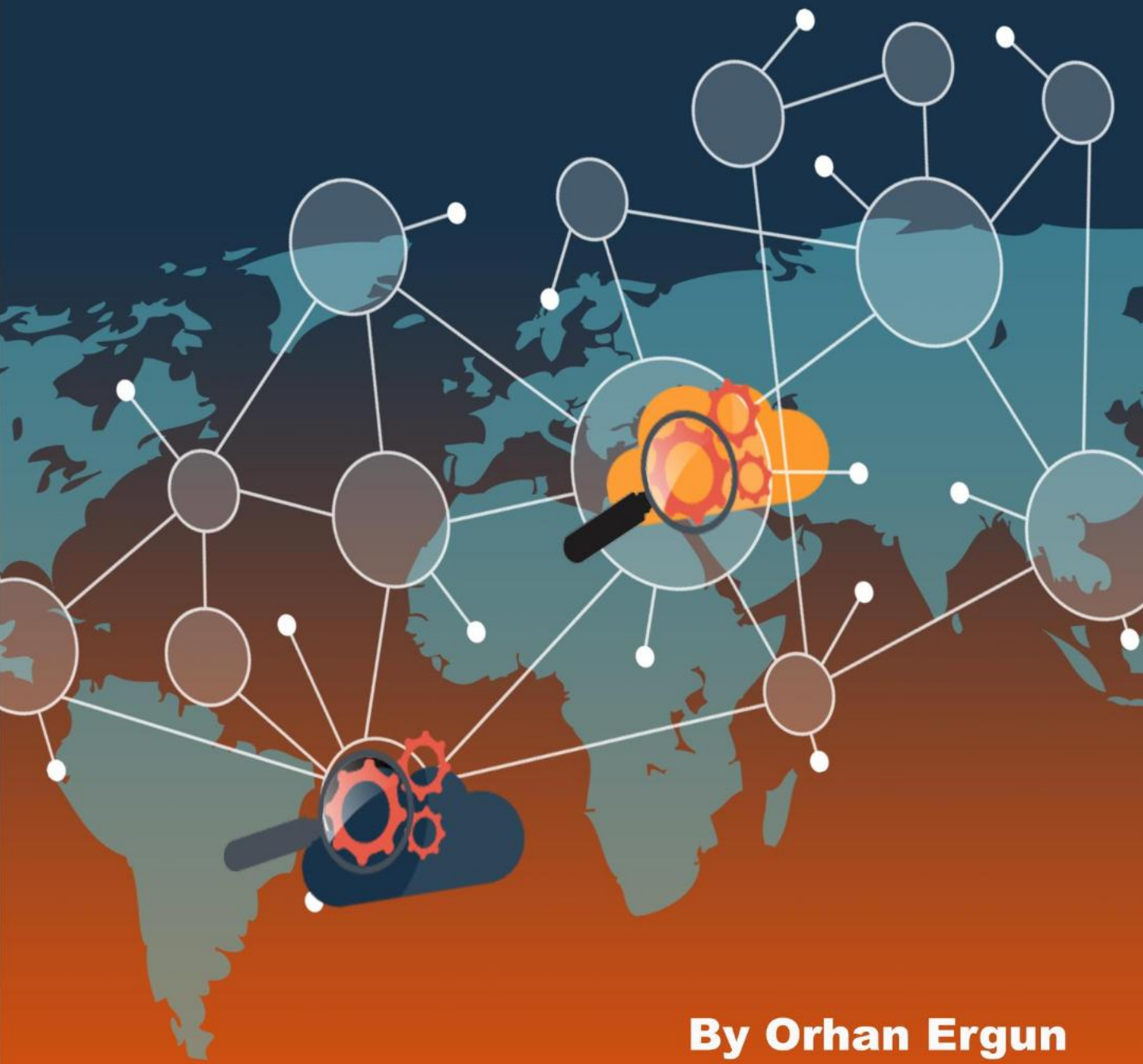
Service Provider Networks

Design and Architecture Perspective



By Orhan Ergun

https://t.me/learningnets

Service Provider Networks

Design and Architecture Perspective

First Edition

by Orhan Ergun

https://t.me/learningnets

Copyright

Orhan Ergun © 2019

All rights reserved.

No part of this publication may be copied, reproduced in any format, by any means, electronic or otherwise, without prior consent from the copyright owner and publisher of this book.

However, quoting for reviews, teaching, or for books, videos, or articles about writing is encouraged and requires no compensation to, or request from, the author or publisher.

Orhan Ergun

https://t.me/learningnets

About the Author

Orhan Ergun, CCIE/CCDE Trainer, Author, Network Design Advisor and Cisco Champion 2019.
Orhan Ergun is award winning Computer Network Architect, CCDE Trainer and Author. Orhan has well known industry certificates CCIE #26567 and CCDE #20140017.
Orhan has more than 17 years of networking experience and has been working on many medium and large-scale network design and deployment projects for Enterprise and Service Provider networks. He has been providing consultancy services to African, Middle East and some Turkish Service Providers and Mobile Operators for many years. Orhan has been providing Cisco network design training such as CCDE, Pre-CCDE, Service Provider Design and many advanced technologies for many years, and created best CCDE Training Program to share his network design experience and knowledge with the networking community. Orhan is sharing his articles and thoughts on his blog www.orhanergun.net. All the training and consultancy services related information can be found from his website. Orhan has a Training and Consultancy company located in Istanbul, Turkey

https://t.me/learningnets

About the Technical Reviewers

Batur Genc has 18 years of experience in telco and service provider industry mainly focused IP/MPLS Network and Datacenter Switching architectures. During his career, he worked in different engineering roles in operation and planning teams and managing roles in planning and architecture teams. He is currently head of IP planning in a leading telecom operator at EMEA region, Turkcell. Beside daily activities, he is focusing on new technologies such as Segment Routing, Next-Gen Datacenter architectures and Virtualization. Also, he is interested about Digital Transformation strategies and execution.
He has BSc. and MSc. degrees in electronics and Telecommunication Engineering from Istanbul Technical University. Also, held an honor degree in Executive MBA from Bahcesehir University

Seyed Hojat Fadavi is a Senior Consultant specialized in Network Architecture and Design. He has 15 years of experience in Network Technologies, Customer Interaction and Networking Products; his focus is on Network Design and Architecture. His long time experience was working in Planning, Design, Team Leading and Troubleshooting Large Scale Networks in the Service Provider, Enterprise and Data Center section. He has worked with Orhan Ergun as a Consultant and also as a Team Leader.

Rogerio Mariano is Network Planning Director at Azion Technologies and is the current Chair of BPF (Brazil Peering Forum) and has been the Chair of LACNOG (Latin American and Caribbean Network Operators Group). In ICANN (Internet Corporation for Assigned Names and Numbers), he was a Fellow at ICANN 54 in Ireland and is now an advisory member of ICANN's Latin America Strategy Group. He was a student of EGI.br and South School Internet Governance in Washington, also holds an MBA in Communications. He has 19 years of experience in Edge & Deep-Edge Caching, Network-Scale, specializes in Centralized-TE (PCE, BGP-LS and SR) and Submarine Cables and the Dark Arts of Interconnections.

Ruben Fonte is Senior Network Architect at Telecom Italia Mobile Brazil and he is the current part of Subject Matter Expert for Cisco CCIE SP program. He is Engineer and holds CCIE Routing & Switching and Service Provider certifications. He has more than 15 years of experience in Complex Service Provider Networks (Fixed Broadband and Mobile broadband), that’s included MPLS, IGP, BGP, IPv6, SDN and NFV.

https://t.me/learningnets

Dedication

I would like to dedicate this book to my two children, Efe, Amine and my lovely wife for continuously supporting me. I had to come home very late and couldn't see my lovely family for many months.

Alhamdulillah Allah gave me this opportunity to share my knowledge and experience with the people.

https://t.me/learningnets

Acknowledgements

I would like to thank some people who encourage me writing this book. Hojat Fadavi and Shirin Sobhani were the first people and since I started writing this book, they supported me in many ways.

