Architectural Considerations for Backhaul of 2G/3G and Long Term Evolution Networks

White Paper

Contents

What You Will Learn

European mobile service providers are seeing unprecedented increases in the mobile backhaul capacity to support their existing service offerings. The latest generation of smart devices and USB data modems support bandwidth-intensive applications such as mobile e-mail, video downloading, gaming, and so on. Mobile broadband through High Speed Packet Access (HSPA) has led to a significant rise in the number of subscribers and increased bandwidth usage per subscriber. The volume of data traffic carried by mobile networks has already exceeded that of voice traffic. The worldwide mobile broadband subscriber numbers have already surpassed fixed broadband numbers, with the total number forecasted to be about 1.5 billion by 2014. The expenditure associated with providing this increasing bandwidth does not linearly match revenue growth. This bandwidth-scaling issue will become even more significant with the evolution towards HSPA+ and Long Term Evolution (LTE). European mobile service providers are deploying IP/Ethernet in their mobile backhaul to increase overall available bandwidth while reducing the overall cost.

This white paper will initially discuss the migration trends that Cisco sees in Europe in relation to the transport options for second-generation (2G) and third-generation (3G) traffic. A more detailed analysis will then examine the options available for transporting Global System for Mobile Communications (GSM)/2G (time-division multiplexing [TDM] based) and Universal Mobile Telecommunications Service (UMTS)/3G traffic (ATM and Ethernet based) over a converged packet network. Finally, a detailed description will discuss the LTE evolution and the feature requirements it makes on the underlying transport network. A conclusion summarizes the advantages and disadvantages of different transport options (Layer 2 VPN and Layer 3 VPN) under consideration today.

European Mobile Migration Trends

European mobile service providers are striving to reduce operating costs, and converging all networks into a single one is a major factor in achieving savings. The challenge is that the mobile architectures originated with different technologies (TDM, Frame Relay, and ATM) since their inception. The progression towards an all-IP vision, seen today with the IP NodeBs and IP Radio Network Controllers (RNCs) for UMTS and outlined by the LTE/EPC (Evolved Packet Core) architecture, is assisting the progression towards a single packet-based network.

The current GSM and UMTS networks are predominately based on SDH as the convergence technology, using 2-Mbps transport connections dedicated to either 2G or 3G mobile radio access network (RAN) traffic. In Europe, a high percentage of this SDH offering rests on microwave technology (in most cases owned by the mobile service provider). The primary issue with the SDH-based architecture is a lack of scalability. Initial LTE bandwidth calculations show a requirement for 40 Mbps on average to each cell site location that could grow to 100 Mbps over time. Cisco has seen the effect of high data traffic volumes over the last 12 to 18 months with the HSDPA (High-Speed Downlink Packet Access) evolution and the future eHSPA (Evolved High Speed Packet Access). LTE RAN requirements will further highlight the scaling issues. Many European operators have had to consider HSDPA offload solutions, using such technologies as DSL and fibre. The reason for this approach lies in the underlying SDH infrastructure, which cannot fulfil the increased data traffic volumes efficiently. The service providers in Nordic countries seem to have led in IP NodeB deployments, with the rest of Europe following this trend. A prominent European radio vendor is actively providing business-case justification for migrating UMTS NodeBs to Ethernet when the backhaul bandwidth exceeds 6 Mbps.

Initially people thought that GSM would be phased out quickly, but recently it seems that GSM will remain for the next five years or more (some countries state support until 2020.) and that at least a portion of the frequency will be reused for other technologies (that is, UMTS and LTE) in the future. There is little expectancy for large traffic growth with GSM (single-digit increases still prevail in some countries), because GSM will primarily be used for voice capabilities. Therefore, no immediate need dictates support of IP/Ethernet on these base stations. Current GSM base stations usually require two to three E1 circuits. It is worth noting that some radio vendors support Ethernet

interfaces on their base transceiver stations (BTS) for GSM, but there have not been any large-scale rollouts. A growing support for software-configurable, flexible radio adoption (radio supporting 2G, 3G, and 4G at the same time) has arisen when cell sites need attention because of bandwidth or obsolete equipment issues. In such cases, transport convergence at the access layer, in addition to the backhaul layer, leads to lower operating expenses and bandwidth efficiency.

In relation to UMTS support, we have seen in the last 6 to 12 months that ATM transport comprises of a smaller percentage of the overall UMTS transport rollout and is declining at a faster rate than initially indicated. In Europe, there are operators where up to 40 percent of their UMTS rollout is now Ethernet-based as opposed to ATM-based.

The evolution outlined above leaves many options for operators to decide how to migrate their existing legacy networks, as highlighted below:

- Migrate all GSM (TDM) and UMTS (ATM) traffic onto a packet-based infrastructure without any change to the radio equipment. Technologies such as TDM over Multi-Protocol Label Switching (MPLS), including Structure Agnostic TDM over Packet (SAToP) and Circuit Emulation Service over Packet Switched Network (CESoPSN), in addition to ATM over MPLS (ATM Pseudowire Emulation Edge-to-Edge) must support the traditional interfaces.
- GSM and UMTS traffic remains on the existing SDH-based transport. All new traffic (HSDPA and High-Speed Uplink Packet Access [HSUPA]), eHSPA, and LTE) will be on a packet-based infrastructure. Existing equipment could be migrated to IP as part of an end-of-life program or where capacity needs dictate it. In the short term, HSDPA offload options over DSL or fibre address data traffic increases in the UMTS.
- GSM traffic remains on the existing SDH-based transport and places all other traffic (3G ATM [through PWE3], UMTS High Speed Packet Access [HSxPA], eHSPA, and LTE) on new packet-based infrastructure. This plan achieves good statistical multiplexing gains. When UMTS NodeBs are upgraded to IP NodeBs, their backhaul technology will migrate from pseudowire to pure IP over Ethernet on the same network.
- Upgrade all GSM Base stations and UMTS NodeBs and controllers to support native IP and migrate onto a packet infrastructure. This eliminates the use and complexity of PWE3 but remains unlikely because of financial and logistical issues. As stated above, there is little bandwidth justification for going to IP/Ethernet on GSM base stations.

In summary, the mobile migration option that you choose depends on the installed radio equipment base. Mobile specifications are evolving towards an all-IP architecture, but there must be cost justification for upgrading existing legacy radio. This could include: traffic increases (unlikely in GSM), the necessary replacing of existing equipment at the end of its life, or the operational savings of supporting one converged network for all radio access. The support of software-configurable, flexible radio technology (radio supporting 2G, 3G, and 4G at the same time) spurs this evolution.

Most European operators have indicated that the most likely outcome is that new or upgraded radio nodes for UMTS, eHSPA, or LTE will employ IP transport over Ethernet interfaces. Existing UMTS sites that have high HSxPA utilisation are also candidates for an upgrade. The GSM base stations and low-capacity UMTS NodeBs should remain unaltered and therefore continue to require TDM or ATM transport services for the near future.

Clearly, there will be differences among operators in how to support the requirement for TDM and ATM transport services. Operators that are a subsidiary of an incumbent operator often leave these services on the existing SDH infrastructure, because this practice represents little cost to them. The incumbent operator will need to leave this underlying SDH network in place in the foreseeable future for other business purposes, including wholesale, regulated services, and so on. Mobile operators, whom the incumbents see as challengers, often lease the underlying transport requirement from incumbent players at a cost. In Europe, this cost is not declining, and many

challengers plan to support their TDM, ATM, and Ethernet requirements over a converged IP/MPLS network and move as quickly as possible to an all-IP vision.

