彎曲評論

科技 · 人物 · 潮流

思科QuantumFlow处理器及其战略研究

陈怀临，首席科学家

《弯曲评论》

www.tektalk.cn

huailin@tektalk.cn

0.前言

本文着重介绍了思科的边缘路由器ASR1000系列的重要组成部分QuantumFlow网络服务处理器（简称QFP），并从各种不同的观点（Viewpoint）来观察和试图剖析QFP的结构。这些观点包括处理器观点，互联观点，软件观点，报文观点，系统观点等。QFP是思科在WAN接入和智能化边缘设备解决方案中的重要技术支持。对QFP的理解对掌握现代高端网络系统体系结构是非常有意义的。

此文不允许被转帖，下载和应用在任何商业途径和（或）商业公司研发中；对于教育和非赢利性的目的，可以自由转帖，下载和修改。本文遵守自由软件GNU Free Document License的文档条款。关于GFDL的细节，可参阅GNU站点GFDL条款。

1. QuantumFlow

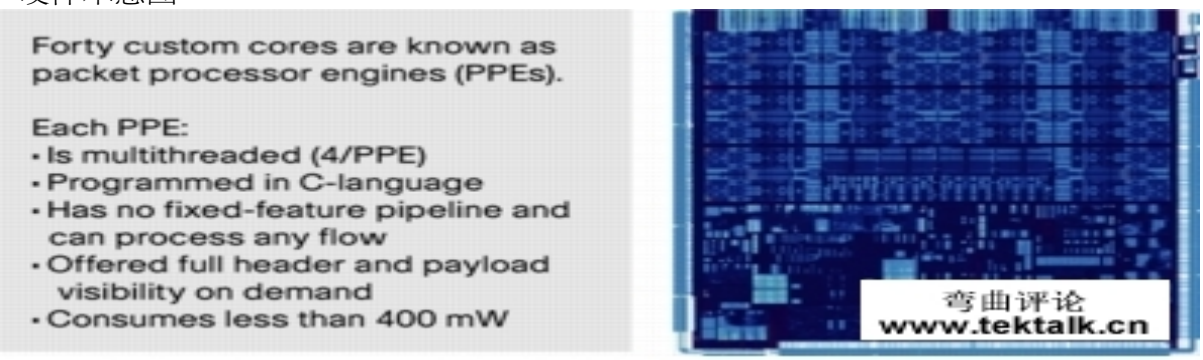
2008年2月25日，思科发布业界新闻，发布其最新一代的，也是世界上最强大的，网络处理器QuantumFlow。其新闻稿的英文原文可参阅如下链接：思科新闻稿
SAN JOSE, Calif., February 25, 2008 - Cisco® today introduced the Cisco QuantumFlow Processor, the most advanced piece of networking silicon in the world and the industry's first fully integrated and programmable networking chipset. More than half a decade in the making, the Cisco QuantumFlow Processor consists of 40 cores on a single chip and can perform up to 160 simultaneous processes, making it uniquely geared for today's network environments and several generations beyond what is currently available in network processors.
The Cisco QuantumFlow Processor was designed by a team of more than 100 Cisco engineers and has led to more than 40 patent submissions. Many of the same engineers who developed the Cisco Silicon Packet Processor (SPP) for the Cisco Carrier Routing System (CRS-1), which

debuted in 2004, also worked on the Cisco QuantumFlow Processor. Continued advancements in technology, design and expertise enabled the team to increase the transistor density on the chip from a then networking-industry-leading 185 million on the Cisco SPP to more than 800 million on the Cisco QuantumFlow processor. Such density puts it in the tier of some of the most advanced processors developed by leading semiconductor companies.
下面是笔者尝试做的中文翻译：
思科公司今天发布其QuantumFlow处理器，工业界最强大的，第一个完全集成的，可编程的网络处理芯片。经过5年多的工程开发，Cisco的 QuantumFlow处理器在一个芯片内部含有40个CPU处理器的核，可以同时做160个数据处理。其超强的计算能力使得QuantumFlow处理 器成为当今网络计算的宠儿，并领先业界中其他的网络处理器数代之遥。
思科的QuantumFlow处理器的设计团队有100多个工程师组成。其设计研发过程中，团队共提交了40多个专利申请。QuantumFlow 的研发工程师许多来自思科2004年发布的多核网络处理器SPP（Silicon Packet Processor）设计团队。SPP被思科用在业界最高端的路由器CRS-1（CRS：CarrierRouting System）上。新技术的不断发展，持续的设计改进和优秀的技术能力使得QuantumFlow的研发团队在该芯片上的晶体管密度提高到8亿个晶体管，而原来SPP芯片为1亿8千5百万个晶体管。QuantumFlow的高密度集成使得其成为半导体业界最复杂的芯片之一。

2. 五年之寒

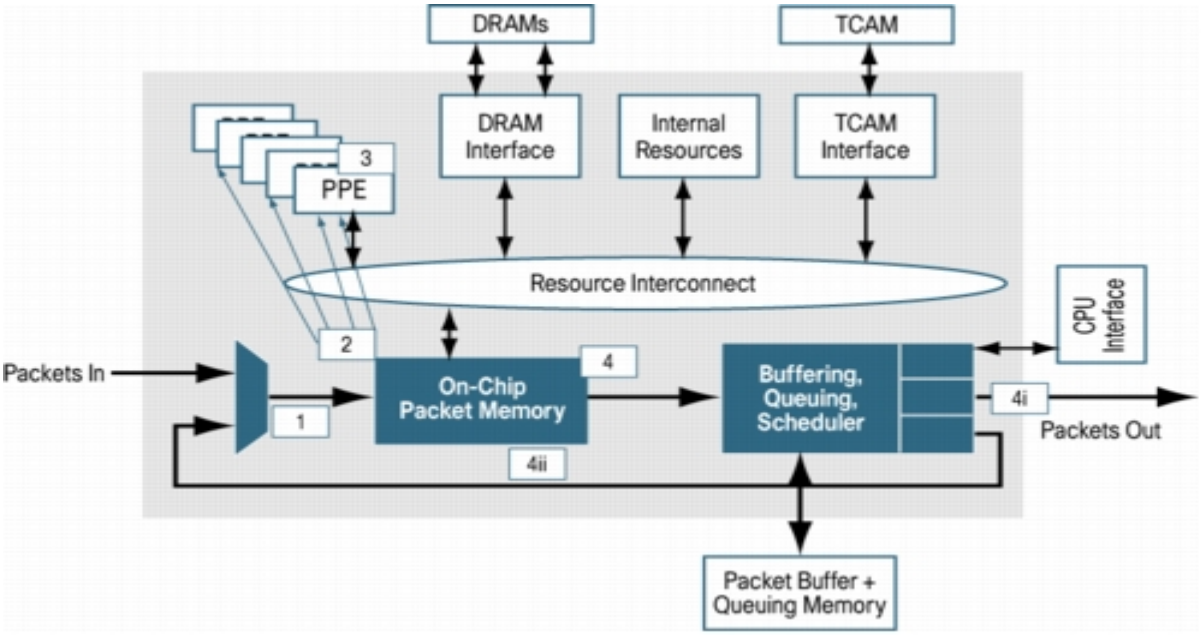
下面是笔者从各方面公开文章中收集整理到的QuantumFlow的一些信息：
× 项目启动时间：2002年Q3或Q4【笔者注：SPP是2002年流片的。CRS-1是2004年推出的。】
×研发耗资：1亿美金
×研发领导：Will Eatherton，Distinguished Engineer and Director of Engineering
×芯片主要定位：边缘（Edge）路由器，企业路由器。
×芯片解决问题：Stateful Service与转发（Forwarding)合一。 【笔者注：如Voice, Video, 防火墙，深度检测等】
×首发系统：ASR1000【笔者注：第二个应用系统是ASR9000。发布时间为2008年11月11日】
×主频：1.2GHZ【笔者注：ESP-5G:900MHz. ESP-10G:900MHz.ESP-20G:1.2GHz】
×晶体管数目：8亿
×（PFE）内存：DDR2。 【笔者注：RLDRAM 】
【笔者注1：两个On-Chip内存控制器（Memory Controller)】
【笔者注2：ESP-5G:256MDRAM. ESP-10G:512MDRAM.ESP-20G:1GDRAM】
×数据报文内存（Packet Buffer）：ESP-5G:64M.ESP-10G:128M.ESP-20G:256m
*CAM:外挂TCAM 【笔者注：ESP-5G:10M. ESP-10G:10M.ESP-20G:40M】
×功耗：80瓦
×多核：40，4 Way-Thread。来自Tensica的Xtensa。
【笔者注1：应该不是Xtensa-7和Xtensa LX2】
【笔者注2：ESP-5G:20Core. ESP-10G:20Core.ESP-20G:40Core】

×片内互联（Interconnect）：Crossbar Switch
×片外互联：ESI【笔者注：在将来的新QFPzhong，将是Interlaken】
×数据报文接口：4个10GBPS SPI4.2 。
×工艺：90nm
×流片：德州仪器
×硬件示意图：



