

现代通信网络技术丛书

网络处理器

体系结构、协议与平台

○ [美] Panos C. Lekkas 著
○ 赵有健 刘 丹 等 译

Mc
Graw
Hill

 人民邮电出版社
POSTS & TELECOM PRESS

现代通信网络技术丛书

网 络 处 理 器

体系结构、协议与平台

[美]Panos C. Lekkas 著
赵有健 刘 丹 等 译

人 民 邮 电 出 版 社

图书在版编目 (CIP) 数据

网络处理器体系结构、协议与平台 / (美) 里卡斯 (Lekkas, P.C.) 著; 赵有健等译.

—北京: 人民邮电出版社, 2006.10

(现代通信网络技术丛书)

ISBN 7-115-14343-9

I. 网... II. ①里... ②赵... III. 计算机网络—微处理器 IV. TP393.02

中国版本图书馆 CIP 数据核字 (2006) 第 088595 号

内 容 提 要

网络现在已成为社会各界和人们工作与生活中信息传递的重要工具之一, 而网络处理器是网络中不可或缺的设备之一。本书就是一本专门介绍网络处理器的专著, 主要内容包括 5 部分, 分别介绍了网络处理器的发展和现状、基本情况和分组处理过程; 网络处理器的体系结构和网络处理的方案; 网络处理器的外围支撑芯片, 主要有: 存储处理器、分类处理器、查找引擎、交换结构、流量管理器; 网络处理器的应用; 安全协处理器。

本书内容丰富、详实, 图文并茂, 理论联系实际, 有较强的理论指导和工程实践意义, 可供从事网络和通信工作的技术人员和管理人员阅读参考。也可作为大专院校网络专业的师生专业课的学习参考书使用。

现代通信网络技术丛书

网络处理器体系结构、协议与平台

- ◆ 著 [美] Panos C. Lekkas
译 赵有健 刘 丹 等
责任编辑 王晓明
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
新华书店总店北京发行所经销
- ◆ 开本: 787×1092 1/16
印张: 22.25
字数: 526 千字 2006 年 10 月第 1 版
印数: 1—3 000 册 2006 年 10 月北京第 1 次印刷
著作权合同登记号 图字: 01-2006-3981 号

ISBN 7-115-14343-9/TN · 2666

定价: 48.00 元

读者服务热线: (010) 67129264 印装质量热线: (010) 67129223

前 言

作者写书的时候，通常最后才写前言。通常前言会详细介绍写书的动机、主题和整本书的结构。当我准备写本书前言的时候，想到一个好方法可以解释人们关于网络处理器概念上的混乱。我想起多年前在 IBM 工作时老板讲的一个笑话，他曾经是一个受人尊敬的海军军官。我觉得这个笑话很合适。

一个大雾弥漫的冬日，天上下着雨，一艘前往北海的全副武装的核战舰航行在英吉利海峡上，这段航线非常凶险。大浪经风一吹变成了层层薄雾，严重影响了舰上的雷达系统，甲板上的水手们因此非常惊慌。因为能见度太低，加上这段凶险的海峡上船只很多，舰上值班的人员变得非常紧张。就在这时，他们借助望远镜看见正前方出现了一个很昏暗的亮点，于是立即向舰长汇报了这一情况。舰长命令发报员立即利用国际通用的应急频率向对方发出信息，信息的正文是：“你好，我是舰长。我们在同一航线上相对行驶，请你们立刻调整航线，向北偏 1/10 度行驶。”

战舰很快就收到了对方的回复：“你好，我是老海盗。我看见了你们的舰艇，请你们立刻调整航线，向南偏 1/10 度行驶。”

发报员看了后非常困惑，立即将回复告诉舰长。舰长皱起了眉头说：“这个人肯定疯了，如果他不尽快更改航线的话，我们会撞在一起的。”于是他又拿起望远镜向正前方看了看，发现那个亮点变得越来越近。

舰长命令发报员立刻发出第二条信息，为了强制对方驶离航线，这次的言辞变得更严厉了：“重复一遍，我是战舰的舰长。请你们立刻调整航线，向北偏 1/10 度行驶。”很快，在同一频道，又收到对方的回复。“重复一遍，我是老海盗。你们必须立刻调整航线，向南偏 1/10 度行驶。”

这回发报员惊慌失措，赶紧将电报交给舰长。舰长看了看前方的亮光，想到对方竟然敢一再无视自己的警告，于是他对发报员吼到：“再给对方发电报，这次的措词必须更强硬！”这次电报的内容是：“我是一艘装有核导弹舰艇的舰长。你们必须立刻调整航线，向北偏 1/10 度行驶。”

