Distributed Intelligent Switching Solution

Today's Switched Network Structure

With the quick development of technologies, extensive application of ASIC chips and network processors, and the large scale prevalence of optical fibers, the Layer 3 switch changes from network edges application to network center application. The traditional Local Area Network (LAN) switching technology has become the mainstream networking technology for campus network and Metro Area Network (WAN). Under such technology background, the architecture mode of today's enterprise network basically includes core switches and edge switches. The core devices are two or more high end Layer 3 switches which connect to the middle and low end switches at the network edge with the 1000M fiber as the backbone. The edge switches connect to the desktop system through 10/100M link.

Enterprises facing the severe competition of the markets, regardless of its scale, take information processing and network communication system as the basic facilities and means of production to improve the production efficiency, thereby promoting the core competitiveness of the enterprise. In the enterprise operation mode which relies greatly on the network, how to ensure the network running normally? How to solve the problem quickly? How to extend the network with the extension of services? These all demand the networks of enterprises with high reliability, high management and flexible scalability. At present, enterprise network provides solutions through the following methods:

Network Center

Network center emphasizes high switching capability and high reliability of devices, so network center mostly consists of two chassis switches with high performance and high reliability. The two devices are redundant backup for each other, and implement the gateway backup with edge switches through VRRP technology. The chassis switch itself has good scalability and can meet the extension needs only by plugging in the card.

Network Edge

Enterprise network edge is mainly to implement the terminal access function. What important is the port density and scalability, while the reliability cannot be ignored either.

The box switch is generally adopted in consideration of cost. You can extend the port through stacking or cascading, improve the reliability of network links through STP/RSTP/MSTP, increase the uplink bandwidth and enhance performance by using link aggregation technology, and improve the connection reliability between the center and edges.

Network Management

Now, network management is mostly implemented through the SNMP-based NMS software. Most small-medium-sized enterprises use TELNET or WEB to manage devices directly. Larger enterprises adopt separate NMS software or platform to manage the whole network and devices.

Disadvantages of Traditional Networking

Reliability

In current networking, the reliability is implemented through redundant devices and links. Technically, the reliability of the devices and links in the network is increased by adopting VRRP, STP/RSTP/MSTP (generally referred to as STP) and link aggregation.

In most cases, although VRRP and STP technologies can meet the reliability demands, but they still has some disadvantages. This is mainly because these technology implementations do not consider the devices (which are backups for each other) as an entity. These technologies involve the master slave mode, an active/passive method emphasizing redundancy excessively, thereby causing disadvantage in load sharing.

In VRRP technology, all network functions of each device participated in VRRP run independently. Only when being the gateway of a VLAN, devices are divided into master device and slave device. Normally, only the master device provides forwarding service for the data of this VLAN, while the slave device is in idle state. This not only causes some devices are extremely busy while some are idle, but also wastes slave device.This issue is solved through network design in many traditional solutions. e.g. core switches A and B as VRRP, where A acts as the primary gateway of VLAN1 and B act as backup gateway of the VLAN2; while in VLAN2, B acts as the primary gateway while A acts as the backup gateway. In this way, data flow can be distributed across two core switches manually for load balancing. Of course, disadvantages of this load balancing mode are obvious: Firstly, the prerequisite of there have to be multiple VLANs. Secondly, if data flow size of each VLAN is different, load of core switches is unbalanced even if there are multiple VLANs.

STP has same problem. Normally, only the active link is transmitting data and the backup link is completely idle. Therefore it is impossible to achieve load balancing. Though LACP can perform load-balancing function, multiple physical links can only be connected with a same switch. This only allows link load-balancing instead of equipment load-balancing, and equipment itself might become single point of failure.

Management

In today's networking, you can manage network devices by using various methods. From the perspective of bandwidth, it includes in-band management and out-band management. in-band management includes methods such as WEB, NM software (SNMP), remote TELNET; in-band management includes management through the control port of the device, and so on.

Because in-band management can be implemented remotely, it is used more frequently. But it demands the device have a unique IP address. In traditional network, except some stack methods, most approaches demand the managed device configured with IP address. It may not be a big issue for the small number of core switches; while for large number of network edge switches, it not only complicates the configuration, but also occupies more address space. Even for switches (with one IP address) running in the stack mode, only individual switch stacked together are displayed on the management interface, not a real entity. The specific configuration is to configure the individual switch separately.