Legacy ATM and TDM Transport over Converged Packet Networks

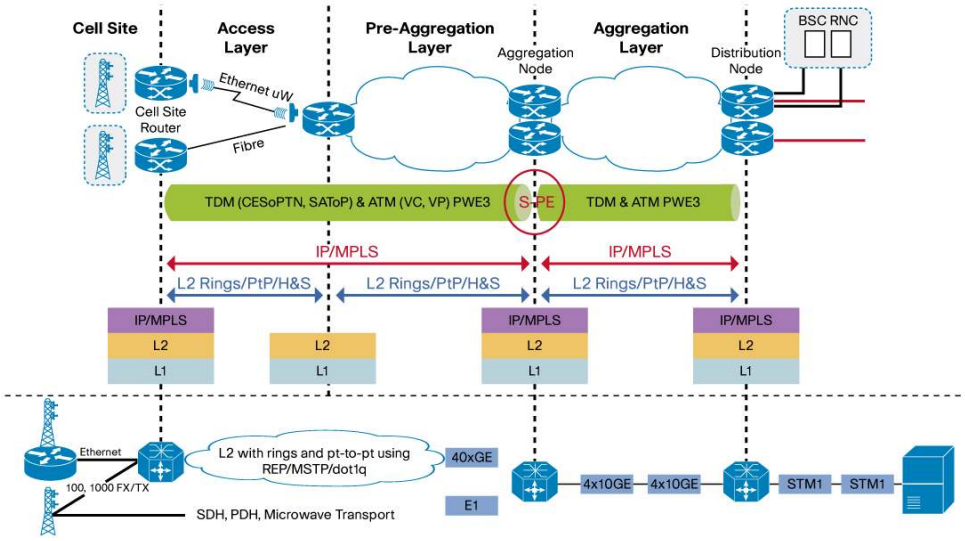
Transport convergence at a packet layer allows a flexible and scalable approach and achieves the desired single-network requirement. The statistical multiplexing gains can significantly reduce the aggregate bandwidth demands that are especially important for the economics of HSDPA, eHSPA, and LTE deployment. However, a move from SDH towards an IP/Ethernet-based architecture must address the bandwidth issue. In addition, the transition requires extra consideration in the areas of resiliency, OAM, QoS, synchronization, and so on. If legacy interfaces (TDM or ATM) must be supported, across a packet-based network, then technologies like TDM over MPLS (SAToP, CESoPSN) and ATM over MPLS (ATM PWE3) will need to be considered. This will incur extra costs, because legacy interfaces will need support on the packet platforms that have been primarily designed for Ethernet support. The pseudowire set of standards will also introduce bandwidth inefficiencies, because with any form of emulation, there is a large transport overhead. Using ATM PWE3 circuits for UMTS does enable statistical multiplexing gains, but using SAToP or CESoPSN for GSM does not. CESoPSN does allow a more efficient use of transport resources than SAToP, because only preconfigured 64-kbps timeslots are transported.

In Europe, some operators see cost justification in transporting TDM or ATM services over a converged network. This could be part of a long-term strategy or may be a transitional stage towards an all-IP infrastructure. The Cisco® Unified RAN Backhaul (C-URB) achieves this goal by extending Cisco IP network intelligence from the current core network out to the edge through transport of all RAN traffic over pseudowires (PWE3). On closer examination of the Cisco Unified RAN Backhaul solution, it is seen that the 2G/3G cell sites connect through a cell site gateway (CSG) up to a pseudowire head-end router and creates an MPLS RAN that uses PWE3 (CESoPSN, SAToP, ATM virtual channel and virtual path) transport circuits with active or standby pseudowire protection. For GSM base stations, the aggregation network implements CESoPSN and SAToP pseudo wires. For ATM-based UMTS NodeBs, the aggregation network implements virtual path or virtual channel AAL0 and AAL5 pseudowire (ATM Adaption Layer and ATM PWE3) transport.

In Europe, often a number of points of presence (POPs) exist between the cell site router and the pseudowire head-end (the router that terminates the pseudowire and hands off the traffic to the RNC or base station controller [BSC]). Architectural choices depend on whether the intermediate POPs are Layer 3/MPLS aware or Layer 2 aware. If the intermediate hops are Layer 3/MPLS aware, the network will use MPLS switching all the way from the cell site router to the head-end, with normal Layer 3/MPLS convergence and resilience techniques (MPLS fast reroute Interior Gateway Protocol [IGP] fast convergence, Border Gateway Protocol Pre-fix Independent Convergence [BGP PIC], IGP Loop-Free Alternates [LFAs]). When the intermediate hops are only Layer 2 aware, the next steps depend on whether hub-and-spoke, point-to-point connections, or ring topologies exist. For Layer 2 rings, intelligence is required to prevent Layer 2 loops (that is, Multiple Spanning Tree [MST] protocol or Cisco Resilient Ethernet Protocol).

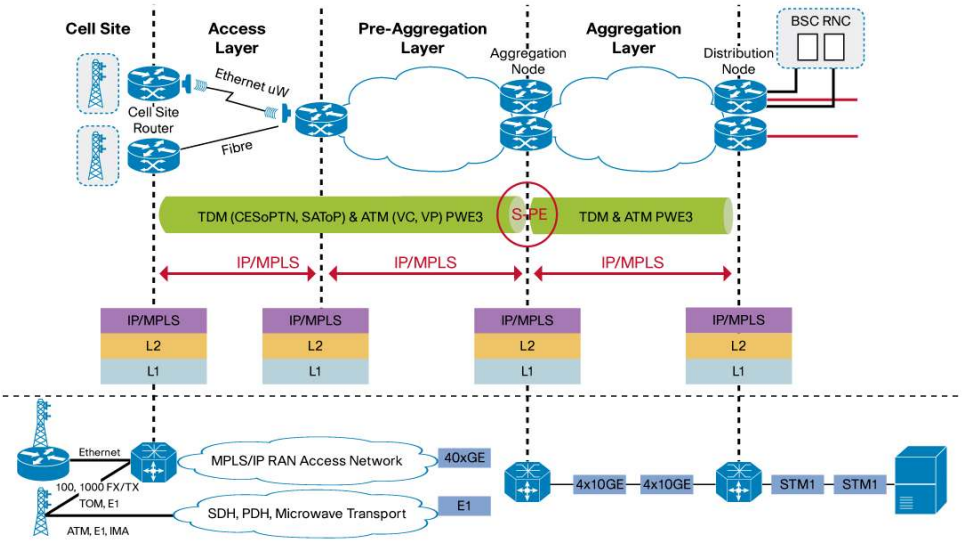
In Figure 1 there is an assumed intermediate Layer 2 POP, and there are rings in the pre-aggregation part of the network. If we consider the Cisco Resilient Ethernet Protocol topology control in the pre-aggregation layer, the MPLS/IP RAN built over Ethernet-bridged infrastructures (physical fibre rings or microwave rings) can rely solely on the Layer 2 topology protection. This means that the Layer 2 protection scheme will automatically provide protection for the Label Switched Path (LSP) and ATM/TDM pseudowires. Because we are relying on Layer 2 convergence techniques, we can build the MPLS/IP RAN on static routes between cell site routers and the aggregation nodes. Static routes are only required between the cell site routers and the aggregation nodes to enable the MPLS LSPs and the PWE3 segments.