这次对方又很快回复了：“我是灯塔。你们必须立刻调整航线，向南偏 1/10 度行驶。”

我觉得这个笑话能够很好地反映网络处理器领域的现态。我接触了工业界和学术界的很多人，网络处理器没有一个明确的标准。来自不同专业、不同背景的人，从完全不同的角度和视角来看待网络处理器，他们的兴趣和目标都不尽相同。我很惊奇地发现，作为一项新技术，网络处理器具有不同的特点和需求。

在刚刚过去的 10 年里，信息流量增长异常迅速，推动了超高速路由和交换网络的快速发展，同时对它们的复杂度和多样性提出了很高的要求。各类在线应用和互连，成为流量增长的主要来源。十几年前，Internet 还只是实验室里面的新生儿，但现在已经是无所不在了，它还是商业和娱乐必不可少的工具。一些专用的计算平台和体系结构需要承担每秒几十亿次的运算任务，这需要强有力的交换机和路由器的支持。

现在这些平台通称为网络处理器，本书将解释、定义和重新定义这些术语。它们都是专用的高性能、高度可编程芯片，每秒能够处理数吉比特的流量，当需要实现一个新功能时，可以通过方便高效地编程即可，这是网络发展的必然趋势和要求。当然，可以利用 QoS 和服务层协定（SLA）来处理流量，购买网络处理器的客户（例如网络设备设计商）可以在设备中设计灵活、方便修改的功能。这样一来，购买这些设备的客户（例如运营商、ISP 和其他企业）就可以使投资保值，因为新设备的寿命越来越长。

虽然有的网络处理器已经推出了好几代，但是仍然存在很多的“误解”。虽然有很多人听说过网络处理器，但是更多的人还没有听过这个名词。有些人只是了解网络处理器的一部分内容，有些人能够全面掌握网络处理器。有的人对网络处理器有所了解，但是他们不知道的东西更多。有的人承认根本不了解网络处理器，也不知道它们是干什么用的。令人感到讽刺的是，一位资深的执行官员甚至让我解释网络处理器能做什么 Intel 奔腾芯片不能做的。有的人虽然了解网络处理器，但是他们还不是很清楚如何让一个网络处理器系统正常运行。“它是交换网络么？它到底是什么？”“它是流量管理器么？还是搜索引擎？它需要和网络接口配合使用么？因特网泡沫已经存在很长时间了。”工业界的这些反应实在让人感到惊讶。网络处理器的概念存在很多的疑雾。

但是这并不意味着找不到关于网络处理器的信息，有很多关于网络处理器的有用信息，其中的一些还相当好。但是，有一个问题，这些信息非常琐碎，可能散布在各个厂商的产品说明书中，或者一些零散的文章和不相关的白皮书，也可能是一些出版社出版的文集，它们可能很不错，也有可能是骗人的东西。一些重要的信息可以从同研究相关的论文中得到，但是工程师很少有时间和精力研读这些论文，也没有时间搜索各类文献，或者静下心来研读。一些技术分析和市场调研的杂志上面登出了很好的文章，覆盖了工业界，搜集了大量的技术报告，通常一年发行几期。但是这也存在一个问题，要解决一个问题时，往往需要大量的技术报告，花费几千美元的费用。这样一来预算会很大，一些企业可能负担得起，但是对于工程师和学生来说就存在问题了。

我觉得很有必要写一本关于网络处理器的书，让那些对网络处理器感兴趣，想深入了解的人能够阅读。我调研了一下发现，现在还没有这样一本专业书。我决定写这样一本书，可以帮助那些使用网络处理器，但不一定是专家的读者。这本书应该涉及整个网络处理器的领域，同时还要揭示网络处理器是如何工作的。附录中的参考文献提供了大量的有用信息，致力于网络处理器的读者可以从中获得更多、更深入的信息。

本书主要涉及网络处理器存在的原因，它们是如何出现的，它们用于实现哪类解决方案，它们适用于解决哪类问题，如何权衡利弊，已经要使用网络处理器实现一个实际的系统时所需的开销。最后一个要求说明除了网络处理器外，还需要一些其他的技术，例如：交换网络，流量管理器等等，这些都必须考虑到。这将涉及一个非常全面的领域。

