1. Introduction and Value Proposition

Service Providers are facing the challenge to provide intelligent services that can quickly adapt to the market needs. New services such as cloud transport require unprecedented flexibility and elasticity from the network. Increasing bandwidth demands and decreasing ARPU put pressure on reducing network cost. At the same time, services need to be deployed faster and more cost effectively to stay competitive.

Carrier Ethernet solutions have evolved from native Ethernet, to static MPLS, then to Unified MPLS to address the above challenges. The Unified MPLS based Carrier Ethernet solution provides a single converged network infrastructure with a unified operational model. It has great advantages in terms of network convergence, high scale, high availability, and optimized forwarding. However, it is still quite challenging to manage, especially on large-scale networks, because of the large number of distributed network protocols involved. Service providers are looking for the evolution of the Carrier Ethernet network solution that can offer all the values of the Unified MPLS solution with operational simplicity and full programmability.

This paper introduces a new SDN-enabled Carrier Ethernet – the ACE (Agile Carrier Ethernet) solution. The ACE solution evolves traditional Carrier Ethernet networks to an SDN enabled programmable network to be capable of delivering all services (Residential, Business, Mobile Backhauling) on the promises of simplicity, programmability, and cloud integration, with guaranteed service level agreements.

The ACE solution will bring tremendous value to the Service Providers:

- **Agile service deployment and rapid time to market** through fully automated service provisioning and service lifetime adjustments
- **Reduce operation complexity** by minimizing network protocols and enable centralized intelligence on the controller
• **Autonomic self-protected network infrastructure** with autonomic segment routing and topology independent fast-reroute software features

• **Application-engineered Routing** through centralized orchestration and path computation

• **Smooth migration towards SDN-enabled network** fully backward-compatible with existing network protocols and services

2. **ACE Architecture Overview**

The ACE architecture seeks the right balance between distributed intelligence and centralized optimization and programming. ACE combines the minimal, but sufficient distributed intelligence, leveraging routing protocols, with centralized optimization and programming provided by an SDN controller. The result is an architecture capable of delivering Carrier Ethernet Services in a simple, flexible and scalable fashion, using a reduced set of network protocols.

The principle of the ACE architecture is network disaggregation: moving away from integrated HW and SW, vendor specific, individual network devices toward network-as-a-platform with disaggregated open components.

![Figure 1: The principle of the ACE – the network dis-aggregation](image-url)

In the ACE architecture, the router will keep the minimal SW features and protocols locally to build up the optimized and self-protected IP/MPLS transport (also called underlay). The service (also called overlay) provisioning and service control plane protocols are centralized on the controller layer. The controller uses open APIs to provision the service features and program the service forwarding tables. More specifically, for IP/MPLS transport the network nodes run distributed IGP protocols (ISIS or OSPF) with required extensions for MPLS Segment Routing locally.

The ACE architecture is made up of three layers of components. Each component can be used independently, but the greatest value is gained from combining them together:

• **Autonomic self-protected network transport with distributed segment routing**: Provides the simple and optimized end-to-end MPLS transport with self-protection

• **Consolidated VPN services with centralized BGP control plane**: EVPN for L2 and/or L3 VPN services, plus IP-VPN for L3
- **SDN Controller and Orchestrator:** for centralized service provisioning and management, as well as centralized service control plane and inter-domain segment routing

The ACE network includes two types of network nodes, one is an access node, and the other is the service node. Access nodes only have a limited set of transport features to be simple. The service nodes will have a full set of transport and service features enabled as in the traditional IP/MPLS network architecture. In this document, we will also call the service node the “Gateway” node.

![Figure 2: The ACE Reference Architecture](image)

As shown in the Figure 2, the network is divided into 2 levels of the ACE domains: the core domain and the leaf domain. For a super large-scale network, the leaf domain can be further divided into sub-leaf domains (please refer to: IETF draft-filsfils-spring-large-scale-interconnect). The core domain usually represents the core network and/or the aggregation network. The leaf domain usually represents a particular metro network. The border routers (called gateways in this paper) participate in the core domain and one or more leaf domains. Each domain will run its own IGP process. These IGP process are isolated to each other to be simple.

Each node in the ACE domain is represented by a pair of SIDs (Segment routing ID): [gateway SID, node SID] based on the direction of traffic flow across the core and leaf domains. For example, the node A1 in ACE domain 1 is represented by SID-list: [16001, 18001]. Node A2 in the ACE leaf domain 2 is represented by SID-list: [16002, 18001]. The SID in the core domain is globally significant network-wide. The SID in the leaf domain is significant within the local leaf domain. In other words, the SID can be re-used across different leaf domains.

