SOLUTION DESIGN GUIDE

Ruckus Networks Multi-Chassis Trunking (MCT) Essentials

Supporting FastIron 08.0.90

Part Number: 53-1005591-01
Publication Date: 14 February 2019

Copyright, Trademark and Proprietary Rights Information

Export Restrictions

These products and associated technical data (in print or electronic form) may be subject to export control laws of the United States of America. It is your responsibility to determine the applicable regulations and to comply with them. The following notice is applicable for all products or technology subject to export control:

These items are controlled by the U.S. Government and authorized for export only to the country of ultimate destination for use by the ultimate consignee or end-user(s) herein identified. They may not be resold, transferred, or otherwise disposed of, to any other country or to any person other than the authorized ultimate consignee or end-user(s), either in their original form or after being incorporated into other items, without first obtaining approval from the U.S. government or as otherwise authorized by U.S. law and regulations.

Disclaimer

THIS CONTENT AND ASSOCIATED PRODUCTS OR SERVICES ("MATERIALS"), ARE PROVIDED "AS IS" AND WITHOUT WARRANTIES OF ANY KIND, WHETHER EXPRESS OR IMPLIED. TO THE FULLEST EXTENT PERMISSIBLE PURSUANT TO APPLICABLE LAW, ARRIS DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, TITLE, NON-INFRINGEMENT, FREEDOM FROM COMPUTER VIRUS, AND WARRANTIES ARISING FROM COURSE OF DEALING OR COURSE OF PERFORMANCE. ARRIS does not represent or warrant that the functions described or contained in the Materials will be uninterrupted or error-free, that defects will be corrected, or are free of viruses or other harmful components. ARRIS does not make any warranties or representations regarding the use of the Materials in terms of their completeness, correctness, accuracy, adequacy, usefulness, timeliness, reliability or otherwise. As a condition of your use of the Materials, you warrant to ARRIS that you will not make use thereof for any purpose that is unlawful or prohibited by their associated terms of use.

Limitation of Liability

IN NO EVENT SHALL ARRIS, ARRIS AFFILIATES, OR THEIR OFFICERS, DIRECTORS, EMPLOYEES, AGENTS, SUPPLIERS, LICENSORS AND THIRD PARTY PARTNERS, BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL, PUNITIVE, INCIDENTAL, EXEMPLARY OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER, EVEN IF ARRIS HAS BEEN PREVIOUSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES, WHETHER IN AN ACTION UNDER CONTRACT, TORT, OR ANY OTHER THEORY ARISING FROM YOUR ACCESS TO, OR USE OF, THE MATERIALS. Because some jurisdictions do not allow limitations on how long an implied warranty lasts, or the exclusion or limitation of liability for consequential or incidental damages, some of the above limitations may not apply to you.

Trademarks

ARRIS, the ARRIS logo, Ruckus, Ruckus Wireless, Ruckus Networks, Ruckus logo, the Big Dog design, BeamFlex, ChannelFly, EdgeIron, FastIron, HyperEdge, ICX, IronPoint, OPENG, SmartCell, Unleashed, Xclaim, ZoneFlex are trademarks of ARRIS International plc and/or its affiliates. Wi-Fi Alliance, Wi-Fi, the Wi-Fi logo, the Wi-Fi CERTIFIED logo, Wi-Fi Protected Access (WPA), the Wi-Fi Protected Setup logo, and WMM are registered trademarks of Wi-Fi Alliance. Wi-Fi Protected Setup™, Wi-Fi Multimedia™, and WPA2™ are trademarks of Wi-Fi Alliance. All other trademarks are the property of their respective owners.

Contents

Introduction

Introduction

Today's enterprise network connects an ever-increasing number of devices that are constantly sending data back and forth. Users around the globe demand 100 percent network uptime. Customers need constant access to data stored in the cloud. An effective solution to address these pressures will possess the qualities of *high scalability* and *fault tolerance.* Ruckus Networks provides multiple solutions to address these growing needs.

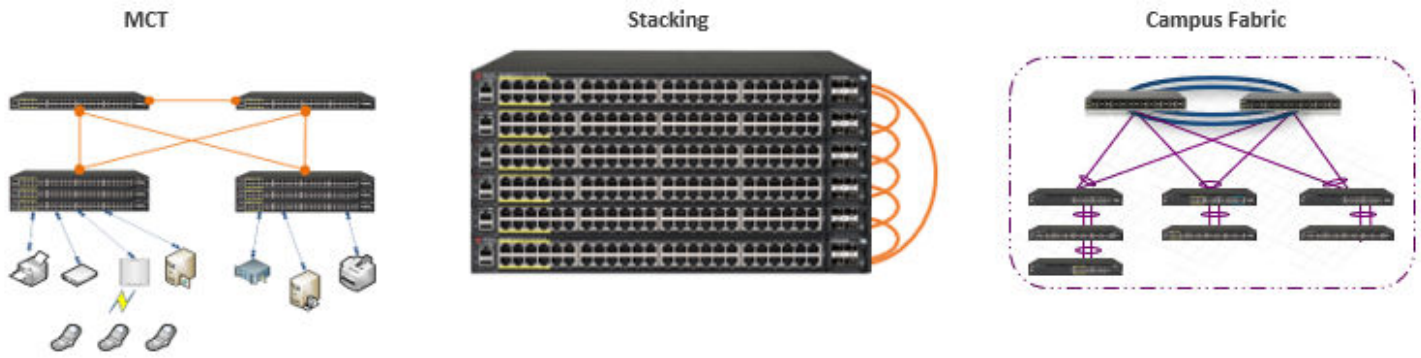
Overview

Ruckus Networks provides various design choices to tackle the needs of a growing enterprise network. The following three choices provide high scalability and fault tolerance with reduced complexity and user interference:

- Multi-Chassis Trunking (MCT)
- Stacking
- Campus Fabric

Let us briefly compare them with respect to ease of deployment and manageability, fault tolerance, and scalability.