在撰写本书的过程中，我知道还有其他几本关于网络处理器的书即将出版。包括 Douglas E.Comer 的《Networks Systems Design Using Network Processors》（Upper Saddle River, New Jersey: Prentice-Hall,2003）（参见 www.npbook.cs.purdue.edu），Patrick Crowley 的《Network Processor Design: Issues and Practices》（San Francisco: Morgan-Kaufmann, 2002）和 Bill Carlson 的《Intel Internet Exchange Architecture: A Practical Guide to Intel's Network Processors》（Intel Press, 2003）。由一位受人尊敬的作者撰写一本新的教科书，只能起到指导和参考的作用。我

能预测其他几本书的内容，我敢保证，这些书不但不会竞争冲突，而且会互为补充。我十分尊敬 Comer 教授，看过他的书稿后，感到很有意义。Comer 教授在网络协议上具有丰富的经验，他设计了一个专用的系统平台（基于特定的网络处理器架构（Intel）），基于该平台，Comer 教授详细讲解了如何实现网络交换、路由协议。其他的两本书都有特定的读者群。其中一本的内容很有趣，但只是基于 Intel 的观点。另一本书针对那些对网络处理器设计感兴趣的读者，介绍了该领域的研究现状。

我的这本书和他们都不一样。Intel 毫无疑问是一家令人羡慕的公司，它提供了优秀的系统平台，但是它不是网络系统设计人员的惟一选择。因此，本书介绍了基于多家生产厂商现有网络处理器架构的多种工业平台，其中一些已经在业界得到广泛的应用，另一些虽然刚起步但是充满希望。接下来，将介绍几项可选技术方案，例如网络特定用途集成电路（Net ASIC），或自定义单芯片系统（SOC）架构（能够进一步演变成用于网络处理的功能强大的 ASIC）。但是，不是每个人都选择所有的情况。有的公司需要产品能够快速、低成本投入市场。其他的公司则可能需要一个优化的最大性能架构，可能有技术、预算和能力自行设计功能强大的芯片。这些都很重要，本书将向读者展示所有可能的方案。讨论完网络处理器的例子后，本书将逐一介绍交换机/路由器领域各个方面的内容，包括：交换网络、流量管理器、分类协处理器、搜索引擎和内容可寻址存储器（CAM）。本书还将介绍存储网络处理器（SNP），在两个单独的章节分别介绍 TCP 减负引擎和安全协处理器。

每介绍一部分内容，我都会概述如何在各种方案间取舍和它们各自的优缺点，包括它们的费用，特别是那些不公开的方面，还有生产厂商不愿意提及的内容都将逐一揭示。如果用户不了解这些缺点，就可能做出错误的、感情用事的决定，最终付出昂贵的代价。本书将通过一个具体的大型多服务交换机/路由器的设计实例来进行分析。

这里不讨论协议的详细实现过程，因为这不在本书讨论的范畴内。但是，本书将介绍编写、维护和升级用于 NPU 平台的代码带来的开销。不同的方案，无论看上去多么花哨或者浮华，它们的实现方案不尽相同。

对于必须学会如何在 NPU 上编写协议的学生和那些想学习 Intel 网络处理架构的初级软件工程师，他们应该选读 Comer 的书和 Carlson 的专著。但是，对于那些考虑转向网络、路由和交换系统的经理、咨询人员和实践工程师，应该选择本书，因为本书可以让他们迅速找到切入点并快速进入角色。

在和业界朋友的讨论中，我遇到一些经验丰富的软件工程师，他们实现过非常优秀的协议。但是他们不清楚调度、反压、交换网络和分类这些概念。他们经常抱怨说他们不知道自己在大型项目中所处的位置，无法了解那些负责硬件和系统的同事们面临的困难，反过来也是一样。类似地，我也遇到过一流的 ASIC 硬件设计人员，他们精通每秒吉比特级的线速处理通路，但是却不了解如何对字符串进行深度分类，也不知道如何编程实现协议，对于机柜内部大量的计算资源也不知道如何分配。我还遇到过一些交换网络的设计人员，他们精通交换结构，在吃饭、喝水、生活，甚至呼吸的时候，都在考虑调度和仲裁，但是他们对于安全协处理器却一窍不通。他们中对于如何建立稳固的安全体系以及系统设计中的实时加密技术的细节了解的人就更少了。

这些拥有不同背景的人肯定能够从本书的不同章节中获得收益。有的读者喜欢书中关于当前最有希望架构的简介和最新介绍，有的读者喜欢书中关于常见问题的讨论，这些问题通

常在读者购买了 5、6 本不同的书后才能自信地提出。还有一些读者可能想知道某个厂商的产品的体系结构和平台。我不打算只介绍某个公司的产品，也不会倾向于某个公司的产品，我会客观地向读者介绍它们各自的优缺点，相信对那些新涉足该领域的读者会有帮助。