×

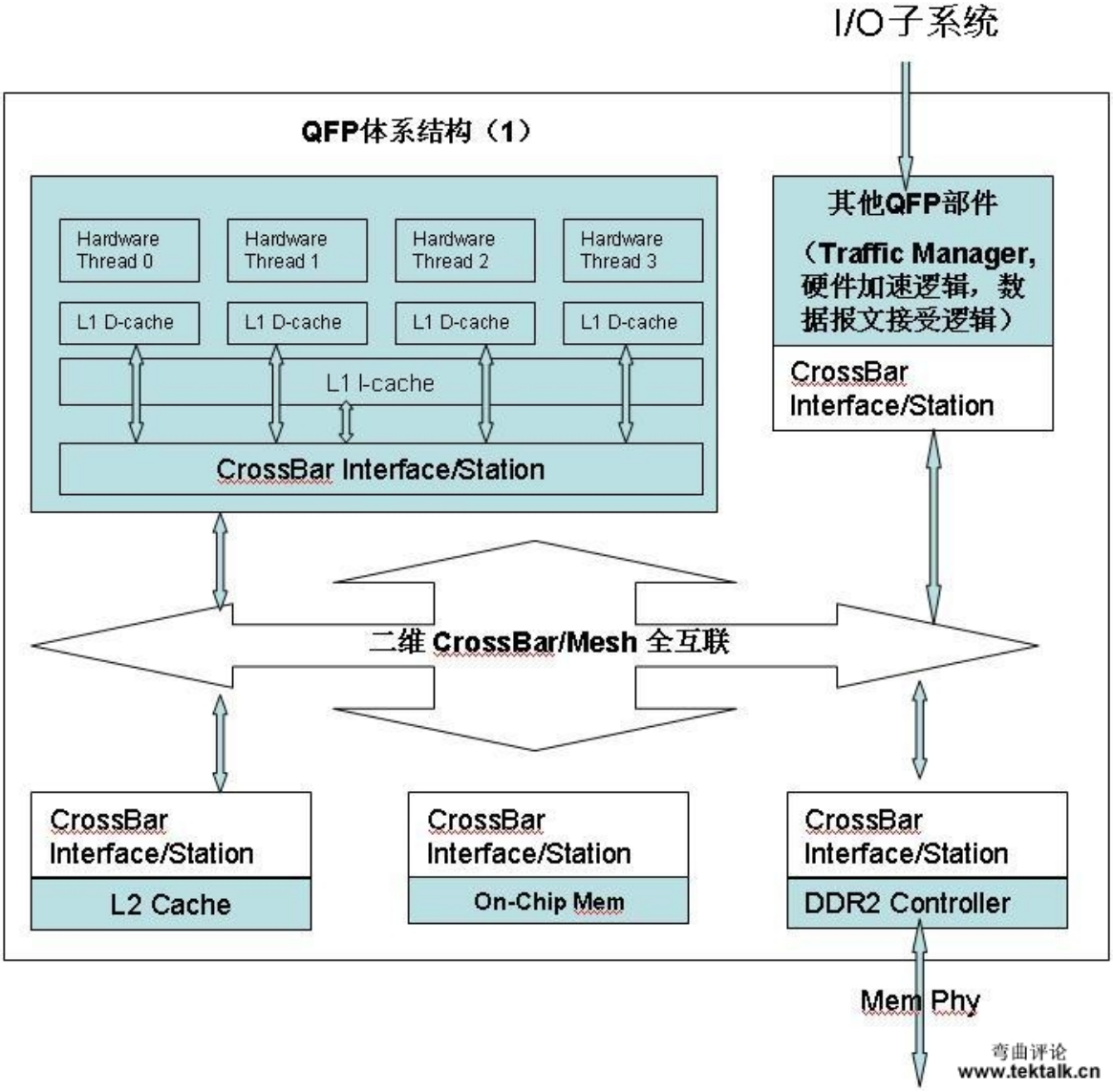
体系结构示意图：



3. 新闻报道

× ASR 1000 QuantumFlow Video
× Strategic Drivers for Development of the Cisco ASR 1000
×Service Provider Market Drivers for the Cisco ASR 1000
×Cisco ASR 1000 Series Aggregation Services Routers
×The Cisco QuantumFlow Processor: Cisco's Next Generation Network Processor
×Cisco puts high-end silicon on the edge：New service router packs custom 40-core CPU

4. 体系结构--处理器观点



观察思科的QFP芯片，可以从不同的观点（Viewpoints）。QFP是一个高端的网络处理器，里面有40个CPU核心，大量的片内和片外内存，高速的互连网络，大量的网络报文引擎，硬件加速器等。从不同的角度来观察QFP，得到的图像是略有不同的。

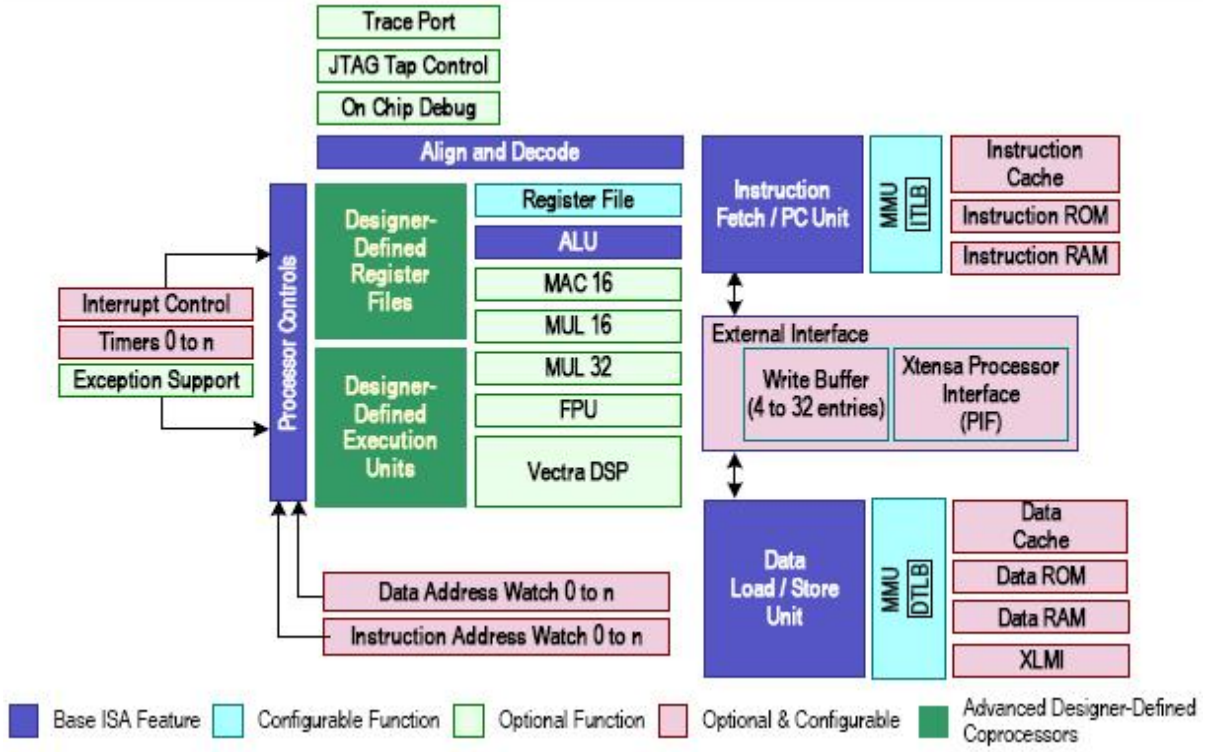
上述图示是笔者从CPU的角度来观察QFP的逻辑。我们可以暂且称之为QFP-CPU。

从QFP-CPU角度，其是一个多核系统。一共有40个核。每个核有4个硬件线程。其4个线程的调度关系应该是FMT的架构，从而最大程度的利用 Mesh互联的带宽。从思科的公共资料可以得知，对于ESP-20BPS的QFP，其CPU核的指令集是Tensilica的Xtensa的ISA（MIPS的变种）。其时钟在1.2GHz。对于Mesh互联的Clock Domain的细节，目前不可知。

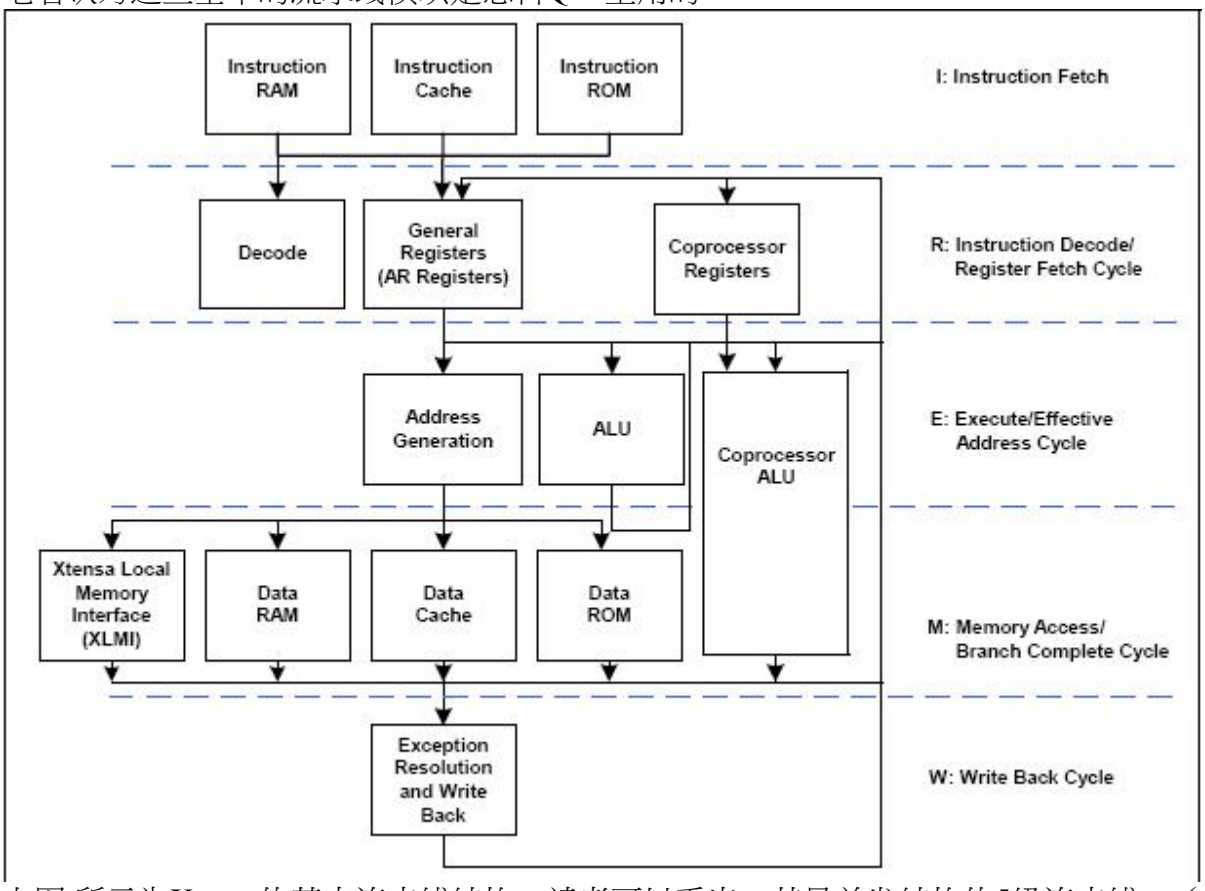
每个核都是32位指令集。换言之，GPR是32bit的。

每个CPU核的硬件线程有自己的L1数据缓存；
一个CPU核的4个硬件线程共享L1指令缓存；
QFP的40个核共享一个L2缓存。（这个L2是单纯的指令缓存，或是Combined指令和数据缓存，目前没有明确的资料可以确认。但笔者倾向于 QFP的L2是一个单纯的指令缓存。
QFP的HT在L1数据缓存Miss后，是直接通过Crossbar的接口读取片外的RLDRAM。笔者分析的原因 是：从QFP-Packet的观点，数据报文都在片内的内存中。所以QFP的CPU核对数据读写应该绝大多数落在片内报文内存上。而指令流的性能对于采用 SIMD并行计算模式的QFP多核系统是最重要的。
读者们要注意，在现代多核系统设计中，总线的概念和机制基本上已经消失。所以从QFP-CPU的角落来观察QFP是，L2缓存，DDR内存控制器等，其实都是挂在Mesh上的一个节点。与其他40个核节点构成一个2维Mesh全联通的图。
另外，从QFP-CPU的角度，其他所有的数据报文（Packet）硬件逻辑可以简单的理解为“I/O子系统 ”。