Figure 1. TDM and ATM PWE3 Backhaul with Layer 2 Interworking



Some service providers have little experience of Layer 2 Ethernet technologies or believe there is increased operational complexity caused by the Layer 2 control protocols or more specifically, its integration with MPLS/IP in the aggregation nodes. For such providers, MPLS IP RAN redundancy can rely on IGP/LDP or MPLS traffic engineering (see Figure 2). The MPLS/IP RAN IGP (for example, OSPF from the aggregation router to the cell site router) can be configured with Bidirectional Forwarding Detection (BFD) support to provide end-to-end failure recovery in this scenario. In Europe also, operators want to support native MPLS switching on all POPs from the CSG to the aggregation router as seen in Figure 2. This architecture allows a common convergence mechanism from end to end, including pseudo-wire redundancy, MPLS fast reroute (FRR), IGP, and LDP fast convergence. Many mobile service providers prefer this design, because they have a lot of knowledge of MPLS and Layer 3 deployments and have not implemented or required Layer 2 Ethernet technologies as seen in wire-line environments today.

Figure 2. TDM and ATM PWE3 Backhaul over MPLS



The networks can also be segmented in terms of the IGP/LDP domains by using the switching provider edge (S-PE) capability on the aggregation nodes. In essence, this technique implements multi-segment pseudowires (MS-PW). This design allows static routing in the RAN access for simplicity, while using the dynamic IGP capabilities in the

core MPLS/IP domain. This design also allows different IGPs to be used across the Radio Access and MPLS/IP networks, permitting better overall scalability. In addition, dynamic IGP helps in failure segmentation and isolation, especially considering several RAN infrastructure aspects that can result in IGP instability with technologies such as xDSL or TDM/Ethernet microwave.

UMTS IP NodeB Transport over Converged Packet Network

The initial UMTS NodeBs made use of legacy ATM interfaces only. The initial evolution towards the all-IP vision started with NodeBs supporting an on-board pseudowire capability. This technique did not gather much traction in the European marketplace, and radio vendors are currently not pushing this solution. There is also support for a hybrid mode on the Node Bs, where the HSDPA traffic is offloaded through an IP/Ethernet interface, and the remaining traffic traverses the ATM interface. Over the last 12 to 18 months, some prominent European radio vendors have deployed IP NodeBs and IP RNCs with all traffic traversing the Ethernet interface only. Importantly, each NodeB is an IP host (statically configured IP address). The relationship with the RNC is still very much a connection-oriented and one-to-one relationship. This evolution is completely different from a wire-line environment where the IP Digital Subscriber Line Access Multiplexers (DSLAMs) and Multi-Service Access Nodes (MSANs) are Layer 2 devices and switch Layer 2 packets. In the European market, Ethernet microware deployment seems to accompany IP NodeB deployments and provides a point-to-point Ethernet (or optionally, ring) access network.

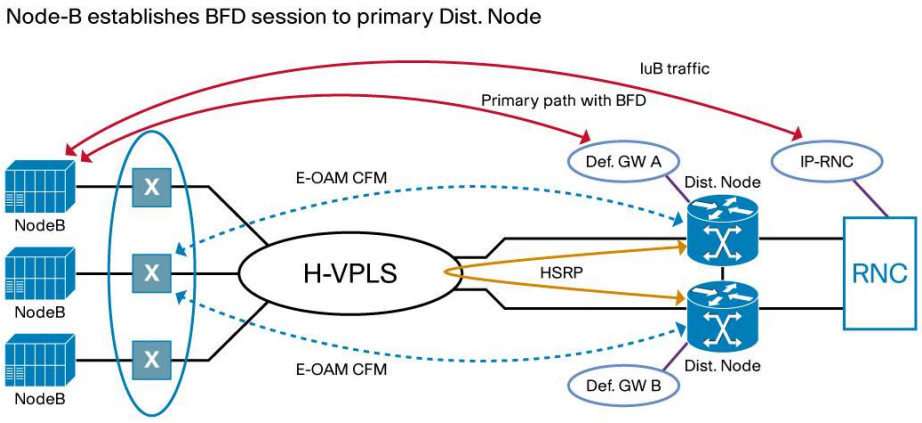
Initial discussions and testing have indicated that there are no plans to support dynamic routing protocols or MPLS on the IP NodeBs in the short to medium term. Limitations on the number of IP addresses and static routes supported should improve in later releases. Support for triggering mechanisms like BFD is challenging and needed for end-to-end resiliency. There is little support for EOAM (Ethernet Operations, Administration, and Maintenance) capability (CFM [connectivity fault management], 802.3ah, and Y.1731) for fault isolation.

In relation to deployment options that have been used for IP NodeB, there are two main technical operating models considered. The first model is based on Layer 2 VPN technologies and could be either an E-Line (point-to-point), E-tree (point-to-multipoint) or E-LAN (multipoint-to-multipoint) service. The second model uses Layer 3 or MPLS VPN. We will discuss both models in more detail in the next section.

Layer 2 VPN Deployment Model

Initially, European operators often used Layer 2 VPNs for connectivity between the IP NodeB and the RNC. The connection between the eNodeB and the RNC acted as a point-to-point connection, and an appropriate solution was a simple Ethernet pseudowire (E-line). As stated previously, NodeBs tested so far have only been able to support a single static route or single default gateway. The static route in this case points at the distribution layer (See Figure 3). Because this is a centralized solution, it is important that there are redundancy options at the distribution layer.

Figure 3. IP NodeB Layer 2 VPN Deployment Option



Currently, Hot Standby Router Protocol (HSRP) and Virtual Router Redundancy Protocol (VRRP) seem preferable, offering node redundancy with the static route configured on the IP NodeB using the HSRP/VRRP Virtual IP address. This solution mandates that an E-tree or E-LAN service (Virtual Private LAN Service [VPLS] or Hierarchical Virtual Private LAN Service [H-VPLS]) is required, because a Layer 2 path must exist between the distribution nodes in order for HSRP/VRRP to function correctly. Another option is providing a single E-line service from the IP NodeB to each distribution node, but an additional Layer 2 path must connect the distribution nodes. This approach addresses resiliency in the uplink direction, but we must consider the downlink direction also.

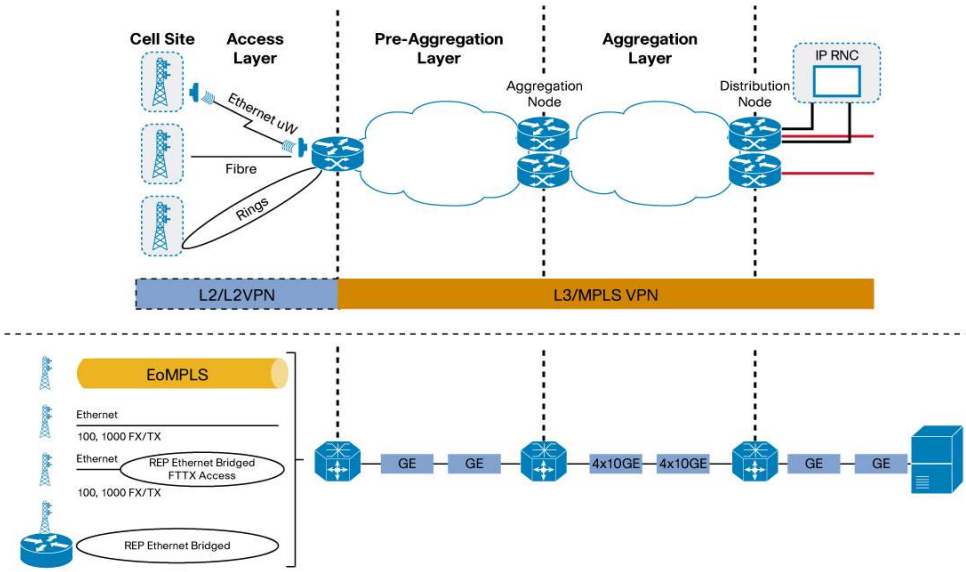
In the downlink direction, we must consider end-to-end resiliency. Certain outage types in the Layer 2 VPN domain may not be relayed quickly to the distribution nodes, which can result in traffic black holing (lost packets) in the downlink direction. Mechanisms such as BFD can help identify a wide range of end-to-end issues and trigger a forwarding change in the distribution node. Ideally, it is best to implement this fast detection mechanism down to the IP NodeB, but this has not been possible in all cases. A trigger for convergence, including some OAM features such as CFM, is a possible answer. Figure 3 highlights the resiliency mechanisms currently needed in these environments.