我尽力在书中融入了大量有用的信息，使它能够介绍某个新兴的产业，该产业的发展速度可能超出想像，接下来会因为某些本书没有提及的原因而停滞不前，但是可怕的并购不可避免。有的厂商可能获得成功，有的厂商也可能无法逃避倒闭的命运。尽管如此，仍然会有一些想法得以保留下来并不断发展。我希望这项令人鼓舞的技术能够被更多的人关注。如果我的这个想法只有部分得以实现，我也会感到非常的欣慰，至少自己没有浪费时间。

本书的内容分为 5 个部分。

第一部分由第 1~3 章构成，第 1 章介绍了网络计算的演变历程。第 2 章讨论了网络处理器技术和商业的正名。第 3 章介绍了分组处理的基本概念。对于新涉足本领域的人员和那些想了解该领域的学生，本书给出了基本的术语和概念。

第二部分由第 4~10 章构成，介绍了几类最重要的网络处理器架构。第 4 章介绍了 IBM 公司在网络处理器上的解决方案。第 5 章讨论 Intel 的平台。第 6 章介绍了 AMCC 公司在网络处理器上的解决方案。第 7 章概述了 Agere 公司的网络处理器平台。第 8 章简述了摩托罗拉公司的网络处理器。第 9 章介绍其他一些厂商的网络处理器产品和架构，例如：Bay Microsystems、Cognigine、Ezchip、Silicon Access、Vitesse、Wintegra 和 Xelerated。第 10 章讨论如何利用网络 ASIC 设计网络处理系统，以及基于 IP Core 的自定义处理器，这里使用了可配置技术和大量的并行技术。本书通过具体的案例，介绍了该领域各公司最有前途的技术，例如：MIPS、Tensilica、Improv Systems、ARC 和 ClearSpeed Technologies。

因此，第二部分是一个概述，介绍了各个厂商的网络处理器产品，可以作为一个快速参考手册，借此比较那些有供货关系和新认识的厂商。这一部分还介绍了同线速分组处理相关的大计算量问题的各类解决方案。有的公司利用可扩展的方式解决这个问题，有的公司则给出了非常精巧的方案。

第三部分由第 11~15 章构成，详细介绍了利用网络处理器构建一个最新的多服务交换机/路由器所需的其他部件。第 11 章“存储网络处理器 (SNP)”讨论 SNP，这是一类在快速发展的存储网络技术下，将协议终端和最常见的 NPU 功能集成在一起的芯片。第 12 章介绍查找引擎和 CAM。第 13 章介绍分类协处理器，在某些架构中，它们位于网络处理器之外，是单独的芯片。第 14 章基于 IBM 和 Agere Systems 两家公司具有代表性的两类不同的架构，介绍了交换结构。第 15 章介绍流量管理器件，这是 QoS 领域不可缺少的器件。在撰写最后两章时，我有很多的想法，例如，是否应该包括这两个领域常用的算法。最后我决定还是不包括这些内容，因为这有悖于本书的宗旨，还会大大增加本书的篇幅。我主要集中在实际芯片的介绍，该主题我只是简单介绍一下，书中给出了大量的参考文献，感兴趣的读者可以从中学更深入地理解交换内部的机理。

第四部分由第 16 章构成。这一章回顾了前面几章的内容，介绍了几类架构，以及它们如何满足网络处理环境下的计算复杂度需求。同时，这一章还将介绍一些前面几章没有包含的内容，例如存储子系统。在网络处理系统中，需要用到几种不同的存储技术，每一种技术都提出了不同的性能和预算要求。除了少数设计人员知道如何选择合适的设计线速应用所需的系统存储，很多的人面临这些问题时都不知所措。因此，本书给出的概述和结论，能

让一些读者建立清晰的思路，消除一些误解。

第 16 章还从效率和隐藏开销的角度分析了软件演变过程中的关键环节。这一章以一个实际应用中的大型可扩展边缘多服务交换机/路由器为例，帮助读者从这个层次来理解前面的内容，并且了解它们在一个庞大的网络系统设计中所处的位置。我尽力向读者揭示系统架构设计人员在做出系统设计决定时，如何满足各类需求，当出现冲突时如何做出权衡，以及约束条件等。

