄存器分别是：AX、BX、CX、DX、SI、DI、BP、SP，而AX、BX、CX和DX可以分别拆成两个8位AH和AL、BH和BL、CH和CL、DH和DL使用，8086的内部结构如图2.3所示：

图2.3

8086的内存寻址采用的是分段模型。由于它的地址总线是20位，最多可以寻址1M大小的内存，但是段寄存器只有16位，所以在进行寻址时要将段寄存器向左移4位，再加上16位的段内偏移，最终构成段地址：偏移量的组合。由于段寄存器是16位的，在段不重叠的情况下，最多可以将1MB的内存分成65536个段，段地址分别是0000H、0001H、0002H … ,一直到FFFFH。同样在不允许段之间重叠的情况下，每个段的最大长度是64KB，因为偏移地址是16位，从0000H到FFFFH。在这种情况下，1MB的内存最多能划分成16个段，每段长64KB，段地址分别是0000H、0001H、2000H、3000H、…、一直到F000H。

2.2.2 Intel 80386

Intel 80386 发布于1985年，是以1978年8086处理器为基础发展起来的。Intel 80386是第一款具有32位的保护模式的CPU，其将PC由16位的时代带入到了32位的时代。

80386具有三种工作模式：实模式（Real mode）、保护模式（Protected mode）、虚拟86模式（V86）。其中的保护模式是本设计内核工作的主要模式。所谓的保护模式是一种与80286系列和之后的x86兼容的CPU的运行模式。保护模式有大量的新特性，如存储器保护、标签页系统以及硬件支持的虚拟内存，能够增强多任务处理和系统稳定度，现今大部分的x86操作系统都在保护模式下运行，包括Linux、FreeBSD、以及微软Windows 2.0和之后的版本。

在80386中，它对8086的一部分寄存器进行了拓展，由16位寄存器拓展成为32位寄存器同时增加了两个段寄存器GS和FS。它将AX、BX、CX、DX、SI、DI、BP、SP分别拓展成32位的EAX、EBX、ECX、EDX、ESI、EDI、EBP、ESP。段寄存器位数保持不变，当其中保存的不再是段机制，而是保存与全局描述符表（Global descriptor table，GDT）或者局部描述符表（Local descriptor table，LDT）相关的一个索引及属性，被称为段选择子（Selector）。

此外，每个段寄存器还包括一个64位不可见的部分，称为描述符高速缓冲器。Intel 80386的内部结构如图2.4所示：



图2.4

80386的地址线有32根，而在段描述符中段基址也是32位，所以理论上来说80386对内存的访问不再需要分段，但是为了向后兼容8086等x86架构的处理器，80386处理器还是需要以分段的方式访问内存。但是它也提供了一种变通的方式，即把整个4G内存看成一个段，段基址是00000H。在这种情况下可视为不分段，即平坦模型（Flat Mode）。

80386，以及后续的32位处理器都兼容实模式，可以运行实模式下的8086程序。在计算机一开始加电的时候，CPU的状态还是处于实模式下，此时它相当于一个非常快速的8086 处理器。在经过一系列的设置后，CPU才从实模式跳入到保护模式。

其他一些新特性将在后面的微内核的实现的章节再细究。

2.3 NASM

Netwide Assembler 是一款基于Intel x86架构的汇编与反汇编工具。它可以用来编写16位、32位（IA-32）和64位（x86-64）的程序，其所采用的是x86汇编语法的变体而不是AT&A语法。

NASM也被用来编译本毕业设计的内核汇编代码，所采用的的版本是 2.10.07。

2.4 GCC

GNU编译器套装（GNU compiler collection，GCC）是一套编程语言编译器，以及GPL及LGPL许可证所发行的自由软件，也是GNU计划的关键部分，也是GNU工具链的主要组成部分之一。GCC原名为GNU C语言编译器（GNU C compiler），因为它原本只能处理C语言，后经过拓展，变得可以处理C++、Fortran、Pascal、Objective-C、Java等一系列主流的编程语言。有别于局限于特定系统与运行环境的编译器，GCC在所有平台上都使用同一个前端处理程序，产生一样的中介嘛，因此该中介码在各个其他平台上使用GCC编译，有很大机会可得到正确无误的输出程序。