### 3. ACE Transport Architecture with Segment Routing

#### 3.1 Transport Architecture Baseline: Intra-domain Segment Routing

The intra-domain architecture in an ACE network is drastically simplified by the use of distributed segment routing on the routers: only an IGP (ISIS, OSPF) protocol (with segment routing extensions) is required to build the transport connectivity. Each node is identified by its own node segment identifier (called SID in this paper) and any node to any node transport connectivity within the same IGP/SR domain, is achieved by using the IGP shortest path. The path is dynamically built by the distributed routing protocol, without requiring any external controller.
Advantages of the segment routing based MPLS transport:

- **Simplicity**: the network nodes only need to run a single protocol, an IGP. It doesn’t need any other label distribution protocols like LDP, RSVP-TE or BGP-LU. In addition, with IP unnumbered interface support for segment routing, operators can ease the pain of the IP address management. For a fast growing dynamic Carrier Ethernet network, this is very helpful. Operators can insert or remove the network nodes without requiring IP address changes on the adjacent links.

- **Self-protected**: segment routing supports topology-independent fast reroute (TI-LFA) for both link and node protection. Compared with traditional RSVP-TE FRR, segment routing TI-FRR has two major advantages. One is simplicity, two lines of the interface configuration will achieve link and node protection instead of creating an RSVP-TE tunnel and manipulating the tunnel protection. The other advantage is the optimization, segment routing TI-LFA will pre-calculate the backup path based on the post-converged topology. So it’s optimized. There is not even a temporarily sub-optimal path during network re-convergence.

- **Anycast SID for simple service node redundancy**: the redundant network nodes (such as the service nodes) can share the same anycast loopback address and SID. This simple anycast SID configuration allows for automatic service node redundancy, without requiring some complex features such as PW redundancy on the access nodes.

### 3.2 Transport Architecture: Inter-domain Segment Routing

One of the major advantages of the Unified MPLS solution is its high scale for large size networks. In the Unified MPLS Architecture the network is divided into multiple standalone IGP/LDP domains. For large networks, hierarchical transport LSP’s, as defined in RFC3107, are used to provide inter-domain connectivity. Hierarchical LSP’s are built using two MPLS labels: an IGP label to provide reachability within each IGP domain, and a BGP label for the reachability across domains.

The hierarchical LSP with BGP is seen as too complex by some operators, thus the ACE solution has been designed in a way to address this complexity issue. The ACE solution simplifies inter-domain MPLS transport by removing the need for the BGP protocol and BGP/IGP label hierarchy: this can be achieved by using an SDN controller to provision the SR label stack (or SID-list) at the ingress node.

The SDN controller learns about the network topology either by static configuration or by using a standard protocol like BGP-LS.
There are two types of inter-domain paths:
- Shortest path with anycast gateway SID
- Application-Engineered path with SR-TE

### 3.2.1 Shortest path with anycast gateway SID

As mentioned earlier, each node in the ACE domain will be represented by a pair of SID’s: [gateway SID, node SID]. For example, the node A1 in the ACE domain 1 is represented by: [16001, 18001]. Since all network nodes register with the network controller or orchestration system, the controller or orchestrator will have the full table of information on node/gateway SIDs.

For the inter-domain shortest-path, the orchestrator can easily provision it based on its local node information. For example, in Figure 4, the path from A1 to A2 will use SID-list [16001, 16002, 18002]. The source router A1 will push this SID-list as an MPLS label stack into the packet header. The rest of the network will just do simple MPLS forwarding based on the top label.

![Figure 4: Inter-Domain Segment Routing Transport with anycast gateway SID](image)

The packet is forwarded to the local gateway using anycast label 16001 based on the IGP shortest path. The local gateway will forward the packet to the remote gateway based on the remaining top label 16002 in the core domain. The remote gateway will forward the packet to the node A2 based on its node label 18002. Packet forwarding in each domain is per IGP shortest-path. Each domain will have its local link and node protection based on TI-LFA independently. The anycast gateway SID provides a simple and efficient way to achieve node redundancy. For the unified MPLS solution, it requires a hierarchical FIB with re-cursive lookup on the router. Advanced software features such as BGP PIC (prefix independent convergence) need to be developed to handle the fast convergence for network failures. For the ACE solution, the label and FIB table is flat. Router only look up the top label for the forwarding. No recursive lookup is required. Convergence only needs to happen at the IGP level using TI-LFA, which is independent of the service and service scale.