Cisco has seen the use of Layer 2 VPNs in operators that have had to use third-party networks for connectivity in the RAN. The third-party network could include an incumbent parent company or in fact, complete outsourcing. In these cases, operators often do not want the third-party supplier to interwork with their routing setup, as would be the case with Layer 3 or MPLS VPNs. Instead, these operators often prefer tunnelling the traffic with Layer 2 VPNs and providing the routing capabilities in their own site to keep overall control. In some deployment, the operator already supports TDM and ATM PWE3 and would prefer to use Ethernet pseudowires for IP NodeB backhaul as well. Ethernet pseudowires will be either implemented in an existing cell-site router or on the pre-aggregation node.

Layer 3 MPLS VPN Deployment Model

The second option, gaining approval in the last 6 to 12 months, makes optimal use of the IP NodeB acting as an IP host supporting static routing. The solution, outlined in Figure 4, distributes the IP/MPLS capabilities out to the edge of the network. This will allow full dynamic routing capabilities out into the pre-aggregation and aggregation layers.

Figure 4. IP NodeB Layer 3/MPLS VPN Deployment Option



This design gives the additional redundancy capabilities of MPLS or IP with similar convergence techniques from end to end. This solution also offers the possibility of supporting MPLS VPNs where multiple different traffic types dictate a need for virtualization or isolation. The solution is also an efficient backhaul option with little transport

overhead, because there is no emulation. The transport overhead seen with the PWE3 technologies can produce inefficiencies in the order of 200 to 300 percent with small packet sizes. The redundancy options available at the pre-aggregation edge could include HSRP/VRRP, offering node redundancy. Many European customers employ a single non-redundant circuit in the access (70 to 80 percent of all access circuits). This would be a single point of failure in any case, and therefore node redundancy at the pre-aggregation level may be unnecessary, making the design easy to provision and monitor.

Also, if a fast convergence triggering mechanism such as BFD is required, the design will scale better in distributed environments than in centralised environments. The support of time-based triggering mechanisms like BFD or CFM will always present a scaling issue in a centralised environment because of high CPU utilisation.

There are some examples in Europe where MPLS is extended to the cell site, normally because a cell site router initiates TDM and ATM PWE3s for GSM and UMTS traffic.

IP NodeB Deployment Conclusions

Initially, operators were in favour of Layer 2 VPN solutions and specifically E-line (point-to-point pseudowires) for connectivity from IP NodeBs to the IP RNC, as they believed they needed point-to-point connectivity. However, when matters such as redundancy and scalability were considered, E-tree and E-LAN services were actually required, meaning that the solution was becoming more complex and less controlled than initially thought. Furthermore, for full resiliency end to end, a triggering mechanism such as BFD was required, and there can be issues with scaling such solutions in large deployments. Overall, the level of complexity has increased in the Layer 2 VPN deployments in order to support better resiliency and greater scale. Recent deployments favour using MPLS VPN as far into the RAN as possible as possible, because MPLS VPN offers common convergence and resilience techniques, good virtualisation and isolation, and simple integration from access (access links are a single point of failure in most cases) and better scale for triggering mechanisms when compared with centralised implementations.

LTE/EPC Transport over Converged Packet Network

The LTE/EPC evolution is about evolving the radio and core networks towards an all-IP architecture.

Figure 5. LTE/EPC Reference Architecture

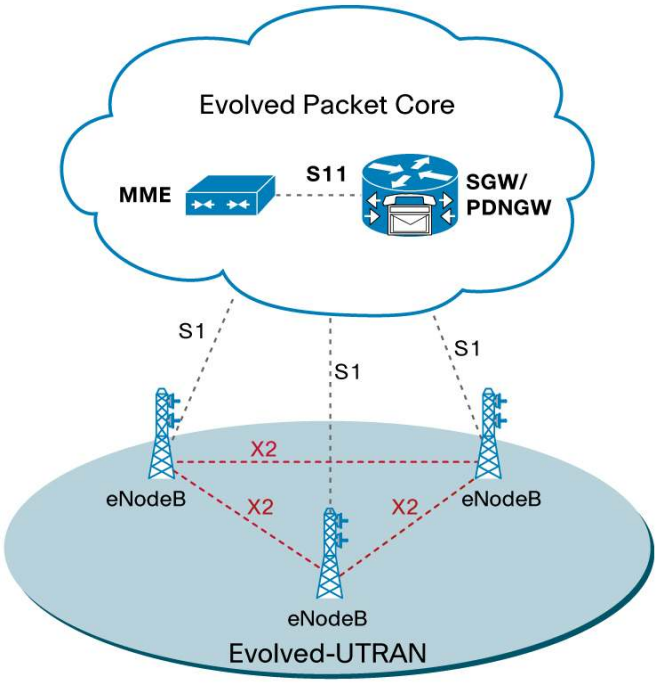


Figure 5 highlights the 3GPP-based reference architecture. The radio technology will change from Wideband Code Division Multiple Access (WCDMA) to Orthogonal Frequency Division Multiple Access (OFDMA), which will result in greater bandwidth and speeds. The flattening of the architecture (removal of the RNC) will result in greater intelligence in the eNodeB. Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) is the official 3GPP name for the radio access network of LTE. The X2 interface between eNodeBs will carry control plane (X2-c) and user plane (X2-u) traffic. The core network is now less hierarchical and will contain control plane elements (mobility management entities [MME]) with S1 control plane (S1-c) traffic and user plane gateways (serving gateways [SGW]) with S1 user plane (S1-u) traffic.

LTE Factors for Consideration with Underlying Transport Network

The LTE architecture introduces additional requirements on the underlying transport network as highlighted in the following sections.

Flattened Mobile Architecture

The traditional mobile infrastructure is very hierarchical with connection-oriented service requirements and one-to-one relationships (that is, IP NodeB has a one-to-one relationship with the RNC). The LTE enhanced NodeB (eNB), now part of the IP infrastructure will have a one-to-many relationship with the core gateways, SGWs, and MMEs. This implies that the underlying infrastructure must offer this capability in a scalable and secure manner.

X2 Interface

The X2 interface is a direct communication between the eNodeBs. There was never direct communication between radio base stations (BTS, NodeB) prior to LTE. This interface will be used for control plane and bursts of user plane traffic during handover. There is also provision for an S1-based handover but is only seen as a fallback option when the X2 interface is not available. Current estimates indicate that the combined X2-c and X2-u traffic could be between 4 and 10 percent of the core-facing bandwidth (S1-u) and the delay should be less then 30ms. This traffic is of the utmost importance and from future releases (LTE Advanced), it is apparent that more user plane traffic will traverse this interface. Also in Release 10, there will be stringent latency requirements necessary to implement features such as collaborative Multiple Input Multiple Output (MIMO). Figures in the region of 10ms are currently being considered.