Hojat deserves special recognition as he has been the Technical Reviewer and also helping for proofreading, drawing many of the figures for me in the book and was always passionate about this book.

I would like to thank to the companies which I provide a consultancy service. We have become a friend with you guys by the time and your encouragement will be probably helpful for many other people through this book and any future version of it.

I would like to thank my "Service Provider Network Design" course students as they asked me many times to publish the course content as a book to help many other people to understand the Service Provider business, technologies and interactions with other networks.

Though many of my social media followers don't know that I have been working on this book, their continuous encouragement and expectation from me give me confidence to produce new tools such as courses, books, videos and others. Hope you guys will like this effort as well.

Thanks to Ammar Hanon on creating a very nice front and back cover of this book.

I would like to thank my CCDE students to help me finding a name (So many suggestions we had, but choose the one Hari suggested, thank you again Hari and for continuously asking the book publication time.

Not sure if it is okay to write in the Acknowledgment section but I am sorry father, you have been sick since I started writing this book and I couldn't visit you enough during past several months but hopefully I will spend more time with you anymore.

https://t.me/learningnets

About the Book

This book will give you a High Level of overview of the Service Provider Network Design and Architecture.
It talks about the unique aspects of Service Provider networks, different types of Service Providers and the business relationships between them.
It covers the Service Providers services, different access last mile offerings and transport networks, and their subscribers and services.
Technical explanation about different types of Fixed and Mobile network services and the service provider physical locations are also explained.
You will see the Big Picture of service Provider Networks.
After understanding the Service Provider Concepts and Technologies, a fictitious National Service Provider network, named ATELCO will be introduced, to give you a more view of the technologies, protocols, services and end to end traffic flow in great detail.
And at last the Evolving Technologies in the Service Providers and Massively Scale Datacenters will be seen.

https://t.me/learningnets

Introduction

Service Provider Networks in many ways are unique networks. Many services might be serving to millions of customers, so there might be many paths between different parts of the network as you will see in the book. There are many different types of Service Providers but there is very little information about some of them. For example, Internet Service Providers, Broadband Service Providers, Transit Service Providers and Backbone Service Providers.

I have been teaching the unique aspects of Service Provider networks, explaining the services, many different access last mile offerings and transport network in my Service Provider Design Workshop courses. I have been encouraged several times by the students to write a book about the topics which was covered during the classes and this book effort started last year.

This book is organized in 9 Chapters.

Chapter 1 will start with explaining different types of Service Providers. Without going into technical details, it will explain the business relationship between different types of Service Providers and their subscribers and services.

Chapter 2 will be little bit more technical and will explain different types of Fixed and Mobile network services such as XDSL, FTTX, Cable Broadband, Fixed and Mobile Satellite, Wireless Internet Service and Mobile Broadband LTE (Long Term Evolution).

Chapter 3 will be covering the different types of Transport network fundamentals. Information in this chapters will be used in the next chapters. Fiber optic, Microwave, Comparison of Fiber and Microwave, SONET/SDH, WDM and Dark fiber will be covered. Also terrestrial and Sub Marine/Undersea Cable Systems and the components of these systems will be introduced.

Chapter 4 will be covering the physical locations where mainly Service Providers use to keep their servers, networking devices and security systems. Locations and the terminology which are used for them are unique to the Service Provider networks. POP, Meet-me room and Carrier Hotel are some examples to those places.

Chapter 5 will show the big picture of a Service Provider. Many information which was covered in the previous chapters will be helpful to demonstrate an end-to-end topology of a sample Broadband/Internet Service Provider network. The sample Service Provider in this chapter will provide XDSL Access, FTTX Access, Cable Access, Mobile Broadband, Fixed Broadband Wireless, WiMAX. In this chapter, these services will not be explained again. In this chapter you will understand how those services fit in to the end to end Service Provider network architecture.

Chapter 6 was the first topic when this book was started to be prepared. Interconnection between the networks. Service Providers have business relationship with many different types of companies. In these business relationship, they mostly connect to other Service Provider networks, Content Provider and Content Delivery Networks. These business relationships can be both

https://t.me/learningnets

Settlement based and Settlement Free Based. Many different types of Service Provider business models will be introduced in this chapter and will go into some technical details as well.

Chapter 7 A Service Provider network will be built from scratch. Services, Technologies, Protocols which you can see in the Access and Transit Internet Service Providers and LTE networks will be explained briefly in this chapter. ATELCO is a fictitious National Service Provider which has 11 million customers from Residential and Corporate segments.

Chapter 8 is explaining the Service Provider Network which was built in the Chapter 7 in detail. Presenting the alternative methods for ATELCO and explaining the technologies, protocol, services and end to end traffic flow in great details. For better understanding Chapters 7 and 8, you should first read the previous Chapters of the book.

Chapter 9 is a quick introduction to the technologies which are evolving in the Service Providers and Massively Scale Datacenters. Segment Routing, TI-LFA, EVPN, NFV, BGP in Massively Scale Datacenter Usage and Multicast BIER are the topics of this Chapter. This Chapter already gave me many ideas for the upcoming edition of this book and many other technologies which are emerging in Service Provider networks. The detail explanations for the ones in this book will be covered in the future version of this book based on the readers feedback.

You can reach me out by sending an email to orhan@orhanergun.net

https://t.me/learningnets

Contents at a Glance

Part I Service Provider Design, Architecture and Services

Chapter 1 Service Provider Types
Chapter 2 Introduction to Service Providers Network and Services
Chapter 3 Service Provider Physical Connectivity & Transport Network
Chapter 4 Service Provider Physical Locations
Chapter 5 Service Providers Modules
Chapter 6 Service Provider Interconnections and Peering's

Part II ATELCO National Service Provider Network

Chapter 7 ATELCO National Internet Service Provider Design
Chapter 8 ATELCO Network Detail Design Explanation

Part III Service Provider Evolving Technologies
Chapter 9 Evolving Technologies in the Service Provider Networks

https://t.me/learningnets

Contents

Chapter-1 Service Provider Types 1
Introduction 1
Broadband Service Provider 2
Transit Service Provider 2
Access Service Provider 3
Backbone Service Provider 4
Regional ISP 5
National ISP 6
Content Providers 7
Over the Top Providers (OTT) 8
Content Delivery Networks 9
Cloud Providers 11
Edge Computing Providers 13
Cable Access Providers 14
Mobile Operators 15
Wireless Internet Service Providers 18
Satellite Service Providers 19
Summary 21
Chapter-2 Introduction to Service Providers Network and Services 22
Introduction 22
Broadband Services 23
Fixed Broadband Service Technologies 24
DSL 24
FTTX 28
Cable Broadband 35
Fixed Wireless Service 38
Satellite Broadband 42
Mobile Service Technologies 51
LTE 52

https://t.me/learningnets

Summary 56
Chapter-3 Service Provider Physical Connectivity and Transport Network 58
Introduction 58
Fiber Optic 58
Total Internal Reflection 59
Fiber Optic Cable Installation 60
Fiber Optic Cable Types 61
Microwave 62
Microwave or Fiber, which one is faster? 63
SDH/SONET 65
WDM 66
DWDM 68
IP Transport Evolution on Wide Area Network 68
Dark Fiber 69
Purchasing and Leasing Capacity on Fiber Links 69
Indefeasible right of use (IRU) 69
IRU vs. Leasing a Fiber 70
Should smaller companies purchase an IRU based fiber? 70
Carrying Network Traffic between Countries 70
Terrestrial Fiber Optic Cables 70
Submarine Fiber Optic Cable Systems 71
Major route concept in sub marine fiber optic cable 72
Who builds sub marine fiber cables? 72
Who uses submarine cables? 73
Submarine Cable Types 73
Cable Landing Point 74
Beach manhole 76
Chapter-4 Service Provider Physical Locations 78
Introduction 78
CO (Central Office)/Telephony Exchange 79