Networking Cost

In today's switched networking, the center is chassis switch and the edge is box switch. Existing technologies cannot fully solve the problem of load sharing while implementing redundancy. So when users buy devices in the initial stage, the performance of a single center switch must meet the needs of forwarding data stream in the whole network. Considering the future development, demands for device performance is much higher than the actual demands when networking. Therefore, center devices have to purchase the whole chassis, increasing the networking cost. In point of reliability, purchasing the same device to be a backup will reduce the efficiency of investment. If the performance of core devices cannot meet the development requirements in the future, we have to purchase new core devices again. Thus we might encounter such an embarrassing situation: performance of the existing equipment is inadequate when used at network core but is excessive when used at network edge.. The investment cannot be protected well.

Solving Problems

From the above-mentioned problems, we can see that whether the core switch or edge switch in the current enterprise network does things in its own way. For example, two switches in the core, whether to forward data packet or process routing protocol, they both operate independently rather than form an entity. The connection between two switches is just to backup with each other through VRRP or STP technologies. Only one core switch works when two core switches process the data packets for the same network segment. So device resources and link resources are greatly wasted. It is actually a networking method with centralized load.

For this problem, we hope there is such a technology, which can compose multiple switches into a logical entity, and all switches can sharing the load. If there is a failure in one device or link, other devices or links can share the load of the failed device without affecting the normal running of services. Extending the network only needs add a switch to the logical entity. In management, logical entity fully behaves as a single device. This technology will greatly improve the reliability, scalability and management of the network, and reduce the networking cost in initial stage simultaneously.

This is the IRF distributed networking technology provided by Huawei-3Com.

IRF Distributed Networking Solution

Introduction to IRF Technology

IRF means **Intelligent Resilient Framework**. It is a completely new software solution. Multiple devices that support IRF can be interconnected to form a “Joint Device”, which can be called a Fabric and each device that forms the Fabric is called a Unit. Multiple units become an entity in terms of management and use after forming a Fabric. It is possible to expand port number and switching capability of the device by adding additional Units at any time or enhance equipment reliability through mutual backup between multiple Units. In addition, user management can be simplified greatly since the entire Fabric can be managed as a single device.

Simply speaking, by interconnecting multiple units, IRF equipment provides a manageable, scalable and highly reliable solution longed for users. It is a kind of device with completely new network concept, which is totally different from that of all the existing networking equipment.

IRF technology consists of the following three parts:

- **Distributed Device Management (DDM)**: Control system of the IRF technology. It is responsible for releasing varies of management and control information to the IRF distributed switching architecture.
- **Distributed Resilient Routing (DRR)**: It allows multiple interconnected switches in an IRF distributed switching architecture to act as a single active routing entity. In addition, it can assign routing load intelligently among all the switches to maximize networking routing performance.
- **Distributed Link Aggregation (DLA)**: It enables full-meshed interconnection between network core devices and edge devices.

IRF provides a number of benefits including high availability, high performance, easy-managing and IT budge optimization, etc. In addition, switches that support IRF technology can interoperate with the existing switches that do not support IRF technology. Switches that do not support IRF technology will not be used as a part of IRF distributed switching center. However, they can be managed as an independent entity through link aggregation, Spanning Tree Protocol or link redundancy technology. Moreover redundant configuration still effects

IRF technology can be used to create network core with high availability and scalability. Its performance, configuration and scalability can grow with network to avoid the big investment at the beginning and physical restriction typically associated with centralized network core devices. Therefore, IRF technology can help enterprises reduce total cost of ownership through a ""Purchase on Demand and Scale Gradually" policy.

IRF technology represents seamless combination of distribution and centralization. Members of the Fabric handle Layer 2 and 3 data forwarding, Layer 2 protocol and routing state independently, which is quite useful for ensuing reliability and integrity. While viewing from outside, these members seem like a single entity with respect to routing protocol, Layer 3 message forwarding and management. Finally, they perform configuration and log output centrally.