FIGURE 1 Design Choices



Multi-Chassis Trunking (MCT)

Multi-Chassis Trunking (MCT) allows users to connect two Ruckus ICX 7850-32Q, ICX 7850-48F, or ICX 7850-48FS switches to form a single logical unit known as a "cluster." Users can then physically split the ports of a single Link Aggregation Group (LAG) on a third device (client) and connect them to the two Ruckus ICX 7850 switches that are part of the cluster. Configured in this way, the cluster appears to be a single unit to the client and thus gives device-level fault tolerance and high network resiliency (explained in detail in MCT Conceptualized on page 5). In a scaled cluster, users can connect up to 68 clients per cluster and achieve ~3000+ port count for endpoint connections, which is highly desirable in a large-scale deployment. MCT is based on spine and leaf architecture, in which the spine is a high-end device such as the Ruckus ICX 7850-32Q and the leaf comprises other ICX devices. This gives users a cost-effective, highly scaled solution.

Stacking

Stacking allows users to connect from 2 through 12 ICX devices of the same type to form a single logical unit known as a "stack." Users can manage and configure the entire stack using just one IP address, which highly simplifies deployment and

manageability. Stacking is based on the "Active-Standby" architecture wherein at any point in time there is only one device which actively controls the entire stack. If the active device fails, a standby unit becomes the new active unit and provides fault tolerance. A completely scaled stack of Ruckus ICX 7850-48F switches gives users a maximum of ~500 (1-, 10-, or 25-Gbps ports) and ~70 (100-Gbps ports) for endpoint connections. This solution is suitable in a small-scale network.

Campus Fabric

Campus Fabric is similar to stacking, but users have the flexibility to connect different types of ICX devices and can scale to a maximum of 36 units to form a single logical unit known as a "fabric." In a scaled fabric, users can deploy multiple Ruckus ICX 7750 switches or ICX 7650 switches and connect them to lower-end devices such as the Ruckus ICX 7150, ICX 7250, and ICX 7450 to achieve a cost-effective solution. This scaled fabric can give users ~1700 ports for endpoint connections, which is suitable for a medium-scale network. This would give users a lot of 1-Gbps or 10-Gbps connectivity towards the end devices and 40-Gbps connectivity towards the core network.

MCT Conceptualized

Multi-Chassis Trunking (MCT) allows users to build a redundant, highly available, load balanced, and highly resilient Active-Active network at the distribution and core layers. MCT is supported on the Ruckus ICX 7850 and ICX 7750 switches. Using Ruckus ICX 7850 switches in an MCT configuration, users get all the benefits of a chassis device and more.

In the MCT configuration, any ICX device can physically connect to a pair of core switches, such as the Ruckus ICX 7850, using the ports from the same Link Aggregation Group (LAG). By doing this, users can reap benefits such as load sharing and redundancy from using a LAG and two separate devices. The pair of core switches are connected to each other to provide multiple data paths in case of device-level failure. This link is called an Inter-Chassis Link (ICL). MCT overcomes well-known limitations found in an STP-deployed traditional network. Such topologies require all the devices to run the Spanning Tree Protocol (STP) to avoid any Layer 2 network loops.

Known inherent STP issues:

- Ports in a blocking state means wasted bandwidth
- Slow convergence—6 to 15 seconds of traffic disruption in case of link or device failure
- No load balancing because there is only one active path for data flow at any given time

MCT as a solution:

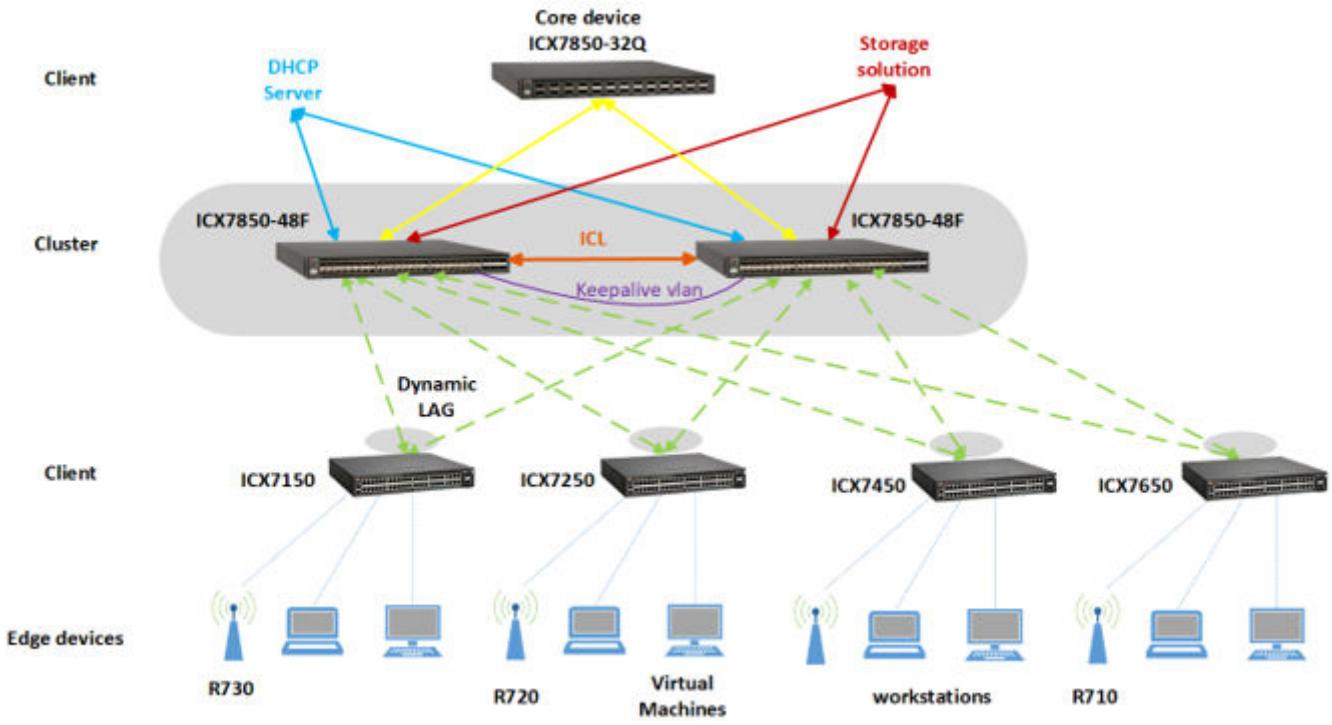
- All the ports in an MCT topology are forwarding and thus provides efficient usage of network bandwidth
- MCT inherits all the benefits of a LAG by providing multiple physical links to act as a single logical link; therefore traffic disruption is in subseconds in case of link or device failure
- Because MCT links are part of a LAG, data is load shared across all the member links, which makes the network balanced and maintains the same level of resiliency with redundant paths available at all times

