Multi-Service IronWare Operating System overview

IronWare® operating system software provides the intelligence behind the high-performance switches, routers and application load balancers from Foundry Networks. Leveraging over 8 years of experience in powering global networks, IronWare has incorporated continuous innovations to ensure non-stop operation of networks.

This white paper provides an overview of the new Multi-Service IronWare operating system that powers Foundry's new line of routers and switches[1]. This operating system (OS) is designed to efficiently address the diverse needs of today's and tomorrow's infrastructures, while providing a flexible framework for powering multiple Foundry platforms. In addition to an overview of the overall architecture of the Multi-Service IronWare OS, this paper also discusses its advanced capabilities such as high availability, modularity, fault tolerance, network monitoring, security and its role in the inevitable migration to IPv6.

Need for non-stop operation

The increasing role of the IP network in powering a variety of end-user applications in both an enterprise and a service provider network, makes network uptime and, by extension, uptime of individual nodes extremely critical. The experience of several network administrators has shown that software plays a very vital role in ensuring network uptime. Non-stop operation capabilities in an operating system (OS), therefore, act as insurance against the unexpected and ensure that failures in a component of a node are not only quickly detected but also swiftly addressed via a corrective action.

Key Characteristics

Multi-Service IronWare operating system incorporates several capabilities that make it ideal for high performance and high availability:

- Distributed operating system with a very high level of distribution for maximum efficiency
- Modular operating system that prevents corruption/interference of different modules within Multi-Service IronWare
- Multi-threaded operating system with different software modules running as light-weight threads
- Pre-emptive operating system for predictable performance by individual threads
- High availability capabilities to ensure non-stop operation of the system
- Advanced protective checks in the operating system to ensure reliable operation of the different software modules

Modular architecture

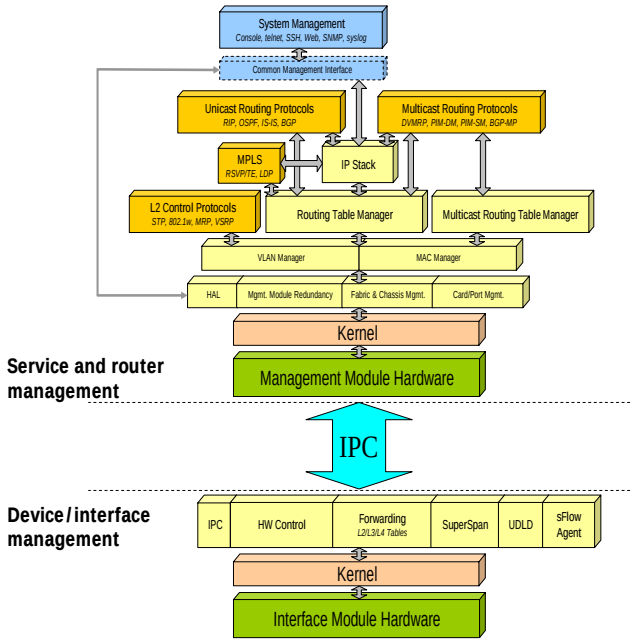

Figure 1: Components in Multi-Service IronWare

The modular architecture of the Multi-Service IronWare OS incorporates a clear separation between different modules and logical components in the operating system. For example, an abstraction layer hides the underlying hardware layer from the upper layers, such as the protocol layer. The underlying modular design of the OS greatly enhances the robustness of the system and makes it easily portable to new platforms. To a network operator, this gives a common look and feel to all systems running

[1] Although the same Multi-service IronWare is leveraged across a variety of platforms, the different capabilities of IronWare are fine-tuned to the specific underlying product.

Multi-Service IronWare, thereby decreasing the operational costs of running a network.

Figure 1 shows the key components in Multi-Service IronWare.

The modular and distributed architecture is also reflected in the industry-leading quick cold-restart time for routers and switches powered by this OS. For example, even high-end, feature-rich systems such as BigIron RX switches and NetIron XMR routers have a cold restart time of less than 60 seconds from cold-start to service activation, even after extensive diagnostic tests that are run during system bring-up.

A few things to note in Figure 1 are:

- Distribution of Layer 2/3/4 forwarding tables to the interface module. This allows packets to be forwarded by the hardware on the interface module, without any processing by software.
- Distributed sFlow agent on every interface module. This distribution allows Foundry Networks switches and routers to provide very high performance traffic monitoring even when sFlow is enabled on all ports concurrently.
- UDLD (Uni-Directional Link Detection) is performed completely on the interface module for speedy detection of link failures and resulting corrective action.
- SuperSpan is also performed by the interface module for scalable and high performance tunneling of customer Spanning Tree BPDUs.[2]

Clear separation between control and data plane

Multi-Service IronWare is architected with a clear separation between the control plane and data plane, which ensures a high availability system at all times. Control messages between the interface module and management module are sent on a separate out-of-band redundant link that is distinct from the path used for data traffic. Consequently, even during times of very high utilization, control messages, which are crucial for correct system operation, can still be exchanged reliably between the management and interface modules, without any loss.