Distributed Architecture

The LTE architecture, compared to other architectures, provides a simpler, less hierarchical model with the capability of simplistically distributing the core gateways. In Europe, there has been much interest in distributing the user plane gateways (SGWs and PGWs) for a number of reasons:

- Bandwidth: Some mobile service providers have determined that the bandwidth increases introduced in LTE will massively increase their core bandwidth. In one example, the core bandwidth requirements will increase to 130 Gbps in 2012, based on estimates (Current core bandwidth requirements are less then 40 Gbps). In this example, distributing to 12 sites from the previous four core sites avoids upgrading the underlying optical network.
- Traffic Offload: Some operators are examining the capability of offloading specific traffic types as early as possible in the backhaul infrastructure (also referred as Selected IP traffic offload in 3GPP). Operators do not see a value in carrying specific traffic types across core bandwidth. In fact, the operators may be adding little value and so want to hand the traffic over to a third party as soon as possible.
- Video Optimisation: Some operators are carrying large amounts of unicast video, and this accounts for a high percentage of their total traffic, even 70 percent. The distribution of the gateways allows operators to use technologies such as caching, offload, and local insertion to save on core transport costs. It is worth noting

that the degree of distribution is very important. An example of this would be with very distributed caching that can result in lower cache hit ratio and hence requires larger caching capacity.

Traffic Types

There are several types of traffic supported from the eNodeB. Each could have different transport, connectivity, and security requirements and will be directed toward different parts of the network. The types of traffic include:

- S1-u traffic destined for the SGW
- S1-c traffic destined for the MME
- X2-u and X2-c traffic destined for other eNodeBs
- OSS (operations support system) traffic destined for core applications that provide fault, configuration, and performance management
- Network Synchronisation traffic

Network Security and Authentication

LTE/EPC is about evolving to an all-IP architecture, and this change provides many advantages in the areas of scalability, availability, flexibility, and less hierarchy with direct connection from the radio nodes to the core components. This evolution does introduce some security issues, because now breaches and infiltrations may be possible from the access network. These breaches that were never seen in previous mobile architectures and could affect the core gateways directly. For this reason it is very important that a mutual authentication scheme is in place to make sure that the eNodeBs are legitimate and also that the network to which eNodeB connects is legitimate (hence mutual). Importantly, the backhaul network is now a carrier Ethernet environment with hundreds or thousands of end users (eNodeBs) who may have varying levels of security. While the network may be private, it is essential to implement all network security features as if building a public network and to choose a transport technology that is most suitable to fulfil this requirement. It is important that the transport technology chosen provides the maximum security possible between eNodeBs. Placing a lot number of eNodeBs in a large L2 domain has already resulted in Distributed Denial of Service (DDoS) attacks. Current investigations explore the capability to extract the IP address of neighbouring cell sites through Automatic Neighbour Relation (ANR) messages for use on dynamic ACLs that will only allow communication between defined neighbouring cell sites.

IPsec Requirements

Prior to LTE, end-user traffic would only be decrypted in the core components (RNC for 3G or SGSN for 2G) of the network, which means that all traffic was encrypted when traversing the less secure or third-party networks (unless roaming). In an LTE deployment, the user equipment-to-MME signalling traffic is encrypted. In the 3GPP standard (33.401 Section 11 and 12), there is a requirement to encrypt both signalling and data traffic from the eNodeB (towards the core gateways such as the SGW and MME) when using an untrusted network. However, there is a provision not to provide encryption when the network is considered secure. Similar requirements apply to X2 (control and user). In Europe, an untrusted network is deemed to include such technologies as SDH, PDH or Ethernet Microwave, third-party fibre, hosted or managed last-mile connectivity. This requirement could mean that a security gateway may need to be positioned within the transport network for X2 and S1 traffic. The security gateway concept has led to other areas of discussion including, location of gateways, integrated or standalone gateways, network resiliency options with IPsec, scale and number of IPsec tunnels, key management and IPsec overhead.

IPv6 Requirements

The LTE 3GPP standards contain very detailed information on the support of IPv6 and IPv4 from both host and transport points of view, with a full array of tunnelling options as well (IPv4 over IPv6 and IPv6 over IPv4). There is little doubt that IPv6 will become a major design consideration during the lifetime of LTE/EPC deployments. Transitional technologies will need to address the period of time when both IPv4 and IPv6 coexist. A 3GPP study

item (TR 23.975) is looking at IPv6 migration guidelines. While the core gateways (PDN gateways) will need to support some of the advanced v6 capabilities (Gateway-initiated dual-stack lite), the underlying network will also need to support both IPv4 and IPv6. There will possibly be a need for carrier-grade Network Address Translation (NAT) capabilities for this transition, and their location in the network will depend on whether a centralised or distributed architecture is deployed.

QoS Requirements

In existing 3G networks, the RAN backhaul presents a challenge for congestion avoidance and the differential treatment of different traffic types or user sessions. The LTE evolution does introduce new concepts, including:

- QoS Class Identifier (QCI): Scalar that controls bearer level QoS treatment; the current specifications have defined 9 QCI values (3GPP TS 23.203).
- Guaranteed Bit Rate (GBR): Bit rate that a GBR bearer is expected to provide
- Maximum Bit Rate (MBR) Limits the bit rate that a GBR bearer is expected to provide
- Allocation and Retention Priority (ARP): Controls how a bearer establishment or modification request can be accepted when resources are constrained

Each QCI corresponds to different traffic types (voice, video, and so on) and will be categorised with a different resource type (GBR or non-GBR). LTE allows the identification of different traffic types, identification of priority, and the decision about whether to reject the bearer request during resource constraint and then treat traffic in a differential manner. While the LTE standards have made improvements from the previous releases by simplifying the overall QoS mechanism, there are still area's that need addressing which include:

- The standards assume that the underlying network is not contended, which is a major issue with IP/Ethernet deployments. Today's networks are very dynamic and the available bandwidth is changing (consider Adaptive Modulation and Coding [AMC] with Ethernet microwave).
- Feedback mechanisms are available to inform the mobile packet core when there is congestion in the radio network. There are no such mechanisms to inform the transport network of issues and hence packets will continue to be forwarded by the transport network to the eNodeB even under heavy radio congestion. The transport network could prioritise and selectively buffer or drop traffic if there was awareness of the congestion. HSPA CAC (Cell Access Control) includes transport congestion in its mechanism (studied in 25.902 and defined in HSDPA) but this isn't defined in LTE.
- Issues occur with mapping of QCI parameters (nine values) in Layer 2 environments where there are insufficient 802.1p bits. While the standards define nine values, most likely more values will be needed for unspecified traffic types (synchronisation, OAM and so on).

The underlying transport will need to support traffic prioritisation, dual-priority and low-latency queues for 3GPP compliance. Hierarchical QoS (H-QoS) is needed to support the GBR and MBR classification types and also so that important traffic types can be prioritised for multiple different cell sites under congestion conditions. H-QoS is important to manage contention in the last mile, by representing last-mile available bandwidth at the aggregation and distribution level. Work is ongoing in relation to the bandwidth feedback mechanisms, and protocols such as Access Node Control Protocol (ANCP) are under consideration.