GCC被用于编译本毕业设计的内核的C语言代码，所采用的版本是4.8.5。

2.5 Bochs虚拟机

Bochs 是一款开源的，能够用来模拟x86硬件平台的虚拟机。Bochs的典型应用是提供x86 PC的完整仿真，包括x86处理器（包括保护模式）、硬件设备、和存储器。这种特点使得能够在主机上模拟运行和调试操作系统的内核。

Bochs用于模拟运行和调试本毕业设计的内核，所采用的的版本是2.6.8。Bochs的配置文件bochsrc如图2.5所示：

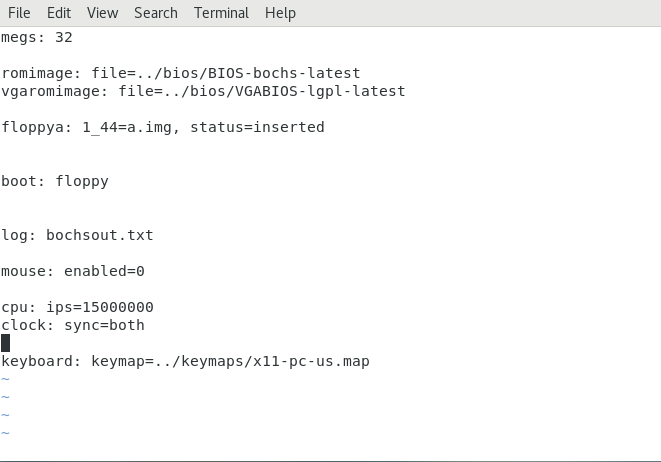


图2.5

3 微内核进程间的通信模型

4 x86架构中内核的总体设计和实现

5 对SBEA算法的改进

6 结论

参考文献

[1] William Stallings. Operating System: Internals and Design Principles(6th edition)[M]. Beijing:

Publishing House of Electronics Industry, 2010:36

[2] 汤小丹, 梁红兵, 哲凤屏, 汤子瀛. 计算机操作系统(4th edition)[M]. 西安: 西安电子科技大学出版社, 2014: 24

[3] 彭星海. 基于x86架构的微内核操作系统的研究与实现[D]. 电子科技大学, 2020

[4] 王宽卿. 微内核进程间通信的研究[D]. 浙江大学, 2010

[5] 中文维基百科. L3微内核[EB/OL]. (2021-2-7) [2021-3-6]. htttps://www.wiki-wiki.top/wiki/L3

[6] 中文维基百科. L4微内核[EB/OL]. (2021-2-26) [2021-3-6]. <https://www.wiki-wiki.top/baike-L4>微内核系列

[7]

致 谢

在本文完成之际，谨向我的导师×××副教授致以衷心的感谢，本论文是在他的精心指导和关怀下完成的，从论文的选题、方案设计，到论文的撰写和修改，都倾注了×××的心血和汗水，在学习期间，他的言传身教将使我终生受益，他认真严谨的治学态度、豁达宽广的胸怀、平易近人的处事风格是我一生的楷模，值此提交论文之时，在此向×××导师表达衷心的感谢！

样式：一级标题

字体： 黑体

字号：小三

段落：居中，段前0.5行，段后0.5行

阅后删除此文本框

附录 内容名称

注意：（1）附录与附录内容之间留有一个空格；（2）附录为可选项，根据实际需要确定，若无，将此页删除

一级标题

字体： 黑体

字号：小三

段落：居中，段前0.5行，段后0.5行

阅后删除此文本框

以下内容可放在附录之内：

(1) 正文内过于冗长的公式推导；

(2) 方便他人阅读所需的辅助性数学工具或表格；

(3) 重复性数据和图表；

(4) 论文使用的主要符号的意义和单位；

(5) 程序说明和程序全文。

字体：宋体

字号：小四

段落：两端对齐，首行缩进2个字符，1.5倍行距

阅后删除此文本框