https://t.me/learningnets

POP – Point of Presence 79
POP Interconnections 82
Colocation Centers 82
Carrier Hotel 85
Meet-me Room 86
Summary 87
Chapter-5 Service Provider Modules - The Big Picture 88
Introduction 88
Core Layer Module 90
Datacenter and Server Farm Modules 92
Border/IGW Module 99
XDSL Service Module 101
FTTX Service Module 104
Cable Broadband Service Module 106
Mobile Broadband Service Module 107
Fixed Wireless Service Module 109
WIMAX Service Module 110
National Peering Module 112
International Peering and Transit Module 115
Business/Corporate Customer Module 117
Chapter-6 Service Provider Interconnections and Peering 123
Introduction 123
Settlement Free Peering 124
Private BGP peering 125
Public BGP Peering 126
Bilateral Peering 127
Multilateral Peering 127
Benefits of Settlement Free Peering 128
Peering Requirements by IXP 130
Peering Requirements by Participants 130

https://t.me/learningnets

Peering Policies....131
Peering Rules....131
What is IXP (Internet Exchange Point)?....132
IXP Best Practices....133
Why Networks Peer at the IXP?....133
Where are the Internet Exchange Points?....133
IXP Membership vs. Commercial Models....134
What is Carrier Neutrality....135
European IXPs....135
IXPs around the World....136
What is Local IXP....136
What is Regional IXP....136
Who are the Internet Hubs in the World?....137
What are Tier 1, Tier 2 and Tier 3 Internet Service Providers....137
Flat Internet vs. Hierarchical Internet....140
What is Remote Peering....145
IP Transit Service....146
Summary....147
Chapter-7 ATELCO National Internet Service Provider Design....148
Introduction....148
Regional Connectivities of ATELCO Network....149
ATELCO's Logical Network (End to End)....155
IP/MPLS Multi Service Network....156
WCL (Worldwide Connectivity Layer)....160
Internet Gateway (IGW) and Shared Services Layer....163
Content Provider/Over the Top Networks in ATELCO....169
Services Provided by ATELCO....170
Residential Fixed Services....170
Residential Mobile Service....175
Business/Corporate Customers Access Connections....177

https://t.me/learningnets

Protocols/Technologies used in the three layers of ATELCO Network 179
ATELCO IP/MPLS Network IGP and BGP Design 182
ATELCO IP/MPLS IGP and BGP Design for Mobile Service 182
ATELCO IP/MPLS IGP and BGP Design for Fixed Service 186
Convergence Mechanisms 189
IP Addressing in ATELCO Network 190
ATELCO IPv6 Design 190
MTU and Neighbor Discovery 192
Synchronization in Mobile Service 192
Security Policy 193
Multicast in ATELCO Network 193
Summary 193
Chapter-8 ATELCO Network - Design Detail Explanations 194
Introduction 194
ATELCO Physical Network 195
Regional Connectivities of ATELCO Network 195
ATELCO Intra Region Physical Connections 203
ATELCO Network – Logical Architecture 207
IP/MPLS Multi Service Network 209
World Wide Connectivity Layer 211
Internet Gateway/Shared Services Layer 216
Services in ATELCO Network 218
Distributed vs. Centralized BNG 219
Design Decisions in the Three Parts of ATELCO's Network 223
Design Decisions in the WCL (World-Wide Connectivity) Layer 223
Design Decisions in the IGW Layer 223
Design Decisions in the IP/MPLS Network 235
ATELCO's Seamless MPLS design 242
Mobile Service IGP, BGP and MPLS Design 242
Fixed Service IGP, BGP and MPLS Design 244

https://t.me/learningnets

ATELCO Security Policy 246
Summary 251
Chapter-9 Evolving Technologies in the Service Provider Networks 252
Introduction 252
Service Provider Design Using Segment Routing 253
Segment Routing Introduction 253
Traffic Engineering using SR 254
PCEP and Segment Routing 255
END to END Segment routing (Single BGP-AS) 256
TI-LFA with Segment Routing 259
Egress Peer Engineering 260
Modern – Better way of EPE 262
Modern EPE Requirements 263
Segment routing and LDP Internetworking 264
Mapping Server 264
Ethernet VPN (EVPN) 265
ARP Suppression 267
Provider Backbone Bridging Ethernet VPN (PBB-EVPN) 268
NFV – Network Function Virtualization 269
NFV Platforms 271
NFV Use Cases 273
Using BGP for Routing in Large-Scale Data Centers 275
BGP in the Datacenter as IGP – Why not other IGPs? 275
CLOS Topology 276
Counter Arguments for using BGP in Datacenter as Routing Protocol 277
Counter Arguments for BGP in DC – It is for WAN 277
Counter Arguments for BGP in DC – It is Slow 277
Lack of BGP Neighbor Auto-Discovery 278
BGP Path Hunting 279
ASN Numbering Schema when EBGP is used inside Datacenter 281

https://t.me/learningnets

Recommended BGP ASN Allocation – Using 2 Byte Private ASN 282
Multicast BIER (RFC8279) 283
ACRONYMS AND ABBREVIATION 284

https://t.me/learningnets

Chapter 1

Chapter-1 Service Provider Types

Introduction

In the first chapter of the book, we will explain different Service Provider types and their businesses. When many people hear the “Service Provider” term, they immediately think about “Internet Service Provider”.

As of 2019, there are so many Service Providers which provide Internet service, but there are many other types of Service Providers that don’t provide Internet service to organizations. For example, in this chapter we will cover Content Providers; they provide content to end users/eyeballs. Also, the CDN – Content Delivery Network Provider business will be explained. They provide a distribution network to the Content Providers. Content Providers and CDN Providers don’t sell Internet service to end users or corporations. They use Internet as an underlay infrastructure to distribute the content.

New computing paradigms are emerging; these are Cloud Computing, Fog Computing and Edge Computing. We will look at Cloud Providers. Their business is not to sell Internet access to end users or corporates.
We will have a look at Edge Computing providers which provide WAF, Edge Applications, Serverless Computing, DDos Protection, Edge Firewall etc.

Some Internet Service Providers sell Internet access to other Internet Service Providers. After finishing this chapter, you will understand Backbone, Transit and Access Internet Service Providers and the business model between these providers.
There are definitely other Service Provider businesses in the IT industry but in the first version of the book, current and common Service Provider types are covered. Let the journey begin!

https://t.me/learningnets

Broadband Service Provider

This Service Provider provides broadband services to the residential and corporate customers. Different types of broadband services such as Cable Broadband, FTTX, XDSL, BPL (Broadband over Power Line), WiMAX,3G, LTE can be provided by the same Broadband Service Provider company.

A Cable Broadband company such as Comcast has millions of Cable Broadband customers in U.S. There are also Mobile Broadband Service Provider companies such as Vodafone and AT&T which provide DSL, FTTx, which also provide mobile broadband services through 3G, LTE etc. Companies generally provide more than one type of broadband access to their customers.

Access Service Providers are mostly providing Broadband services. With broadband connection, customers can receive Internet service. They can use Internet service to access the Internet and can also create a Virtual Private Network between their offices, HQ and Datacenters by using broadband technology.

Transit Service Provider

A company which provides an Internet access to the whole Internet region is considered as a Transit Service Provider. It's also known as IP Transit Service Provider. Transit is the service of allowing traffic from a network to cross or "transit" the provider's network, usually used to connect a smaller Internet Service Provider (ISP) to the rest of the Internet.

