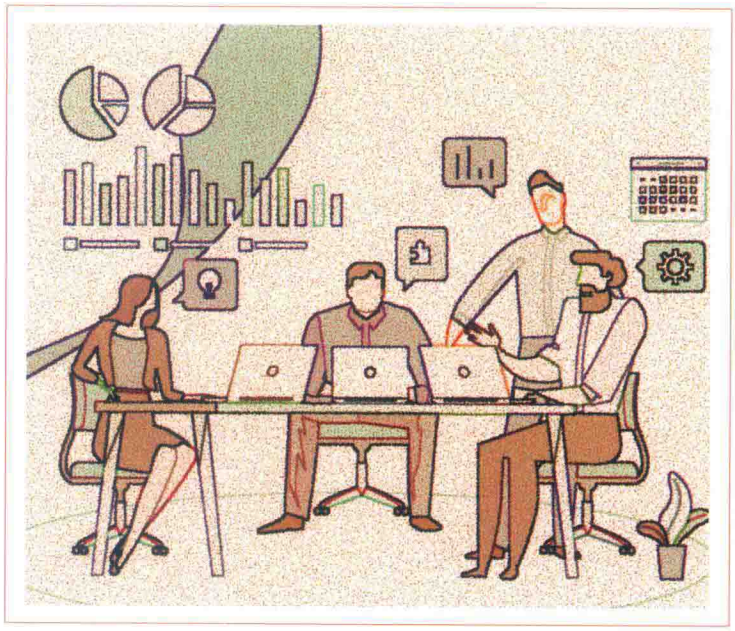
Broadview®
www.broadview.com.cn

Linux开源网络技术大图谱

DPDK/OpenFlow/SDN/OpenDaylight/OpenStack/容器/
Kubernetes网络/Service Mesh等，一应俱全

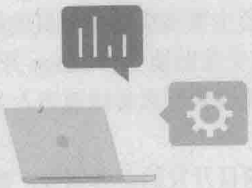


Linux开源网络全栈详解

从DPDK到OpenFlow

英特尔亚太研发有限公司　编著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn



Linux开源网络全栈详解

从DPDK到OpenFlow

英特尔亚太研发有限公司 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书基于 Linux 基金会划分的开源网络技术层次框架，对处于主导地位的、较为流行的开源网络项目进行阐述，包括 DPDK、OpenDaylight、Tungsten Fabric、OpenStack Neutron、容器网络、ONAP、OPNFV 等。本书内容主要围绕各个项目的起源与发展、实现原理与框架、要解决的网络问题等方面展开讨论，致力于帮助读者对 Linux 开源网络技术的实现与发展形成完整、清晰的认识。本书语言通俗易懂，能够带领读者快速走入 Linux 开源网络的世界并做出自己的贡献。

本书适合参与 Linux 开源网络项目开发的读者阅读，也适合互联网应用的开发者、架构师和创业者参考。

图书在版编目（CIP）数据

Linux 开源网络全栈详解：从 DPDK 到 OpenFlow / 英特尔亚太研发有限公司编著. —北京：电子工业出版社，2019.7

ISBN 978-7-121-36786-1

Ⅰ. ①L… Ⅱ. ①英… Ⅲ. ①Linux 操作系统 Ⅳ. ①TP316.85

中国版本图书馆 CIP 数据核字（2019）第 108126 号

责任编辑：孙学瑛
文字编辑：宋亚东
印　　刷：三河市良远印务有限公司
装　　订：三河市良远印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：720×1000　1/16　　印张：16.75　　字数：429 千字
版　　次：2019 年 7 月第 1 版
印　　次：2019 年 7 月第 1 次印刷
定　　价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

推荐序一

Network functions are rapidly transforming from being delivered on proprietary, purpose-built hardware to capabilities running on intelligent and composable infrastructure. We are transforming the network from a statically configured and inflexible offering, to a network which can be provisioned to specific end users, specific verticals and specific needs on a standard server-based infrastructure. Greater customer value is being derived from this flexibility and programmability.