Networking Diagram

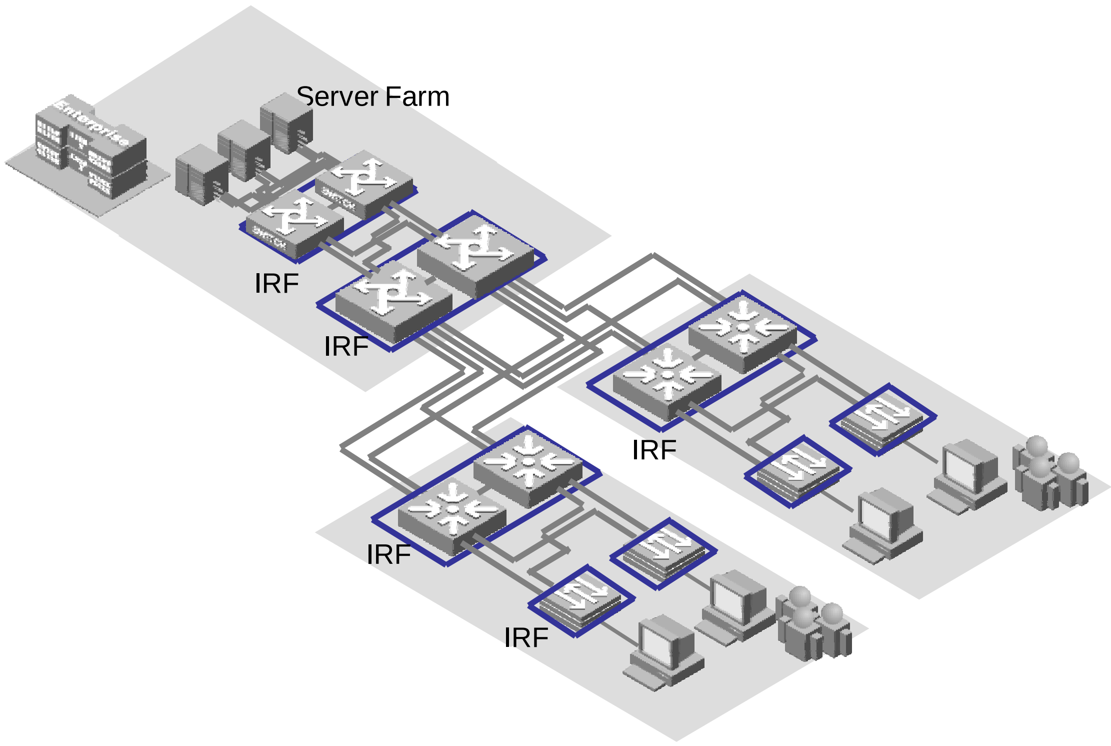

Figure 1 IRF whole network solution

Figure 1 is the whole network solution diagram of IRF, in which all switches support IRF function. In this figure, two core switches, server access switches, distribution layer switches and access layer switches are all in IRF Fabrics. Each hierarchy is connected with others through dual-homed links.

In this figure, IRF networking and traditional networking is similar in topology, except that there are more connections between each hierarchy. But these seemingly complicated connections can improve the reliability and overall forwarding to a new level, and is easier to manage than traditional network.

In Figure 1, network devices in different hierarchy are connected through dual-homed links. Although there are many physical lines, but the devices compose a LACP group according to DLA technology of IRF. Therefore, the network with IRF fabric improves the forwarding capability of the whole network and guarantees the reliability of the network. With the future development of services, you can add new IRF switches to each IRF Fabric to improve performance of the network core and edges as required. This is also networking method in IT industry, which first benefits the reliability, scalability and performance of the network through the same technology.

DLA connection not only avoids single point failure on the line, but also allows one end of LACP to be connected to different physical units in Fabric. This also avoids the risk of all link breakdown caused by the single point failure of device, and taking advantages of LACP link load balancing.

Excellent network also needs excellent management. In network management, IRF switch networking has unparalleled advantages. Its DDM feature combines distribution and centralization organically, and manages the Fabric composed of individual IRF switch as a single device. You can only see a device whether managing the switches through NM software, WEB, TELNET or control port. You only need to configure the Fabric without needing to configure each device individually. Simultaneously, Fabric also has centralized log output. In the logical NM view, each network hierarchy in Figure 1 only has one switch. The management interface is simplified greatly, as shown in Figure 3. Each Fabric only needs a management IP address. In fact, even managing through the control ports of different Units in each Fabric, you can only see the management interface of the same device. Without

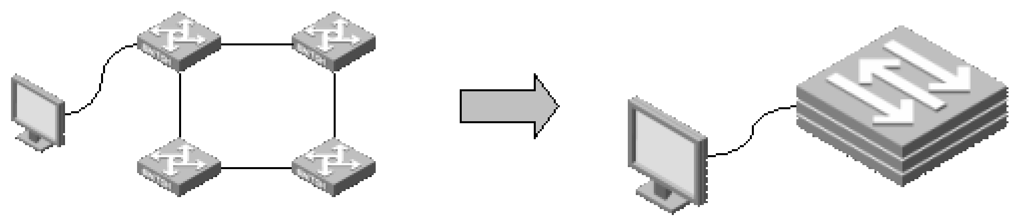

Figure 2 Distributed fabric and centralized management

question, in this mode, not only the topology structure of network is simplified, but also the number of the devices to be configured is reduced greatly. It benefits the initial installation and the future maintenance of the network greatly.