### 3.2.2 Application-Engineered path with SR-TE

Operators want to fully monetize their network infrastructure by offering differentiated services. Traffic engineering tunnels are used to provide different paths for different applications. The classic
RSVP-TE tunnel requires signalling along the path for tunnel setup or tear down, and all nodes in the path need to maintain the state. This won’t work well for cloud applications which have hyper scale and nature of the elasticity.

Segment Routing provides a very simple and scalable way to define an end-to-end application-aware traffic engineering path. As shown in the Figure 5, the service orchestrator simply provisions the service with a Service Level Agreement (SLA) tag, which is requested by the application via northbound API. The network node will trigger the Path Computation Element Protocol (PCEP) request to the controller. The controller will return the path information based on the SLA tag, and will program the SR tunnel via PCEP dynamically. If there is any topology change in the network, the controller will re-optimize the tunnels and re-program the routers via PCEP dynamically. If the service is deactivated, the associated SR tunnels will be removed as well.

To simplify the operation further, user doesn’t need to configure the traditional traffic engineering tunnel interface. The SR tunnel here is nothing but the SR SID-list which will be programmed into the forwarding table directly.

4. ACE Service Architecture

The principle of the ACE service architecture is to disaggregate the service control plane from the network node. The network controller has the intelligence to program the service label to the network node. The network node doesn’t need to run any network protocols for the service layer. The service label could be a static PW label, or a dynamic VPN label learned from BGP, or anything which can identify the service, for example, the network service header. Three possible modes of operations are available:

- Option 0: Distributed service control plane on the routers
- Option 1: Static VPN label provisioning
- Option 2: Centralized service control plane on SDN controller

With SR transport mentioned in the section 3, the traditional distributed control plane protocols for L2VPN and L3VPN service, such as BGP and T-LDP, work transparently as normal. No change is required. So this baseline option is not described in this paper.
4.1 Static VPN label provisioning

With the advent of SDN, NFV, and DC virtualization, more and more advanced services are being moved to the cloud. Carrier Ethernet networks will retain the function of providing the intelligent pipe between customers and the cloud. The traffic pattern will be quite straightforward; most of the traffic flow is between the customer site and the Data Center site or the service node with distributed Data Centers. With this new trend, the Carrier Ethernet service can be simplified by using a new model based on static service labels provisioned by a service orchestrator.

As shown in the Figure 6, all services in the ACE leaf domain will use Pseudowires (PW) to backhaul traffic from the access node to the service node. The service node will terminate the packets into different services, such as VRF for L3 VPN, bridge-domain for L2 VPN, or any other advanced service. The PWs will provide simple point-to-point connectivity. The Service Node doesn’t need to run any control plane protocol for the PW label distribution because the service orchestrator will provision the PW label statically. The service orchestrator will use an anycast PW label for the service which could terminate on multiple service nodes. Together with the segment routing anycast SID, it provides a very simple solution for service node redundancy.

Static PW label provisioning in the ACE solution is similar to some existing solutions like MPLS-TP where a network management system (NMS) is used to provision the MPLS-TP tunnel and PW label statically. While the two approaches have some common attributes for the network node, like no service control plane running on the network node, the ACE solution offers several key advantages compared to MPLS-TP due to the dynamic segment routing transport, such as simple plug-n-play for dynamic network topology changes, shortest path routing, equal cost multiple paths, and anycast service node redundancy.

4.2 Centralized service control plane on SDN controller
While static PW label provisioning is simple and works well for hub-and-spoke types of traffic flows, some applications may require more optimized forwarding between the access nodes directly. For example, mobile LTE/LTE-A/5G backhaul requires optimized forwarding due to requirements for low network latency. For these types of use cases, ACE will use a network controller to program the VPN label on the nodes dynamically to achieve optimized forwarding.

![Figure 7: ACE service architecture: Centralized service control plane on SDN controller](image)

As shown in Figure 7, the ACE network controller has virtual XRv as the key component. XRv runs the traditional network control protocols like BGP for the L2VPN and L3VPN services. XRv will peer with the external network nodes outside of the ACE system. XRv will exchange the VPN route with its peer devices, and then use open APIs to program the VPN labels on the network nodes within the ACE network.