第五部分由第 17 章构成。这一章详细讨论安全协处理器的各个方面，范围涉及很广，从加密基础到系统架构，介绍该协处理器如何同传统的网络处理器协同工作。因为要掌握这些芯片，需要理解加密的知识和安全处理的常识，所以本章包括了加密的基础知识和通信安全知识。我知道这部分内容对于一些读者非常熟悉。为了避免读者抱怨，我决定在书的结尾单独起一章。如果有人需要了解安全协处理，可以随时翻查该部分的内容。如果读者对这部分的内容已经很熟悉，可以跳过这部分的内容。

在本书的最后一部分，列出了该领域大量的缩略语，可以方便读者翻查，附录以表格的方式总结了一些网络处理器平台的发展现状，另外还给出了网络处理器工业的发展历程。就像我在前面提到的，这是一个新兴的行业，很多方面还处于摸索的状态，需要制定相应的市场规范。

虽然一些读者打算逐章阅读本书，但这不符合我写本书的宗旨。本书按模块化的方式进行组织，读者可以直接跳过那些不感兴趣的章节，阅读那些他们感兴趣的和有用的章节。我觉得只有这样才能发挥这样一本专业书的作用。这样一来，对交换结构、分类、硬件、传输控制协议（TCP）减负或者安全协处理的读者，可以选读他们感兴趣的章节。如果读者遇到不熟悉的术语，可以借助索引跳到相应的章节来查看，如果读者想找到具体的案例，可以跳转到包含相应平台的章节，查看该领域各个部件如何配合工作。

作者简介

Panos C. Lekkas 已经在工业界活跃了 20 多年,他现在是麻萨诸塞州波士顿 Xstream 技术有限公司的创始人和董事长。Lekkas 致力于多个领域前沿技术和商业开发,包括:网络处理器、宽带光纤、射频无线通信和网络(3G、Wi-Fi、Wi-Max 等)、高级信号处理、光记录、通信安全和神经计算等。Lekkas 是政府部门、高科技公司、投资银行和风险投资公司在技术方面的专家顾问。他在技术和商业开发上拥有丰富的经验,涉及的方面很广,包括:新技术的引进,为客户提供相应的支持和技术评估,公司的估价和出售,为那些有希望的公司咨询者提供技术支持。他的公司可以对一批专利和知识产权进行授权,同时也提供下述业务的整套解决方案,包括:通信系统的分析、仿真和开发,超大规模集成电路和系统电路的架构设计,FPGA 原型设计以及嵌入式软件的开发。

Lekkas 最初在 Silvar-Lisco 公司担当 VLSI 工程师,后来擢升负责监督公司在欧洲的应用工程组。20 世纪 80 年代初,他加盟了 IBM 公司。在 IBM 公司任职期间,他担任过美国德克萨斯州奥斯汀的首席架构、系统设计师,负责过处理器和存储器的管理架构,他还参与了 IBM 公司举世闻名的 RISC 架构的研发,该架构最后成为著名的 IBM RS/6000 超级计算机的核心。Lekkas 在 IBM 工作期间,曾经在美国和欧洲担任过高级技术开发和技术市场管理的职位。

离开 IBM 公司后,在成立 Xstream Technologies 公司之前,他在 Mass 万宝路的 THLC 公司担任了技术主管和技术部门总经理,致力于高速通信安全。通过和 IBM 微电子部门的合作,他通过一系列的合并,在 THLC 公司建立了工程组,引导了该公司高度复杂 ASIC 产品的开发。他在 Mass 伯灵顿投资建立了 WirelessEncryption.com 有限公司,并担任该公司的主席和 CEO,在这里他发明了一项关于流通信安全技术的专利。他担任了 Mass 哈得逊 ACI 公司的工程副总裁,负责设计高级通信 ASIC 工作。他还担任 Mass 康科德 TCC 公司的商业开发主管,在这里他进一步建立了内部密码组,加入了工业标准制定组织,引导了军用和包括链路与协议敏感的工业加密器的智能通信系统产品的定义。在 Galileo 公司,作为国际技术销售和应用工程部门的总管,领导了 WDM(波分复用)技术和光纤放大器通信技术的引进,两年内该公司在日本的业绩增加了 10 倍。在 Galileo 公司,他领导建立了公司高级电光技术,该技术成为平版印刷术、电子显微扫描、半导体设备厂商的表面分析的事实标准,他为基于光纤的一致成像商业应用的多样化作出了贡献,使电子设备显示进入到医疗设备领域,为微观外科手术带来了革命性的变化。