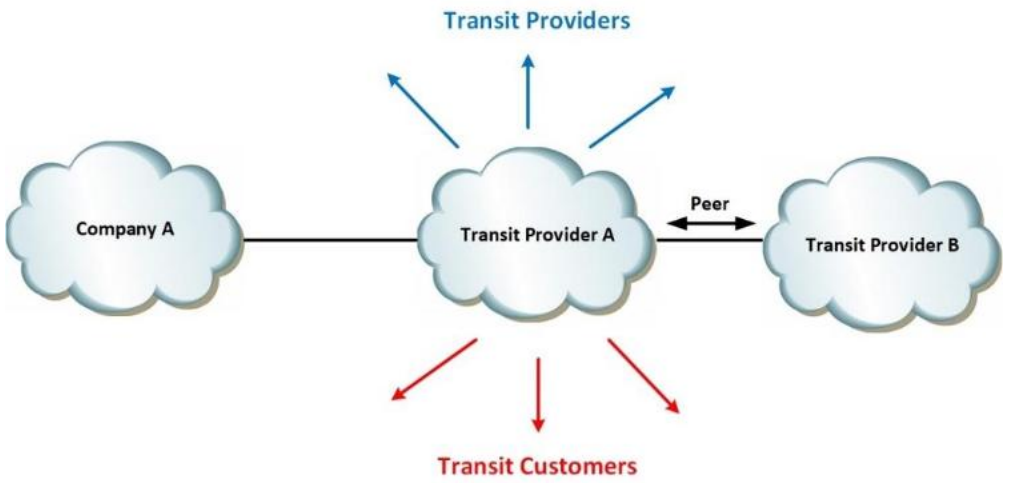


Figure 1-1 *Transit Service Provider and its connectivities*

https://t.me/learningnets

In figure 1-1, Provider A is the Transit Provider for Company A, as it also allows Company A to access the entire Internet. In this figure, Peering connection is shown between Provider A and Provider B. This Peering connection is a Settlement Free Peering which is one of the important Interconnection models and will be explained in detail in this chapter.

Last but not least, in figure 1-1, Provider A has its own Transit Providers as well. In the Service Provider connectivity, all Service Providers have their own Transit Providers. The only exception is the Tier 1 Internet Service Provider. Tier 1 Service Providers don't receive Transit Service from any other Provider. Different Tiers and the meaning of each one will be explained in this chapter.

Transit Service Providers are the Wholesale Service Providers. They provide Internet Access to other Service Providers. A Transit Service Provider might be providing Access to the customer which will be explained next.

IP Transit, which is also commonly known as "Internet Transit" is a simple service from the customer perspective. All you have to do is pay for the Internet Transit Service and all traffic sent to the upstream Internet Service Provider is delivered to the Internet. Internet Transit is typically a measured service. The more you send or receive, the more you pay.

Internet Transit has commits and discounts. Upstream Service Providers generally offer volume discounts based on negotiated confirmation levels. So, if you commit to 10Gbps of traffic per month, you will probably get a better unit price than if you compromise with only 1Gbps of traffic per month. However, you must pay (at least) the value of the level of confirmation of traffic, regardless of how much traffic you send.

Transit contracts over the Internet have a duration and deadline. Internet Transit prices drop every year.

Access Service Provider

This type of service provider, provides last mile access to the customers. What is the definition of last mile and first mile? This is an important telecommunication term which is used in all broadband communication methods. In fact, last mile is the same as first mile. From the Service Providers perspective, the link between the Service Providers and end users is often called "last mile". From an end user's perspective, this link is called "first mile".

In any of the broadband access technologies, such as xDSL, CATV (Cable Broadband), FTTx, BPL (Broadband over Powerline), Satellite, Fixed Wireless or Mobile Broadband, the term last mile is used extensively. Last mile is the part of access network. In the last mile, we have Customer CPE (Router, switch, PC etc.), DSL modem, twisted pair copper cable and DSLAM. DSLAM is the rack which keeps so many DSL modems at the Service Provider location. Each customer side DSL modem is terminated on the modem in the DSLAM.

https://t.me/learningnets

In xDSL networks, the link between the customer Modem and the DSLAM is the last mile. Access service providers provide last/first mile connections and the necessary access network equipment's to the customers.

As it was mentioned above, Service Providers might provide many services at the same time. The Access Service Provider Company might be providing Transit Service at the same time. AT&T in U.S or Turkish Telecom in Turkey are Access and Transit Service Providers at the same time.

Backbone Service Provider

ISP Tiers will be explained in detail later in the chapter, but in general Tier 1 ISPs are considered to be the Backbone of the Internet. Backbone ISPs can provide full Internet access to other ISPs or Corporate customers. They generally don't provide access services to the residential users. (Though some Tier 1 providers provide both of these services, they are not only Tier 1 ISPs, but are also an Access ISP).

A Tier 1 ISP is an ISP that has access to the entire Internet Region, solely via its Settlement Free Peering relationship. Settlement Free Peering concept will be explained later in the book. Tier 1 ISPs don't pay money to other ISPs to reach the Global Internet. Tier 1 ISPs only peer with other Tier 1 ISPs. They don't have any Transit ISP as they are the top tier ISP.

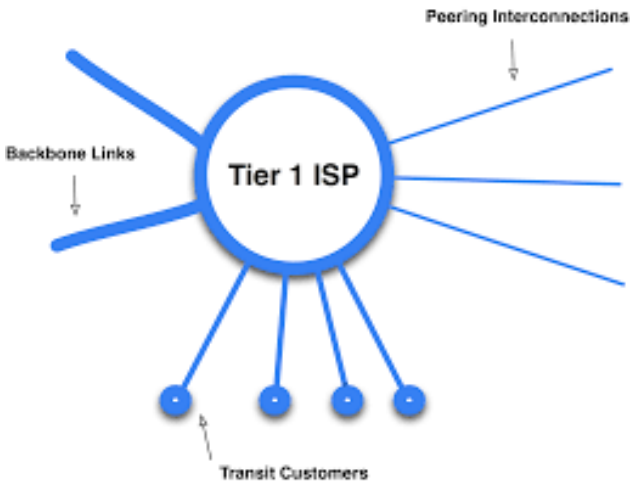


Figure 1-2 *Customers and Peers of Tier 1/Backbone Service Provider*

Tier 1/Backbone ISPs can have the following connections:

- Backbone links
- Peering Interconnections
- Transit customers

https://t.me/learningnets

There are currently 13 ISPs which are listed in the baker's dozen list as Global Tier 1 ISPs. None of these ISPs receive Transit Service from other Service Providers. Baker's Dozen is considered as the Tier 1 ISP list and every year the list is updated with the ISP ranking. The list is provided by measuring the Transit IP Space of each ISP. Unfortunately, the list has not been updated since 2016 and there are some changes.

For example, Deutsche Telecom and KPN are Tier 1 Operators but are not in the list. Also, recently, CenturyLink acquired Level 3 and became the largest Tier 1 Operator in the world. (Based on the number of AS customers). CenturyLink provides residential broadband service as well.

Some Transit ISPs are Tier 1 ISPs. But not every Transit ISP is a Backbone ISP. Transit ISP can be a Tier 2 ISP; thus, they may need to pay money to some Tier 1 ISPs to reach the Global Internet. (Full Internet Routing Table – Default Free Zone)

Regional ISP

The ISP which provides services in one or some parts part of the individual country is considered as Regional ISP. Geographically, their network is deployed in more than one city but not in the entire country. Regional ISPs today, might start providing services region wide and can become National ISPs later. This kind of regional ISPs are mostly common at North America and China.

For Regional ISPs, other terminologies are used in different parts of the world. In some continents, the definition of a Regional ISP is an ISP which provides service in a region of several countries. These ISPs are larger than National ISPs but not as large as Global or Tier-1 ISPs. That's why they are mostly defined as Tier-2 ISPs.

As an example for a Tier-2 Regional ISP; OTE is a National ISP in Greece, which also provides services to several countries at south east Europe as a Regional ISP. Also, Orange (France) and Neterra (Romania) are other examples of Regional ISPs. Regional ISPs might be providing Internet access to eyeballs (end users) or selling Transit Service to other ISPs or even providing Internet or VPN services to Enterprise companies. They shouldn't be considered only as Transit ISP, they might be providing services similar to Access ISPs.

https://t.me/learningnets

National ISP

The ISP which provides services in the entire country is considered as National ISP. Geographically, their network is deployed in the entire country. In practice, National ISPs don't have a presence in each and every city in the country but, they have presence in urban and sub urban areas. Rural areas are not economically feasible for them to build a network infrastructure.