思科的QFP的40个CPU核采用的是Tensilica的Xtensa的ISA。从思科公开的新闻资料中，思科声称是只采用了Xtensa的指令集结构(ISA)，而其他部件或子系统的逻辑设计，加上后端设计和封装，都是思科自己研发队伍完成的。
从这些声明，对思科QFP有兴趣的读者可能会误以为思科只是购买了Xtensa的指令集（Instruction Set)。这是不精确的。Xtensa ISA的含义不仅仅包括指令集，也包含Xtensa的一些基本的的微结构，例如基本的流水线结构等。否则，思科没有必要去购买一个非主流的指令集，而从新 做一个CPU，用OpenRISC的ISA就可以了。
Tensilica的Xtensa是一个SoC软核，从而第三方可以进行定制和裁剪做出其自己的SoC。也可以同过Tensilica提供的集成环境调试，增加新的指令等等。下图是Xtensa的体系结构略图:



从图中所示，可以得知蓝色模块是Xtensa体系结构ISA中的基本模块。
笔者认为这些基本的流水线模块是思科QFP重用的。

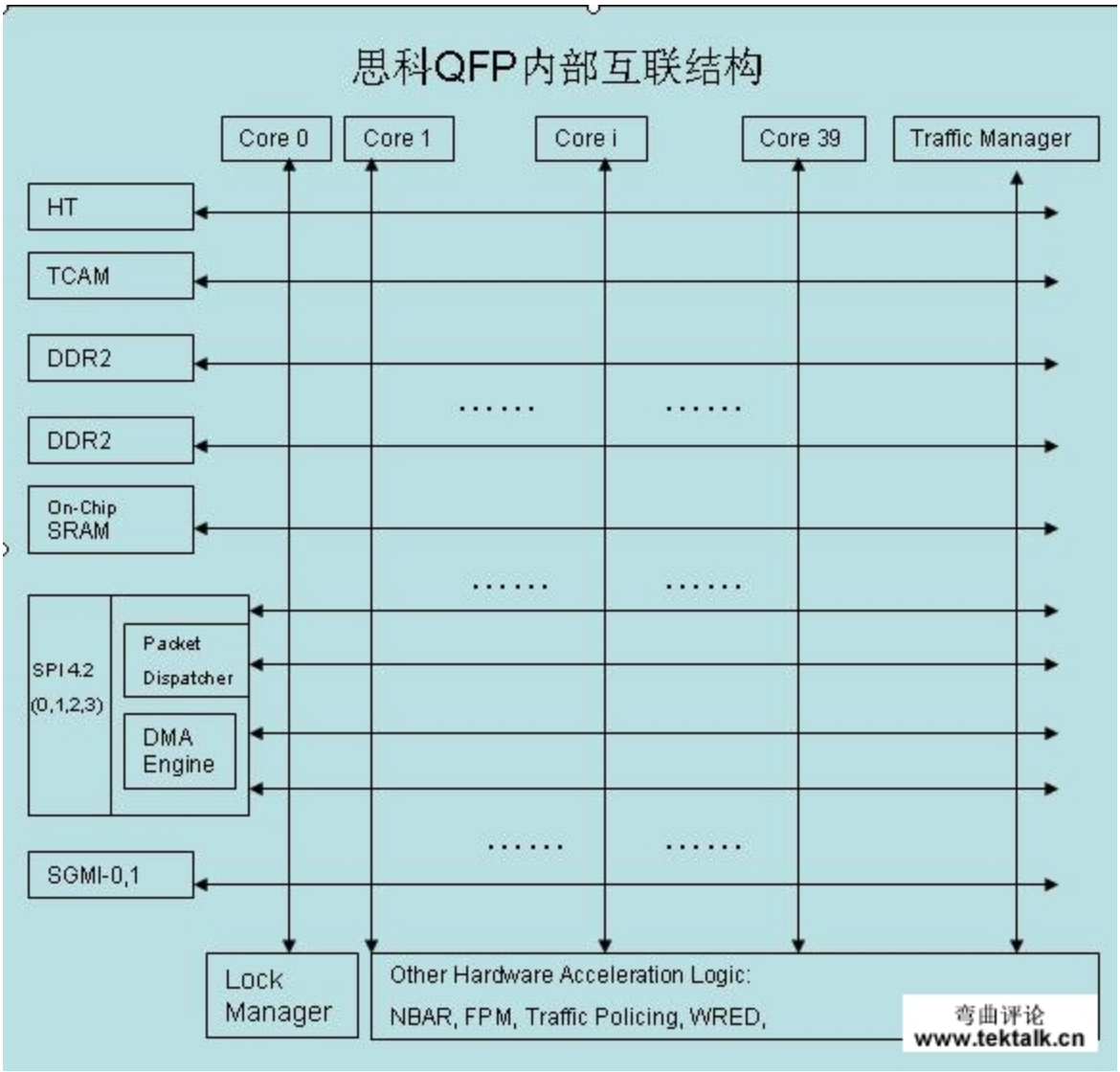


上图 所示为Xtensa的基本流水线结构。读者可以看出，其是单发结构的5级流水线。（思科声称QFP是3发射）。另外，因为Xtensa的核心并没有实现复 杂的Out of Order的执行和超标量结构，没有Interlock的处理，因此，在设计内存操作的指令时，仍然会有delay slot等的发生。
从上述Xtensa的ISA微结构，然后比较思科QFP的微结构，特别是每个核有4个硬件线程的结构。我们可以知道思科确实是需要在逻辑和物理设计 方面做许多特定设计，才能达到QFP的设计目标。例如，基于FMT的4个硬件线程的微结构的设计需要增加几套寄存器，局部总线（Local Bus）和裁决逻辑（Arbitor）等。
另外，在MMU方面，QFP估计也是基本上会重用Xtensa的结构。但感觉在缓存结构上和内存总线接口方面，思科的QFP研发团队需要对Xtensa做许多修改。例如必须将Xtensa挂到QFP的2维CrossBar或Mesh互连结构上。这都需要巨大的研发工作。
从思科发布的QFP资料，其核的主频可以做到1.2GHz。为了能做到1.2GHz，大量的后端设计和手工调试需要开展，而非简单的流片。笔者曾经在5年前用过Xtensa定制过一款SoC的软核。记得当时的体会是，在缓存方面，只要稍微一加大，主频立刻急剧的下来。
一般而言，思科的QFP里面一定会通过Xtensa的全套工具，加入QFP自己的指令集扩充，并且在gcc tool-chain上直接支持。
除了是可定制的SoC软核，编译器和工具链的强大支持是Xtensa之所以能够10年之久还生存的重要原因之一。

5. 体系结构--互联观点

考察一个高级网络处理器时，其互联结构（Interconnect）是一个非常重要的一环。对于一个SoC芯片而已，CPU核本身，作为一个计算 单元，只是众多的计算逻辑的一个子部分（Sub-System）。系统中的许多其他部件，如硬件加速器，网络报文处理逻辑等等，都是非常重要的组成部分。 因此，如果将这些处理部件有机的互联起来，从而达到快速，低功耗，和其他诸多电气要求，就成为一个现代高级芯片设计的关键。

不同的网络处理器设计公司通常 采用不同的方法。各有优缺点。例如，RMI的XLR系列是通过FMN（Fast Message Network）的环结构（Ring），从而使得系统的CPU Core，报文处理单元PDE，加密单元Cypto引擎，多处理器互联的HT接口（HyperTransport），PCI-X等等，都互联在一起。值得 注意的是XLR的内存模块并没有挂在FMN的环上。这也是RMI XLR体系结构的一个特点--通过将FMN和MDI（内存互联结构）分拆，从而达到，数据通道通过FMN，而通常比较慢的内存访问通过MDI互联。另外一 款著名的网络处理器是Cavium的Octane系列。Octane芯片的内部互联相对而言略简单一些--仍然通过经典的支持Cache Coherence的共享内存总线（Memory Bus）的方式将CPU核与内存互联。其它部件，例如网络加速部件，DFA查询部件，SPI端口等等都是通过一个I/O总线的方式互联，然后通过一个接口 转换桥（Bridge）挂在内存总线上。关于RMI的XLR系列和Cavium的Octane系列的体系结构，笔者会在将来的文章中做更详细介绍。



思科的QFP是一个inhouse的网络处理器，因此基本上不可能有公开的资料显示其内部详细结构，特别是其互联结构。上图所示，是笔者从其公共的一些产品文档中多QFP的一些介绍，对其互联结构做出的一些基于个人技术观点的推测。

总体而言，QFP的互联是一个基于二维的Crossbar的互联。在这个互联上，每个逻辑部件对于互联逻辑而言，都是一个节点（node）。这个节 点除了实现自己的计算逻辑之外，实现一个与这个互联网络协议的接口（interface），通常而言，是一个命令（command）协议和一个数据 （data）协议。这个互联协议的重点是通过在这个互联网络上发送命令和数据单元，从而可以从一个节点抵达任何另外一个节点。

如图所示，40个基于Tensilica的Xtensa ISA的CPU核都是这个互联上的节点。QFP的另外一大部件Traffic Manager也是一个节点。在这个互联上，还拥有许多其他的网络加速，内存，和为了与QFP芯片外部互联的逻辑部件。