"Multi-Chassis Trunking (MCT) is a better alternative to traditional redundancy protocols"

In Figure 2, two Ruckus ICX 7850 switches are interconnected using a single link or a LAG known as an Inter-Chassis Link (ICL). A Ruckus Networks proprietary protocol known as Cluster Communication Protocol (CCP) runs on the ICL and establishes a TCP session to form a single logical device known as a "cluster." The ICL is configured to be a tagged member of a dedicated VLAN known as a "session VLAN," which provides a secure control path for all control packet exchanges between the two Ruckus ICX 7850 switches. After the cluster is established, users can scale this network by adding any device known as a "client." A client can be any ICX switch, server, storage solution, or third-party switch. Member ports of the dynamic LAG on the client are physically split and connected to both the Ruckus ICX 7850 switches in the cluster, which results in two highly available, load balanced, and redundant data paths to the cluster. In case a link between a client and one of theRuckus ICX 7850 switches in the cluster fails,

connectivity is maintained through the other equally good and readily available link. Data is forwarded to the rest of the network as if there was no network failure.

FIGURE 2 MCT Cluster-Client Topology



Similar to the ICL link, there is another physically connected link between the cluster devices known as a "keepalive." This backup link plays a key role in maintaining proper functioning of the cluster in case of an ICL failure. If the ICL goes down, the two cluster devices perform a per-client master/slave negotiation and all the physical ports on the slave device are administratively brought down. This ensures continued network connectivity from the edge device to the rest of the network even during an ICL failure. Keepalive link configuration is allowed only when the cluster is deployed in "Loose mode." Loose mode is described in

Consider the following guidelines while deploying a cluster and adding clients:

- The ICL can be a single or multiport static LAG only.
- A device can be a member of only one cluster at a time.
- Clients are connected to the cluster using only dynamic LAGs.
- One or multiple clients can be part of a VLAN known as the MCT VLAN.
- The ICL must be a tagged member of a session VLAN.
- Maintain a separate physical link between the two cluster devices known as a "keepalive," which is a member of a dedicated VLAN known as the "keepalive VLAN."
- Control packets to synchronize all MAC entries between the two switches in a cluster are exchanged over the ICL on a dedicated session VLAN. In case of an ICL failure, the keepalive VLAN maintains proper cluster functionality. The keepalive VLAN is active only when the two cluster devices are not reachable over the session VLAN, and it does not preform any MAC table packet exchanges.

Efficient Data Forwarding Using MCT

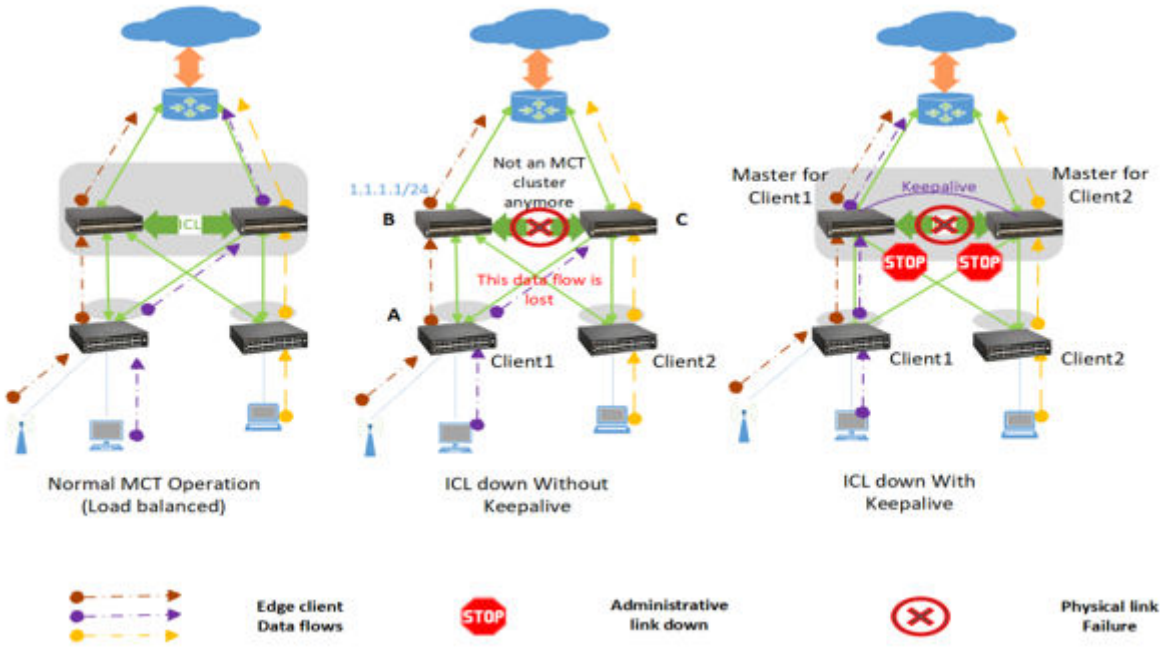
Client Isolation Modes

An MCT cluster can be deployed in two different modes: Loose and Strict. These two modes differ in the way the client connection to the cluster is handled in case of an ICL failure. By default, MCT is deployed in Loose mode.

Loose Mode

In Loose mode, the presence of a keepalive link determines how the cluster-client communication is handled when an ICL fails.