The foundation of the 5G network requires an intelligent infrastructure built on NFV and SDN-based architecture that takes advantage of server volume economics, virtualization and cloud technologies. This enables new services to be deployed more quickly and cost effectively. Intel has a growing portfolio of products and technologies that deliver solutions to help network transformation, bringing advanced performance and intelligence from the network edge to the core of the data center.

Open source software is one key part of the portfolio, which unlocks the platform capabilities of the network for packet processing. Intel invented the Data Plane Development Kit (DPDK), later co-founded as an open source project, and currently leads its growth with the community, helping make DPDK a de-facto standard for packet processing. In addition to DPDK, Intel has significantly contributed to other important open source projects, including Open Virtual Switch

(OVS), FD.io, Vector Packet Processing (VPP) for network stack, Tungsten Fabric (TF) for virtual router, and HYPERSCAN for pattern matching.

We have a skilled and committed team in China who have contributed to these open source projects over the years, and continue to collaborate with our partners in these communities to solve challenging network problems. As a good linkage with the Chinese ecosystem, it is with great pride that the team presents this book as a resource to help contributors who want to get involved, influence communities, and drive continued innovation.

Sandra Rivera

Senior Vice President and General Manager

Network Platforms Group

Intel 5G Executive Sponsor

推荐序二

The rise of Cloud and Edge computing has brought about a shift in networking technology. Increasingly, purpose-built physical systems are being replaced with flexible, adaptable solutions built on open source technology. Open software is driving the innovation that powers this evolution, with Software-Defined Networking (SDN) and network function virtualization paving the way for tremendous growth in connected services.

Intel is a strong contributor to open source software across technologies and market segments. Intel has been a leader in advancing the open networking ecosystem, contributing to projects such as the Data Plane Development Kit (DPDK) for data plane acceleration; Open Virtual Switch (OVS) and Vector Packet Processing (VPP) for virtual switch; OpenStack Neutron, OpenDaylight (ODL) and Tungsten Fabric (TF) for SDN; the Open Networking Automation Platform (ONAP) for orchestration and automation; and Open Platform Network Function Virtualization (OPNFV).

This book draws from the deep experience of Intel software engineers who work in open source networking communities and discusses how these projects fit as part of a complete stack for cloud infrastructure. We are proud to offer this resource to help contributors who want to get involved, influence communities, and drive continued innovation.

Imad Sousou

Corporate Vice President, Intel Corporation

General Manager, Intel System Software Products

前 言

自 1991 年 Linux 诞生，时间已经走过了接近三个十年。即将而立之年的 Linux 早已没有了初生时的稚气，正在各个领域展示自己成熟的魅力。

以 Linux 为基础，也衍生出了各种开源生态，比如网络，比如存储。而生态离不开形形色色的开源项目，在人人谈开源的今天，一个又一个知名的开源项目正在全球野蛮生长。当然，本书的主题仅限于 Linux 开源网络生态，面对其中一个又一个扑面而来而又快速更迭的新项目新名词，我们会有一定的紧迫感想去了解这些他们背后的故事，也会有一定的动力去踏上 Linux 开源网络世界之旅。面对这样的一段旅途，我们心底浮现的最为愉悦的开场白或许应该是“说实话，我学习的热情从来都没有低落过。Just for Fun.”，正如 Linus 在自己的自传《Just for Fun》中所希望的那样。

面对 Linux 开源网络这么一个庞大而又杂乱的世界，让人最为惴惴不安的问题或许是：我该如何更快更好的适应这个全新的世界？人工智能与机器学习领域里研究的一个很重要的问题是“为什么我们小时候有人牵一匹马告诉我们那是马，于是之后我们看到其他的马就知道那是马了？”针对这个问题的一个结论是：我们头脑里形成了一个生物关系的拓扑，我们所认知的各种生物都会放进这个拓扑的结构里，而我们随着年纪不断成长的过程就是形成并完善各种各样或树形或环形等拓扑的过程，并以此来认知我们所面对的各种新事物。

