Increasing performance and resiliency with deterministic link aggregation

White paper

This white paper explores the significant role that the Link Aggregation Group (LAG) concept can play in optimizing network performance and providing high availability in modern network infrastructures. A LAG, also known as link bundling, allows multiple physical links to be combined to create a logical link that appears as a single link to higher layers. This increases bandwidth beyond what one link can offer on its own and provides redundancy in case of link failure. Link aggregation implementations traditionally either split traffic uniformly across all links using a device-embedded hardware algorithm hashing on packet header fields or make some links active while others are in standby.

This white paper delves into the advantages of load balancing, fault tolerance and seamless failover provided by link aggregation, highlighting how it enhances network resiliency and minimizes potential points of failure.

The paper also presents recently standardized link aggregation enhancements: Per-Service Frame Distribution (PSFD), which enables the operator to control which link of a LAG frame can be forwarded, and the revertive or non-revertive mode of operation that provides for automatic return to the normal state.

Finally, the paper explores the concept of Multi-Chassis LAG (MC-LAG), which provides additional protection against switch failures by allowing two switches to terminate one end of a LAG, even if they are in different geographical locations. This concept has been standardized a posteriori after many vendor implementations.

By exploring real-world use cases, this paper serves as a comprehensive resource for network administrators and engineers seeking to deploy link aggregation in their environments and consider recent standard enhancements. Nokia is at the forefront of implementing these new enhancements. The insights shared in this paper empower organizations to employ link aggregation effectively, unlocking the full potential of their network infrastructure for enhanced performance, scalability and reliability.

Contents

1 Introduction

Link aggregation was created to provide increased capacity and load balancing across multiple links, as well as protection against link or switch failure, although it was not designed as an ITU-T protection protocol. It allows two or more links to be aggregated together to form a Link Aggregation Group (LAG) so that the link aggregation client can treat these links as if they were a single link. Link aggregation uses the Link Aggregation Control Protocol (LACP) to allow the two ends of the LAG to identify their capabilities for compatibility and smooth operation.

Link aggregation was originally standardized in IEEE Std 802.3ad-2000 [1], subsequently incorporated in IEEE Std 802.3-2000 [2]. It was then moved under the responsibility of the IEEE 802.1 Working Group starting with IEEE Std 802.1AX-2008 [3]. The IEEE 802.1AX standard has since been revised twice: IEEE Std 802.1AX-2014 and IEEE Std 802.1AX-2020. The 2014 edition built on the 2011 publication of MEF 32 [4], specifying requirements for service protection across external interfaces.

The IEEE 802.1AX standard defines a Media Access Control (MAC)-independent link aggregation technique. The 2014 edition built on the 2011 publication of MEF 32, specifying requirements for service protection across external interfaces. This resulted in IEEE 802.1AX-2014 [5] standardizing what is commonly known and deployed as Multi-Chassis LAG (MC-LAG). It also standardized a new form of load balancing, known as Per-Service Frame Distribution (PSFD). PSFD can be used as a standardized replacement for device-embedded hardware hashing algorithms. MEF standards utilize PSFD in MEF 10.3.2 [6] (later incorporated in MEF 10.4 [7]) as a foundational technique to enhance User Network Interface (UNI) and External Network Network Interface (ENNI) resiliency, respectively. The 2020 edition of IEEE 802.1AX [8] introduced refinements to the mechanisms defined within the 2014 edition and introduced revertive or non-revertive modes of operation.

These and other techniques are at the core of the Subscriber UNI Link Aggregation service attribute in MEF 10.4 and the Operator UNI Link Aggregation service attribute in MEF 26.2 [9]. These service attributes can be set to different values to determine how to protect against link failures:

- **Not Applicable** or **None** means that switches at each end of the external interface (UNI or ENNI) are interconnected without using a LAG.
- **2-Link Active/Standby** or more simply **Active/Standby** means that the LAG has one link active and one link standing by in case the active link fails.
- **All Active** means the LAG may comprise two or more active links.
- **Other** is for resiliency solutions other than those above. For example, the Nokia 1830 Photonic Service Switch (PSS) can have one group of active links and another group of standby links.

This white paper discusses LAG-based techniques for traditional (device-embedded hardware hashing algorithm) and novel (PSFD) load balancing, as well as for resiliency protection against link failure (active/standby) and/or switch failure (MC-LAG). It also describes and contrasts these load balancing and resiliency techniques, noting their benefits and drawbacks, and summarizes their applicability in optimizing network performance and reliability.

2 Enhanced resiliency and load balancing techniques

Over time, link aggregation has evolved to support four main techniques:

1. **Device-embedded hardware hashing** is the original load balancing technique. It relies on hardware to hash the header of layer 2 (L2) frames to choose which LAG link to use to forward a given frame. It only applies when link aggregation is used for capacity increases in load-sharing mode.

2. **PSFD** is the most recently standardized load balancing technique and can be based on specified Virtual Local Area Network (VLAN) Identifier (ID) values. It allows the operator to control which link a given frame follows to predictably determine frame distribution and collection.

3. **Active/standby** is a resiliency technique that protects against link failure involving two or more links. It can enable the operator to control when operation returns to the normal state. In terms of implementation, this can be realized using two so-called "LAG subgroups," with one in active mode and the other in standby mode. The standby LAG subgroup switches to active mode when the number of failed links in the active LAG subgroup crosses a specified threshold.

4. **MC-LAG** is a resiliency technique that provides protection against link and switch failures. In this case, one end of the LAG may comprise two switches while the other end is oblivious to this fact and sees a single partner.

Switches at both ends of a LAG rely on LACP so they can agree on the load balancing or resiliency technique to use and synchronize their state machines. These techniques are discussed in additional detail in the rest of this section.

2.1 Device-embedded hardware hashing

Device-embedded hardware hashing is the original load balancing technique over a LAG. A hashing algorithm spreads the client traffic across all links of the LAG when the MEF 10.4 Subscriber UNI Link Aggregation service attribute value or the MEF 26.2 Operator UNI Link Aggregation service attribute value is All Active. The frame distribution decision takes place in the transmit direction, while the receive direction is passive and simply recombines traffic from the different links.