FIGURE 3 Loose Mode Operation



In Figure 3, if there is no keepalive link configured in the cluster, there is no way to exchange information between the cluster devices. As a result, the two devices in the cluster start behaving like two separate devices and the users do not get any advantages of MCT. This results in scenarios where data is sent to one of the cluster switches that has no further uplink connectivity and, as a result, the packets are dropped. For example, in Figure 3, the next hop for switch A traffic is 1.1.1.1/24, which is the IP address configured on switch B. ARP on switch A is resolved to reach the next hop using the LAG interface (physically connected to switch B and switch C). Traffic flows from switch A towards the cluster is load balanced on all the member ports of the LAG. So, some of the traffic flows from switch A are forwarded to switch C with the next hop set as 1.1.1.1/24. But the cluster switches are disconnected, and as a result switch C has no link to send this data towards switch B and therefore all these packets are dropped on switch C. This is highly undesirable, and Ruckus Networks recommends configuring a keepalive VLAN to avoid such network behavior.

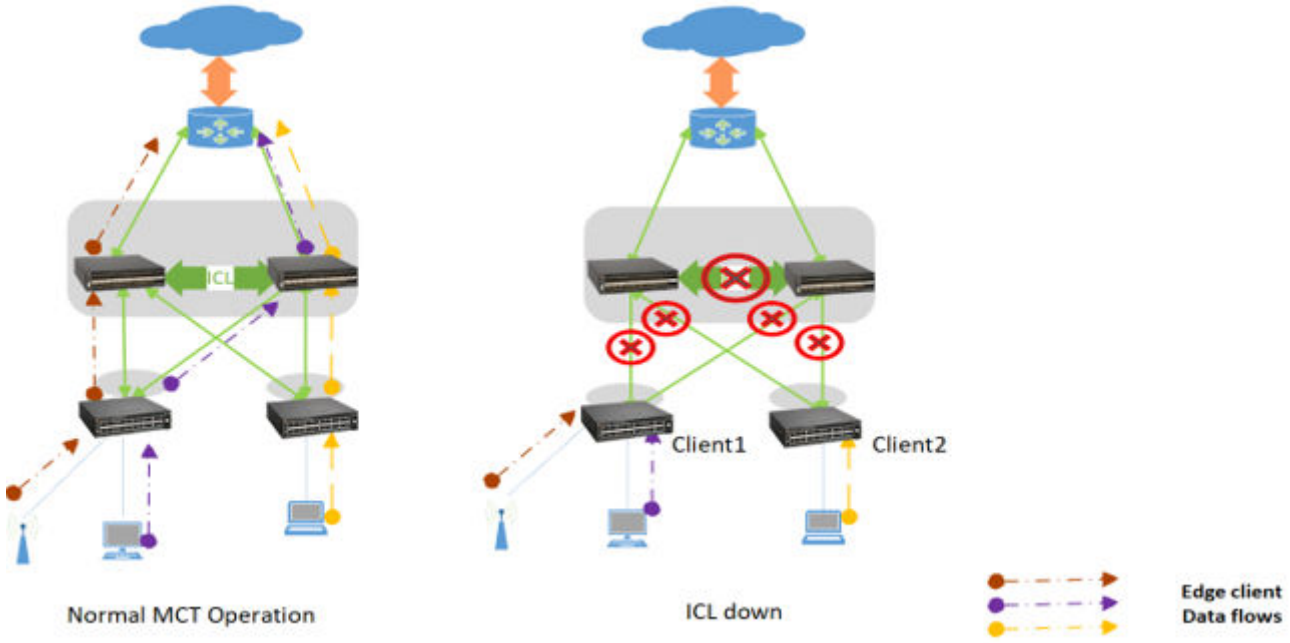
If the ICL goes down in the presence of a keepalive link, the two cluster devices perform a per-client master/slave negotiation. For each of the clients, one of the cluster devices is the master and the other is a slave. All the ports connected from the slave device to that particular client are brought down administratively, ensuring that the traffic from the client is always forwarded to the master cluster device and beyond. This sort of a behavior ensures that the network behaves in a known manner in case of any

failure. As a result, the user has more control over the network and can rectify a problem while still maintaining an active data flow path.

Strict Mode

In Strict mode, when the ICL goes down, all the client-connected interfaces are brought down on both the cluster devices and the clients are completely isolated from the entire network. There is no data flow in the network until the ICL is restored. Also, there is no concept of a keepalive link in Strict mode. Once the ICL is restored, all the ports are enabled automatically and go back to behaving as a normal MCT network. This is a more conservative effort and the user must manually configure the cluster to be in Strict mode before deploying.

FIGURE 4 Strict Mode Operation



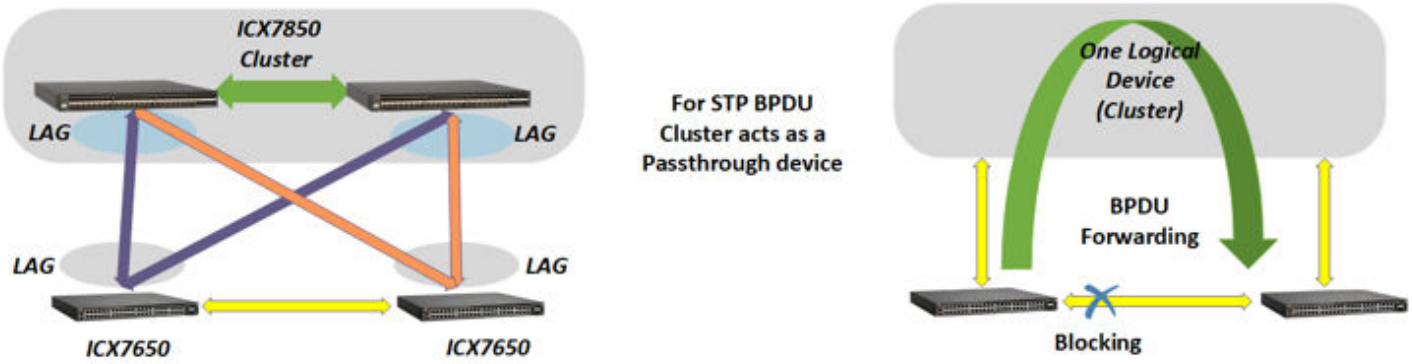
Mode	Advantages	Considerations
Loose	• Needs no user configuration • There is always a path from the client to the cluster and beyond • In a well-designed network, a cluster can work in master/slave mode, carrying the entire traffic load even when the ICL goes down • User can perform a live network upgrade with minimal traffic disruption when MCT is configured along with VRRP-E and short-path forwarding	Based on the network design, there may be traffic loss when the ICL goes down

Mode	Advantages	Considerations
Strict	Conservative approach, no risk of any untoward behavior in the network due to loss in ICL	• User must manually configure • The clients are completely isolated from the network when the ICL goes down • User cannot perform a live network upgrade without causing a complete traffic disruption

MCT in a Traditional Network

MCT is an alternate solution to the Spanning Tree Protocol (STP) which provides Active-Active connectivity. In a traditional network, which needs client-client physical connection apart from an MCT cluster, STP is configured to avoid network loops. MCT must work along with STP to best suit the network. By default, STP is disabled on MCT, but STP BPDUs are hardware forwarded and the required loop-free topology can be achieved.