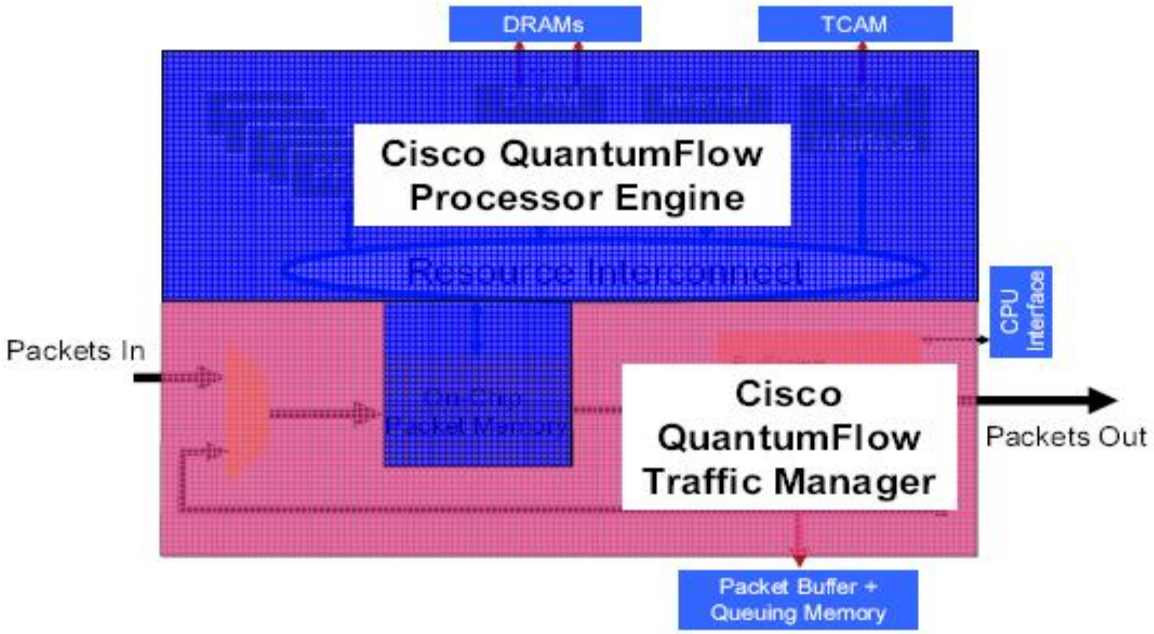
HT: HyperTransport接口。这是用来使得ASR1000上的ESP（Embedded Service Processor）模块（板）上的主控CPU与QFP互联的主要接口。通过HT互联，从主控CPU的角度，QFP就是一个设备，从而可以使得主控CPU 可以通过HT来操作QFP，例如通过读写控制寄存器的方式等。

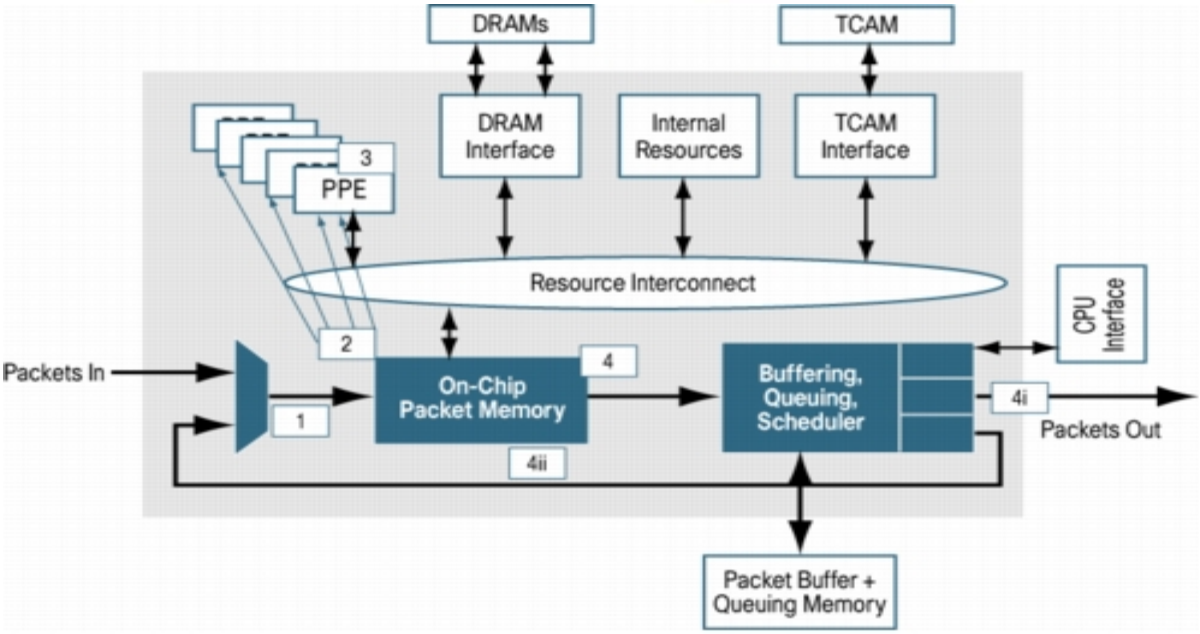
TCAM：这是QFP做数据报文查询（lookup）的外挂 CAM存储器。
DDR2：QFP应该是内含有两个DDR2的控制器。一个是给40个CPU核要用的RLDRAM的控制器；一个是给Traffic Manager的DRAM。
On-Chip SRAM：QFP的报文，当通过入口逻辑（Ingress）抵达之后，是被全部（报文头和报文数据）都被放到了QFP内部的报文内存中，而非在外挂的 DRAM。这一点是QFP有高性能的重要技术亮点之一。许多其他类似的网络处理器都是将数据报头可以放到芯片的内含逻辑中（如寄存器，或L2 Cache里），但数据包的数据（Payload）通常是通过DMA引擎传送到外挂的DRAM里。
SPI-4.2的Packet Dispatcher：QFP的Ingress是支持4个SPI 4.2的接口。也就是说，可以支持40G的线速。这4个SPI在QFP在ASR1000上其实是有分工的。从不完全的资料显示，应该是2个SPI是 QFP，通过外部系统互联，链接线卡（SIP）。其他2个SPI是链接其他的ESP/QFP（支持HA）和控制平面RP（Routing Processor）的。值得注意的是，QFP芯片本身不内含有加密逻辑部件。加密逻辑部件是外挂的，并且通过SPI接口与QFP互相连接和通信。
SGMI：QFP应该还支持2个1G的以太端口，作为控制使用。
QFP 除了 上述主要互联部件之外，还支持许多叫做Internal Resources的硬件加速部件，并且作为节点挂在这个二维的互联结构上。其中一个特别值得注意的是硬件锁机制。笔者相信这个部件是提供了一系列的 spinlock，mutex等逻辑，从而可以使得40×4=160个硬件线程可以通过高速的锁机制，从而达到并行计算中所必须的同步机制。
除了硬件锁机制，QFP还提供了许多其他的硬件加速模块，例如NBAR, FRM，Traffic Policing， WRED等等。

6. 体系结构--报文观点

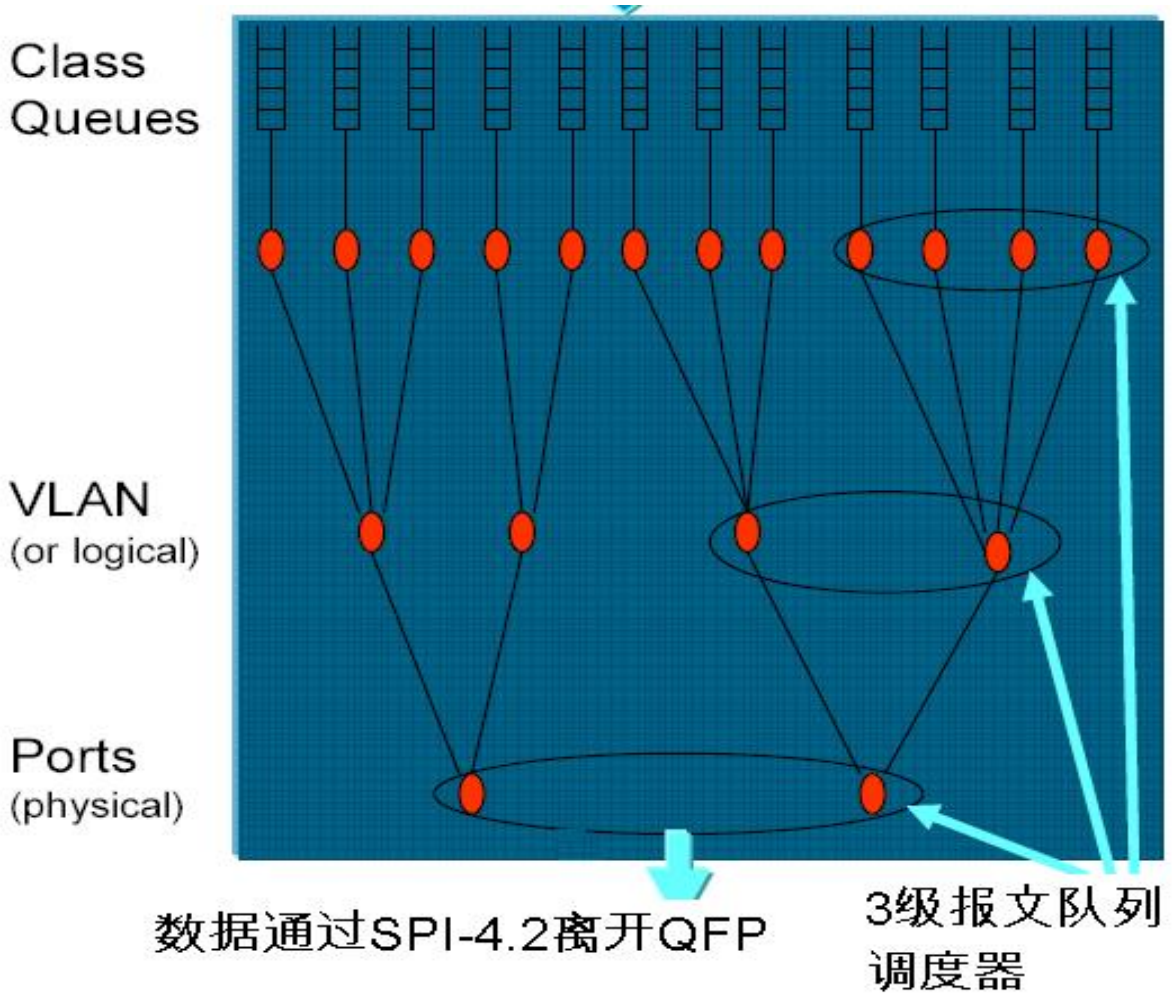
从网络数据报文的观点来观察思科的QFP，QFP就是一个数据报文的从层2一直到层7的数据处理与转发的引擎。在基于QFP的思科ASR1000系 统中，读者要非常值得注意的一个观点是：QFP扮演的是一个集中式数据处理的角色。换言之，系统中所有的数据的，从线卡（SPA-->SIP）和控 制平面卡（RP），都是通过系统的背板Backplane互联ESI，而进入QFP。QFP处理后，决定是应该转发给某个线卡并发送出去，或者是应该转发 个控制平面卡。QFP扮演着一个集中式数据控制和处理的角色。具有HA的ASR1000系列，具有两个ESP卡。美国ESP卡上有一个QFP。这两个 QFP/ESP的关系是一个Active，另外一个Standby。所有的数据报文都是进出当前的主QFP处理器。辅QFP，通过一个专用的SPI4.2 10Gbps的通道，与主QFP通信，做状态备份。