If you haven't heard these definitions before, the definitions such as Urban, Sub-urban, Rural areas are used heavily in the ISP, Carrier and Fiber Operator industries. Broadband network designers always take these definitions into an account while they do their design.

- **Urban Area:** Typical Urban Areas have high population and large settlements. Crowded city centers can be given as an example.
- **Suburban Area:** Less population and less human density based on geography compared to Urban Area. A Crowded town can be given as an example.
- **Rural Area:** In general, a Rural Area or countryside is a geographic area that is located outside of towns and cities. Typical Rural Areas have low population densities and small settlements. Whatever is not Urban Area, is considered as Rural Area, though some people use the term of Suburban Area as less population than Urban Area but more population than Rural Area. Villages can be given as an example.
- **Underserved Area**: These areas, are areas that have no good network coverage (Broadband, Voice or any other data types).
- **Unserved Area**: Unserved Areas, are areas where there is no network coverage at all.

For example, if a mobile operator will place a cell site in an Urban Area, since the population density will be too high, they will consider to place more cell sites than if they place those cell sites in a Rural Area.

FTTx planers consider to change their ODN (Optical Distribution Network) design entirely based on whether they are doing FTTx deployment in Urban or Rural Areas. FTTX and ODN will be explained later in the chapter.

In general, having a fiber access to Rural Areas is not economically considered as a good idea, thus in Rural Areas, either Mobile Broadband or WISPs (Wireless Internet Service Provider) with unlicensed spectrum can provide services to customers. That's why, National Service Providers may not extend their fiber network infrastructure to the Rural Areas.

https://t.me/learningnets

Content Providers

Content Providers are defined as companies that provide actual content to consumers. There are two types of Internet sources: Eyeballs and Content. These two terms are used in the networking communities in the standard bodies (IETF, IEEE etc.) and at the events such as NOG, RIPE and IETF meetings.

- **Eyeballs**: refers to actual users
- **Content**: refers to data which the users are interested in

Search companies (Bing, Google, Yandex, Baidu), TV stations (ABC News, BBC, CNN), video providers (YouTube, Netflix), online libraries and E-Commerce websites all are Content Providers. Content Providers are commonly referred as OTT (Over the Top) Providers.

Content Providers have a direct relationship with billions of customers. Customers pay for ISPs and Content Provider's services. Content Providers are not affected by the regulations. This is a big debate between Service Providers and Content Providers. All of these regulations lead the Content Providers to become the largest companies in the world.

Public / Private Internet Companies, Ranked by Market Valuation (5/29/18)

Rank			Market Value ($B)	
2018	Company	Region	5/29/13	5/29/18
1)	Apple	USA	$418	$924
2)	Amazon	USA	121	783
3)	Microsoft	USA	291	753
4)	Google / Alphabet	USA	288	739
5)	Facebook	USA	56	538
6)	Alibaba	China	--	509
7)	Tencent	China	71	483
8)	Netflix	USA	13	152
9)	Ant Financial	China	--	150
10)	eBay + PayPal*	USA	71	133

Figure 1-3 *6 out of top 10 largest companies in the world are Content Providers*

In the list shown in figure 1-3; Apple, Amazon, Microsoft, Google, Facebook and Netflix are the Content Providers/Internet companies. Content Providers distribute their content via their own CDN and/or within third party CDN Providers, which will be discussed throughout this chapter. Content Providers not only distribute their content via other CDN companies, but they have their own CDN networks as well.

Google, Netflix, Facebook, Microsoft and almost all other big Content Providers have their own CDN networks and deploy their cache engines widely inside ISP's and/or IXP networks to be closer to their customers. Large Content Providers have global networks. If Google were an ISP, it would be the second largest carrier in the planet!

https://t.me/learningnets

Over the Top Providers (OTT)

Over the Top is a term used to refer to Content Providers. So, when you hear Over the Top Providers, they are Content Providers. Content can be any application, any service such as Instant messaging services (Skype, WhatsApp), streaming video services (YouTube, Netflix, Amazon Prime), voice over IP and many other voice or video content type.

An Over-the-Top (OTT) provider provides a content over the Internet and bypasses traditional private networks. Some OTT Providers distribute their content over their CDN over their private networks though (Google, YouTube, Akamai).

OTT delivers the content over traditional ISP networks. The creation of OTT applications has created a conflict between companies that offer similar or overlapping services. The traditional ISPs and Telco's have had to anticipate challenges related to third-party firms that offer OTT applications and the services.

For example, the conflict between a Content Provider company such as Netflix and a Cable Access Provider Company such as Comcast, which consumers still pay the cable company for having access to the Internet, but they might want to get rid of their cable TV service in favor of heaper streaming video over the Internet.

While the cable company wants to offer fast downloads, there is an inherent conflict of interest in not supporting a competitor, such as Netflix, that bypasses cable's traditional distribution channel.

Conflict between the ISPs and the OTT Providers had lead the Net Neutrality discussion. Net Neutrality is the principle that data should be treated equally by ISPs and without favoring or blocking particular content or websites. Those who are in favor of Net Neutrality argue that ISPs should not be able to block access to a website own by their competitor or offer "fast lanes" to deliver data more efficiently for additional cost. Net Neutrality will be discussed later in the book.

OTT services such as Skype and WhatsApp are banned in some Middle East countries by some Operators, as OTT applications taking some part of their revenue. For example, in 2016, social media applications such as Snapchat, WhatsApp and Viber were blocked by the two UAE telecoms companies, Du and Etisalat. They claimed that these services are against the country VOIP regulations.

In fact, UAE is not the only country for blocking access to some OTT applications and the services. Many countries in Middle East have followed the same model. They either completely blocked access to some OTT applications or throttled them, so the voice conversation over these services became near to impossible.

https://t.me/learningnets

Content Delivery Networks

Content Delivery Network companies replicate content caches close to large user population. They don't provide Internet access or Transit Service to customers or ISPs, but distribute the content of the Content Providers. Today, many Internet Service Providers started their own CDN business as well.

An example is Level 3. Level 3 provides their CDN services from their POP locations which are spread all over the World.

Content Distribution Networks reduce latency and increase service resilience (Content is replicated to more than one location). More popular contents are cached locally and the least popular ones can be served from the origin.

Before CDNs, the content were served from the source locations which increased latency, thus reduced throughput. Content were delivered from the central site. User requests were reaching to the central site where the source was located.

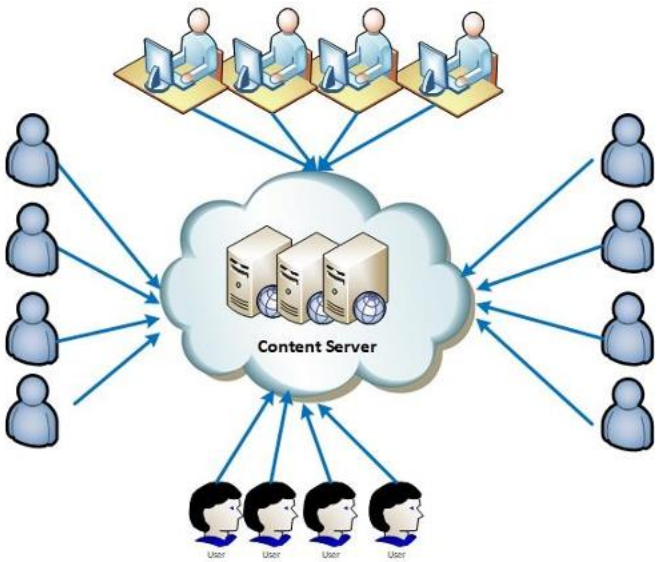


Figure 1-4 *Before CDN*

With CDN Technology, the Content are distributed to the local sites.