FIGURE 5 MCT with STP Running on Client Devices



In Figure 5, if all the Ruckus ICX 7650 links are part of the same VLAN, then there is a Layer 2 loop in the network (as shown on the right side). To avoid any Layer 2 loop, STP is run on the ICX 7650 switches. The MCT cluster acts as passthrough for the STP BPDUs and one of the client ports is moved to blocking mode.

MCT with Layer 3 Protocols

IP routing protocols such as BGP and OSPF are supported over a Virtual Routing and Forwarding (VRF) instance in an MCT cluster. Beginning with FastIron 08.0.70, user-defined VRFs are supported on MCT. Prior to FastIron 08.0.70, MCT supported routing only on the default VRF. Presently, there is no support for IPv6 routing over an MCT network. VRF allows multiple instances of routing tables to coexist. A service provider can cater to multiple clients by keeping the routing information separate for each client, and different clients can use similar or overlapping IP addresses without fear of information being sent out to devices other than their own.

Layer 3 traffic destined to MCT clients follows normal IP routing but requires a dynamic trunk (LACP) to be configured on the MCT client. The client routes traffic towards its next hop, which can be either one of the MCT cluster devices. If ECMP is deployed on the client, each MCT device can be a possible next hop and provide Layer 3 load balancing. Because the link on the MCT client is already a dynamic LAG, the traffic is subjected to Layer 2 load balancing at the port level and for some of the streams the traffic sent out with next hop as one of the MCT devices can reach it directly or through the cluster peer. Almost 50 percent of the traffic

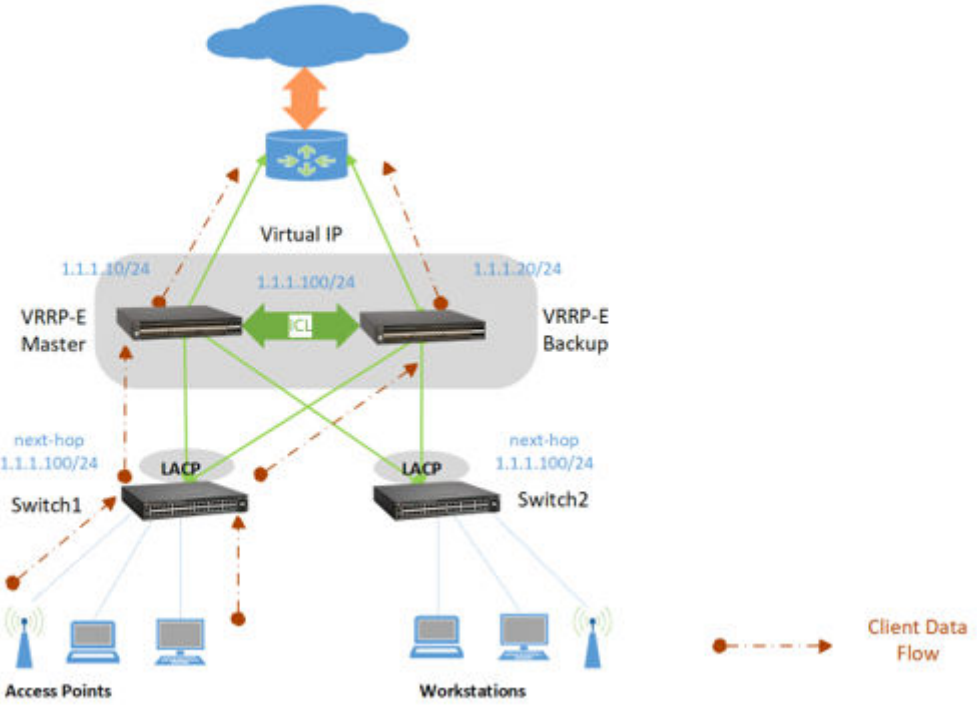
forwarded from the MCT client can pass through the ICL. Users must consider this fact while designing the network because the ICL can become a bottleneck for the entire network.

In case one of the MCT devices fails, Layer 3 traffic is not lossless. This is because each MCT cluster device forms its own adjacency. When one of the devices goes down, Layer 3 reconvergence is required, which results in traffic loss.

VRRP-E over MCT to Achieve In-Service Software Upgrade (ISSU)

While MCT provides Layer 2 redundancy, Ruckus recommends using VRRP-E for Layer 3 redundancy. MCT, along with VRRP-E, helps solve the issue of a single point of failure for both Layer 2 and Layer 3 traffic.

FIGURE 6 MCT with VRRP-E



In Figure 6, VRRP-E is enabled on both the cluster devices. Based on predefined parameters, one of the peers acts as a master and the other as a backup. In a simple VRRP deployment, the traffic that reaches the backup device is switched to the master over the ICL and the master routes it back to the backup device which is then forwarded to the router upstream. But this data flow over the ICL creates a bottleneck in the network. To overcome this inefficient behavior, Ruckus recommends enabling short-path forwarding on both the master as well as the backup devices. With short-path forwarding, the backup devices can directly route the traffic upstream instead of switching the traffic over the ICL.

For example, in Figure 6, a data stream from two different end devices reaches Switch1. The link between Switch1 and the cluster being a LAG balances the load on its physical link and forwards one of the streams to the master and the other to the backup. Because short-path forwarding is enabled, both the peers can route the traffic to the upstream device provided they both have the route already established for that subnet. This ensures an efficient, load balanced Layer 2 and Layer 3 traffic forwarding.