思科QFP数据报文流程





如上图所示，从报文的观点来观察，QFP的逻辑分为两大部分。第一个部分是QFP-Processor Engine。第二部分是QFP-Traffic。第一个部分主要就是那40个Tensilica的Xtensa ISA的处理器单元。第二部分是由一些数据缓存，队列（Queue）和相应的调度算法逻辑组成。下图所示为Traffic Manager的一个逻辑结构图。



思科QFP的Traffic Manager结构图

1. 当一个数据报文通过一个相应的SPI-4.2数据通道抵达后，QFP的报文分发部件（Dispatcher）会将报文层2的Frame所以数据都传送到 QFP的内部报文缓存里（On-Chip Packet Memory）。也就是说，不仅仅是报文的头（Header），是包括数据（Payload）都存放在QFP芯片内部的缓存中处理。这部分的功能还包含一 些基本的数据报文的处理和分析工作

2. QFP的报文分发部件将这个新的数据报文分配给一个计算单元（一个CPU核的一个硬件线程）。这个线程将从头到尾的负责这个数据报文的处理序列，其中包括：

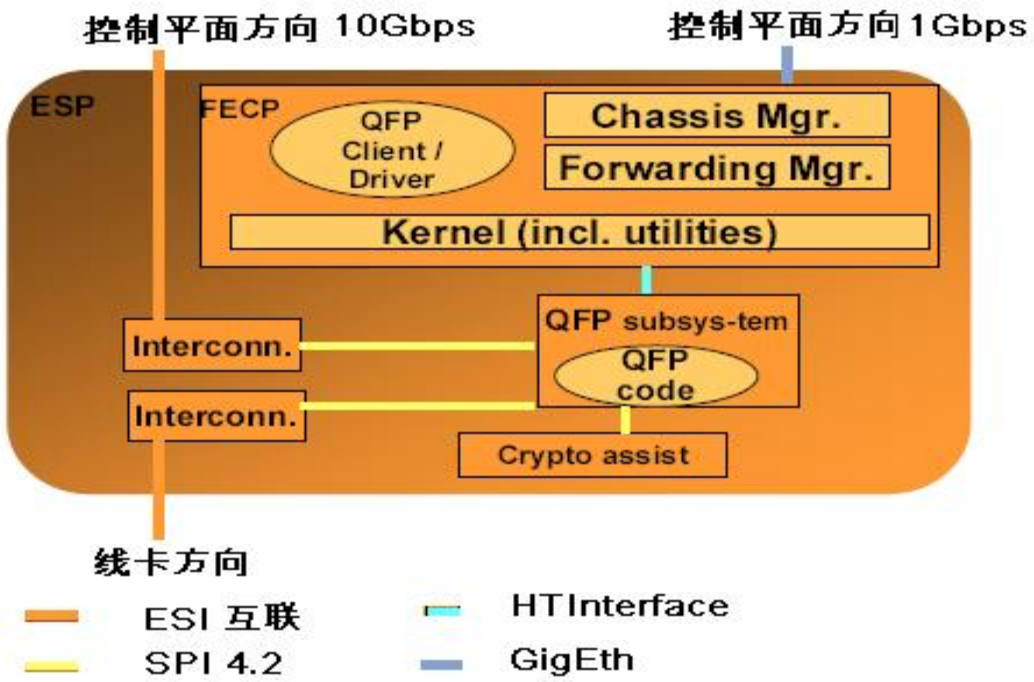
进入时：要进行Netflow，MQC/NBAR Classify, Firewall, RPF, mark/Police, NAT, WCCP, Deep Inspection, 等等。

转发功能：QFP当然首先是一个路由器的数据平面的一个引擎。所以要处理这个报文该往哪里转发的问题。因此QFP的Processor Engine要完成如下工作： IPV4的FIB，MPLS, Multicast等等。

当要离开QFP Processor Engine时，还需要 Netflow，MQC/NBAR Classify, Firewall, NAT, Police/Mark和Crypto 等功能。为什么要考虑加密呢，因为，有可能是VPN tunnel的数据报文。
这时，一个数据报文就已经完成所有的软件处理可以被释放给QFP的Traffic Manager做最后的调度并发出离开QFP了。这个数据报文会被传送到相应的Traffic Manager的队列中。这个过程与这个报文是一个Trhough Traffic（要去另外一个线卡端口），或者是一个去控制平面（RP）的报文，或者是一个HA报文，有关系。不同的报文类型将被放到不同的队列中。从而 Traffic Manager的队列调度器可以通过不同的调度算法去相应的发送一个报文。
3. QFP的Traffic Manager的队列调度功能将决定一个报文的去向。目前，Traffic Manager可以支持128K个队列。强大的队列调度器可以应用许多QoS算法在系统的数据报文上。值得注意的两点是：如果一个数据报文被一个 Processor Engine的线程处理完之后，还需要再来一遍，Traffic Manager负责相应的队列中的数据报文再次转发到On-chip数据报文缓存中；如果一个数据报文需要加密，Traffic Manager会通过相应的SPI通道启动QFP外部（在ESP板子上）的加密部件。

7. 体系结构--软件观点

从思科发布的资料可以得知，ASR1000系统的整个软件操作系统叫做IOS-XE。IOS-XE是一个基于Linux的IOS操作系统。如何理解 这个“基于Linux的IOS操作系统”呢？笔者会在将来单独对思科的操作系统的演化和战略做专题分析。对于拥有QFP芯片 做转发和服务处理的 ASR1000系统，IOS-XE的基本特点是：
--在控制平面上，可以支持单卡（RE）的两个IOS的运行。从而可以支持单RE（控制平面）的HA，例如，ISSU。这是非常重要的一个亮点。思 科就是利用Linux 2.6之后的KVM虚拟技术，从而在Linux上运行了两个IOS。每个IOS作为一个Linux的用户进程来运行。
--在ESP板子上，或者说，数据处理卡上，通过一个主控CPU（一个PowerPC），运行一个Linux Kernel和相应的Chasis Manager， Forwarding Manager等管理进程，并与RE的IOS和相应的进程通过标准的IPC进行通信，从而达到控制平面和数据平面的各种路由，ACL，Config， Policy，动态log数据采集，错误汇报等同步工作。另外，这个主控CPU起到一个控制QFP芯片的作用。换言之，QFP的所有软件都是这个主控 CPU来安装，启动，运行的。QFP在这个层面上，扮演的是一个（专门处理数据报文的）协处理器的角色。
--在线卡（SIP)上，也通过一个主控CPU（一个PowerPC），运行一个Linux Kernel和相应的Chasis Manager， Interface Manager等管理进程，并与RE的IOS和相应的进程通过标准的IPC进行通信，从而达到控制平面和线卡的各种同步工作 。
在这篇文章里，笔者不详细讨论控制平面RE和线卡的软件系统，而是专注于数据处理卡ESP和QFP的软件结构。
ESP和QFP的软件结构图基本上如下图所示：



思科QFP软件结构（1）

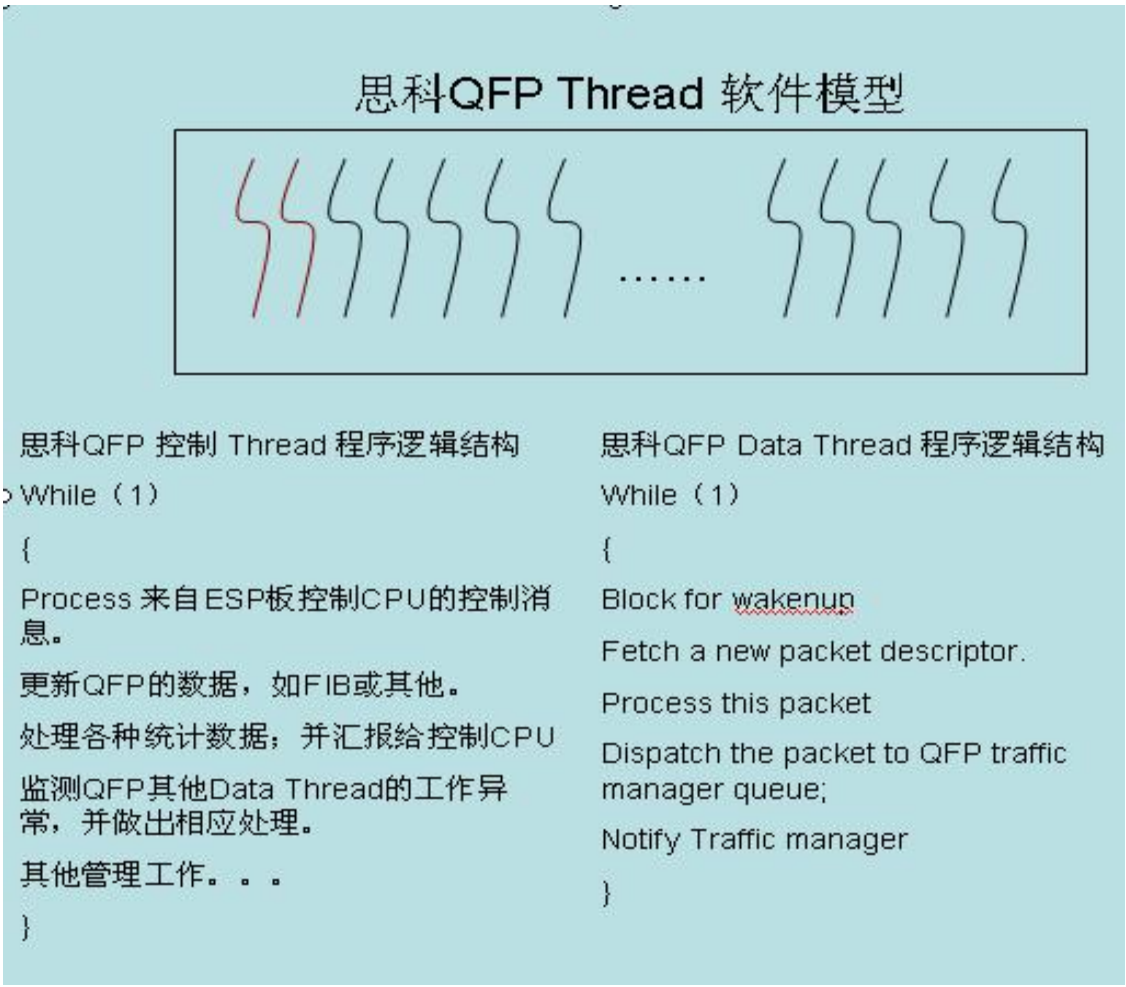
从上图可知，QFP软件子系统分为三个部分：位于主控CPU上的Linux核心上的QFP Client进程；位于主控CPU上的Linux核心里的QFP Driver（驱动）代码函数库，和位于QFP芯片40个Xtensa核上的QFP代码。QFP上的代码通过HyperTransport Interface与主控CPU上的QFP Client进行通信。通过SPI 4.2来获取报文数据，启动加密引擎，和做HA的处理（与另外的ESP板子上的QFP芯片）。主控CPU上的Linux环境中的Chaisis Manager和Forwarding Manager通过一个1Gbps的以太端口与RE（控制平面）通过IPC进行通信。