Lekkas 在德克萨斯州休斯顿的赖斯大学就读研究生期间,从事量子电子学的研究,他的导师是诺贝尔奖获得者 Robert Curl 教授。他拥有两个研究生学位,一个是超大规模集成电路和半导体专业,另一个是无线通信领域的射频微电子和微波天线专业。他在比利时鲁文大学修完了公司金融方面的 MBA 学业,他还是欧盟的专业工程师。他在通信传输、编码、安全等领域获得了多项美国和其他国家的专利,他是 IEEE 的会员,美国数学学会(American Mathematical Society)成员和美国数学协会(Mathematical Association of America)成员。

Lekkas 同现在华盛顿大学的 Randall Nichols 教授(Lekkas 以前研究与试验发展组的关键成

员)合著了一本名为《无线安全》的书,该书由 McGraw-Hill 出版社于 2001 年 12 月在世界范围内出版发行。该书的前言由前美国国家安全局总负责人 Admiral Michael McConnell 撰写。

Lekkas 在北欧、日本、亚洲、中东和拉丁美洲工作过,他能够利用十几种欧洲、亚洲和中东的语言流利地会话和写作。

致 谢

在准备该书的过程中，我得到了来自工业界和学术界多人的帮助（太多了，这里不能一一列出）。朋友和同事们花时间来帮我，他们提供了大量的信息、经验。我特别感激那些审稿人，他们和我一直保持通信来往，有时候即便在深夜也毫无抱怨。但是，我还是要列出一些人，他们给予我莫大的帮助和支持。

首先，我要感谢 Marjorie Spencer，她是我在 McGraw-Hill 出版社的咨询编辑，Jessica Hornick 是她不知疲倦的助手，还有其他一些出色的编辑出版人员。还要特别感谢 Steve Chapman（McGraw-Hill 出版社的总编辑）和 Pamela Pelton（McGraw-Hill 出版社的出版负责人）。还有 MacAllister 出版服务的 Beth Brown 和她出色的团队，正是他们使得这本书能够最终出版。另外，特别感谢 Jeanne Henning、Angela Isner、Nonie Ratcliff 和 Joann Woy。

这本书得到工业界很多人士的帮助，我在这里要感谢 IBM 公司的 Jeff Pauza、Gilles Garcia 和 Rene Glaise，Agere Systems 公司的 Rob Muñoz、Wei Li 和 Venkata Krishna Mallampati，原在 Tensilica 公司的 George Taglieri（现在回到 Synopsys），Ensim 公司的 Srinivasan Keshav 博士，ClearSpeed 公司的 Simon McIntosh-Smith，EZchip 公司的 Amir Eyal 和 Kosta Sidopoulos。尽管得到了这些人的帮助，但是我还是难免在书中犯错。

借此机会，我要感谢我母亲的支持和我父亲的记忆力。正是他们赋予我对知识的渴望和勇于探索的精神。我希望我能把这些优点遗传给我的孩子们。

最后，我要感谢我的妻子 Régine 和我的 4 个孩子。Jean-Chrysostome 在研究中帮助我，还替我搜集资料；Marie 承担了校正书稿和编辑的工作；Marina 替我绘制了很多复杂的插图；我的小 Nicholas，他让我保持开朗的心情，在我感到太寂静的时候，他在我旁边玩耍，和我聊河马、鲨鱼、海象、巨蜥、拖车、锄耕机的拖斗。我要特别感谢我的妻子，她是我终生的朋友和搭档，在我写书的过程中，遇到了很多的困难，正是她矢志不渝的爱让我度过这段日子。为了写书，无数个夜晚和周末我都无法陪伴在我的家人身边，但他们没有丝毫的抱怨，给予了我很大的支持。

目 录

第一部分 概 况

第 1 章 网络技术的革命——分布运算和网络汇聚	3
1.1 开始阶段	3
1.2 工作组计算机代替了大型机的继续发展	4
1.3 第一个局域网 (LAN, Local Area Network)	4
1.4 把大型机放到桌面上: PC 和工作站	5
1.5 客户端/服务器 (C/S, Client/Server) 模式	6
1.6 分组交换与电路交换网络的对比	7
1.7 因特网、路由和相关的 Web 技术	8
1.8 网络管理	10
1.9 交换局域网、快速以太网和光纤分布式数据接口 (FDDI)	11
1.10 IP 网络: Intranet 和 Extranet	11
1.11 IP 电话	12
1.12 ATM、局域网仿真、异步传输模式多协议 (MPOA) 和 IP over ATM	13
1.13 无线和移动网络	14
1.14 1Gbit/s 和 10Gbit/s 以太网	15
1.15 存储网络	15
1.16 网络汇聚	16
1.17 光网络突飞猛进	16
1.18 处理器: RISC、数字信号处理器 (DSP) 和片上系统 (SOC) 的集成	17
1.19 带宽和 QoS 需求	19
1.20 交换革命: 从二层交换到路由器到三层交换	19
1.21 MPLS、λ交换和波长路由器	20
1.22 VPN	21
1.23 协同处理器的安全性	22
1.24 流量引擎 (TE)	22
1.25 QoS	23
1.26 性能限制对通信网络设备的影响	23
1.27 总结	24
第 2 章 网络处理器的基本内容	26
2.1 什么是网络处理器	26
2.2 网络设备功能模块	26
2.2.1 物理层接口	27