由此可见，或许我们认知 Linux 开源网络世界最快也最为自然的方式就是努力在

脑海里形成它的拓扑，并不断的进行细化。比如这个生态里包括了什么样的层次，每个层次里又有什么样的项目去实现，各个项目又实现了哪些服务以及功能，这些功能又是以什么样的方式实现的，等等，对于我们感兴趣的项目又可以更为细致的去勾勒它其中的脉络。就好似我们头脑里形成的有关一个城市的地图，它有哪些区，区里又有哪些标志建筑以及街道，对于我们熟悉的地方可以将它的周围进行放大细化，甚至于一个微不足道的角落。

本书的组织形式

本书的内容组织正是为了尽一切能力帮助读者能够形成有关 Linux 开源网络世界比较细致的拓扑。首先是前两章，对 Linux 开源网络的生态以及 Linux 本身对网络的支持与实现进行了阐述，希望能够帮助读者对 Linux 开源网络有个全面的基本的认识和了解。

第 1 章主要基于 Linux 基金会划分的开源网络技术层次框架，对 Linux 开源网络生态进行整体的介绍。此外，也介绍了网络有关的开源组织与标准架构。

第 2 章详尽的介绍了 Linux 虚拟网络的实现，包括 Linux 环境下一些网络设备的虚拟化形式，以及组建虚拟化网络时涉及到的主要技术，为更深一步对 Linux 开源网络生态下的开源项目展开讨论打下基础。

然后第 3 ~ 7 章的内容对 Linux 开源网络生态各个层次中处于主导地位的、较为流行的项目进行介绍。按照认识的发展规律，通过前面两章的介绍我们已经对 Linux 开源网络世界有了全局的认识和了解，接下来就可以以兴趣或工作需要为导向，选择一个项目进行深入的钻研和分析。这些章节的内容也是希望能够尽量帮助读者形成对相应项目的比较细致的拓扑，并不求对所有实现细节的详尽分析。

网络数据平面的性能开销复杂多样且彼此关联，第 3 章即对相关的优化技术与项目进行讨论，包括 DPDK、OVS-DPDK、FD.IO 等。

第 4 章讨论网络的控制面，并介绍主要开源 SDN（软件定义网络）控制器，包括 OpenDaylight 与 Tungsten Fabric 等。

第 5 章与第 6 章分别对 OpenStack 与 Kubernetes 两种主要云平台中的网络支持进

行讨论。没有网络，任何虚拟机或者容器都将只是这个虚拟世界中的孤岛，不知道自己生存的价值。

第 7 章讨论网络世界中的大脑——编排器。内容主要涵盖两种开源的编排器，包括 ONAP 与 OPNFV。

感谢

作为英特尔的开源技术中心，参与各个 Linux 开源网络项目的开发与推广是再为自然不过的事情。除了为各个开源项目的完善与稳定贡献更多的思考和代码，我们也希望能通过这本书让更多的人更快捷的融入 Linux 开源网络世界的大家庭。

如果没有 Sandra Rivera（英特尔高级副总裁兼网络平台事业部总经理）、Imad Sousou（英特尔公司副总裁兼系统软件产品部总经理）、Mark Skarpness（英特尔系统软件产品部副总裁兼数据中心系统软件总经理）、Timmy Labatte（网络平台事业部副总裁兼软件工程总经理）、练丽萍（英特尔系统软件产品部网络与存储研发总监）、冯晓焰（英特尔系统软件产品部安卓系统工程研发总监）、周林（网络平台事业部中国区软件开发总监）、梁冰（英特尔系统软件产品部市场总监）、王庆（英特尔系统软件产品部网络与存储研发经理）的支持，这本书不可能完成，谨在此感谢他们的关怀与帮助。

也要感谢本书的编辑孙学瑛老师与宋亚东老师，从选题到最后的定稿，整个过程中，都给予我们无私的帮助和指导。

然后要感谢参与各章内容编写的各位同事，他们是郭瑞景、陆连浩、秦凯伦、徐琛杰、应若愚、丁亮、朱礼波、黄海滨、任桥伟、梁存铭、胡雪焜、胡嘉瑜、王潇、何少鹏、姚磊、倪红军、吴菁菁、陈兆彦。为了本书的顺利完成，他们付出了很多努力。