对于QFP的160(40 * 4)个线程，其软件结构（环境）大致是这样的：

×QFP上没有一个宿主操作系统。没有Linux kernel，当然更不存在IOS等。QFP上多核部分的数据报文处理逻辑应该是运行在一个裸机环境下。或者一个非常简单的硬件抽象层上（HAL）。这个 HAL做一些简单的Tensilica Xtensa ISA的初始化，TLB的设置，中断的处理，内存的划分等等。从目前思科所以公开资料，看不出来其运行了任何OS和微内核。

×QFP的启动， 数据报文处理软件的下载，安装，运行，都是由主控CPU，通过HT接口，来控制的，例如，对于主控CPU而已，QFP的控制寄存器，就是主控CPU下可看 见的一批Memory Mapped I/O地址。可以被读，写等。QFP Driver逻辑就是完成QFP驱动的Linux核心模块。

×QFP的线程作为引擎的角色，运行数据报文处理软件。

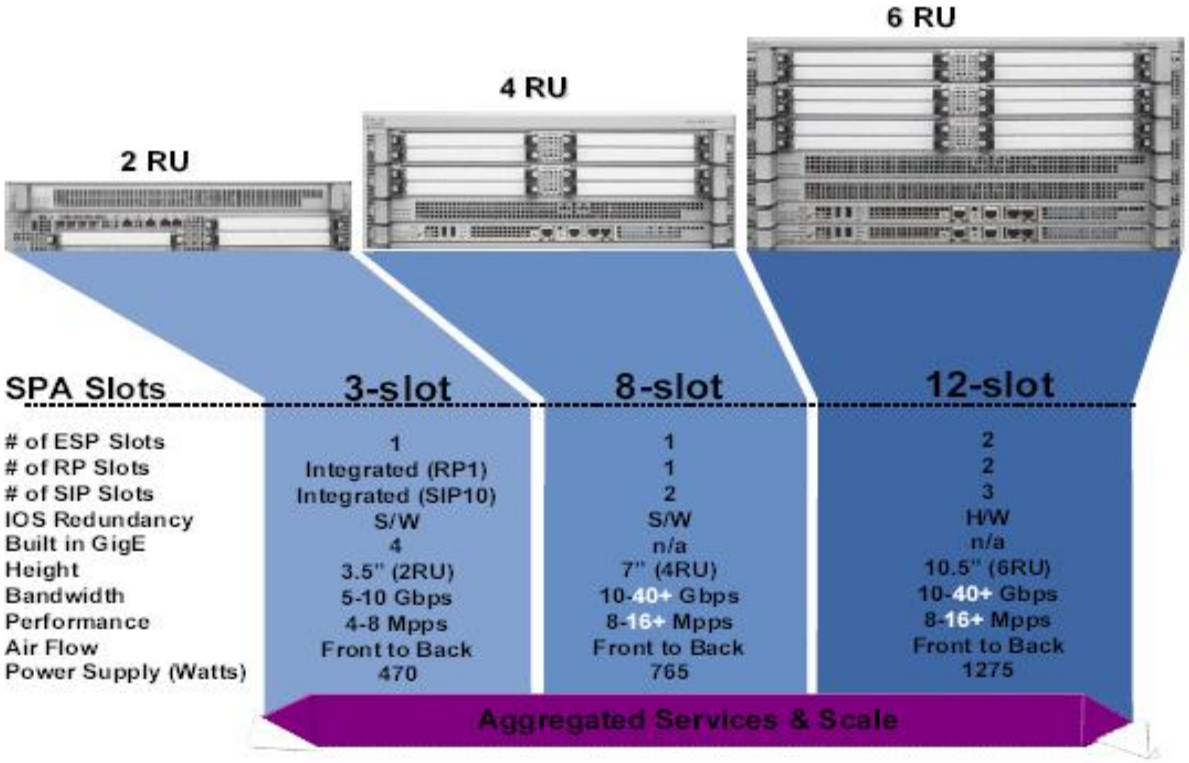


从上图可见，笔者将QFP的160个线程做了一个功能划分--控制线程和数据线程。
控制线程的功能是在QFP的多核上提供与ESP板子上主控CPU（一个PowerPC的处理器）进行通信，例如，接收来自主控CPU通过HT接口发 来的各种控制消息；将QFP的各种状态数据，Log数据等等发回给主控CPU。读者需要注意到的是：FIB表的任何更新路径是：控制平面（RE）通知并发 送给ESP的主控CPU。主控CPU通知并发送给QFP的控制线程。控制线程通过IPC将FIB存储在其能访问的内存中，例如RLDRAM。
数据线程的功能就比较直接和单纯。做纯粹的数据报文处理。
基本上，思科是不可能透露其QFP多核上软件结构的细节。但是，笔者认为，其结构应该大致为上述分析的结果。
在控制线程方面，另外一个工程选择可以是通过中断处理例程来实现，从而可以避免将160个线程中的几个单独拿出来做控制用，从而损害数据处理的能 力。但是通过中断例程的方法的缺点也是比较明显，例如，对线程的cache缓存的打断和侵入，从而导致数据报文处理的性能受到影响。还有的不确定性在 于，QFP的多核的中断处理逻辑的发送机制是如何设计的。是每个Xtensa的核都有一个中断逻辑，还是整个40个核拥有一个中断逻辑。
不同的QFP的微结构设计，都会导致软件系统设计的取舍不一。
另外，笔者感觉QFP上的40个核上的软件逻辑的稳定性和可扩展性非常有可能存在许多潜在的问题。如果没有一个良好的轻量级的微内核的支撑，上述的 “run to complete"的死循

环（while（1）...)代码结构是有许多问题的。当然，从这个角度，读者应该可以认识到，QFP里提供的硬件锁逻辑（提供 许多spinlock，mutex等）是非常重要的。对于性能和编程模型，都是很关键的。
另外，从目前分析的QFP多核的软件结构来看，对于一个QFP的线程处理了崩溃，对其他线程的影响是什么 ？ESP主控CPU如何对待这种异常，是重新启动QFP全部系统，还是做而且只是重新启动那个死掉的线程？从笔者的技术观点，应该是，一旦主控CPU发现 异常，HA的Active/Passive就会切换。这个当前的QFP全部重新启动，从主控CPU拿到QFP的代码，并且从新运行。这个也是非常可以理解 的，例如ISSU的Upgrade等。
总之，在QFP的多核上构筑一个良好的软件系统是不容易的。目前来看，思科的ASR1000上的QFP软件也是比较简单的。
这里面原因有二：
×报文处理引擎，简单就是美。简单就是高性能。
×非常不容易设计和把握。

8. 体系结构--系统观点

思科的QFP芯片，是作为思科的边缘路由器ASR1000的重要组成部分而粉墨登场的。ASR1000产品系列为三个：ASR1002，ASR1004和ASR1006。其不同的配置和结构的差别，可参阅下图所示。



思科ASR1000产品系列

在从系统的观点考察QFP和其所在的系统时，首先介绍一下ASP1000产品的一些命名约定。

CPP: Cisco Packet Processor。在ASR1000里，就是这个QuantumFlow Processor（QFP）

CC：CarrierCard。即线卡模块板。在ASR1000里，叫做SIP（SPA Interface Processor)。

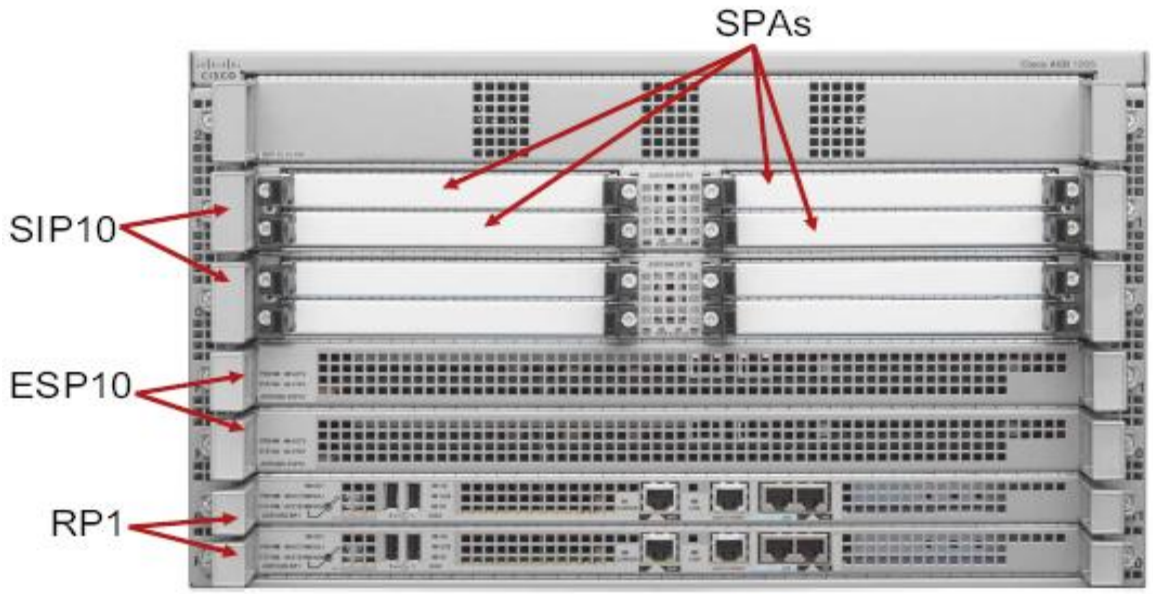
SPA（Shared Port Adaptor)为线卡。 一个SIP模块上可以插入1-4个SPA线卡。支持Ethernet，ATM，POS等等报文格式端口。