The transmit algorithm uses a hash function to determine which link to use for forwarding a given packet. This function relies on a subset of the packet header field values, including:

- Source MAC address
- Destination MAC address
- VLAN ID
- EtherType
- MPLS labels
- Source IP address
- Destination IP address
- Source Port
- Destination Port.

The goal is to evenly distribute the traffic across all links. However, this goal is almost never completely achieved. Initially, LAG implementations solely implemented L2 hash criteria, which typically resulted in a 70 percent/30 percent split on a two-link LAG. However, advanced load balancing techniques implemented in modern packet switching hardware can employ a layer 4 (L4) hash (e.g., using TCP/UDP port numbers), layer 3 (L3) hash (e.g., based on the IP address) or L2 hash (e.g., based on MAC addresses) to make the distribution among the links

in the LAG as close as possible to even. This avoids the fairness problem with early LAG implementations. For example, it allows different L4 flows between two clients to use different physical links. Switches may rely on other advanced techniques to achieve the goal of even distribution when they handle diverse traffic patterns. In general, packet switching hardware employs different hashing criteria for IP and non-IP frames.

2.2 PSFD

PSFD puts the operator in control of load balancing. With PSFD, the operator assigns the link of a LAG to use for forwarding a predetermined set of service frames. This is done while the MEF 10.4 Subscriber UNI Link Aggregation service attribute value or the MEF 26.2 Operator UNI Link Aggregation service attribute remains All Active, as for device-embedded hardware hashing.

The operator controls PSFD by setting the MEF 10.4 Port Conversation ID to Aggregation Link Map service attribute for a given UNI or the MEF 26.2 ENNI Port Conversation ID to Aggregation Link Map common attribute. These attributes rely on a table, called the "PSFD configuration table" in this paper. This is illustrated in Table 1 for the UNI case, which is the focus for the rest of this white paper. Table 1 lists the Port Conversation IDs, which identify service frames destined to a preferred link of the LAG, and the Link Selection Priority Lists. The first link in the first row of the Link Selection Priority List, i.e., link 1 (where 1 is the Link Number ID associated with this link), is the preferred link, and the last link, i.e., link 3, is the least preferred. In this example, the Port Conversation ID is the Customer VLAN (C-VLAN) ID of a C-VLAN tagged frame or the value 0 when a frame is untagged or priority tagged.

With PSFD, service frames with Port Conversation ID 0, 3 or 7 are forwarded on link 1 by default. If this link fails (i.e., becomes nonoperational), these service frames are forwarded on link 4 (if it is operational). If the latter fails or was nonoperational in the first place, the frames are forwarded on link 3. If links 1 and 4 recover from failure, the service frames are again forwarded on link 1 because it has the highest priority.

Table 1. Example of how PSFD can map Port Conversation IDs to links of a LAG at a UNI

Port Conversation ID	Link Selection Priority List
0, 3, 7	1, 4, 3
6	2, 1, 3
8	4, 2, 1
2000	2, 3
All other values	

Link 3 is not the first entry in any of the Link Selection Priority Lists, so it is used only if links 1, 2 or 4 fail. Furthermore, all service frames with VLAN ID values other than those explicitly set in Table 1 are not transported on any link and are dropped. The implications should be noted: PSFD is not fully transparent to service frames presented at the UNI when compared to a device-embedded hardware hashing algorithm, which does not put any constraints on the UNI (or ENNI) demarcation point and leaves the definition of the filtering criteria at the level of the EVC End Point (EP). PSFD forces the operator to explicitly list the Port Conversation IDs (C-VLAN IDs here) that are allowed and assigned to individual LAG member links. Any oversight will result in traffic being dropped.

With PSFD the receive direction is no longer passive. The received frame is always checked against expected Port Conversation IDs to decide whether to pass or drop the frame.

When applied at a UNI, MEF 10.4 requires that PSFD be used according to IEEE 802.1AX-2014 with traffic distribution per C-VLAN ID. When applied at an ENNI, MEF 26.2 requires that PSFD be used according to IEEE 802.1AX-2014 with traffic distribution per Service VLAN (S-VLAN) ID. PSFD is not meant to be used on an NNI within the same service provider network (i.e., an Internal NNI).

The sections below describe four examples of how PSFD can be used. In all the following examples, an identical copy of the PSFD configuration table must exist at both ends of the LAG, that is, at the Subscriber Network (SN) end and Service Provider Network (SP Network) end.

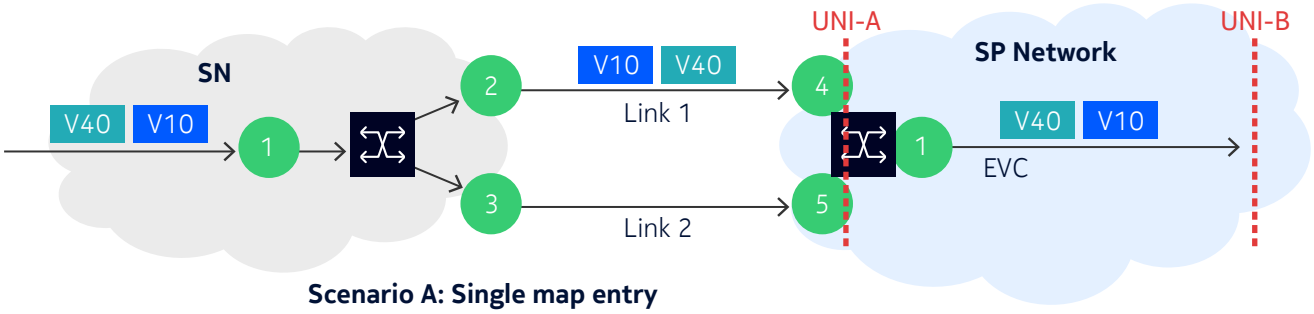
2.2.1 Single EVC with the EVC EP Map service attribute set to List

Figure 1 shows two scenarios where a SN is connected via a LAG to an SP Network that has a single Ethernet Virtual Connection (EVC) between UNI-A and UNI-B. The value of the MEF 10.4 EVC EP Map service attribute is assumed to be set to List, meaning that PSFD only handles service frames with C-VLAN ID values from 1 to 4094. Each numbered circle represents a physical port on a packet card.