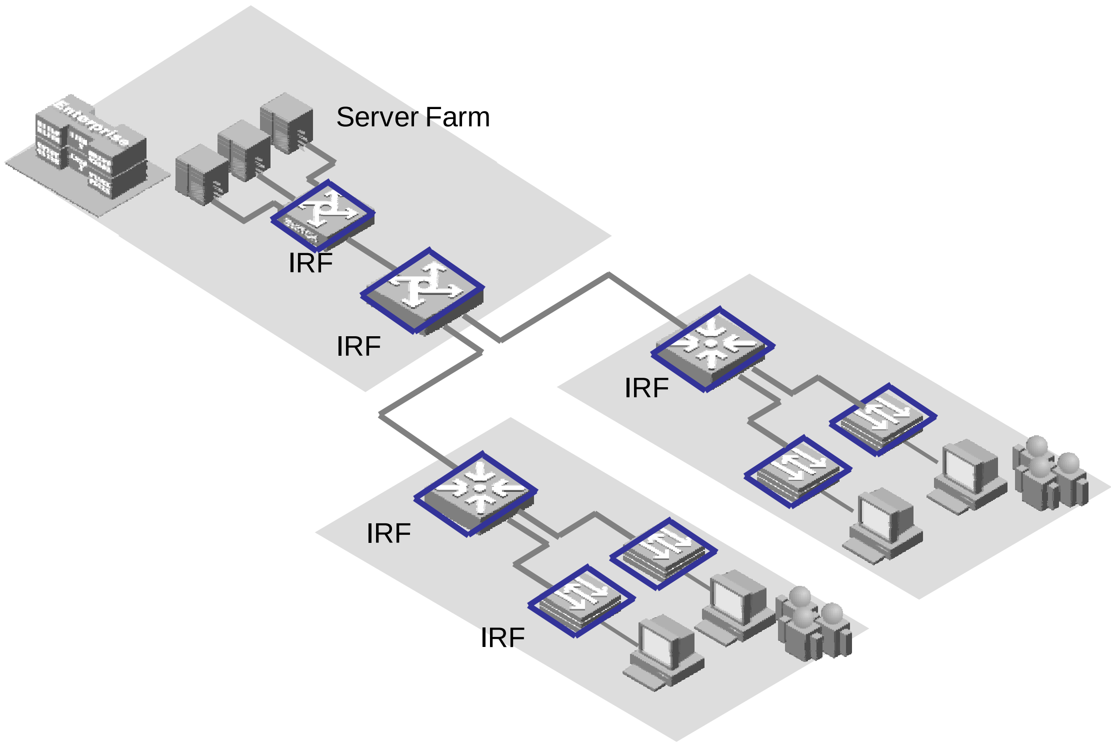

Figure 3 Management view of IRF whole network solution

Compared with traditional networking, IRF networking technology gains an advantage in networking cost. Its networking principle is "“Purchase on Demand and Scale Gradually”. Increase of Units in IRF Fabric will improve the overall data forwarding capability of Fabric. In the initial stage of networking, you do not need to purchase expensive core switches beyond current needs only for future development. Choose the IRF-enabled core switches so as to add new IRF switch dynamically to form a Fabric to enhance the whole performance. IRF technology allows the Units in each Fabric to join or leave (similar to the hot swap of the chassis switch), and it can implement the seamless upgrading of the network. This networking technology which only needs to pay for the current requirements will save an amount of device investment for users.

Advantages of IRF Distributed Networking

Distributed Features

When distributed devices are combined to form a Fabric through IRF technology, they can appear as a completely new device, and IRF switches can be added into IRF Fabric seamlessly as required at any moment. Thus a new networking method with high reliability and scalability is provided.

1. Device units in Fabric are distributed physically

Each switch in the IRF fabric is independent. These switches interact with each other through special link and protocol. Because the optical fiber is allowed for connection, you can connect multiple IRF switches remotely to form a Fabric. Therefore, in some applications such as remote backup, although the switches in backup center and information center are not in the same place, they can compose a Fabric and be managed as a same device.