最后感谢所有对 Linux 开源网络技术抱有兴趣或从事各个 Linux 开源网络项目工作的人，没有你们的源码与大量技术资料，本书便会成为无源之水。

作　者

目 录

第1章 Linux开源网络

在人人谈开源的今天，看着一个又一个知名的开源项目在全球快速发展，开发者会非常想去了解这些开源项目。囿于本书的主题，我们只会努力去对Linux开源网络道出个一二三来。

1.1 开源网络组织

1.1.1 云计算与三大基金会

在形形色色的开源组织里，有三个巨无霸的角色，就是Linux基金会、OpenStack基金会和Apache基金会。而三大基金会又与盛极一时的云计算有着千丝万缕的关系。

整体而言，云计算的开源体系可以分为硬件、容器/虚拟化与虚拟化管理、跨容器和资源调度的管理和应用。在这几个领域里，Linux基金会关注硬件、容器及资源调度管理，在虚拟化层面，也有KVM和Xen等为人熟知的项目。在容器方面，Linux基金会和Docker联合发起了OCI（Open Container Initiative）；在跨容器和资源调度管理上，Linux基金会和Kubernetes发起了CNCF（Cloud Native Computing Foundation）。相比之下，OpenStack基金会更为聚焦，专注于虚拟化管理。

（1）Linux基金会

Linux基金会的核心目标是推动Linux的发展。我们耳熟能详的Xen、KVM、CNCF

等，都来自 Linux 基金会。

Linux 基金会采用的是会员制，分为银级、金级、白金级三个等级，白金级是最高等级。Linux 基金会的会员数量不胜枚举，不过由于白金级高达 50 万美元的年费门槛，白金级会员却是一份短名单，仅包括思科、富士通、惠普、华为、IBM、Intel、NEC、甲骨文、高通、三星和微软等知名企业。

值得一提的是，作为白金级会员的华为，在 Linux 基金会成功建立了一个项目——OpenSDS，这是首个由我国主导的 Linux 基金会项目。OpenSDS 旨在为不同的云、容器、虚拟化等环境创建一个通用开放的 SDS（Software Defined Storage）解决方案，提供灵活的按需供给的数据存储服务。

另外，2018 年 3 月，由英特尔开源技术中心中国团队主导的车载虚拟化项目 ACRN 也被 Linux 基金会接受并发布。ACRN 是一个专为物联网和嵌入式设备设计的管理程序，目标是创建一个灵活小巧的虚拟机管理系统。通过基于 Linux 的服务操作系统，ACRN 可以同时运行多个客户操作系统，如 Android、Linux 其他发行版或 RTOS，使其成为许多场景的理想选择。

（2）OpenStack 基金会

近些年，在开源的世界，OpenStack 应该是最为红火的面孔之一。OpenStack 基金会就是围绕 OpenStack 项目发展而来的。2012 年 9 月，在 OpenStack 发行了第 6 个版本 Folsom 的时候，非营利组织 OpenStack 基金会成立。OpenStack 基金会最初拥有 24 名成员，共获得了 1000 万美元的赞助基金，由 RackSpace 的 Jonathan Bryce 担任常务董事。OpenStack 社区决定 OpenStack 项目从此以后都由 OpenStack 基金会管理。

OpenStack 基金会的职责为推进 OpenStack 的开发、发布以及能作为云操作系统被采纳，并服务于来自全球的所有 28000 名个人会员。

OpenStack 基金会的目标是为 OpenStack 开发者、用户和整个生态系统提供服务，并通过资源共享，推进 OpenStack 公有云和私有云的发展，辅助技术提供商在 OpenStack 中集成新兴技术，帮助开发者开发出更好的云计算软件。

OpenStack 基金会在成立之初就设立了专门的技术委员会，用来指导 OpenStack 技术相关的工作。对于技术问题讨论、某项技术决策和未来技术展望，技术委员会负