Scenario A shows two Port Conversation ID values, 10 and 40, which correspond to C-VLAN ID values and are mapped to the EVC EP at UNI-A. At UNI-A, the tabulated Port Conversation ID to Aggregation Link Map shows that only service frames with C-VLAN ID value 10 or 40 traverse link 1 if it is operational or link 2 if link 1 is not operational and link 2 is operational. Other service frames are dropped.

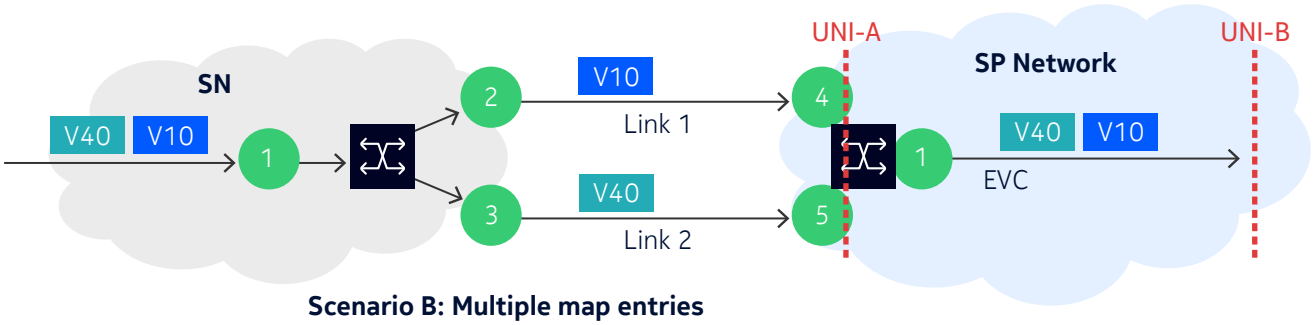
In Scenario B, each Port Conversation ID value has its own mapping. In this case, service frames with C-VLAN ID value 10 use link 1 and service frames with C-VLAN ID value 40 use link 2 if both links are operational. If one of the two links is not operational, service frames with either C-VLAN ID values are carried on the same link. Other service frames are dropped.

Figure 1: Single EVC with the EVC EP Map service attribute set to List



Scenario A: Single map entry

Port conversation ID	Link selection priority list
10, 40	1, 2
All other values	



Scenario B: Multiple map entries

Port conversation ID	Link selection priority list
10	1, 2
40	2,1
All other values	

2.2.2 Single EVC with the EVC EP Map service attribute set to All

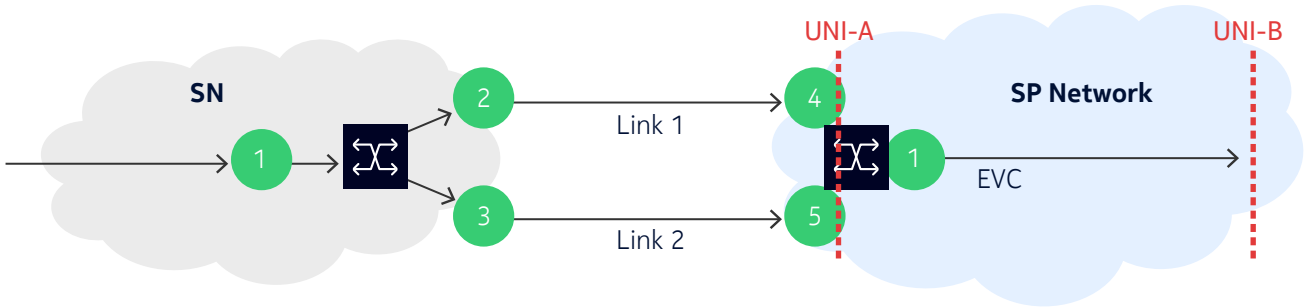
Figure 2 shows the same network as Figure 1 but with the EVC EP Map service attribute set to All, meaning that PSFD handles all service frames, including untagged and priority-tagged frames.

Scenario A shows that all service frames traverse link 1 if it is operational and switch to link 2 when link 1 is not operational and link 2 is operational.

Scenario B shows that untagged and priority-tagged service frames and the service frames with C-VLAN ID value 5, 17, 22 or 200 traverse link 1 and other service frames traverse link 2 when both links are operational. If one of the two links is not operational, the link that remains operational carries all service frames.

Scenario C shows that untagged and priority-tagged service frames and the service frames with C-VLAN ID value 5, 17, 22 or 200 have link 2 standing by if link 1 fails. The other service frames on link 1 are dropped.

