Design Guide

Multicast Design Guidelines—Cisco IOS Software Release 12.1(13)E7 and Subsequent Releases

Introduction

A strategic goal of the Cisco IOS® Multicast Development and Deployment Groups for CY'03 is to simplify the implementation, deployment, and operation of the IP Multicast Solution sets for Enterprise and Service Provider customers. IP Multicast in Cisco IOS Software Release 12.1(13)E and all subsequent releases results from years of multicast development and deployment in the world's largest and most critical IP Multicast networks.

There were some very dramatic enhancements incorporated into the Multicast feature set in Cisco IOS Software Release 12.1(13)E. Cisco dedicated more than eighteen months of development and testing to optimizing the PIM protocol to achieve sub-second convergence in many of the most common redundant topologies without the need to perform any specific tuning configuration tasks. Most of these enhancements were batched together into a single DDTS CSCdw13674 "PIM Scalability/Convergence Enhancements for 12.1E". These images have been exhaustively tested by two of the premier testing organizations within Cisco: the Financial Test Lab (FTL) in RTP and the NSITE Lab (Part of the Neverest Project) in San Jose. Both have conducted exhaustive tests on the most common redundant topologies that the Finance industry uses.

Both organizations dedicate their resources to specifically testing highly redundant "customer-based" topologies and configurations against these new images to ensure the highest level of software and solution quality with the industry's best convergence.

Two additional enhancements were specifically aimed at reducing design complexity and easing the implementation of common requirements across most customers. DDTS CSCdp62690 "Auto-RP Messages do not respect Multicast Boundaries" is intended to offer a more flexible and less complicated alternative to the legacy TTL scoping of RP-Announce and RP-Discovery packets used to implement Administrative Address Scoping. The updated mechanism enables Cisco to eliminate the TTL mechanism and to filter the actual RP-Announce and RP-Discovery packets, based on the multicast addresses that should be propagated into different areas of the network by extending the legacy "IP Multicast Boundary" command with an extension, "filter-autorp". This enables the use of simple and extended Access Control Lists (ACLs) to manage the flow of RP-announce and Discovery messages through the network.

The second enhancement is the first step in a series of new commands, which will eliminate the potential for groups to transition from Sparse Mode into Dense Mode during various failure scenarios. This is a multi-faceted problem that can trigger multiple error situations. Attempts to avoid all of these potential error conditions have created multiple creative design methodologies that attempt to avoid these conditions. This initial DDTS CSCdu66936 "Need Auto-RP to work with Sparse-Mode Interface Config" is the first step towards providing a completely guaranteed solution to this problem. This initial step requires a single global configuration command, "IP Pim Autorp Listener", and allows all interfaces to be configured in Sparse-Mode. It still maintains the availability of Auto-RP as the mechanism for distributing Group-RP mappings, while only but hard-coding the two Auto-RP groups to function as Dense-Mode groups. It disallows any interface in the router to slip into dense mode even if there is a loss of connectivity to the RP.

While the aforementioned Cisco IOS Software changes do not begin to encompass all the enhancements to made between Cisco IOS Software Release 12.1(8)bE14 and 12.1(13)E7, these enhancements are the enablers that have allowed Cisco to provide a more robust and scaleable solution. They have also allowed Cisco to progress towards a simpler configuration and implementation.

Only one other design modification has been recommended below, and that modification is a slightly different method of configuring Anycast RPs, which requires no software enhancements in this initial phase, but is part of a larger solution that will involve several software modifications before Cisco achieves its stated end-state.

The only design modifications that have been proposed are those that represent a component of a software enhancement, or the first in a series of steps involving software enhancements targeted at simplifying overall solution complexity and implementation, or providing enhanced functionality.

Multicast Addressing

With the growing popularity of IP multicast applications, many customers are considering the deployment of, or have already deployed, IP multicast in their networks. The design and documentation of an appropriate multicast addressing scheme are important components of a successful IP multicast deployment.

Some of the common problems that customers encounter during the IP multicast deployment are:

- Simplify administration and troubleshooting of IP multicast
- Control the distribution and the use of IP multicast group addresses within an organization (ie: between Business Units)
- Control the distribution and scope of multicast application data within an organization
- Locate Rendezvous Point with PIM Sparse mode and determine which IP multicast groups each will serve
- Control IP multicast traffic on WAN links so that high rate groups cannot saturate low speed links
- Control who can send IP multicast and who can receive (Security)
- Be ready to deploy next generation IP multicast protocols (ie: Bi-Directional PIM, Source Specific Multicast, [SSM])
- Link to the Internet for Multicast
- Allow for future requirements and expansion so that re-addressing does not have to occur at a later date

A solid addressing policy is the key to solving or simplifying many of these issues lies. Without a correctly planned and scoped IP multicast addressing scheme, customers will encounter more complex configurations, which significantly decreases the control of IP multicast over the network and increases administration and support overhead.