Users can perform ISSU of the network using VRRP-E over MCT byconsulting the following steps:

1. Configure MCT client isolation to operate in Loose mode.

2. Enable routing on interfaces connecting both the MCT peers to the upstream router.
3. Enable VRRP-E with short-path forwarding on both the peers.
4. Upgrade and reload the MCT peer that happens to be the VRRP-E backup device.
5. After a successful upgrade, change the backup priority of this peer to make it a VRRP-E master.
6. Upgrade and reload the other MCT peer.

When one of the MCT peers is upgrading, the locally connected physical ports towards the client are brought down. But because the link between the client and the cluster is a LAG, the traffic is switched to the active ports and sent to the other peer which is still up and running. The active peer has all the required routes because of VRRP-E and is fully capable of forwarding the traffic upstream. As a result, the traffic loss due to ISSU is limited to the duration it takes to switch traffic from the inactive to active client LAG ports. In this way, a user can perform with minimal downtime and traffic disruption in a large-scale network.

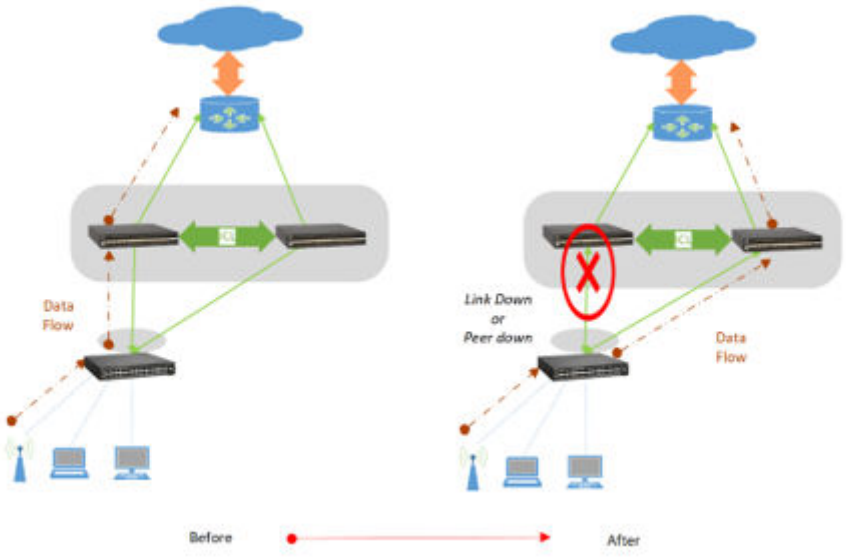
MCT Failover Scenarios

Two types of MCT failover scenarios can occur:

- Client interface on one of the MCT devices goes down or MCT cluster device goes down
- Multiple client link failures

Client Interface on One of the MCT Devices Goes Down or MCT Cluster Device Goes Down

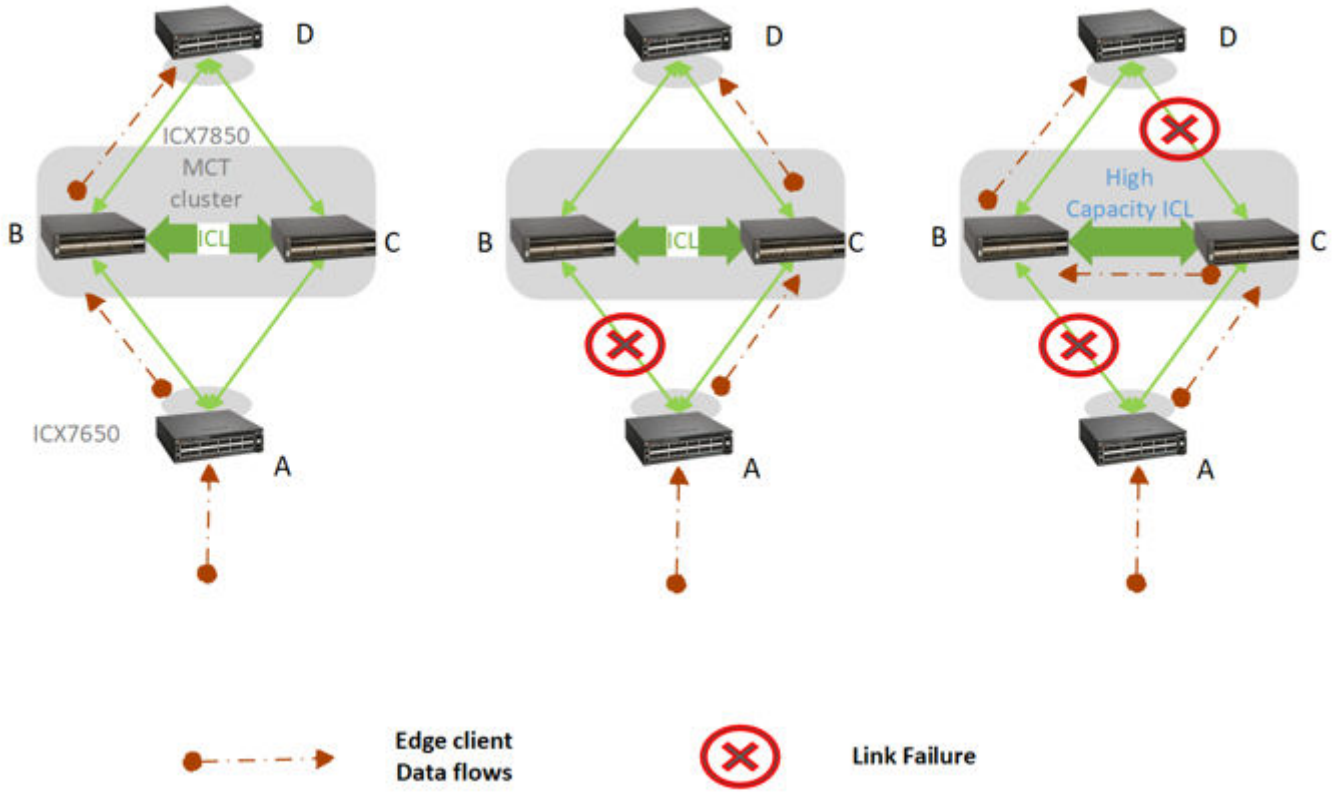
FIGURE 7 MCT Client Interface or MCT Device Goes Down



In this scenario, if one of the client interfaces goes down, the traffic from that client switches to the other cluster device with minimal traffic loss. If the MCT cluster device undergoes a reboot or power failure, then traffic from all the clients is switched to the active peer device. This failover mechanism ensures that the traffic disruption is minimal.