Figure 2: Single EVC with the EVC EP Map service attribute set to All

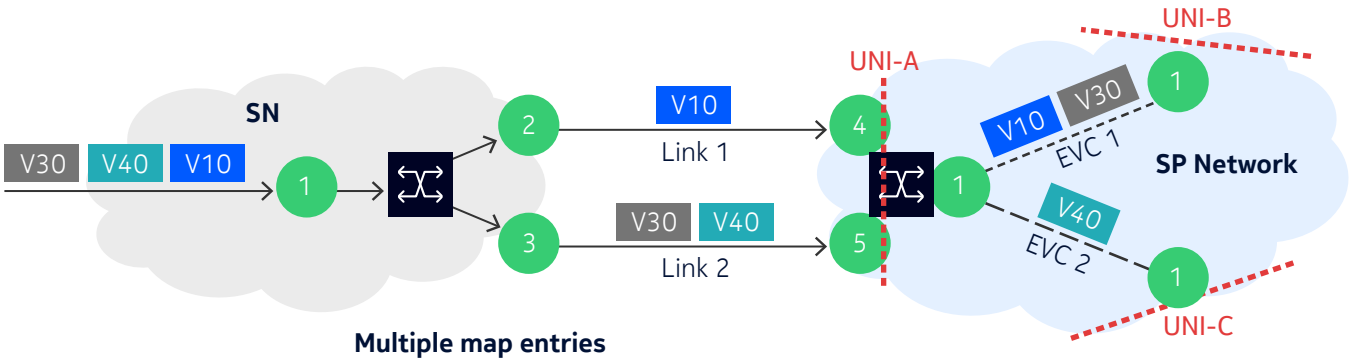


Port conversation ID	Link selection priority list
Scenario A	
All C-Tag VLAN ID values	1, 2
Scenario B	
0, 5, 17, 22, 200	1, 2
All other values	2,1
Scenario C	
0, 5, 17, 22, 200	1, 2
All other values	1

2.2.3 Multiple map entries serving two different EVCs

Figure 3 shows the tabulated Port Conversation ID to Aggregation Link Map for a scenario that has two EVCs. In this case, the EVC EP Map service attribute is set to List. Service frames with C-VLAN ID value 10 (mapped to EVC-1) use link 1. Service frames with C-VLAN ID value 30 (mapped to EVC-1) or 40 (mapped to EVC-2) use link 2 if both links are operational. If one of the two links is not operational, the link that remains operational carries service frames with any of these three C-VLAN ID values. All other service frames are always dropped.

Figure 3: Two EVCs with multiple map entries



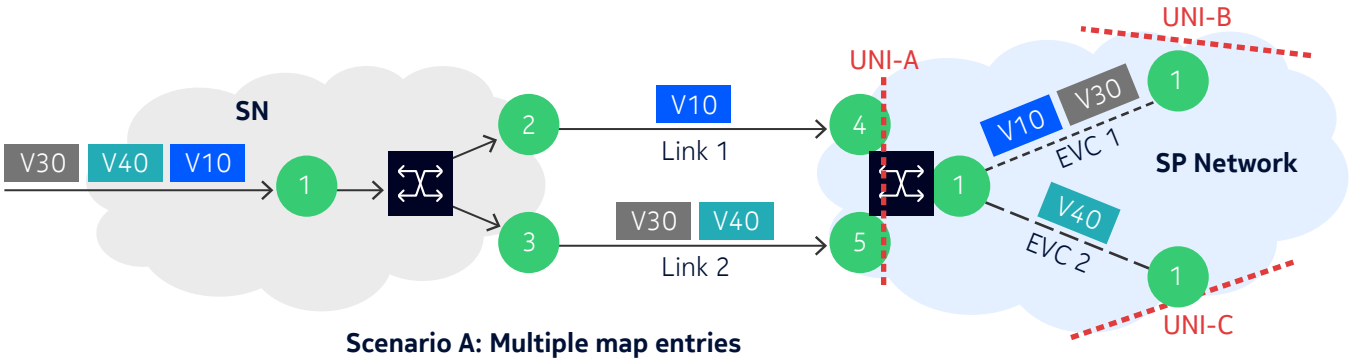
Multiple map entries

Port conversation ID	Link selection priority list
10	1, 2
30, 40	2,1
All other values	

2.2.4 PSFD operation in the presence of a link failure

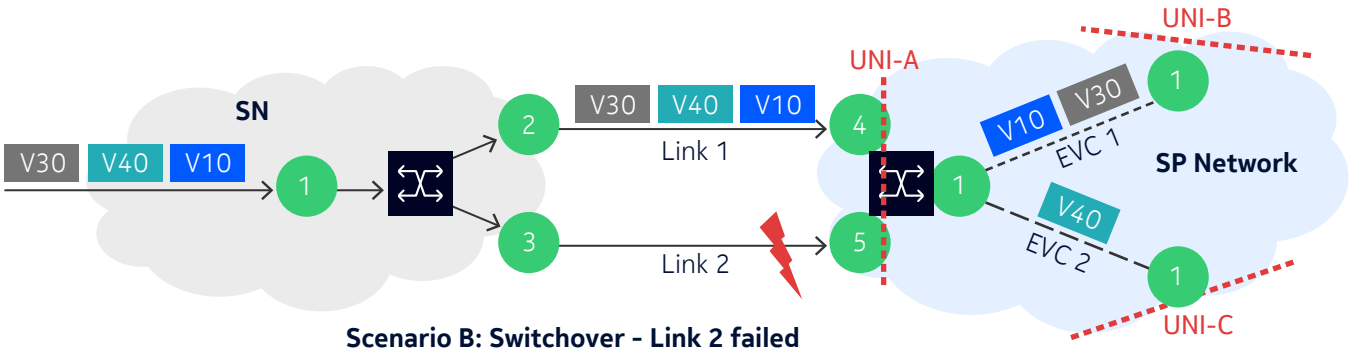
Figure 4 illustrates how PSFD operates when a link fails. Scenario A shows normal operation for a given PSFD configuration. It is the same as the example discussed in section 2.2.3.

Figure 4: PSFD operation in the presence of a link failure



Scenario A: Multiple map entries

Port conversation ID	Link selection priority list
10	**1**, 2
30, 40	**2**,1
All other values	



Scenario B: Switchover - Link 2 failed

Port conversation ID	Link selection priority list
10	**1**, ~~2~~
30, 40	~~2~~,**1**
All other values	

Scenario B shows that link 2 has failed. It also shows how the selected link in the PSFD configuration table changed in response to the failure of link 2. As explained earlier, there are identical copies of the PSFD configuration table at each end of the LAG, one at the SN end and the other at the SP Network end. The PSFD configuration tables allow the same response to a link failure at each end, causing the C-VLAN IDs originally carried on link 2 to now be carried on link 1.