The IPv4 multicast address space is handled differently than its unicast counterpart. Unlike unicast addresses, which are uniquely assigned to organizations by IANA, the multicast address space is openly available for use. This openness could potentially create problems with address collisions, so steps have been taken to minimize this possibility. Mainly, the multicast address space has been divided into some well-known ranges to give guidance to network operators and to facilitate deployment of certain applications. The well-known group address ranges include:

1. Multicast Address range: 224.0.0.0/4 RFC1112
2. Local Scoped range: 224.0.0.0/24 (http://www.iana.org/assignments/multicast-addresses)
3. IANA Assigned range: 224.0.1.0/24 (http://www.iana.org/assignments/multicast-addresses)
4. SSM range: 232.0.0.0/8 IETF: draft-ietf-ssm-arch-01.txt
5. GLOP range: 233.0.0.0/8 RFC2770
6. Administrative Scoped range: 239.0.0.0/8 RFC2365

Administrative Scoped Addresses

RFC 2365 provides guidelines as to how the multicast address space can be divided and used privately by enterprises. The terminology ‘administratively scoped IPv4 multicast space’ relates to the group address range 239.0.0.0 to 239.255.255.255. The “key properties of administratively scoped IP multicast are that:

- Packets addressed to administratively scoped multicast addresses do not cross configured administrative boundaries
- Administratively scoped multicast addresses are locally assigned, so they do not need to be unique across administrative boundaries.

This range of multicast addresses was defined to give autonomous networks a set of private multicast addresses that could be used inside their networks without the fear of address collision from outside entities. It is the equivalent of unicast private addresses as described in RFC1918. In order to maintain the integrity of this address space and prevent leaks of control or data traffic in or out of this boundary, it needs to be scoped at the network edge.

A detailed document on IP Multicast Addressing and Scoping can be found at http://ftp-eng.cisco.com. Since Cisco IOS Software Releases 12.2.12, 12.2.12S and 12.1.13E, there is now a new command option on the “ip multicast boundary” interface command that can automatically filter RP-announce and RP-discovery messages based on the multicast groups the boundary command will allow to pass. The new command option is:

```
ip multicast boundary <acl> [ filter-autorp ]
```

This command is not enabled by default. When the new option is enabled, then filtering for Auto-RP messages will occur in three cases:

- RP-announce or a RP-discovery packet is received from the interface
- RP-announce or a RP-discovery packet received from another interface is forwarded out to this interface
- Internally generated RP-announce or a RP-discovery packet is sent out to the interface;

Default PIM Interface Configuration Mode

The recommended default PIM Interface Configuration Mode is PIM Sparse-Mode. Additionally, there is a new command that enables Auto-RP to function with 224.0.1.40 and 224.0.1.39, which is fixed to operate in Dense-mode for the announcement and discovery of dynamic Group-RP mappings between all routers. This new command is:

```
Ip multicast auto-rp [listener]
```

This is a global configuration command and simply enables the router to have all of its interfaces configured in Sparse-mode, thereby eliminating the potential for dense mode flooding across a topology. It still allows the two groups necessary for Auto-RP to function in Dense-mode for the dissemination of Group-RP mapping information and announcements.

PIM-SM Distribution Trees (Shared or Source)

Recommendations

Two recommendations are valid: Shortest Path Tree (SPT) and Shared Tree (RPT).

RPT is intended for Trading networks and other networks of mission-critical nature where a high degree of deterministic operation is required and the network is fairly static and rigid in its configuration. SPTs are the default recommendation for generic enterprise networks where things can be more fluid in nature and a less rigid configuration is required.

SPT Versus RPT

SPT is the multicast distribution tree based shortest unicast path between an interested receiver and a source for the same group. Significant optimizations have been performed on the control and forwarding planes of most platforms, and the scalability of large amounts of (S,G) state is becoming much less problematic than it was in the past. As a default for Enterprise networks, the SPT is the simplest to configure and requires no additional configuration above and beyond assigning each multicast-enabled interface a PIM mode (Dense, Sparse, or Sparse-Dense). SPTs scale fairly well up to around 5-6000 (S,G)s on a typical routing platform with either hardware or software forwarding support. As hardware limits increase and PIM is further optimized, actual hardware platform limits may increase, but it is still a good idea to keep a manageable limit to the size of a multicast routing table and 5-6000 is an appropriate number to use as a target. Note that it is certainly not a hard limit. With networks above this level, some due diligence should be placed to certify significantly higher state levels, ensuring that steady-state CPU levels and convergence targets are characterized accurately and targets for overall network availability are met.

RPT has been a recommendation for some time in Financial Trading networks where the potential amount of state could clearly surpass the 5-6000 (S,G) point and high availability with deterministic network behaviour is critical. Cisco maintains this recommendations, and expands it to include generic enterprise networks in which address summarization is necessary to manage network state to the stated 5-6000 (S,G) level. RPT networks have been proven to be very deterministic in nature and considerable amounts of system and network testing occurs as part of the release process, in order to ensure that RPTrees work reliably in standard network topologies where a high degree of network resilience and a deterministic nature is required.