https://t.me/learningnets

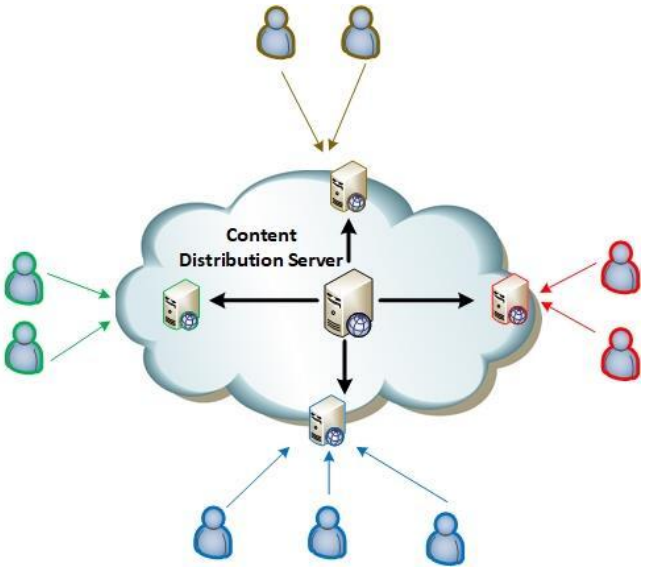


Figure 1-5 *After CDN*

Amazon, Akamai, Limelight, Fastly and Cloudflare are the largest CDN providers which provide services to different Content Providers all over the world. Also, some major Content Providers such as Google, Facebook, Netflix, etc. prefer to build their own CDN infrastructures and become large CDN providers.

CDN providers have servers all around the world. These servers are located Inside the Service Providers networks and the Internet Exchange Points. They have thousands of servers and they serve huge amount of Internet content. CDNs are highly distributed platforms. As mentioned before, Akamai is one of the Content Delivery Networks.

As per 2019, number of servers, number of countries, daily transactions and more information of Akamai's Content Distribution Network are as follow:

- More than 216,000 servers
- Located in 120 countries around the world
- Within more than 1,500 networks around the world.
- Delivers over 2 trillion Internet interactions daily
- Delivers approximately 30% of all Web traffic
- Their customers include: All top 20 global ecommerce sites, top 30 media companies, 7 of the top 10 banks, 9 of the largest newspapers, 9 out of 10 top social media sites

https://t.me/learningnets

Cloud Providers

Cloud Computing service is providing services like Storage, Databases, Servers, Networking and Software's etc. through the Internet. Few Companies offer such computing services, hence named as "Cloud Computing Providers/ Companies". They charge their users for utilizing such services and the charges are based on their usage of services.

Generally, Cloud Computing services are categorized into three types.

1. **Infrastructure as a Service (IaaS):** This service provides the infrastructure, such as Servers, Operating Systems, Virtual Machines, Networks, and Storage etc. on rent basis.
   Example: Amazon Web Service, Microsoft Azure.

2. **Platform as a Service (PaaS):** This service is used in developing, testing and maintaining software's. PaaS is same as IaaS but also provides additional tools such as DBMS, BI services etc.
   Example: Oracle Cloud Platform (OCP), Red Hat OpenShift, Google App Engine

3. **Software as a Service (SaaS):** This service makes the users connect to the applications through the Internet on a subscription basis.
   Example: Google Applications, Salesforce

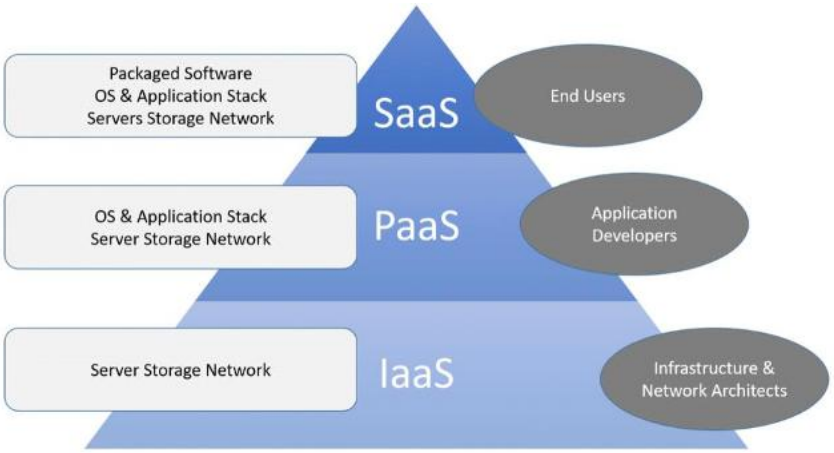


Figure 1-6 *Cloud Service Models*

The large benefit of using a Cloud Service Provider comes in efficiency and economies of scale. Rather than individuals and companies building their own infrastructure to support internal services and applications, the services can be purchased from the Cloud Service Providers (CSP), which provide the services to many customers from a shared infrastructure.

https://t.me/learningnets

Some Examples of Cloud Providers are Amazon Web Services, Microsoft Azure, Google Cloud Platform, Adobe, VMware, IBM Cloud, Rackspace, Red Hat, Salesforce, Oracle Cloud, SAP, Dropbox etc.

There are tradeoffs in the cloud. As enterprises move their applications and infrastructure to the cloud, they also give up control. Reliability and Security are the major concerns. Many CSPs are focusing on providing high levels of service and security, and PaaS and IaaS often come with performance guarantees. A common goal might be 99.9% (3x9s) or 99.99% (4x9s) uptime.

Because the CSP also hosts data storage and applications, customers must be assured that their data will be secure and the data center where the applications or services are hosted meet certain requirements.



Figure 1-7 *Cloud Service Providers – CSPs*

As CSPs have grown rapidly and require new levels of scalability and management, they have had a large effect on computing, storage, and networking technologies. The CSP popularity to a large extent has driven demand for virtualization, in which hardware can be segmented for access by different customers using software techniques. The growth of CSPs over the last ten years has also driven some of the fastest growth in technology segments ranging from servers to switches and business applications.

https://t.me/learningnets

Edge Computing Providers

Edge computing is a networking philosophy focused on bringing computing as close to the source of data as possible, in order to reduce latency and bandwidth usage. In a simpler term, edge computing means running fewer processes in the cloud and moving those processes to local places, such as on a user's computer, an IoT device, or an edge server. Bringing computation to the network's edge minimizes the amount of long-distance communication that has to happen between a client and server.

For Internet devices, the network edge is where the device, or the local network containing the device, communicates with the Internet. The edge may not be a clear term; for example, a user's computer or the processor inside of an IoT camera can be considered the network edge, while the user's router, ISP, or local edge servers are also considered the edge.

It is important to understand that the edge of the network is geographically close to the device, unlike origin servers and cloud servers, which can be very far from the devices they communicate with.

Cloud computing offers significant amount of resources (e.g., processing, memory and storage resources) for the computation requirement of mobile applications. However, gathering all the computation resources in a distant cloud environment started to cause issues for applications that are latency sensitive and bandwidth hungry. The underlying reason is that network traffic has to travel through several routers managed by Internet Service Providers (ISPs), operating at varying tiers. All these routers significantly increase the Round-Trip Time (RTT) that latency-sensitive applications face.

In addition to this, end-to-end routing path delays can change very dynamically due to ISPs and network conditions.

Akamai, CloudFront, CloudFlare and many other Edge Computing Providers provide edge services like WAF, Edge Applications, Serverless Computing, DDos Protection, Edge Firewall etc.

https://t.me/learningnets

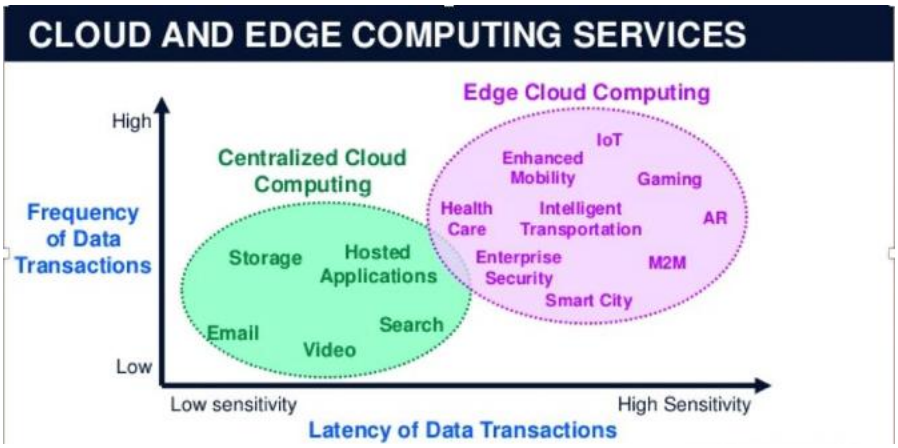


Figure 1-8 *Applications of Cloud and Edge Computing (source: www.NTT.com)*

In figure 1-8, common use cases of Cloud and Edge Computing Services are shown. Many emerging technologies will require Edge computing.