2.3 Active/standby

When used in active/standby mode per MEF 10.4 or MEF 26.2, link aggregation operates with two links in a LAG. However, this can be extended in practice with the following capabilities:

- The ability to leverage more than two links. In this case, the links can be divided into two subgroups. One subgroup is active while the other is standby. To this end, link aggregation allows the operator to flexibly define subgroups by setting priority values for the participating switches and every link in between. The standby subgroup becomes active when the number of failed links in the active subgroup crosses a specified threshold. Nokia 1830 PSS packet cards implement a solution that defines a threshold for the number of failed links that trigger the switching from the active to standby subgroup.
- Support for revertive or non-revertive mode of operation, an LACP feature added in IEEE 802.1AX-2020. The revertive/non-revertive mode of operation is well known in the protection switching mechanism standardized for transport in ITU-T Study Group 15 but was missing in IEEE 802.1AX-2008. IEEE 802.1AX-2014 incorrectly introduced the specification of such a mode, which was removed by IEEE 802.1AX-2014/Cor 1-2017. A properly specified revertive/non-revertive mode of operation was re-introduced as part of LACP in IEEE 802.1AX-2020. In non-revertive mode, the recovered aggregation port can only be active again if it is the only link in the LAG allowed to do so. In revertive mode, which is the default, the recovered aggregation port can again be active once a Wait-To-Restore (WTR) period has expired.

Extending active/standby as stated adds value by increasing capacity and by allowing for an automatic return to the normal state. IEEE 802.1AX-2020 specifies the revertive/non-revertive mode of operation, complete with controls that allow the operator to:

- Decide which aggregation ports in a LAG are revertive or non-revertive
- Configure the duration of the WTR period in seconds.

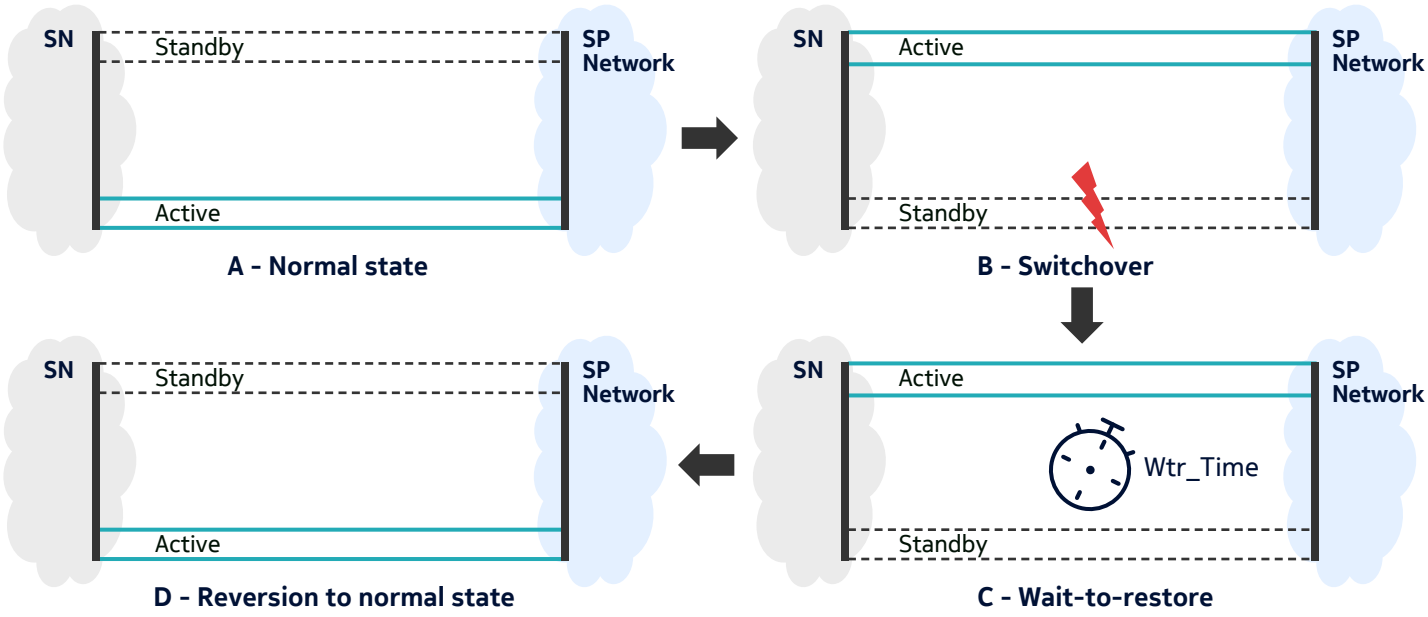
MEF 32 specifies requirements [R26], [R27] and [O1] for the revertive or non-revertive mode of operation at the EVC EP level.

The following sections describe three examples of how the revertive or non-revertive mode of operation can be used when a LAG has more than two links.

2.3.1 Revertive mode of operation

Figure 5 shows a clockwise progression of LAG subgroups reacting to link failures while working in revertive mode of operation. The progression starts with the normal state of a LAG (where two links are active and two links are standing by) and then shows the two subgroups after switching because of link failures. When the failed links have become operational again, the switchover occurs after the WTR period has expired. The LAG then reverts to the normal state.