Multicast Requirement

Many mobile operators are looking at means to deliver multicast services optimally across their existing networks. Mobile standards have not really addressed this area in a scalable fashion. Clearly, other services could use a multicast type delivery model; these include phone patching, security or software downloads, gaming, and so on. LTE and future releases will introduce eMBMS (enhanced Mobile Broadcast Multicast System) with multicast and broadcast modes of operation. Regardless of the modes used, support of Source Specific Mode (SSM) and Internet

Group Management Protocol version 3 (IGMPv3), and Multicast Listener Discovery version 2 (MLDv2) snooping on the backhaul network is needed.

Synchronisation Requirements

The LTE is primarily concerned with positioning an all-IP solution, and the reliance on legacy networks and infrastructure will be minimal. Capabilities such as SyncE (Synchronous Ethernet) and packet-based capabilities such as IEEE 1588 version 2 and Network Time Protocol (NTP) are supported to provide network synchronisation over the existing transport infrastructure. It is important to remember that LTE may introduce stringent parameters and support for both frequency and phase synchronisation may be required. Time Division Duplex (TDD) technologies and LTE Multi-Media Broadcast over a Single Frequency Network (MBSFN) are examples of when phase synchronisation is required. Only certain protocols such as IEEE1588 version 2 have the capability to provide phase synchronisation.

Network Convergence

The LTE standard makes use of the GPRS tunnelling protocol (GTP), along with Stream Control Transmission Protocol (SCTP) for user and control plane connectivity between the LTE/EPC node components (eNodeB, MME, SGW, and PGW). The standard only mandates end-to-end connectivity checks with variable intervals and has not specified how the overall network will converge in an optimal fashion. SCTP has built-in recovery techniques and requires path diversity for switchover at about 700 msec in 3GPP R4 networks. This presents issues when you consider that this protocol needs to be supported at the eNodeB, because there is a high probability that path diversity will not be present. GTP has inherent path management messages and timers (Echo Request Interval/Echo Response Interval), but the intervals are in the order of tens of seconds, which does not allow optimal convergence. The underlying transport network will provide optimal convergence at an IP layer with mechanisms such as VRRP/HSRP, BGP Prefix-Independent Convergence (PIC), MPLS FRR, IGP Fast Convergence, IGP Loop-Free Alternates (LFAs), and BFD.

RAN Sharing

European mobile operators have acknowledged that reducing the cost per bit in their backhaul is now their primary objective. Recent commentary indicates that a means being considered to achieve this objective is by implementing RAN or E-UTRAN sharing between different operators. In LTE, E UTRAN sharing is an agreement between operators and shall be transparent to the user. This Multi-Operator Core Network (MOCN) configuration as defined in TS 23.251 is supported over the S1-c and S1-u reference points. This implies that an E UTRAN UE needs to be able to discriminate between core network operators available in a shared radio access network. An E UTRAN sharing architecture allows the operators to not only share the radio network elements, but may also share the radio resources themselves. European operators are currently considering the sharing of resources down to the cell site. This implies that the underlying transport must be able to identify, isolate and provide secure backhaul for different operator traffic over a single converged network. Cisco is working with operators on a means to provide dynamic service creation based on multiple different first signs of life (FSOL). An example would be where traffic on a specific VLAN would initiate a Radius Request towards an AAA (Authentication, Authorization and Accounting) server that would return information to dynamically setup an Ethernet PW towards the core of the network.

Fault Isolation/Identification and Fast Convergence Triggering

As stated before, the LTE standard only mandates end-to-end connectivity checks with variable intervals. This does not help with fault isolation, identification, and triggering. The underlying network will be responsible for such capabilities, and there have been proposals submitted for per-link and segment checks. The proposals include:

- Layer 1, Ethernet: IEEE 802.3ah OAM
- Layer 2, Ethernet: ITU-T Y.1731/IEEE 802.1ag

- Layer 3, IP: IETF BFD (single hop and multi-hop)

Latency Requirements

Latency is a key requirement of the LTE/EPC architecture with the goal to achieve a 10 to 20 ms one-way delay that is an improvement when compared to 100 to 200 ms in release 99 architectures. The high peak rates and short latency of LTE allow real-time applications such as gaming and IPTV. Latency, jitter, and delay parameters must be set for specific interfaces (X2 when supporting some advanced features such as collaborative MIMO). It is imperative that the overall design accounts for these factors and does not introduce excessive latency due to encapsulation (IPsec or unnecessary tunnelling that result in suboptimal routing).

Traffic Separation and IP Addressing Models at the eNodeB

One of the most important considerations is how the eNodeB will present different traffic types to the backhaul network, because this may be of the utmost importance in determining how the traffic is backhauled to the correct destination. The traffic separation options at the eNodeB are as follows:

- There is no VLAN support, and all traffic is forwarded out the same port.
- Traffic divides into two VLANs at the eNodeB. The first VLAN is for X2 traffic that runs directly between the eNodeBs. The second VLAN carries all traffic destined for core applications. This traffic would include S1-u interface, S1-c interface, OSS traffic, and so on.
- Each traffic type is separated and placed into individual VLANs at the eNodeB.

Radio vendors are showing most interest for the second option (two-VLAN Support) because this method represents a good compromise with minimal segregation, because it places traffic that is destined for a similar part of the network into the same VLAN. It also means that there will be no scaling issues in relation to IPsec or VLANs (a minimal number of IPsec tunnels). This option does produce, however a significant issue, because of the difficulty in differentiating and identifying the different traffic types within the same VLAN. This is important when traffic forwarding goes to the correct end device (SGW, MME, or OSS server}, possibly through different transport types.

The last option requires the separation of traffic types according to VLAN and gives excellent traffic separation and a means of identifying traffic types towards the core. The technique does present an issue when it comes to VLAN scaling in the network, because each eNodeB could require up to five to six VLANs. More importantly, if each traffic type has its own VLAN, each eNodeB must support five or six instances of IPsec. Radio vendors indicate that this could affect the performance of the eNodeB. Also worth consideration is that the security gateway will need to terminate five to six times more IPSec tunnels, which could also affect the scaling of this platform. For these reasons, it is apparent that shared VLANs will need to be considered for further discussion.

If shared VLANs are considered, then we also need to decide how to identify different traffic types. Here are some suggestions about how this could happen:

- Traffic types could be marked by the eNodeB, but this scenario would assume that all traffic belonging to a traffic type would be treated in the same manner. Because we could be traversing some Ethernet domain, then the number of 802.1p bits supported imposes a restriction also.
- Use of the destination IP address to identify traffic can be complex and is prone to security attacks, because a well-known destination IP address can be spoofed.
- Use of the source IP address to identify traffic is also an option but this would require that each traffic type be given a separate IP address by the eNodeB. This method could lead to complex IP address planning and address exhaustion.
- Use of IPSec tunnels or child associations to identify different traffic types is another option.

Some believe that the use of traffic marking or destination IP addresses on their own may not be sufficient to identify traffic types. Currently, all different traffic types would be assigned a different IP subnet (/30 proposed). Depending on the deployment model, IPSec tunnels or child associations could also be used in the security gateway as a means of identification before forwarding traffic into an MPLS VPN.

Backhaul Technology for an LTE-Based Converged Packet Network

To determine the technical merit of each architecture type, there are ongoing discussions with a number of European operators about possible LTE transport models and the different points mentioned in the preceding sections.