Common denial of service attacks in the network often target vulnerabilities in a system when system utilization becomes high. A clear separation between control and data plane greatly diminishes the probability of a system being brought to its knees by a DOS attack.

High Availability—always!

Multi-Service IronWare addresses high availability of a network from multiple angles:

- Protection against faults within the system
- Protection against faults within the network

Just as a chain is only as strong as its weakest link, a network too should protect against faults in the individual nodes and connectivity related faults at the physical and higher network layers.

The root cause for software faults in a system can often be traced to issues such as lost control messages, buffer overflows or poor congestion handling mechanisms in the control plane. It is therefore extremely important for the underlying kernel to provide robust mechanisms and take preventive measures against such conditions.

Protection against faults in the system is achieved using several principles:

- Advanced protection capabilities of the operating system's robust kernel
- Fault-tolerant capabilities to ensure non-stop operation in higher layer applications

Protection against faults in the network is accomplished by incorporating redundant links/paths in the network architecture with the use of appropriate protocols to detect and handle network faults.

Each of these goals is accomplished using several mechanisms that are described in the section below.

Ironclad protection in the kernel

In order to ensure non-stop operation, the OS incorporates advanced protection and fairness mechanisms. These underlying constructs ensure high availability of the system, predictable and

[2] For more information on SuperSpan, please refer to the white paper, "SuperSpan: A Break-Through For Layer 2 Networks", http://www.foundrynet.com/solutions/appNotes/PDFs/SuperSpan.PDF

deterministic performance, and processing prioritization of high-impact events at peak load. Some of these are described below:

- Enforcement of read/write rights by the kernel on shared data structures that are accessed by multiple processes. With this capability, a process can register to access a certain shared data structure with read-only or read-write privileges. Violations of these access privileges can then be spotted immediately by the kernel.
- Detection of memory violations by using separate virtual memory spaces. Key components of the OS such as routing protocols have their databases in a secure virtual memory space to prevent accidental corruption/access of those tables by errant threads.
- Detection of errant threads: The kernel watches for symptoms of errant threads. Examples of monitored symptoms include spotting of stack usage violations, excessive memory usage by a thread, or excessive CPU hogging by a single thread
- Multiple communication paths during inter-process communication (IPC) with tiered priority levels between the interacting threads.
- Lightweight, reliable transport of IPC messages exchanged between processes on different slots.
- In distributed operating systems, it is imperative that the main table on the central management module and the cached copy in the interface module remain in sync. Multi-Service IronWare incorporates an integrity check mechanism wherein the data table integrity is checked periodically between the module hosting the main copy (e.g. management module) and the module hosting the cached copy (e.g. interface module).

Fault-tolerant capabilities

Multi-Service IronWare OS includes several capabilities to enhance the tolerance to faults in both the system as well as the network:

- Hitless Layer 2 and Layer 3 failover ensures that layer 2 protocols and layer 3 protocols running on the primary management module smoothly fail over to the redundant management module with virtually no negative impact on traffic forwarding nor the routing/switching domain throughout the network.
- Graceful restart for BGP and OSPF allows a router to cooperate with its adjacent routers in updating its own routing table (e.g. after a management module failover) without causing network wide disruptions due to routing protocol re-convergence.
- In-service OS upgrade allows new software patches/versions to be downloaded to the system without any loss in traffic.
- Switch fabric / fabric element / single lane failure detection in just a few milliseconds
- Removal of switch fabric module with no loss of traffic, a capability that is extremely useful in planned network upgrades.
- Rapid detection of ECMP failures, irrespective of whether the constituent links are on the same interface module or on different modules
- Rapid detection of trunk failures, irrespective of whether the constituent links in the trunk group are on the same interface module or on different modules.
- Support for protocols such as VRRP / VRRP-E, RSTP, LACP (via IEEE 802.3ad), ECMP, BFD.
- In MPLS networks, Fast re-route and standby LSP Paths are also invaluable in protecting against faults in the network. Multi-Service IronWare provides both capabilities on MPLS-enabled platforms.

The ability to do in-service upgrades is particularly critical in ensuring high availability. Several studies have shown that planned upgrades of the network constitute a significant portion of network down-time. In contrast to the ability of downloadable modules, wherein specific processes can be restarted, in-service upgrades with multiple management modules allows for almost uninterrupted operation of the system during upgrades.

Security in Multi-Service IronWare

Multi-Service IronWare's security features assist in guarding against malicious attacks that left unattended, could eventually lead to a compromised infrastructure. Support for Secure Shell (SSH-v2) and Secure Copy (SCP) ensures that management traffic is encrypted. Similarly, use of SNMPv3, use of MD5 authentication in

routing protocol exchanges ensures that session or protocol exchanges are not exchanged in the clear. User authentication can be done using several mechanisms such as RADIUS or TACACS+. 802.1x authentication is also supported on switch ports with advanced features such as multiple 802.1x client support (including limiting the number of clients that can be authenticated per interface) and MAC port security.

The operating system allows configuration of the underlying hardware to detect and prevent Denial of Service (DoS) attacks such as SYN attacks or Smurf attacks by monitoring interfaces for unusual activity and setting thresholds for various traffic types. When a DoS attack is detected, the port is automatically shut off by the OS and an operator alert is logged.