责提供指导性建议和意见。除此之外，技术委员会还要确保 OpenStack 项目的公开性、透明性、普遍性、融合性和高质量。

一般情况下，OpenStack 技术委员会由 13 位成员组成，他们完全是由 OpenStack 社区中有过代码贡献的开发者投票选举出来的，通常任职 6 个月后需要重选。有趣的是，其中的 6 位成员是在每年秋天选举产生的，另外 7 位是在每年春季选举产生的，通过时间错开保持了该委员会成员的稳定性和延续性。技术委员会成员候选人的唯一条件是，该候选人必须是 OpenStack 基金会的个人成员，除此之外无其他要求。而且，技术委员会成员也可以同时在 OpenStack 基金会其他部门兼任职位。

而随着越来越多的用户在生产环境中使用 OpenStack，以及 OpenStack 生态圈里越来越多的合作伙伴在云中支持 OpenStack，社区指导用户使用和产品发展的使命就变得越来越重要。鉴于此，OpenStack 用户委员会应运而生。

OpenStack 用户委员会的主要任务是收集和归纳用户需求，并向董事会和技术委员会报告；以用户反馈的方式向开发团队提供指导；跟踪 OpenStack 部署和使用，并在用户中分享经验和案例；与各地 OpenStack 用户组一起在全球推广 OpenStack。

（3）Apache 基金会

Apache 基金会简称为 ASF，在它支持的 Apache 项目与子项目中，所发行的软件产品都需要遵循 Apache 许可证。

对于开发者来说，在 Apache 的生态世界中，有“贡献者→提交者→成员”这样的成长路径。积极为 Apache 社区贡献代码、补丁或文档就能成为贡献者。通过会员的指定，能够成为提交者，就会拥有一些“特权”。提交者中的优秀分子可以“毕业”成为成员。

Apache 基金会为孵化项目提供组织、法律和财务方面的支持，目前其已经监管了数百个开源项目，包括 Apache HTTP Server、Apache Hadoop、Apache Tomcat 等。其中，Kylin 就是中国首个 Apache 顶级项目。

1.1.2　LFN

为了解决项目太多、协调性太差，从而导致的整个生态系统不协调的问题，2018 年年初，ONAP、OPNFV、OpenDaylight、FD.IO、PDNA 和 SNAS 等 Linux 基金会

旗下的六大网络开源项目聚集在一起，创立了用于跨项目合作的 LFN（LF Networking Fund）。

LFN 的这六大创始开源项目，覆盖了从数据平面到控制平面、编排、自动化、端到端测试等领域，为跨项目协作提供了一个平台。通过统一的董事会管理，LFN 消除不同项目之间的重叠或冗余，创建更高效的流程，加快开源网络的发展进程。

LFN 仅仅为各个项目之间的合作提供一个平台，其中的每个项目都将继续保持技术独立和发布蓝图，六个项目的技术指导委员会（TSC）保持不变，但是将由一个技术咨询委员会（TAC）监管。此外，还有一个营销顾问委员会（MAC），统一负责六个项目的市场活动。

新的组织结构解决了各个成员项目之间重复收费的问题，在 LFN 成立之前，成员想要加入任何一个项目都需要缴纳会员费，但是 LFN 成立之后只需要缴纳 LFN 的会员费，就可以参加已经加入及未来即将加入的任何 LFN 项目。

1.2 网络标准及架构

1.2.1 OpenFlow

作为 SDN 的主要实现方式，OpenFlow 发展史就是 SDN 的发展史，对整个 SDN 的发展起着功不可没的作用。

1. OpenFlow 起源

OpenFlow 起源于斯坦福大学的 Clean Slate 项目组，Clean Slate 项目的最终目的非常大胆，是要“重新发明因特网（Reinvent the Internet）”，改变被认为已经略显不合时宜且难以进化发展的现有网络基础架构。

Clean Slate 项目的学术主任（Faculty Director）——Nick McKeown 教授，与他的学生 Martin Casado 发现，如果将传统网络设备的数据平面（Data Plane，数据转发）和控制平面（Control Plane，路由控制）相分离，通过集中式的控制器（Controller）以标准化的接口对各种网络设备进行管理和配置，那么将为网络资源的设计、管理和使用提供更多的可能性，从而更容易推动网络的革新与发展。于是，他们于 2008 年 4 月在 ACM Communications Review 发表了题为 *OpenFlow: enabling innovation in*