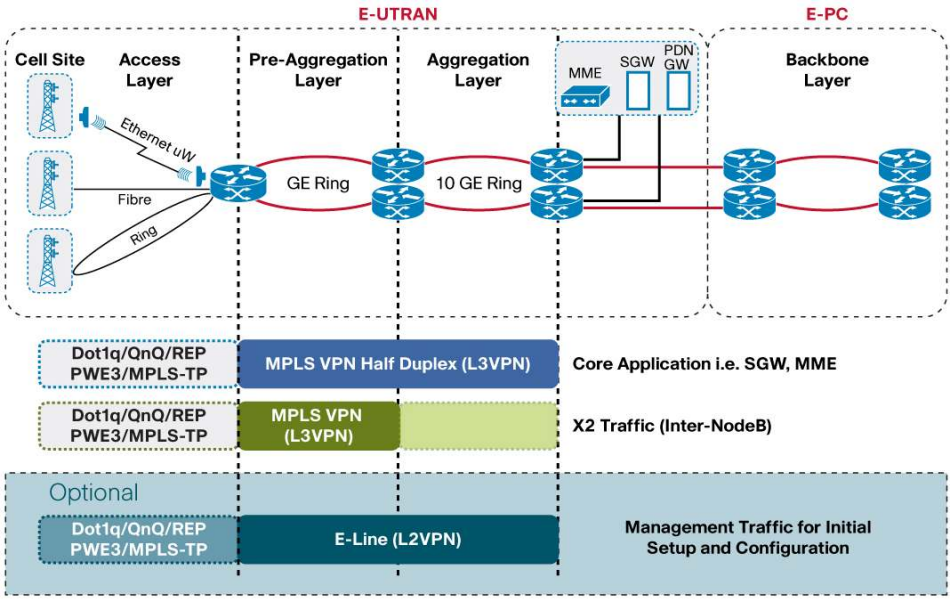
In the European market today, pushing the IP/MPLS control plane out into the RAN and choosing the best possible data plane forwarding technique are becoming a popular option. MPLS VPNs seem to be offering quite an advantage over other forwarding techniques such as Layer 2 VPNs, but the overall positioning does not rule out the use of Layer 2 VPNs when needed. The model described in the next section highlights the fact that most operators may not be able to get MPLS functionality to the cell site; there may be no active equipment on the cell site. The model shows MPLS functionality going as far as the pre-aggregation with the option of Layer 2 (point-to-point or rings)/pseudowire or MPLS Transport Profile in the access.

Various types of traffic presented from the eNodeB need individual treatment. The model described in the next section represents the most basic traffic profile model from Europe, with just three different traffic types. In some cases, there are up to six different traffic types. Other traffic types been considered include synchronisation transport, out-of-band management, and closed-circuit TV or cell site monitoring.

Layer 3/MPLS VPN Model for LTE/EPC Deployments

For the Layer 3/MPLS VPN model as outlined in Figure 6, the eNodeB traffic is separated into two VLANs, one for core applications and the other for X2 traffic. The core application VLAN must be backhauled towards the core nodes. An MPLS VPN (or half-duplex MPLS VPN) can achieve this when extended over the pre-aggregation and aggregation layers. The Cisco IOS® MPLS VPN Half-Duplex VRF (Virtual Routing and Forwarding) feature may be helpful, because some operators want to use a hub-and-spoke configuration initially for a configuration like their current one, with no local “hair pinning” and simplified VPN provisioning across the infrastructure.

Figure 6. LTE/EPC Layer 3VPN Connectivity Options



The advantage to this model for core application traffic is the flexibility of the overall architecture, which can be modified with minimal disruption. If operators can easily insert security gateways for either centralised or distributed IPsec support. This design also offers an advantage to other operators who are looking to distribute some of their core gateways (security gateways or SGW, PGW) in later phases. Cisco also uses common resiliency and availability models right through the pre-aggregation, aggregation, and core networks, which helps overcome some of the resilience issues seen in the Layer 2 VPN deployments, especially the complexity encountered when a Layer 2 VPN service must map to a Layer 3 service. A number of European operators have also determined that the use of a single technology from end -to end without interconnection can reduce operating expenses. An MPLS VPN also offers separation of different traffic types and provides flexible interaction with the security framework. Because MPLS VPN is a Layer 3 service, Layer 3 attributes can identify and forward traffic or apply different services (QoS, security, and so on). The model also provides optimal routing between nodes, which is most important between the eNodeBs; the X2 interface requires direct communication. Features such as collaborative MIMO may place strict latency, jitter, and delay characteristics on this interface in later releases. The introduction of tunnelling in a hub-and-spoke model will incur suboptimal routing and will introduce unnecessary latency (this is also critical when considering the IPsec implementation options).

The X2 traffic is routed through the pre-aggregation layer using MPLS VPNs. The principle advantage of this method is the optimal routing; in this way, the eNodeBs are communicating directly with each other through distributed intelligence. This model optimises latency and increases bandwidth efficiency when compared with a centralised approach. The MPLS approach provides the ability to control and manage accessibility between the eNodeBs through features such as ACLs (access control lists), route summarization, and so on. Current investigations explore the capability to extract the IP address of neighbouring cell sites through Automatic Neighbour Relation (ANR) messages for use on dynamic ACLs that will only allow communication between defined neighbouring cell sites. The MPLS approach will also help support both the direct connectivity model and the model that traverses the IPSec security gateway.

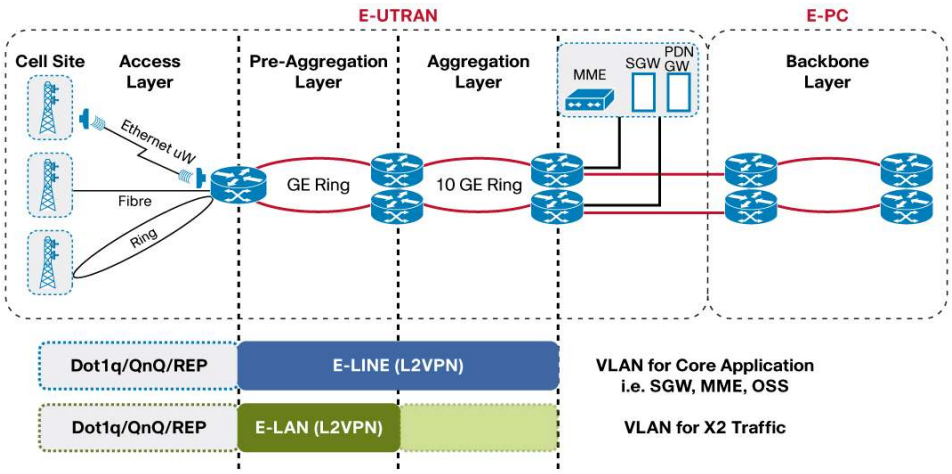
The overall philosophy would be to push the MPLS control plane as far into the RAN as possible and then choose the appropriate data plane for different traffic types. As stated above, this model allows provision of other service types over this converged network when needed. At the bottom of Figure 6, we see from the proposed model that a Layer 2 VPN could support the transport type relating to initial setup, node configuration, and software download. A tunnelled connection from the cell site into the centralised servers is required, with little interaction with the underlying network and no possibility of breaking out. This method provides a level of security and segregation from other traffic types that are classified as more trusted.

Some Operators perceive that pushing IP/MPLS and specifically MPLS VPN capabilities further into the RAN increases the complexity from a configuration and operating expense point of view. The capital expenditure for platforms supporting MPLS VPN would historically have been higher, but more low-end router and switching platforms are now supporting MPLS natively.