Cable Access Providers

Cable Access Providers provide Cable TV and Cable Broadband services to end users. Cable broadband can provide nearly 1Gbps bandwidth today with the recent enhancements of DOCSIS. Cable TV and Internet traffic are carried through HFC infrastructure. Hybrid fiber-coaxial (**HFC**) is a telecommunications industry term for a broadband network that combines optical fiber and coaxial cable. It has been commonly deployed globally by cable television operators since the early 1990s.

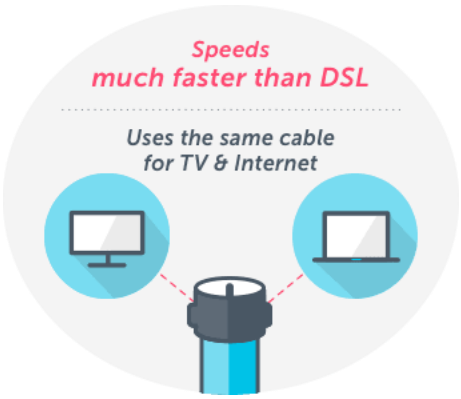


Figure 1-9 *Cable Broadband Marketing*

https://t.me/learningnets

Cable access was promoted to be faster than DSL with an advantage of providing both Internet and Cable TV without having two separate physical infrastructures. Today Cable broadband is still faster than most of the DSL technologies, but not faster than Fiber. Also, DSL is capable to carry IPTV and many Internet Service Providers carry their IPTV service over their DSL infrastructure.

While cable broadband is faster than DSL, transmission speeds vary depending on the type of modem, cable network, and how many people in the neighborhood are using a cable connection.

In cable broadband, the distance between your residence and the cable company will not affect your Internet speed. In DSL, distance between customer location and the telecom operator exchange office will greatly affect the service speed.

HFC, Cable Broadband Architecture and DOCSIS protocol will be explained later in the book. Cable operator is also known as Multiple System Operator (MSO). MSO is an operator of multiple cable or direct-broadcast satellite television systems.

AT&T, Comcast, Verizon, Cox are the U.S based cable MSO companies. Comcast used to be the biggest cable TV company in the U.S and it still is. But now it's something more important: It's the biggest broadband company in the U.S.

Mobile Operators

Mobile Operators provide Mobile broadband service over wireless. In general, broadband wireless networks can be categorized into two types: fixed and mobile wireless.

The broadband fixed wireless network technologies of interest which are covered here are Wireless Fidelity (Wi-Fi), which is an IEEE 802.11 standard and Worldwide Interoperability for Microwave Access (WiMAX), which is also an IEEE 802.16 standard.

Two broadband mobile wireless network technologies are the third Generation (3G) and Fourth Generation (4G) networks which will be further explained in the next chapter.

Mobile operators are also known as Mobile Network Operators (MNO), Mobile Network Carriers, and Cellular Operators. Mobile Operators own or control all the necessary elements of mobile networks, infrastructure, Backhaul Infrastructure and Radio Spectrum Allocations. Mobile Operators have the license of Radio Spectrum.

Mobile Operators can lease their wireless infrastructure to the other Mobile Carriers called MVNO (Mobile Virtual Network Operators). Mobile Operators provide Voice, Video and Data Communications Services. As of 2019, Mobile Operators provide 2G, 3G, LTE and LTE Advanced services to their customers.

https://t.me/learningnets

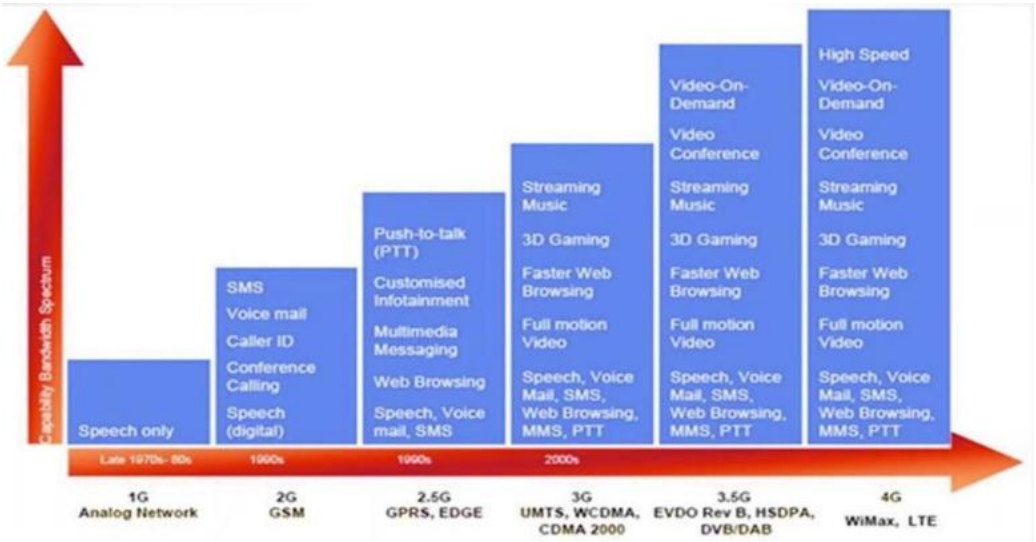


Figure1-10 *Services over Mobile Networks*

3GPP (The 3rd Generation Partnership Project) is the standard organization that works within the scope of ITU to develop 3rd (and future) generation wireless technologies that build upon the base provided by GSM.

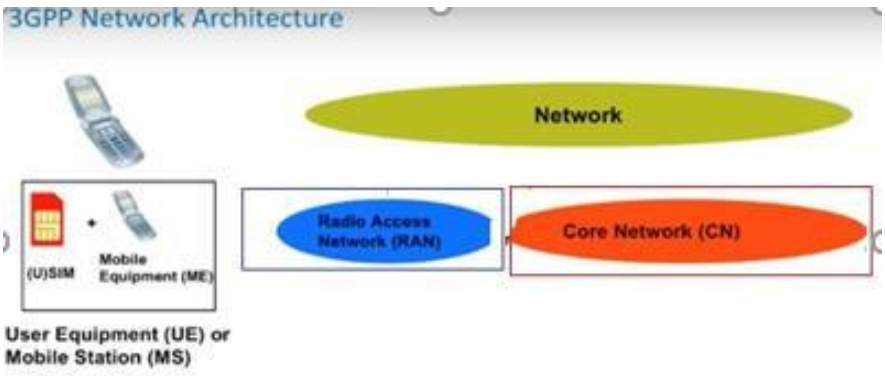


Figure 1-11 *3GPP Network Architecture (source: https://www.3gpp.org)*

Mobile broadband provides broadband service's over Radio Access Networks through air interface to the users. Mobile Broadband Services are good alternatives to provide high speed

https://t.me/learningnets

broadband service to the rural, remote and underserved areas due to high cost of building fiber infrastructure.