2. Distributed Link Aggregation

Traditional link aggregation technology is used to extend the link bandwidth and implement link redundancy. The disadvantage is that each end can only connect with the same switch, thus all links will disconnect when the switch fails. Therefore, traditional LACP technology only stresses links rather than networking. The DLA technology of IRF can solve this problem. Because multiple switches in IRF Fabric behave as a single device logically, LACP can connect to different switches when connecting with the IRF Fabric. DLA supports three aggregation types: manual aggregation group, static aggregation group and dynamic aggregation group.

3. Distributed Layer 2 and Layer 3 Data Forwarding

In traditional stack application of switches, Layer 3 data forwarding is as shown in Figure 4. That is to say, in the stack devices, only the master unit has Layer 3 forwarding capability, other units must send the Layer 3 packet to the master unit (unit1 in Figure 4) for Layer 3 forwarding after receiving the Layer 3 packet.

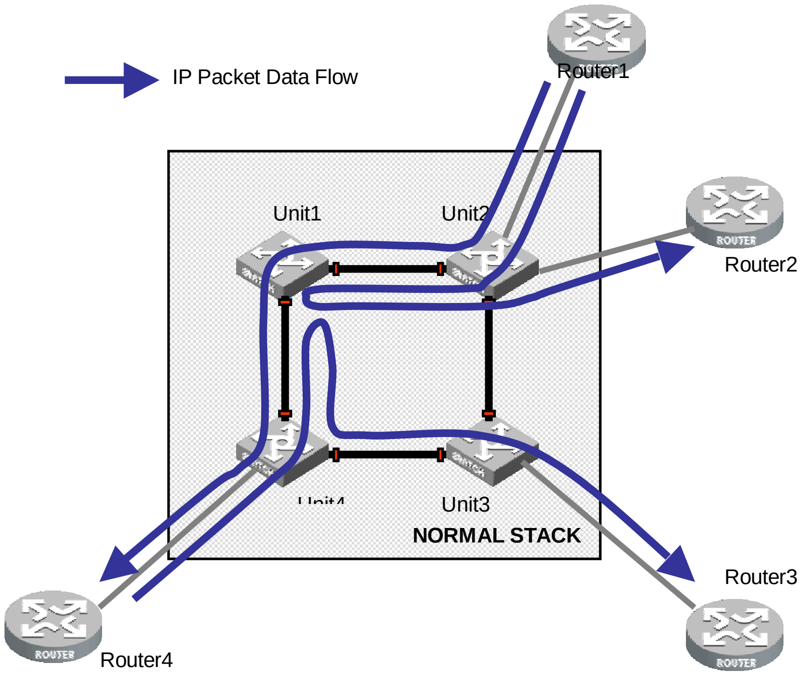

Figure 4 Layer 3 forwarding diagram of normal stack device

In the Fabric composed of IRF switches, because each unit contains the Layer 3 routing table, Layer 3 FIB table and Layer 2 FIB table for the whole Fabric, all units can forward Layer 2 and Layer 3 data. Layer 3 data forwarding mode is as shown in Figure 5:

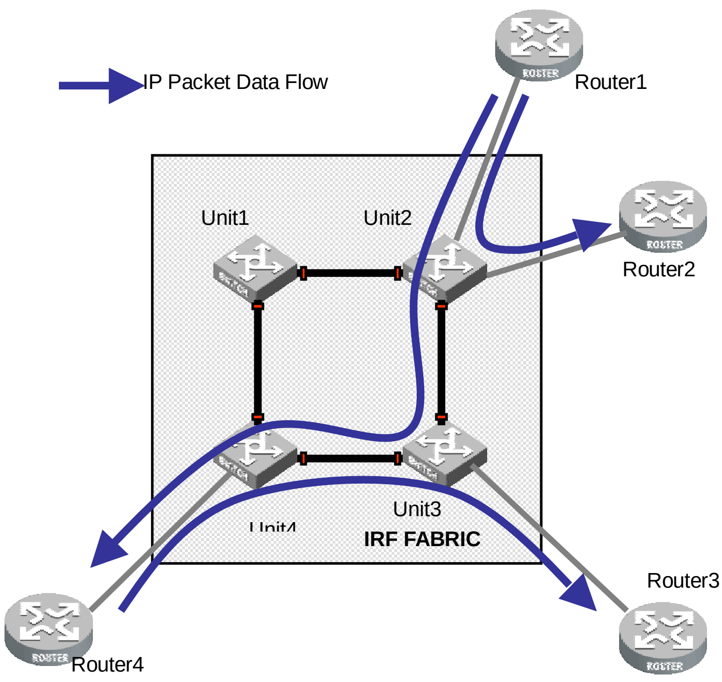

Figure 5 IRF distributed Layer 3 forwarding diagram

Any Unit in the Fabric has complete Layer 3 forwarding ability. When a unit receives the Layer 3 packet to be forwarded, it can search the Layer 3 forwarding table in itself to get the outgoing interface and the next hop, and then send the packet to the correct egress port. This egress port may be on this Unit or on other Units. Sending the packet from one Unit to another Unit is an inner implementation, and is shielded completely from the outside. For Layer 3 packet, regardless of how many Units it has passed through in Fabric, only the number of hops increases one, namely, only one network device has been passed. Combined with link connection of DLA, each switch forwards data locally to make the forwarding capability of Fabric be sum of forwarding capability of each Unit, thereby maximizing the data forwarding capability of Fabric and making full use of the network bandwidth.

4. Distributed Routing Redundancy

DRR technology of IRF can put the routes of multiple IRF Layer 3 switches together and keep routing synchronization in Fabric. If any of the switches fails, its routing will be taken by other switches. If an inner connection or a switch fails, service data traffic can keep running through dual-homed connection.

5. Distributed Layer 2 Protocol

IRF technology can process Layer 2 protocols such as STP, IGMP SNOOPING and LACP in distributed mode. For the outer environment, IRF Fabric run as a protocol entity, while in the Fabric, protocols run on each unit distributed. Each device takes the calculation of its own protocol, different units interact the information with each other. These all improve the reliability of Fabric and the utilization of each unit in Fabric, and reduce the dependency of the protocol among devices.

6. Physically distributed, logically centralized

The unusualness of IRF technology is to combine different physical devices organically. Each switch is equivalent to an interface card on the chassis switch. Multiple switches are combined to form a single entity. Layer 2 and Layer 3 features among different switches can work together well. They appear to be a switch.

All units in IRF Fabric participate in the running of services to share load. Once a device fails, its load will be taken by other units without affecting the normal running of services. Compared with the previous 1+1 and N+1 redundancy mechanisms, IRF does not distinguish master device and slave device when processing data forwarding. All devices are both master devices and slave devices, which is a 1:N redundancy mechanism.

High Reliability

IRF technology will bring great redundancy to the network, and guarantee the smooth running of the network by using the following methods.

- Distributed networking can place devices in different physical locations to avoid domino effect.
- DLA technology can transfer the traffic to other links and the switches located at the core of the network, thereby avoiding the faults of network routing system and switches affect the whole network and reducing breakdown time of the network.
- Distributed redundancy routing can avoid single point failure in the network by allocating the network among multiple independent switches in the distributed switching fabric.
- DDM provides distributed device interfaces for accessing the distributed switching fabric continuously.
- The Rapid Spanning Tree Protocol (IEEE 802.1w) and Link Aggregation Control Protocol (IEEE 802.3ad) working with IRF technology not only provide Layer 2 redundancy, but also support interoperation with the existent basic networks.

Flexible Scalability

High speed IRF technology supports the bandwidth scalability of the fabric core of the whole network. To implement that, the network administrator only needs to add device units to the distributed switching fabric. Each device unit in IRF distributed switching fabric has its own switching and routing engine. Actually, the potential switching capability of the whole distributed switching fabric is the sum of the switching capability of each interconnected switch.

In addition, using DLA technology, the network administrator can configure multiple links (used for interconnecting switches at remote end) in the distributed switching fabric. Therefore, the administrator can support the growing interconnection demands between network edge and network core by using newly added link level redundancy technology.

Convenient Centralized Management

The network administrator can manage the interconnected core switches as a single entity, and IRF supports several management methods such as console, Telnet, Web, SNMP and RMON. The administrator can manage IRF Fabric systematically by using each method, thereby making it easy to manage the whole IRF distributed switching fabric and save valuable IT resources for the enterprise.

Multiple Step Investment Reduces Risk

The scalability of IRF technology enables users to add or remove new device units to or from the network any moment, without purchasing expensive chassis switch whose performance is more than current demands. This not only reduces the upgrading cost, but also protects the investment in existent network. Simultaneously, new technology on the network emerges endlessly; the method of “pay-as-you-grow” also avoids certain technology risks.

Summary

IRF, a distributed network technology, brings us an innovative concept and application method of device networking. For users who hope to avoid expensive costs for chassis switches, the new switch equipped with IRF technology not only solves this problem, but also provides excellent network reliability, scalability, high performance and powerful network control capability.