Foundry's high-end platforms provide a large capacity for ACL entries that are distributed on each individual interface module, which is extremely valuable in ensuring security of the network. In addition to these large numbers of hardware-based ACLs on the platforms, the OS also has extensive support for ACL accounting and ACL logging. With these capabilities, network administrators can track the number of hits to an ACL clause and also log the information in an IP packet header when a hit is encountered. Such functions are completely distributed to the interface modules to increase efficiency.

In high-end platforms, a set of mechanisms called CPU protection is employed. These mechanisms protect the local processor (LP), residing on the interface module, against CPU hogging tasks. Under this framework, more functions that traditionally required CPU processing are delegated to the hardware, like broadcast/multicast/unknown unicast flooding over VPLS. In addition, non-critical traffic requiring some LP processing for completing a specific task, like ACL logging / uRPF logging, is automatically throttled.

The amount of traffic going to the control plane is a common Trojan horse that is employed by hackers to gain control of the system. Multi-Service IronWare OS allows ACL-based hardware traffic policers to be applied to traffic going to the CPUs within the system. Further, multiple priorities are used by the OS in handling control traffic so that higher priority control traffic is given preferential treatment in the event of a congestion.

Multi-Service IronWare also has numerous advanced facilities that can be used to mitigate or prevent malicious man-in-middle (MiM) attacks. These capabilities are implemented with support in the underlying hardware to ensure high performance even when the security features are enabled on the platform. Some of these advanced facilities are:

- BPDU guard to prevent hijacking of networks running spanning tree protocol by an errant switch
- Dynamic ARP inspection to detect malicious ARP packets and false replies to ARP requests
- IP source guard prevents source IP address spoofing by malicious users. With this capability, the OS automatically installs an anti-spoof filter in the underlying platform hardware after learning the IP address on a port.
- With DHCP snooping, the OS automatically inspects DHCP packets, learns and maintains IP address to MAC address bindings.
- DHCP option 82 (relay agent) functionality on Multi-Service IronWare can be used to ensure that DHCP requests from clients across un-trusted ports are forwarded to the DHCP server after attaching information on the circuit-id corresponding to the port.

Easing the IPv4 → v6 migration

As IP-based delivery of services continues its onward march, the demand for IP addresses will continue to increase at a rate that will make migration to IPv6 imperative. Nevertheless, it is essential for such a transition to be seamless. Multi-Service IronWare offers two mechanisms to facilitate this transition:

- Dual-stack routing where routers run dual IPv4 and IPv6 stacks: In this scenario, the backbone can be a dual-stack backbone with both the IPv4 and IPv6 routing protocols and forwarding processes running in parallel. End-systems can also run dual IPv4/v6 stacks in such a network.
- IPv6 over IPv4 tunneling, which allows different IPv6 domains to communicate via an intermediate IPv4 network. The OS supports the ability to create manually configured tunnels, automatic IPv4-compatible IPv6 tunnels, and automatic 6to4 tunnels.

Feature-rich Multi-Service IronWare

Multi-Service IronWare supports a broad range of capabilities that makes it suitable for mission-critical applications in both enterprise and service-provider networks. The ability to support multiple services on the same port simultaneously is one of the key attributes of Multi-Service IronWare. The OS offers extensive support for unicast IPv4 and IPv6 routing protocols, multicast protocols, and great flexibility in configuring ACLs and traffic policers. On core router platforms such as NetIron XMR, MPLS routing protocols such as RSVP, LDP and advanced traffic engineering capabilities permit the implementation of large-scale L2 VPN and L3 VPN networks. Advanced capabilities such as GRE tunnels and virtual routing without MPLS provide additional choices to the network designer for designing a VPN.

For details on the specific feature set that is enabled on a platform, please consult the appropriate product datasheet.

The power of sFlow

sFlow, specified in RFC 3176, is a powerful network monitoring technology that can be used for a variety of purposes such as network anomaly detection, fault management, performance management, service accounting, billing and more. sFlow uses a statistical packet sampling approach to accomplish its objectives and can be used to troubleshoot any L2-L4 flows in the network. Multi-Service IronWare's distributed implementation of the sFlow agent on the target platform allows real-time traffic monitoring on all ports without any performance compromise.

On platforms that support L2 VPNs or L3 VPNs such as NetIron IMR or NetIron XMR, sFlow can also be used on VPN endpoints to provide valuable information on VPN traffic.

Summary

Foundry Networks' new Multi-Service IronWare operating system is a feature-rich, multi-threaded, distributed, operating system with advanced capabilities to ensure secure, non-stop operation and high availability of the network. The versatility of its design makes it an ideal operating system for powering many of the routers and switches available today from Foundry Networks.

Document version 3.1

Foundry Networks, Inc.
Headquarters
2100 Gold Street
P.O. Box 649100
San Jose, CA 95164-9100
U.S. and Canada Toll-free: (888) TURBOLAN
Direct telephone: +1 408.586.1700
Fax: +1 408.586.1900
Email: info@foundrynet.com
Web: http://www.foundrynet.com