2.2.2	交换结构	27
2.2.3	数据包处理	27
2.2.4	主处理	28
2.3	更进一步看数据包处理	28
2.4	采用标准的现有 CPU 进行设计的折中方案	28
2.5	采用 ASICS 进行设计的折中方案	30
2.6	网络处理器的突破	31
2.7	网络处理器的优势	32
2.8	网络处理器的分类	33
2.9	总结	34
第 3 章	数据包处理过程	35
3.1	网络整体脉络: 客户、访问、边缘和核心	35
3.2	网络处理发展的时间序列	39
3.3	对于网络设备压倒一切的需求	40
3.4	数据和控制平面的处理过程	41
3.5	数据包处理操作	42
3.5.1	数据包封装成帧	42
3.5.2	模型搜索和数据包分类	42
3.5.3	CAM (内容寻址存储器)	44
3.5.4	查找引擎	45
3.5.5	数据包解析	45
3.5.6	数据包分类和快速转发	45
3.5.7	修改 (Modification)	47
3.5.8	交换	47
3.5.9	流量管理和其他操作	47
3.6	总结	47

第二部分 网络处理器的体系结构

第 4 章	IBM PowerNP™	51
4.1	IBM PowerNP 的概况	51
4.2	体系结构	52
4.3	NP4GS3 的主要功能模块	54
4.4	专用协处理器和辅助硬件	57
4.5	软件体系结构	59
4.6	围绕 NP4GS3 的系统和软件开发	59
4.7	NP4GX——IBM 的下一代 OC-48 网络处理器	61
4.8	使用 NP4GS3 进行设计时的折中考虑	61
4.9	总结	62
第 5 章	Intel IXA™ 网络处理器	63

5.1	Intel IXA 的概况	63
5.2	体系结构	64
5.3	软件体系结构	68
5.4	围绕 IXA 体系结构 NPU 的软件和系统开发	69
5.5	使用 Intel NPU 进行设计时的系统考虑和折中	69
5.6	总结	76
第 6 章	AMCCnP™ 系列网络处理器	77
6.1	nP™ 的体系结构的概况	77
6.2	nP 系列网络处理器的软件开发	79
6.3	流量管理	80
6.4	交换结构	82
6.5	使用 AMCC nP 系列 NPU 进行设计时的系统考虑	83
6.6	第五代技术	85
6.7	总结	85
第 7 章	Agere PayloadPlus 系列网络处理器	87
7.1	PayloadPlus 体系结构	87
7.2	快速模式处理器 FPP	89
7.3	路由交换处理器 RSP	90
7.4	ASI 芯片	92
7.5	双漏桶算法 DLB	94
7.6	Agere APP750NP (EX-NP10) 和 APP750TM (EX-TM10) 芯片组	94
7.7	Agere APP550 (EX-INP5) 网络处理器	96
7.8	PayloadPlus 系列网络处理器的系统和软件开发	97
7.9	总结	99
第 8 章	Motorola C-Port™ 系列网络处理器	100
8.1	C-Port 体系结构概述	100
8.2	NPU 的体系结构	101
8.3	Q-5 TMC 芯片	104
8.4	C-Port 系列处理器的软件开发	107
8.5	使用 C-Port 系列网络处理器进行系统开发	109
8.6	总结	109
第 9 章	网络处理器的其他结构	110
9.1	Silicon Access Network 公司的 iFLOW™ 芯片组	111
9.2	BAY 微系统公司的 Montego™ 和 InP™ 芯片组	114
9.3	Cognigine	117
9.4	EZchip 公司的 TOPcore™ 芯片	120
9.5	Vitesse IQ™ 系列网络处理器	123
9.6	Wintegra	125
9.7	Xelerated 分组设备	126

9.8 其他方法	127
9.9 总结	127
第 10 章 网络处理的其他方案——Net ASIC 芯片和 IP Core 设计	129
10.1 Net ASIC	129
10.2 使用 IP Core 进行设计	130
10.3 MIPS 技术	131
10.4 ClearSpeed 技术	133
10.5 Tensilica 技术	138
10.6 FLIX: 可配置 VLIW 结构	145
10.7 ARC 核心技术	146
10.8 Improv 系统技术	147
10.9 总结	148