FP: Forwarding Processor。 在ASR1000里，是ESP（Embedded Services Processor)卡。QFP芯片就是被焊接在ESP卡上，并通过SP-4.2接口来获取数据报文的。

RP: Routing Processor。在ASR1000里，仍然是叫做RP。就是思科经典的IOS的控制平面卡的意思。

如果读者不熟悉思科路由器产品命名约定的话，要注意的是，不要把这里的一些“Processor”与真正意义上的CPU处理器相混淆。SIP，ESP和RP等的Processor的意思是一个逻辑概念，其物理上是指相应的硬件处理卡。

下图所示为ASP1006的一个硬件结构图。

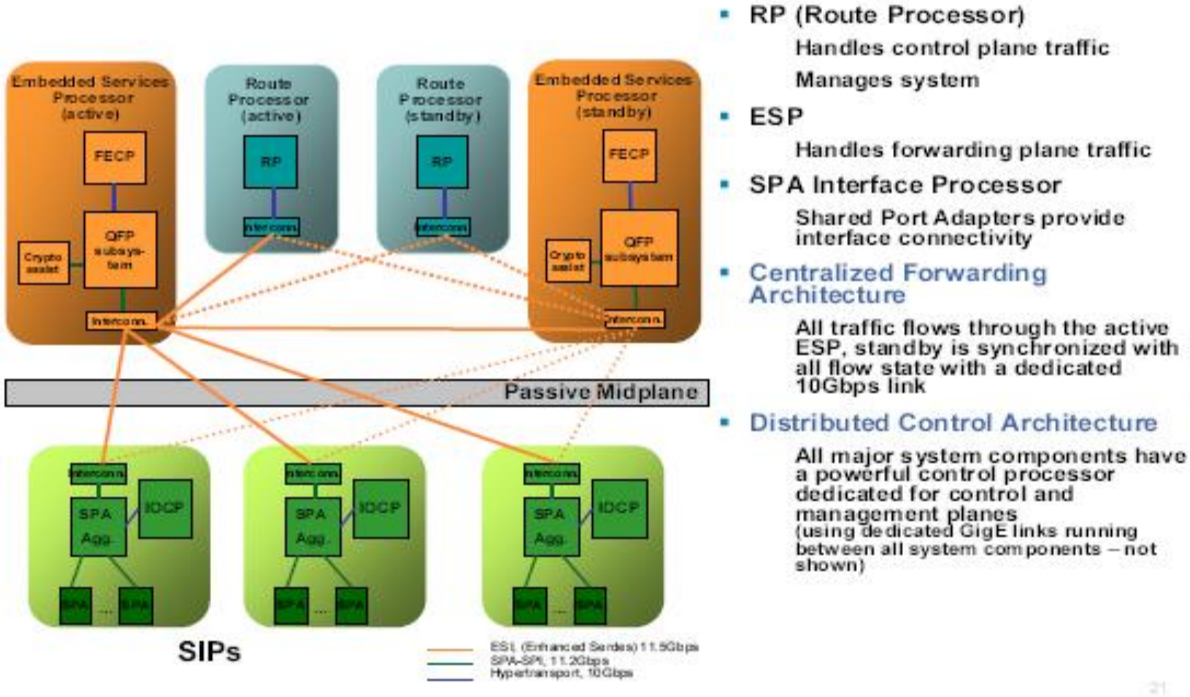


思科ASR1006硬件系统正面图一览

读者可以认识到，ASP1006可以支持两个RP，即控制平面卡，从而提高Dual-RE的HA。（ASP1004可以支持单RE的HA，即两个 IOS运行在Linux的KVM环境上。ASR1006不支持单卡的Dual-IOS结构。）。可以支持两个ESP卡，也就是，两个QF芯片 。这两个 ESP卡的HA关系是Active/Passive。从而为ASR1000提供Statue-ful的服务处理的HA支持。这是ESP10中10的意思是 10bps线速。。ASR1000不同的产品可以支持不同的ESP性能。例如ASR1002的ESP是ESP5，也就是说，是5Gbps的线速。 另外，可以看出ASR1006可以有两个SIP模块。每个模块可以支持1-4个SPA线卡。

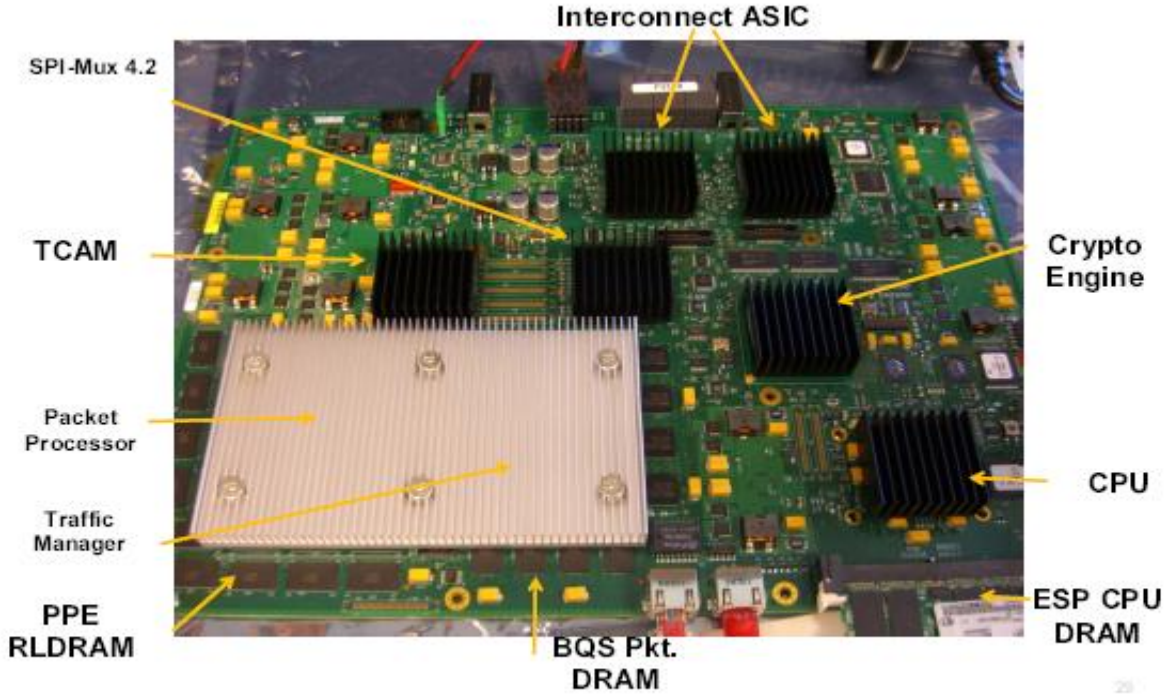
当读者试图把握ASR1000的系统结构时，很重要的一点是：ASR1000是一个分布式结构，但是一个集中式的数据处理。换言之，任何一个数据报 文，都要通过SPA-->SIP-->互联-->ESP，最后进入QFP进行处理。从RE出来的数据报文，例如BGP，OSPF报文， 也是如此。在系统中，QFP起着一个枢纽的角色。

下图所示为QFP，ESP在整个系统中的逻辑结构图。



基于QFP芯片 思科ASR1000逻辑结构图

另外，有兴趣的读者，可以浏览QFP在ESP卡上的物理布板结构。从而可以更好的理解QFP是如何 通过系统级别的互联，然后通过SPI-4.2来收发报文，与主控CPU进行通信，被主控CPU监测，启动，更新等等的硬件结构。



思科QFP/ESP版图一览

从上述图中，读者也可以发现，在ESP板子上有两个“Interconnect ASIC"。这个互联逻辑就是思科目前力挺的ESI--Enhanced Serdes Interconnect。ESI互联结构是ASR1000系统把系统的部件（SIP，ESP，RP）结合起来的通信路径。 换言之，ASR1000的背板是基于ESI的。
那么为什么在ESP上有两个ESI芯片，其原因是为了ESP之间的高可靠性（HA）。在一个ESP或者QFP在Active模式下的时 候，Active与Passive的ESP（或QFP）通过专门的一个ESI连接进行通信，从而完成Stateful的状态备份。实现ESP级别的HA。