Multiple Client Link Failures

FIGURE 8 MCT Handling Multiple Client Link Failures



In this scenario, there is continuous traffic from switch A to switch D. The traffic path is switch A to switch B to switch D. For example, if the link between switch A and switch B fails, then the traffic is switched to the other active member of the LAG and thus it reaches switch C. For a brief moment, the traffic is sent over the ICL to reach switch B and then switch D. But the cluster devices exchange control packets to sync their MAC tables and then the traffic from switch A to switch D will look like switch A to switch C to switch D. Now if there is a failure on the link between switch C and switch D, then the traffic from switch C is sent to switch B over the ICL and switch B will forward it to switch D. Thus, the new flow will look like switch A to switch C to switch B to switch D. In this case, the ICL is carrying all the control packets as well as the data packets for end-to-end communication. Therefore, users must use a higher bandwidth ICL link to accommodate such network failures. This shows that a MCT network can handle multiple link failures at the same time without compromising on traffic forwarding.

Large Campus Networks

Large Campus Networks Overview

Different topologies offer different usages and advantages. The two main topologies are "Single tier" and "Two tier." As the names suggest, there are one or two levels of an MCT cluster, which provide scaling for different requirements.

Single-Tier MCT Architecture

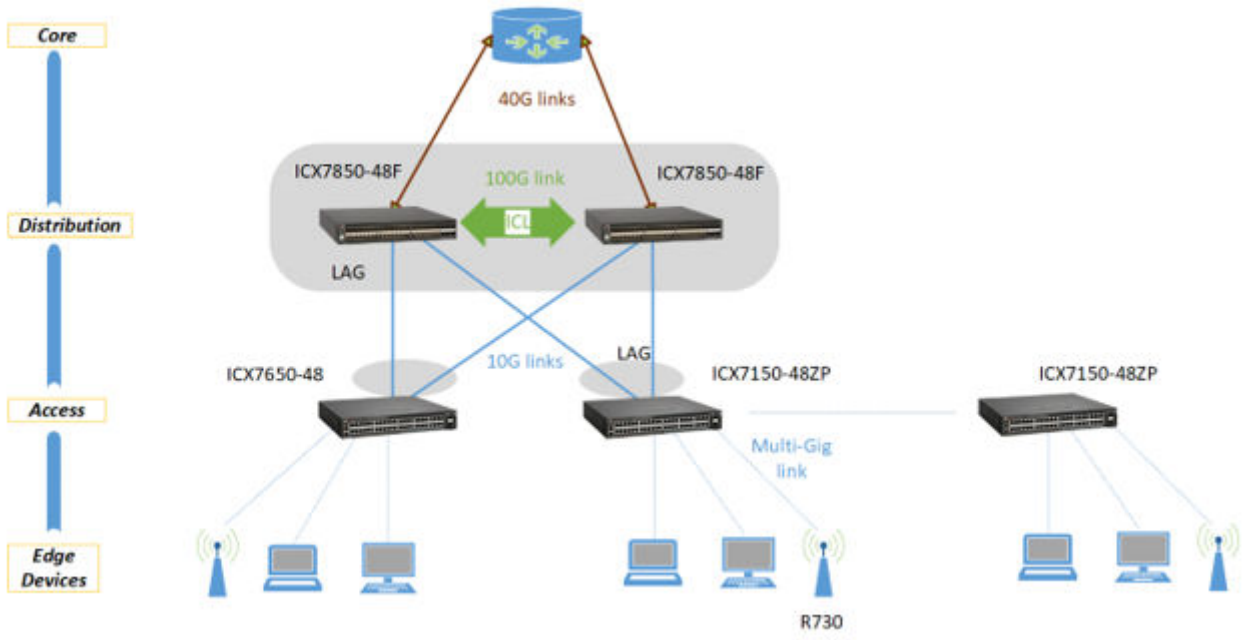
Single-tier MCT deploys a single level of an MCT cluster which connects to multiple MCT clients. These clients provide cost-effective solutions with high port density to connect all the edge devices to the network.

Recommended Configuration

- Use the Ruckus ICX 7850-48F or ICX 7850-48FS as the MCT cluster device.
- Configure a dynamic LAG between the cluster devices and have a keepalive VLAN for the backup link.
- MCT clients can be

  Ruckus ICX 7150, ICX 7250, ICX 7450, ICX 7650, and ICX 7750 switches.
- A client can be a single device or a stack (12 units maximum per stack).
- Multi-Gig client ports are connected to edge devices and each client has a multilink dynamic LAG uplink connection to the Ruckus ICX 7850 switches.
- Use the Ruckus ICX 7850-48FS as the MCT cluster device to provide a secure campus network using MACsec.
- All connections from MCT cluster to the clients, clients to edge device is configured as a Layer 2 network
- The cluster connects to the core device as a Layer 3 link. VRRP-E with short-path forwarding is enabled to provide Layer 3 redundancy.

FIGURE 9 Single-Tier MCT Cluster Topology



Key Advantages

- In case of a single cluster device failure, there is still an equally good path for traffic from the edge device to get to the core.
- Because the link from the client to the cluster is a LAG, a link failure will switch traffic to the active link within subseconds.
- Because all the links are active in an MCT deployment, traffic flows from edge devices are load balanced at the cluster level. This maximizes the overall network bandwidth usage and improves scalability due to efficient usage of port capacity. This maximizes the return on investment for the customer.
- MCT clients with Multi-Gig ports connect to edge devices at speeds of 1 or 2.5 Gbps. These clients have 10-Gbps uplink LAGs to the Ruckus ICX 7850-48F or ICX 7850-48FS.
- To scale this network, connect and configure low-cost MCT clients, which in turn connect to many edge devices.

Two-Tier MCT Architecture

Two-tier MCT deploys two levels of an MCT cluster which connects to multiple MCT clients. Adding another level of MCT future-proofs the network to scale to very large port density. There are multiple paths available for data from end devices to the core network. This ensures that the network is highly available and load balanced.