第三部分 网络处理器的外围支撑芯片——存储处理器、分类处理器、 查找引擎、交换结构、流量管理器

第 11 章 存储网络处理器	153
11.1 存储网络发展历史和背景	153
11.2 存储区域网络技术	155
11.3 光纤通道 (FC)	156
11.4 IP 存储	158
11.4.1 网络接口卡 (NIC)	158
11.4.2 存储主机总线适配器	159
11.4.3 iSCSI 适配器	159
11.5 存储虚拟化	159
11.6 因特网小型计算机系统接口协议 (iSCSI)	161
11.7 IP 光纤通道 (FCIP)	162
11.8 光纤通道与 iSCSI 桥接	163
11.9 存储网络处理器 (SNP) 的典型应用	163
11.10 对存储网络处理器 (SNP) 的要求	163
11.11 TCP 终端引擎或 TCP 卸载引擎 (TOES)	164
11.12 范例分析一: Trebia 网络 SAN 协议处理器 (SPP)	167
11.13 范例分析二: SilverbackSystems 存储网络访问处理器 (iSNAP™)	169
11.14 存储网络安全	171
11.15 安全存储网络处理器的发展趋势	172
11.16 总结	173
第 12 章 查找引擎	174
12.1 查找引擎的包分类背景知识	174
12.2 内容寻址存储器 (CAM)	175

12.3	内容寻址存储器结构	176
12.4	CAM 的查找表管理	178
12.5	CAM 使用 (系统工程观点)	181
12.6	基于查找引擎的 CAM 的缺陷	182
12.7	进一步的发展	184
12.8	查找引擎的其他实现方法	185
12.9	总结	186
第 13 章	分类处理器	188
13.1	两种类型的数据包分类	188
13.2	查找和转发	190
13.3	用于管理查找表更新的算法	194
13.4	支持查找和转发的算法和数据结构	195
13.5	深度数据包分类	197
13.6	基于多个域的分类	198
13.7	实现	200
13.8	分类处理器还是 CAM	201
13.9	分类功能集成还是独立	204
13.10	个案研究: Raqia 的正则表达式分类处理器	205
13.11	总结	208
第 14 章	交换结构	209
14.1	交换结构的定义	209
14.2	交换的基本原理	209
14.3	拥塞	213
14.4	基本交换单元	213
14.5	常见的交换平台	214
14.6	多服务路由/交换设备的发展	215
14.7	背板描述	218
14.8	交换结构的扩展性	219
14.9	交换结构的冗余	220
14.10	路由交换系统的考虑	222
14.11	交换结构的体系结构	223
14.11.1	输入缓存和输出缓存交换	224
14.11.2	带缓存的交叉开关	226
14.11.3	仲裁交叉开关	227
14.11.4	共享内存交换机	228
14.12	多级开关	229
14.12.1	基于 Banyan 网的交换机	229
14.12.2	Batcher-Banyan 交换机	230
14.13	其他例子	231

14.14	一组商业应用	231
14.15	Agere 交换结构	237
14.16	总结	240
第 15 章	流量管理器	241
15.1	流量管理器的定义和目的	241
15.2	流量管理器作为独立芯片	241
15.3	流量管理的基本概念	242
15.4	面向 QoS 的协议	245
15.4.1	RSVP	245
15.4.2	IntServ	246
15.4.3	DiffServ	246
15.5	主要的任务和算法	248
15.6	统计	249
15.7	流量标记、整形和监管	249
15.8	拥塞管理	250
15.9	调度和缓冲管理	253
15.10	总结	256

第四部分 集腋成裘 (Putting Everything Together)

第 16 章	系统工程学问题	259
16.1	存储器子系统	259
16.1.1	DRAM 的特性	260
16.1.2	SRAM 的特性	262
16.1.3	CAM	263
16.2	网络处理单元 (NPU) 体系结构问题	263
16.3	软件开发问题	264
16.4	软件开发成本	266
16.5	个案研究：一台多重服务路由器 (MSR) 的设计	268
16.5.1	任务定义	268
16.5.2	设计方案	269
16.5.3	初步设计概况	269
16.5.4	交换结构	273
16.5.5	系统考虑	274
16.5.6	资源预算	277
16.6	总结	277

第五部分 安全协处理器

第 17 章	安全协处理器	281
---------------	---------------------	------------