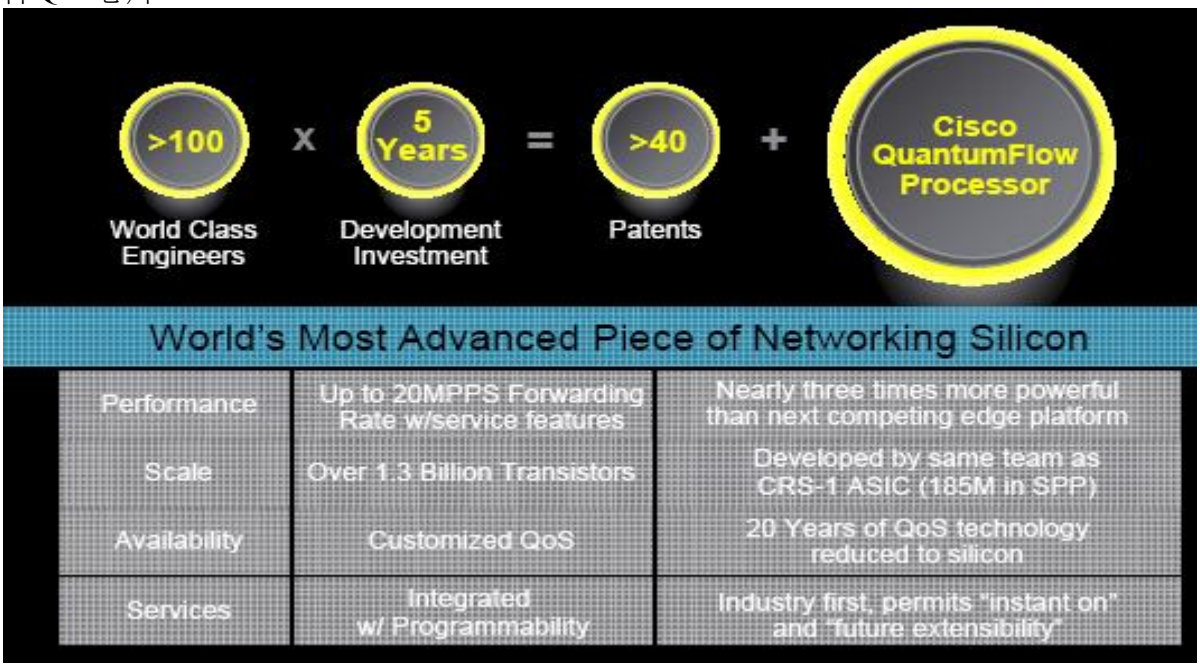
9. 战略规划

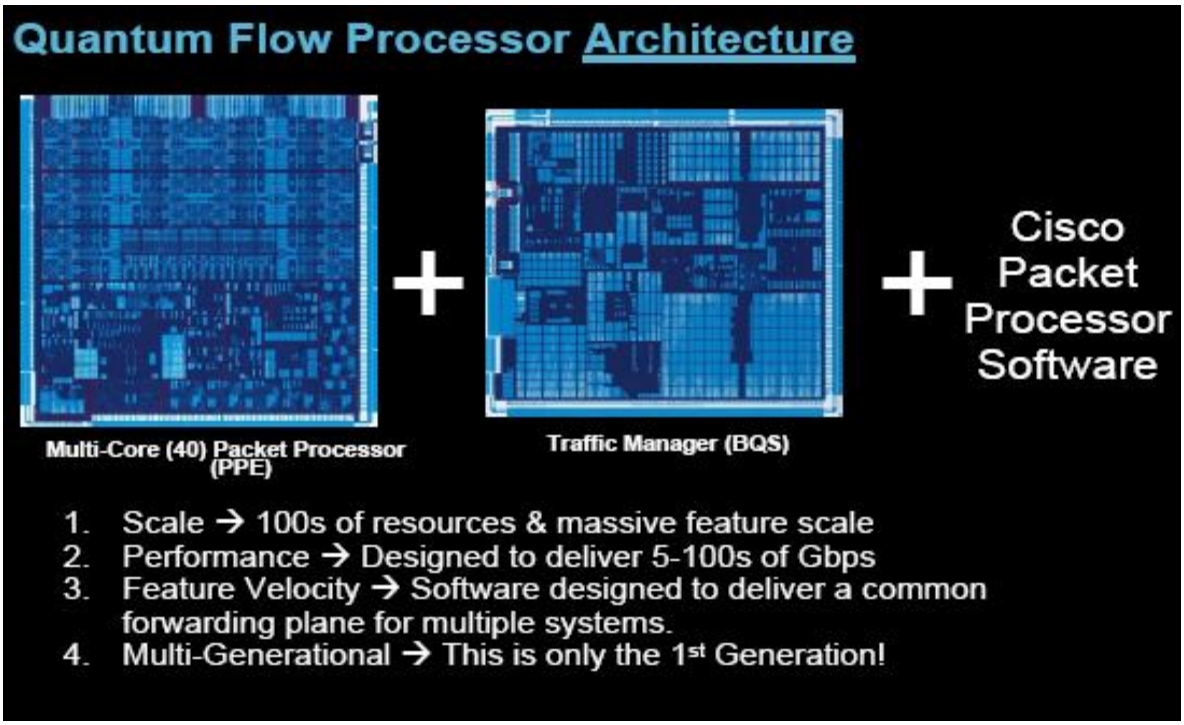
思科的QFP网络服务处理器及其ASR1000产品系列是思科在下一代网络（NGN)战略规划中的重要一环。以笔者的判断，其主导核心思想就是：
×在企业网方面，NGN的发展方向从LAN-Centric朝WAN-Centric的方向发展。 QFP与ASR1000系列就是为了加强思科在企业在WAN接入方面的竞争力。例如，取代其原来的7200系列产品，从各个方面帮助一个企业网的WAN接 入，例如，网络安全，WAN优化，声音与视频接入等。
×在运营商方面，NGN的发展方向是Edge Router的智能化。思科认为这运营方面，竞争的战火将在Edge上，而非在骨干上。Edge的智能化是下一步的发展方向。QFP与ASR1000就是这个战略规划下的产物。
下图所示可以看出ASR系统在思科WAN解决方案中的位置。

从性能参数比较，基本上可以认为，一个ASR1000产品相当于16个7200产品。 可见ASR1000的服务集成度之高，令人惊叹。
在运营商解决方案中，ASR1000的位置为下图所示：

可以看出，ASR1000属于思科智能边缘路由器的重要一环。这个“Intelligent”就是这个QFP所处理的各种网络服务所带来的。

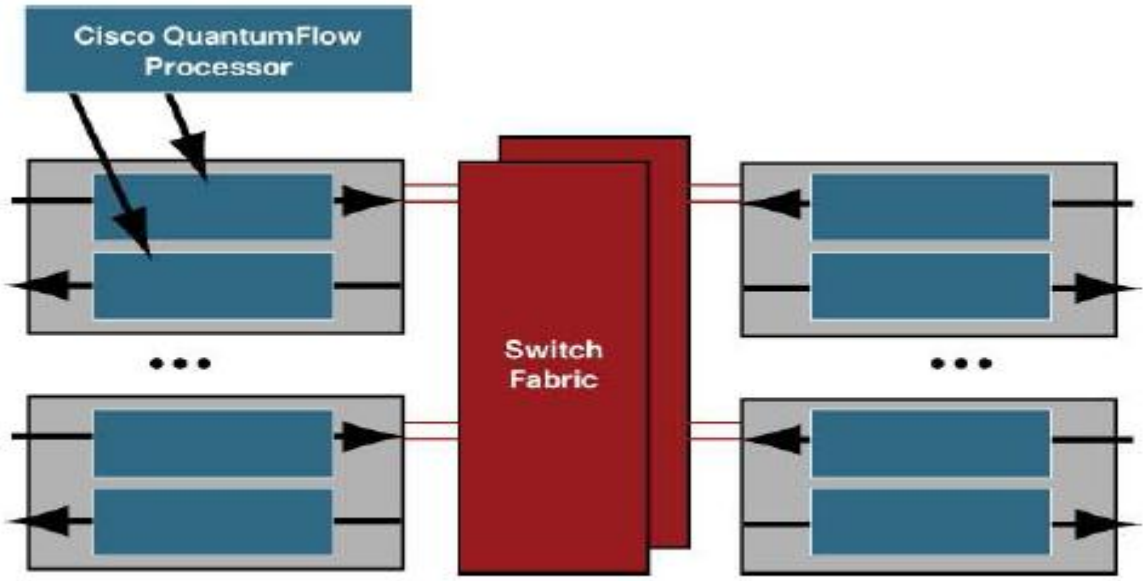
思科 为了实现其“决战在边缘”的战略，投入了大量的力量开发ASR1000系列最重要的部件QFP芯片 。





从思科公开的资料显示可以得知。目前的QFP及相应的系统只是其第一代产品。 在今后的数年，笔者相信思科将会持续推出新一代的QFP芯片发布和相应的系统。例如，20G，40G和100Gbps的QFP。因此，基于相应的QFP的 20G，40G和100Gbps的ESP将会持续的推动ASR系列的性能达到新的台阶。

另外，目前的ASR1000（1002，1004和1006）是采用了集中式的一个QFP（另外一个QFP做备份作用）处理数据报文的体系结构。思 科在将来的系统中，非常有可能将QFP应用在线卡（SIP）上，从而达到一个分布式计算结构。换言之，每个SIP卡模块上支持一个，甚至多个 （10Gbps，20Gbps，40Gbps，或100Gbps的）QFP。



思科未来ASR系统--分布式QFP结构

上述的基于分布式结构的QFP计算模式，将使得思科在ASR系列产品性能跨越式发展。其系统整体聚合性能将是目前ASR1006等的4，6，8或10倍不止。
这种结构的下一代ASR产品必将对NGN的WAN接入和边缘路由器的市场起着重要的重新洗牌的作用。

10. 结束语

思科在2008年3月，历经5年之功，推出了其基于40核的网络处理器QFP的边缘路由器ASR1000系列，1002，1004和1006，意在 取代其7200系列。在2008年11月，又推出其的最新的20Gbps的QFP/ESP数据处理卡。思科对于其边缘路由器服务智能化和WAN接入市场高 性能的重视足见一斑。
在核心路由器市场，思科和Juniper几乎占去绝大多数的份额。但是在边缘路由器市场争夺战中，竞争非常激烈。目前，几乎各大设备厂商都有自己的 边缘路由器产品，如思科的ASR系列、Juniper的MX系列，爱立信/Redback的SmartEdge系列，中兴的ZXR10ISR系列、华为的 NE40、20系列。思科的ASR系列产品必将给其竞争对手的相应产品线极大的技术和市场压力。
思科的ASR1000系列中，其最重要的一个技术突破就是QuantumFlow网络处理器的研发成功和使用。从而思科可以贯彻其高性能WAN接入 和智能化边缘路由器的战略举措。思科耗资1亿多美元、投入5年时间研发的QuantumFlow网络处理器，可以将防火墙、IPSecVPN、DPI、会话边界控制(SBC)等多种应用集成在一个系统中，而要在运营商级和终端用户网络的边缘提供这些功能，大约需要6个设备 。
笔者认为，随着思科围绕着QFP芯片不断升级和系统结构的不断复杂，其在WAN接入和边缘路由器市场将持续保持其主导地位，并且拉开相应竞争对手的距离。

居安思危，笔者希望这篇关于QFP处理器及其战略研究的文章能对读者有所帮助。