To summarize: instead of having each network node running its own control plane protocol for the VPN services, ACE uses a centralized control plane running on the virtualized XR on behalf of all network nodes in the same ACE leaf domain. This will not only reduce the network level protocol complexity, but also reduce the complexity in terms of hardware requirements on the physical network nodes and thus will lead to CAPEX reduction.

5. Controller and Orchestrator Options

The ACE solution has multiple deployment options for the controller and orchestrator, depending on the customer use case and requirements.

Option 1: Cisco NSO as the ACE orchestrator
In this option, NSO is used as the cross-domain orchestrator to provision the end-to-end service. In addition to the service provisioning, NSO with the ACE functional module, will also provision the static PW label and provision the inter-domain segment routing SID-list using anycast gateway SID. Please refer to the previous sections for the details.

For more information about Cisco NSO, please refer to: [http://www.cisco.com/go/nso](http://www.cisco.com/go/nso)

**Option 2: Cisco NSO as the ACE orchestrator, plus IOS-XR PCE server for inter-domain transport**

In option 1, NSO provisions the inter-domain segment routing SID-list by using an anycast gateway SID and shortest path routing. The solution is very simple, however it may not meet some customer’s requirements for application-aware traffic engineering paths, such as low latency paths, disjointed paths for redundancy, etc. Option 2 introduces the IOS-XR PCE server to provision the inter-domain SRTE per service SLA.

In option 2, NSO only provision the service and static PW labels. The service provisioning will trigger the PCEP request to the remote node’s prefix depending on the SLA. The PCE server will calculate the best path based on the SLA tag, then use PCEP to provision the SRTE tunnel on the network node.

With this option, the service provisioning and the transport tunnel provisioning are de-coupled. NSO provisions the service, and the PCE server provisions the inter-domain transport tunnel on-demand.
**Option 3: ACE controller for centralized service management, provisioning and control plane**

The ACE controller is built on top of the open source controller platform OpenDaylight (ODL). XRv is a routing service plug-in for ODL. Besides the routing service plug-in, there are a few other carrier Ethernet service plug-ins, such as underlay manager, service monitoring, service chaining, and telemetry system. The details of the ODL implementation are beyond the scope of this whitepaper.

![Diagram of ACE orchestrator](image)

**Figure 10: ACE orchestrator for centralized service management, provisioning and control plane**

Option 3 uses an ACE controller to manage and provision all the network nodes within the ACE domain. NSO and other management systems will use open APIs to manage the ACE controller, which will use open APIs to provision and program the individual nodes in the ACE domain.

In addition to centralized management and provisioning, the ACE controller also has centralized service control plane protocols running on IOS-XRv. The ACE controller will peer with external routers with control plane protocols such as BGP. The ACE controller learns and advertises the routes, and programs the RIBs on the network nodes via open APIs.

6. **Inter-operability**

The ACE solution is fully compatible with existing IP/MPLS network protocols. It can enable a smooth migration from an existing IP/MPLS based solution to an SDN-enabled Agile Carrier Ethernet solution.

To be more specific, within the ACE domain, the network nodes can run segment routing and/or LDP. Segment routing and LDP are fully inter-operable with each other and can co-exist on the same node or network. It is simple to migrate the network node from LDP to segment routing.

For the inter-domain routing, it can use BGP-LU to inter-connect the ACE metro domain and the traditional IP/MPLS metro domains. The inter-domain transport solutions mentioned in the previous sections are fully inter-operable with the unified MPLS solution.
Conclusion

The industry is transitioning to an Evolved Programmable Network Era that promises more automation, orchestration and control enabling new service experiences for end customers and desired business outcomes for further streamlining of service provider company operations. The pace of combining proven existing technologies with emerging innovations is key to a provider’s success in building future-proof networks prepared to support the Internet of Everything.

The Agile Carrier Ethernet solution combines the power of Segment Routing based transport, consolidated VPN service with BGP EVPN control plane, and SDN control for service provisioning to build an architecture capable of delivering all services (Residential, Business, Mobile Backhauling) on the promises of simplicity, programmability, and cloud integration, with guaranteed service level agreements over an Evolved Programmable Network infrastructure.

Appendix: Terminology

ACE: Agile Carrier Ethernet
BGP-LU: BGP Labelled Unicast
EVPN: Ethernet VPN
NSO: Network Service Orchestrator
ODL: OpenDaylight
PCE: Path Computation Element
PCEP: Path Computation Element (PCE) Communication Protocol
PW: Pseudowire
SLA: Service Level Agreement
SR: Segment Routing
SID: Segment Identifier