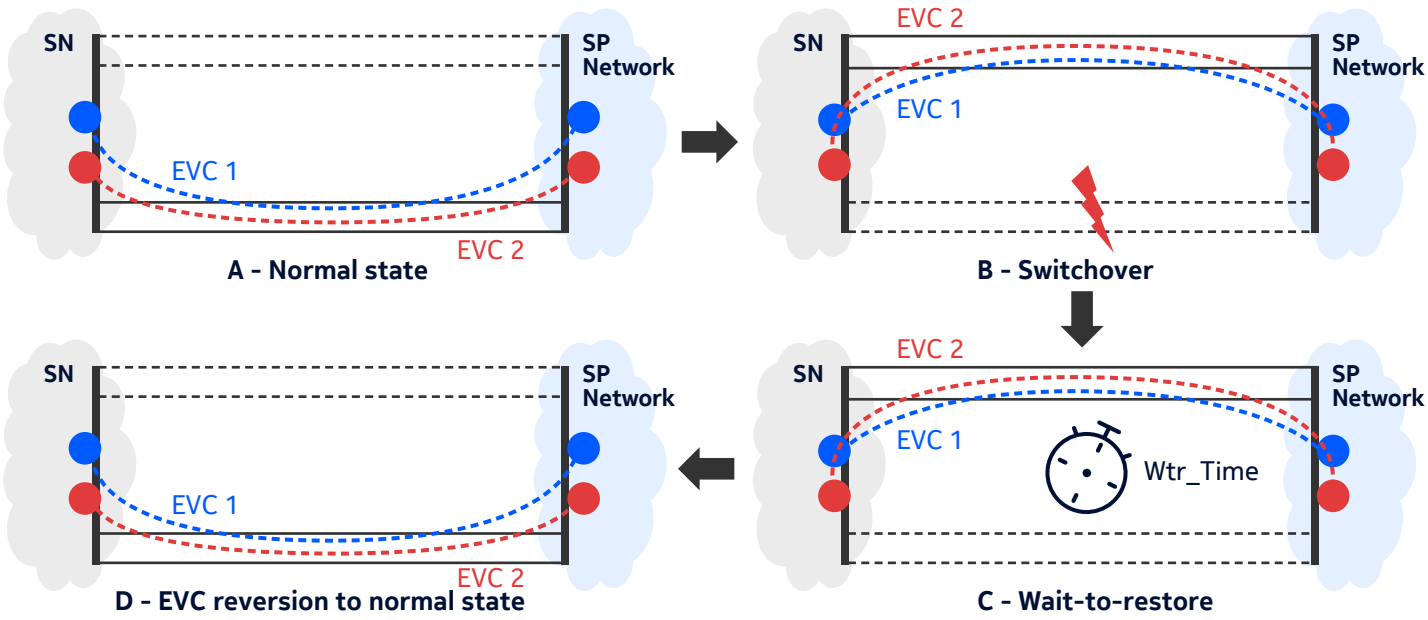
Figure 5: Revertive operation



2.3.2 Revertive mode of operation at EVC EP level

Figure 6 illustrates the revertive mode of operation at the EVC EP level in four steps (clockwise). The progression begins with the normal state, where two EVCs of a LAG are served by the lower subgroup's two active links and the upper subgroup's two links are standing by. The two subgroups then switch because of link failures, resulting in the two EVCs migrating to the other subgroup. When the failed links have become operational again, the switchover of the two EVCs occurs after the WTR period expires. The LAG then reverts back to the normal state and carries both EVCs.

Figure 6: Revertive operation at the EVC EP level

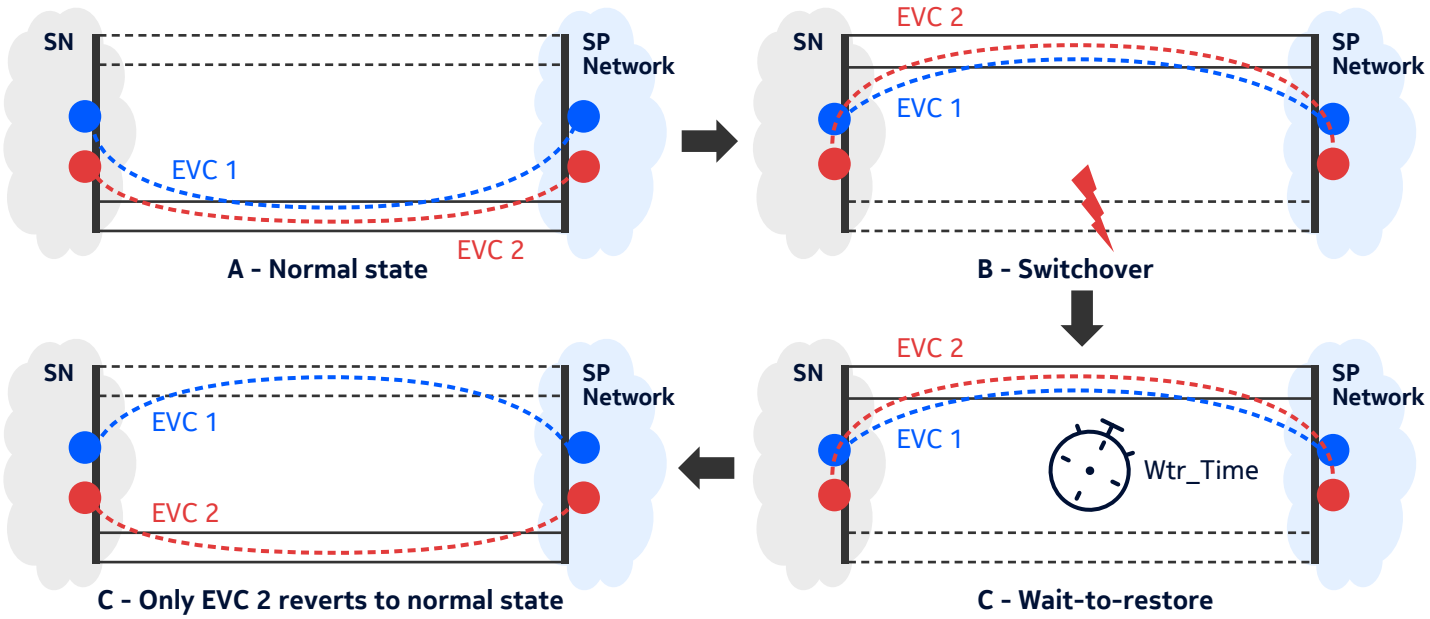


Although it is not shown in Figure 6, the operator can decide which EVC reverts. In this example, both EVCs were chosen to revert. The next section illustrates the capability of reverting only a selected EVC.

2.3.3 Mixed reversion at the EVC EP level

Figure 7 illustrates mixed revertive mode at the EVC EP level. The operation in this scenario is identical to that explained in section 2.3.2. However, in this example, when the failed links become operational again, only EVC 2 is in revertive mode and pending WTR period expiration (not visualized in the figure). After the WTR period expires, the LAG reverts back to the normal state but carries only EVC 2, whereas EVC 1 remains on the upper subgroup in non-revertive mode.