Rendezvous Point Placement

Reccommendation

For ASM Modeled applications and networks, it is assumed that sources can exist anywhere in the network or topology, and that there is no more optimal location for the RP than in the logical center or core of the network.

For one-to-many or Single Source Multicast (SSM) applications, the assumption is that the flow of information is from one source to a number of receivers. Also, the user can choose the SSM forwarding model is chosen that requires no RP, but rather IGMPv3 support, or the RP can exist either close to the source or in the core as is the ASM model.

Group-RP Mapping Mechanism

Recommendation

Cisco recommends Static RP Configurations for networks in which all multicast services are well-known and fairly static, and there is a willingness to modify network configurations if changes need to occur. Examples of these network types might include Multicast Transit ISP networks and Trading Floor networks.

Auto-RP is recommended in the following situations:

- All multicast services are unknown
- New services are expected to be defined over time
- Generally a high degree of change in the networking environment
  - Businesses involved in mergers and acquisitions constantly combining network infrastructures
  - Desire to have a dynamic Group-RP Mapping mechanism that can preclude wider-scale network re-configuration

Comments on Auto-RP

One of the historical issues related to Auto-RP use has been that it required all interfaces to be configured in Sparse-Dense mode, which could allow the interface to fall into Dense-mode operation during a time when communication to an RP was lost for more than three minutes. With this came the potential for Dense-mode flooding between routers, which could negatively impact users and routers if it becomes excessive. Configuring a Default-RP, or RP of last resort, on each router was a work-around to avoid the potential loss of default-RP information and avoid the Dense-mode flooding issue.

With the implementation of the Auto-RP Listener functionality, Cisco has taken the first step toward eliminating the need for the Default-RP configuration as the only failsafe mechanism that can preclude either groups or interfaces from failing over into Dense-Mode. Cisco is planning additional steps that will completely eliminate the need for the Default-RP configuration steps and will be implemented in future Cisco IOS Software releases.

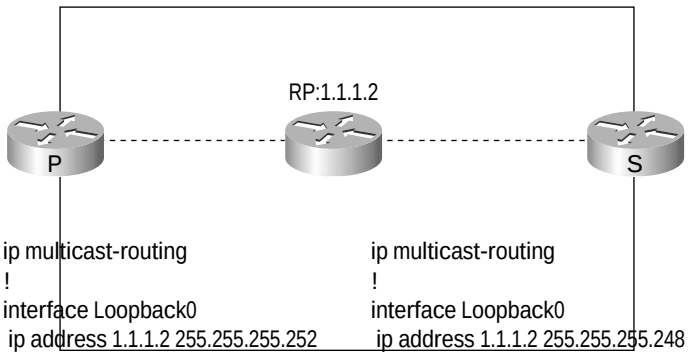
Comments on Static RP

Cisco is not aware of any technical drawbacks related to using Static RP definitions, rather than some other dynamic method. The only potential issue could arise if the network configuration needs to change, requiring some configuration changes to the Group-RP mapping information. Static RP configurations are recommended for well-defined networks in which the potential of change is very low, or the ability and disposition to making manual changes is deemed favorable.

Recommendation

In this situation, primary/secondary relationship work by advertising routes for the RP with different netmasks. It relies on unicast routing longest match algorithms to always pick the primary over the secondary. , Because it announces a longest match route (ie: a /30 route for the RP address), the primary router will always be preferred over the less specific route being announced by the secondary router (ie: a /29 for the same RP address). Figure 1 shows an example of how this can be configured. In this example, the primary router advertises the /32 route of the RP, while the secondary router advertises a route with a shorter mask (a /31 that includes the RP address). Provided that both routes are present (both routers are up and available), unicast routing will choose the longest match and converge to the primary router. The secondary router's advertised route is only chosen when the primary router goes offline or all of its interfaces go down.

Figure 1
Rendezvous Point Redundancy




```
ip multicast-routing                    ip multicast-routing
!                                       !
interface Loopback0                     interface Loopback0
 ip address 1.1.1.2 255.255.255.252      ip address 1.1.1.2 255.255.255.248
 ip pim sparse-mode                      ip pim sparse-mode
 ip ospf network point-to-point          ip ospf network point-to-point
!                                       !
router ospf 11                          router ospf 11
 network 1.1.1.0.0.0.0.3 area 0          network 1.1.1.0.0.0.0.7 area 0
 network 10.1.1.0.0.0.0.255 area 0       network 10.1.1.0.0.0.0.255 area 0
 network 10.1.2.0.0.0.0.255 area 0       network 10.1.2.0.0.0.0.255 area 0
!                                       !
ip pim rp-address 1.1.1.2               ip pim rp-address 1.1.1.2
```

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: 31 0 20 357 1000
Fax: 31 0 20 357 1100

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
Capital Tower
168 Robinson Road
#22-01 to #29-01
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the Cisco Web site at www.cisco.com/go/offices

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

ETMG 203115—SH 07.03