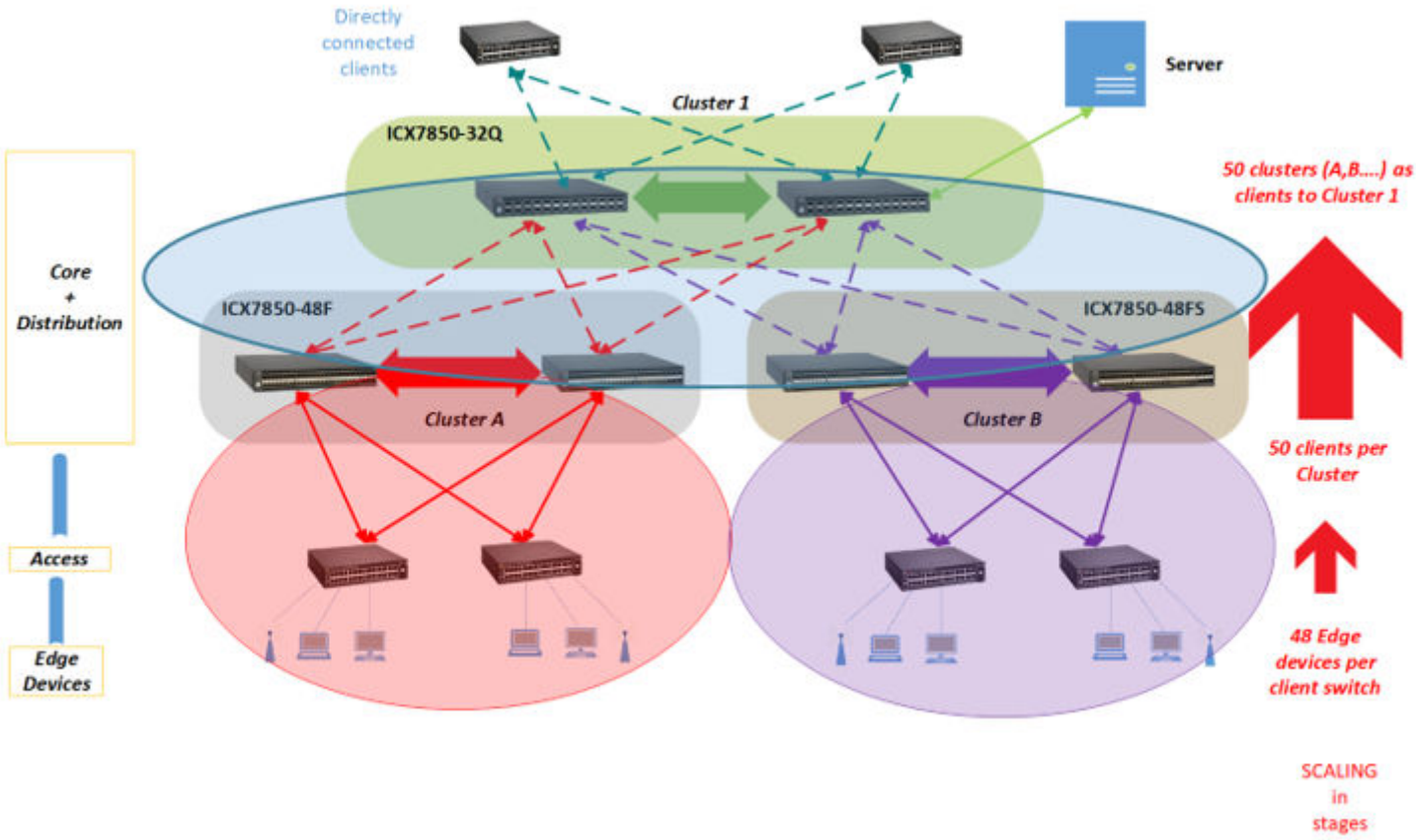
Recommended Configuration

- Use the Ruckus ICX 7850-32Q or ICX 7850-48FS as MCT Cluster 1.
- Use the Ruckus ICX 7850-48F as MCT Cluster A and Cluster B, or use the Ruckus ICX 7750 to lower the cost.
- Clients can be Ruckus

  ICX 7150, ICX 7250, ICX 7450, and ICX 7650 switches.

- Cluster A and Cluster B can be scaled to 50-plus clients each.
- Fifty such Cluster As can be added as clients to Cluster 1.
- Connect multiple storage devices or servers as clients to Cluster 1 with high-capacity links.

FIGURE 10 Two-Tier MCT Cluster Topology



Key Advantages

- Adding the second layer of MCT clusters adds another level of redundancy and load balancing.
- The user can replicate an existing single MCT network and connect it to the top MCT cluster level to scale the network easily.
- Clients can be directly connected to the top level of an MCT cluster.
- Servers can act as MCT clients to the top level and as a result become highly available to the rest of the network as it eliminates a single point of failure.
- Layer 3 redundancy is achieved using VRRP-E, and with short-path forwarding enabled, the entire network can be upgraded easily with minimal traffic disruption.
- There is an increased density of 10-Gbps ports to build large core and distribution networks.

Conclusion

MCT is not merely a feature but a way to architect a large-scale network. MCT increases High Availability (HA) with multiple redundant paths for data forwarding and it is highly resilient against network device or link failures. The Active-Active architecture eliminates bottlenecks in enterprise networks and makes efficient use of network bandwidth, thus maximizing the return on investment for valued customers. Popularly used Layer 3 protocols such as OSPF and BGP are supported over MCT. Layer 3 redundancy is achieved when MCT is paired with VRRP-E. Users can start with any-sized network and can scale easily to large port counts when needed. In case of multi-level redundant architecture, the topology can be expanded by replicating a smaller single-tier cluster. This allows great growth of the network without increasing the footprint as is traditionally done by chassis solutions. Also as a result of MCT benefits, the total cost of ownership is drastically reduced.

Ruckus Wireless, Inc., a wholly owned subsidiary of ARRIS International plc.
350 West Java Dr., Sunnyvale, CA 94089 USA
www.ruckuswireless.com