Figure 7: Mixed reversion at the EVC EP level



2.4 MC-LAG

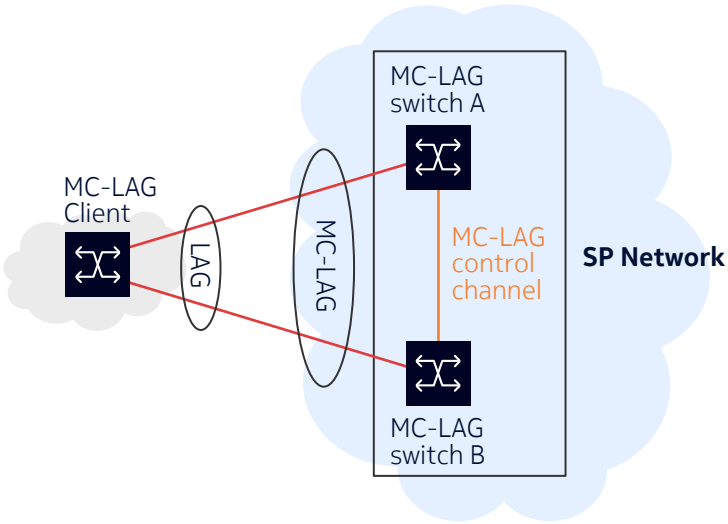
The MC-LAG concept has existed for a long time in proprietary implementations. MC-LAG is a networking technology used to enhance the availability, resilience and bandwidth capacity of network connections by combining two switches or chassis into a single logical switch. As such, MC-LAG allows two switches (or chassis) at one end of a LAG to appear as one to the switch at the other end. This enables physical redundancy against switch failure.

The key components of an MC-LAG configuration typically include:

- **MC-LAG peers**: These are the switches or chassis that are part of the MC-LAG. They work together to form a logical switch and share information to maintain synchronization.
- **Inter-chassis link**: This is a link that connects the MC-LAG peers. It carries control information between MC-LAG peers.
- **MC-LAG client**: This refers to a device connected to the MC-LAG peers. It could be a switch, a server or any other networking device that could benefit from the increased bandwidth and redundancy provided by MC-LAG.
- **Split horizon**: MC-LAG prevents traffic from being sent back to the MC-LAG client by employing a mechanism called split horizon. When the client sends traffic to peer A, peer A forwards the traffic to the desired destinations, including peer B. Split horizon ensures that the traffic received by peer B is not sent back to the client by peer B.

Figure 8 shows an MC-LAG example. A switch in the SN (an MC-LAG client), is connected to two MC-LAG peers (switch A and switch B) in the SP Network. The peers exchange control information through an MC-LAG control channel over the inter-chassis link. This information allows them to coordinate with each other so that the user traffic is appropriately forwarded within the MC-LAG. Peers A and B operate synchronously, presenting themselves as a single logical switch to the MC-LAG client. If A fails, B stays active, and vice versa, ensuring continuous connectivity without interruption. This is the key benefit of this configuration.

Figure 8: MC-LAG example



MC-LAG also supports the LAG features discussed earlier, such as the simultaneous use of both links within the LAG. This allows for load balancing, where traffic can be distributed across the links based on a range of factors, such as a subset of the packet header field values listed in section 2.1. The load balancing algorithm ensures that traffic is evenly distributed across active links, maximizing the utilization of available bandwidth and preventing any single link from becoming congested. In the event of a link failure or maintenance activity, MC-LAG provides seamless failover. If one link within the LAG becomes inactive, the traffic can automatically be redirected to the remaining active link(s), ensuring uninterrupted network connectivity. The simultaneous use of both links in this MC-LAG configuration provides increased throughput, improved network performance and enhanced reliability by utilizing the combined capacity of the aggregated links.

3 Pros and cons of link aggregation techniques

3.1 Device-embedded hardware hashing

Device-embedded hardware hashing has the following benefits:

- It is simpler to deploy compared to PSFD or active/standby LAG and requires fewer configuration steps.
- It provides efficient and fast traffic distribution across the links in the LAG by using advanced hardware-based hash algorithms to evenly distribute traffic based on a range of factors, such as a subset of the packet header field values listed in section 2.1. This helps maximize link utilization, allows for the dynamic and effective utilization of available bandwidth, and helps prevent individual links from becoming overloaded while others are underutilized.

- It can be used to optimize network performance for applications and services as traffic is distributed evenly across available links, which can result in reduced latency and improved response times.

Device-embedded hardware hashing has the following drawbacks:

- The operator has no control over which link a given frame is forwarded because hashing is implemented in hardware and is typically not configurable.
- It is sensitive to traffic types. For example, it can hash unicast traffic and distribute it efficiently among the links of a LAG but it cannot achieve the same result with multicast traffic.

3.2 PSFD

The biggest advantage of PSFD is that the operator is in full control of frame distribution. For service providers that want strict and deterministic control of traffic assignment to LAG member links, manageability is a key requirement. The operator can assign certain service frames to a specific Link Selection Priority List, which can help achieve minimum and constant delay for high-priority traffic. PSFD has a wide range of applications, including business Ethernet and 5G services. For example, traffic types such as enhanced Mobile Broadband (eMBB), massive Machine-Type Communication (mMTC) and Ultra Reliable Low Latency Communication (URLLC) can be assigned different Port Conversation IDs (such as C-VLAN IDs) to help achieve minimum and constant delay. Furthermore, by carefully selecting the sequence in the Link Selection Priority Lists, service providers can protect high-priority traffic with more than one link, thereby increasing availability.