Layer 2 VPN Model for LTE/EPC Deployments

Using L2VPN technology only for backhauling LTE traffic is a possibility, as outlined in Figure 7. The eNodeB traffic is separated into two VLANs, one for core applications and the other for X2 traffic. The core application VLAN needs to be backhauled towards the core nodes in a point-to-point fashion. An E-Line service (Ethernet pseudowire) that can be extended over the pre-aggregation and aggregation layers achieves this backhaul.

Figure 7. LTE/EPC Layer 2 VPN Connectivity Operating Modes



The X2 VLAN will make use of the E-LAN service (VPLS), as the eNodeBs must communicate directly for cell site handover. A critical component of a handover is that the Source eNodeB is able to communicate directly with the Target eNodeB.

While this model presents a very simplistic approach, here are some considerations:

- Supporting the X2 interface by means of an E-LAN service presents an issue, because a mobile user (user equipment) will hand over between different cell sites that must communicate directly with each other. Even if the number of neighbours is low (10 to 15), the issue is that the neighbouring list will change continuously as the user equipment moves from cell to cell. There are two factors that need to be considered: First, the E-LAN domain cannot be so large that it represents a large broadcast domain and hence a security risk; second, different E-LAN domains must communicate with each other to allow handover. Some degree of X2 zoning could be done by connecting the access E-LAN services to pre-aggregation E-LAN services (Hierarchy of E-LAN services). This zoning should be constructed in such a way that cell sites are reachable whenever a cell site handover is possible.
- Using E-LAN services can result in large broadcast domains that present a major security risk, because all eNodeBs in the E-LAN domain could undergo a Distributed Denial of Service (DDOS) breach. Secondly, although the eNodeBs are present in the same E-LAN domain, we only want neighbours to communicate with each other. This segregation in an E-LAN is very difficult to realise and can only be done on a MAC layer through MAC address control access lists, which are operationally complex and not dynamic.
- There could be issues with E-LAN configuration complexity and scaling, because multiple E-LAN services must connect in the access to a hierarchy of E-LAN services in the pre-aggregation layer, to allow handover between cell sites.
- As with deployments seen today involving IP NodeB and IP RNCs, the end-to-end resiliency can present scaling issues and complexity when traversing over an underlying E-LAN, E-tree, and E-Line service. Current investigations involve some OAM protocols and mechanisms such as BFD, but there are still some unresolved scaling issues.
- Early analysis conducted on IPSec gateway placement indicated that these gateways might need to be in the transport network in the pre-aggregation or aggregation locations. IPsec termination requires a Layer 3 presence, and this would have implications on any Layer 2 VPN implementation.
- Early indications favour a more distributed approach for security gateway and PDN gateway placement in the later phases. In Europe, an operator is moving from 4 to 6 centralised sites in Phase 1 to 16 to 20 more distributed sites in Phase 2. This distribution is based solely on bandwidth requirements and an issue around

the scaling of the underlying optical network. This architecture allows the operator to adopt any offload solution that has currently been analysed. This approach would have serious effects on the way the Layer 2 VPN model can work and would result in a major redesign of the underlying transport network.

- Some proposed eNodeB authentication mechanisms, such as 802.1x, would have some issues with Layer 2 environments and will not function if there are multiple Layer 2 hops and bridge domains present within the backhaul network.
- Some proposals that promote the use of E-line services, resulting in connection-oriented and centralised backhaul models, will suffer from suboptimal routing and also the insertion of unnecessary latency, which could affect the performance of some features, such as collaborative MIMO or VoIP, in future releases. It also breaks the requirements of having an any-to-any relationship between the radio nodes and the core nodes as outlined in the 3GPP standards.
- A single VLAN with multiple traffic types will present issues when using the Layer 2 VPN backhaul model, as this service will not be able to interpret any Layer 3 attributes. The core would need to support some routing capability to allow transport towards the correct end devices that will be in different IP address subnets.

LTE/EPC Transport Conclusions

The LTE/EPC evolution is an evolution towards an all-IP architecture and will fundamentally change how mobile backhaul networks are built in the future. The availability of Ethernet-enabled NodeBs and the evolution towards LTE/EPC pushes IP awareness further into the edge of the mobile network. Mobile operators are beginning to view these backhaul networks like carrier ethernet environments offering multiple concurrent services. LTE/EPC will make demands on the underlying transport in areas such as security, IPv6, distributed intelligence, multicast, synchronisation, QoS, fast convergence, instrumentation, and management. The transport technology choices of today will be important for future evolution of the mobile architecture. The LTE/EPC evolution demands a lot of intelligence and flexibility in the underlying network. Cisco recommends a design model to support a distributed, multi-service, MPLS enabled network that offers the flexibility, scalability and intelligence to address current and future needs. This design allows the use of intelligent Layer 3/MPLS VPN technology for optimal routing, security, flexibility, and resiliency and also provides possible support of Layer 2 VPN technologies if deemed necessary for certain traffic types.

Conclusion

European mobile providers are currently experiencing large increases in mobile backhaul capacity to address their current and future service requirements. The costs and expenditures associated with providing this increasing bandwidth has not being linearly matched by revenue growth. The primary objective is to increase the bandwidth while simultaneously reducing the cost per bit. Existing TDM/ATM infrastructure will neither scale to the required bandwidth nor meet the cost reduction requirement. Recent reports[1] have shown that all operators now believe that IP/Ethernet-based backhaul is a mandatory requirement. These reports also show the growing belief that a single, converged, all-IP-based Ethernet backhaul is required, with 85 percent of respondents seeing LTE as a key driver for IP/Ethernet-based backhaul.

While there is a clear effort towards supporting IP/Ethernet backhaul, there are ATM/TDM-based requirements for GSM and 3G that need support. Cisco believes that a converged architecture is essential where the mobile backhaul solution simultaneously supports ATM/TDM and Ethernet requirements. The ATM/TDM requirements can be met through of pseudowire technology (PWE3), and the current Ethernet requirements can be supported by means of Layer 2/Layer 2 VPN or Layer 3/MPLS VPN technologies. The transport solution chosen for the current Ethernet requirements must allow for future scaling, simplistic and optimal resiliency, and optimal support for future technology such as LTE. Current Layer 2 VPN-based deployments for 3G-based IP NodeBs are showing issues

[1] Infonetics Research: IP/Ethernet Mobile Backhaul Strategies: Global Service Provider Survey- March 2010

regarding scale and optimal resiliency. A more distributed Layer 3/MPLS VPN approach is showing better resiliency and scale and better support for the service requirements of the evolving mobile standards.

The LTE/EPC evolution is an evolution towards an all-IP architecture and is seen as one of the important incentive for the adoption of IP/Ethernet in the backhaul. The LTE/EPC evolution will push more intelligence further out into the RAN and onto the eNodeBs with direct interfaces (X2), and requires an any-to-any relationship between the radio and core nodes. These changes make demands on the underlying transport in areas such as security, IPv6, distributed intelligence, multicast, synchronisation, QoS, fast convergence, instrumentation, and management. Cisco recommends a design model that supports a distributed, multiservice, and MPLS-enabled network. This design allows the use of intelligent Layer 3/MPLS VPN technology for optimal routing, security, flexibility, and resiliency but also provides the possibility of supporting Layer 2 VPN technologies if deemed necessary for certain traffic types.

For More Information

For more information, email spmobility_europe@cisco.com or contact your local Cisco account representative.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-613002-00 08/10