The main benefits of PSFD are:

- **Service isolation**: PSFD can distribute each service or flow to a specific link, enabling service isolation. This can be beneficial in scenarios where different services or applications require dedicated bandwidth or quality of service.
- **Enhanced load balancing**: PSFD allows for more granular load balancing by distributing traffic on a per-service or per-flow basis. With careful assignment of Port Conversation IDs, service providers can achieve an optimized utilization of available bandwidth across the aggregated links. This minimizes congestion and optimizes overall network performance.
- **Same link in both transmit and receive directions**: Without care, frames for a given service could be transmitted over one physical link of a LAG while being received by that same node on a different link. Should one of these links fail or degrade, the service could be partially or completely impacted. PSFD can be used to avoid this. By allowing both directions of traffic to use the same physical link, PSFD works well with the many protocols designed with this expectation.
- **Ideal setup for fault management and performance monitoring**: Service frames and attendant operations, administration and maintenance frames (as specified in Recommendation ITU-T G.8013/Y.1731 [10]), including connectivity fault management frames (also specified in IEEE Std 802.1Q [11]), ideally share the same physical link. Since they also share the same service identification, they can use the same Port Conversation ID. PSFD then guarantees that they share that same physical link.
- **Easier ingress metering**: By steering frames for a given service to use a specific physical link, PSFD makes it easy to locate the meter used for the ingress metering of this service.

PSFD has the following drawbacks:

- Link Selection Priority Lists must be carefully selected across the PSFD configuration table. This is to avoid selecting the same link upon failure, which might overload this link and result in traffic loss.

- PSFD is a static configuration, so it may result in less-efficient load distribution and underutilization of some of the LAG member links when traffic in certain services is reduced or absent at a certain moment in time. There is a trade-off between operator control and the ability to dynamically adjust to varying traffic loads.

3.3 Active/standby

Active/standby has the following benefits:

- It provides a simple way to achieve protection against link failures.
- It is relatively straightforward to configure and manage compared to more complex link aggregation schemes such as PSFD or device-embedded hardware hashing because it requires minimal configuration of the backup link. This simplifies network setup and maintenance.

Active/standby has the following drawbacks:

- It does not provide load balancing across both links, so all traffic is impacted by a protection switch rather than only half the traffic in a load-sharing configuration.

3.4 MC-LAG

MC-LAG has existed since the late 2000s in proprietary implementations and is now well established as a technology for improving network redundancy. Its main benefits are that it can protect against switch failure at one end of a LAG and that it can use LAG to protect against inter-chassis link failure. However, MC-LAG has the drawback that it can be more complex to configure and manage than PSFD or device-embedded hardware hashing.

4 Conclusion

Link aggregation has evolved considerably since being specified in IEEE 802.3ad-2000 and later moved under the responsibility of the IEEE 802.1 Working Group to become IEEE 802.1AX, which has seen successive revisions. Originally relying on hardware-based hashing with limited entropy, today's link aggregation enjoys hardware-based hashing with predictable load balancing to ensure fairness. It is enhanced with advanced techniques such as PSFD and revertive or non-revertive mode of operation. Although proprietary implementations were initially implemented, the MC-LAG concept is now standardized. PSFD allows the operator to predictably assign service frames to any desired link member of a LAG, resulting in an unprecedented level of control over link aggregation. Revertive or non-revertive mode of operation provides for automatic return to the normal state. MC-LAG provides protection against link and switch failures.

Understanding the different types and operational modes of link aggregation is essential for network architects and administrators seeking to optimize network performance and reliability. Service providers should choose options that align closely with their priorities and network demands. By leveraging these techniques and modes effectively, a service provider can deliver a robust and resilient network infrastructure that meets present and future traffic demands.

Abbreviations

C-VLAN	Customer Virtual Local Area Network
eMBB	enhanced Mobile Broadband
ENNI	External Network Network Interface
EP	End Point
EVC	Ethernet Virtual Connection
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
L2	Layer 2, known as the Data Link Layer of the OSI Model
L3	Layer 3, known as the Network Layer of the OSI Model.
L4	Layer 4, known as the Transport Layer of the OSI Model
LACP	Link Aggregation Control Protocol
LAG	Link Aggregation Group
MAC	Media Access Control
MC-LAG	Multi-Chassis Link Aggregation Group
MEF	MEF Forum
mMTC	massive Machine-Type Communication
OSI	Open Systems Interconnection
PSFD	Per-Service Frame Distribution
PSS	Photonic Service Switch
SN	Subscriber Network
SP	Service Provider
S-VLAN	Service Virtual Local Area Network
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UNI	User Network Interface
URLLC	Ultra Reliable Low Latency Communication
VLAN ID	VLAN Identifier
VLAN	Virtual Local Area Network
WTR	Wait To Restore

References

[1] “IEEE Std 802.3ad-2000, Amendment to Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications—Aggregation of Multiple Link Segments,” IEEE, 2000.

[2] “IEEE Std 802.3, 2000 Edition, Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications,” IEEE, 2000.

[3] “IEEE Std 802.1AX-2008, IEEE Standard for Local and metropolitan area networks—Link Aggregation,” IEEE, 2008.

[4] “MEF 32, Requirements for Service Protection Across External Interfaces,” MEF Forum, 2011.

[5] “IEEE Std 802.1AX-2014, IEEE Standard for Local and metropolitan area networks—Link Aggregation,” IEEE, 2014.

[6] “MEF 10.3.2, Amendment to MEF 10.3 - UNI Resiliency Enhancement,” MEF, 2015.

[7] “MEF 10.4, Subscriber Ethernet Service Attributes,” MEF Forum, 2018.

[8] “IEEE Std 802.1AX-2020, IEEE Standard for Local and metropolitan area networks—Link Aggregation,” IEEE, 2020.

[9] “MEF 26.2, External Network Network Interface (ENNI) and Operator Service Attributes,” MEF Forum, 2016.

[10] “ITU-T G.8013/Y.1731, Operation, administration and maintenance (OAM) functions and mechanisms for Ethernet-based networks,” ITU-T.

[11] “IEEE Std 802.1Q, IEEE Standard for Local and metropolitan area networks—Bridges and Bridged Networks,” IEEE.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: (November) CID213657