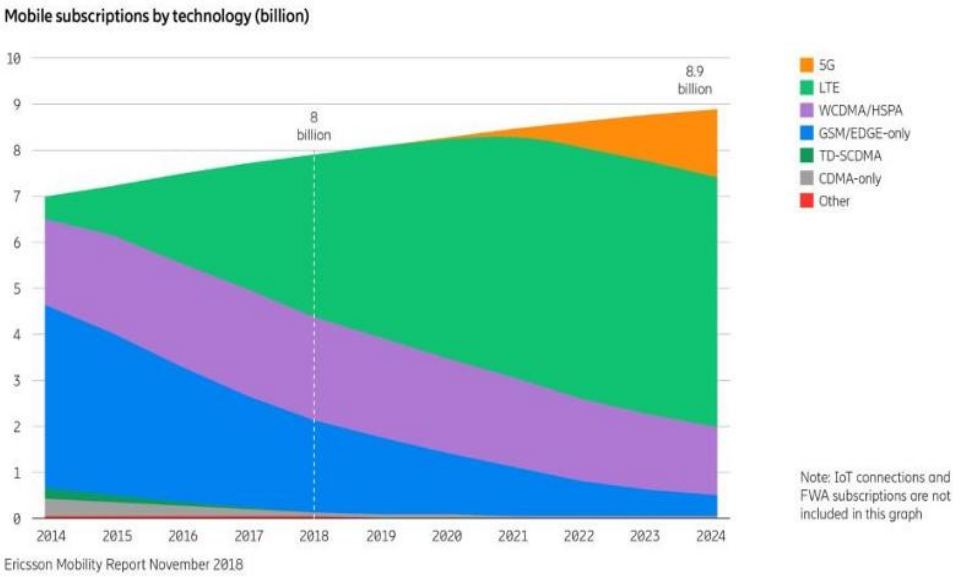


Figure 1-12 *3G and LTE are the most dominant mobile technologies worldwide (source: www.Ericsson.com)*

Due to undeveloped and developing countries population, 2G was the most dominant mobile technology until 2015, but as of 2019, 3G and LTE are the most dominant mobile technologies used worldwide.

According to FCC (U.S Communication Regulatory body), a mobile wireless broadband provider is considered “facilities-based”, if it provides service to a mobile wireless broadband subscriber using the provider’s own facilities and spectrum for which it holds a license, manages, or for which it has obtained the right to use via a spectrum leasing arrangement. Note that the facilities-based provider may—or may not—sell the Internet access service that is delivered over that broadband connection directly to the end user.

A broadband “end user” is a residential, business, institutional, or government entity that uses broadband services for its own purposes and does not resell such services to other entities. For the purposes of this form, an Internet Service Provider (ISP) is not an end user of a broadband connection.

https://t.me/learningnets

Wireless Internet Service Providers

WISP (Wireless Internet Service Provider) mainly provides an Internet Service to the rural areas, though some WISPs serve in the urban areas as well. The first WISP in the world was LARIAT, which was founded in 1992 and served in Albany, U.S. It was a non-profit organization until 2003. WISPs, similar to many other ISPs, can provide voice and VPN services too.

Wireless Internet Service Providers (WISP) provides broadband wireless Internet connections wherever traditional ADSL, cable or satellite services are either unavailable or uncompetitive.

Most WISPs offer tiered service levels, charging higher fees for faster speeds and more bandwidth. Similar to Telco's, cable companies, and other ISPs, WISPs typically require you to commit to a one or two-year contract, and they charge an installation or activation fee. WISPs commonly use unlicensed wireless spectrum.

Wireless towers (Radio towers, water towers, tall buildings) are connected to each other via one of the traditional backhaul technologies (Fiber, TDM, Wireless).

Like other ISPs, some WISPs limit how much data you can use per month (Data cap), but these limits are generally more generous than what Cell Providers, Satellite Providers and even some Cable Providers offer. Some WISPs though offer pretty good amount of bandwidth without data caps which is not generally the case with the Mobile Operators.

WISP is commonly known as a fixed wireless access (FWA) or broadband wireless access (BWA) "last mile" service, meaning it serves fixed locations rather than mobile devices.



Figure 1-13 *Wireless Internet Service Provider (source: www.pcworld.com)*

Due to very expensive spectrum auctions, WISPs are limited to use unlicensed spectrum which has higher interference. Wireless frequency spectrum is sold by auction. WISPs can attend these auctions but the spectrum is so expensive.

https://t.me/learningnets

WISPs are generally initiated by the small business owners who cannot afford the expensive spectrum costs (100s of Millions of Dollar). WISPs provide fixed solutions, so if customers require mobility for their Internet services, WISP cannot be a good option as a broadband service for them. In this case, cellular mobile technology such as 3G, 4G, LTE and upcoming 5G are the options.

Satellite Service Providers

Satellite Broadband is provided through communication satellites. Access to high speed Internet has become very important for any type of business today. For the companies located in more remote parts of the world where terrestrial infrastructure is not an option, satellite broadband service provides an excellent alternative.

Compared to terrestrial systems, remote sites/rural areas can be deployed very quickly with satellite, which is also independent of any terrestrial system. Over the Satellite, companies can enable VPN connections between their sites. They can connect their remote sites to the other sites which can be a terrestrial based infrastructure.

Satellite systems generally suffer two conditions. Weather situations and Latency. These two aspects and more will be explained in detail in the next chapter. Three commonly used Satellite broadband bands are Ku, Ka and C bands. Based on the band that the satellite operates, bandwidth, signal strength and tolerance to interference will be different. Different Satellite broadband providers have different Satellite band systems.

Also, based on the Satellite's Orbit, the latency and number of Satellite equipment's at the user location will vary. Satellite Bands and the Orbits will be explained in the next chapter. For many businesses like mining, oil and gas, geology etc., satellite Internet might be the only available Internet option.

Inmarsat, Intelsat, Eutelsat, Iridium and Globalstar are some important Satellite Service Providers in the world. Inmarsat works with GEO (Geostationary Earth Orbit) satellites. Iridium and Globalstar work with LEO (Low Earth Orbit) satellites.

There are 3 types of Satellite services; broadcast Satellite Service (BSS), Fixed Satellite Service (FSS) and Mobile Satellite Service.

https://t.me/learningnets

Broadcasting, Fixed and Mobile Satellite Services

Broadcasting Satellite Services (BSS) is a radio communication service in which signals transmitted or retransmitted by space stations are intended for direct reception by the general public. Examples of BSS applications are satellite direct-to-home (DTH) for services such as satellite televisions, as well as satellite radio.

Mobile Satellite Services (MSS) refers to networks of communication satellites intended for use with mobile and portable wireless devices. Typical applications for MSS are satellite phones capable of voice and data. Another example of an MSS application is the Broadband Global Area Network (BGAN), operated by Inmarsat. BGAN uses small mobile terminals, with the size of a laptop to provide broadband Internet access via satellite.

Fixed Satellite Services (FSS) uses geosynchronous satellites for broadcasting purposes such as TV and radio, telecommunications, and satellite communication that are used by government, military organizations, small and large enterprises, and other end-users. The satellites used for FSS generally have a low power output and require large dish-style antennas for reception. They have less power than broadcasting satellite services. There are several FSS applications, including broadband Internet over satellite, information gathering, videoconferencing, distance learning and backhaul.

Mobile systems have smaller antennas, lower hardware costs and broader coverage. But the cost per minute of use is much higher than fixed satellite systems, and the throughput rates are far lower than those for fixed satellite systems.

Satellite communication is not only necessary for Rural/Underserved Areas. It can also be used in Urban Areas which are high population density locations or city centers. As was mentioned in the chapter, in Rural Areas there may not be available terrestrial infrastructure, so using satellite communication can be the only choice.

In Urban Areas, which also have various types of communication technologies, Satellite Communication can be used as a backup connection. Failure of terrestrial infrastructure through natural disaster can be seen, that’s why having a satellite technology when terrestrial systems are partly or fully unavailable is critical.

ISPs partner with Satellite communication companies to provide Internet access to the consumers and the corporate companies.

According to Morgan Stanley, the estimation of the global space industry could generate revenue of $1.1 trillion or more in 2040, up from $350 billion, currently.

https://t.me/learningnets

Summary

In the first chapter of the book, different types of Service Provider businesses was introduced. We have seen that Service Provider business is not limited to Internet Service Providers. In fact, many Service Providers that were discussed in this chapter were not providing Internet service, but they provide the content or applications over the Internet Service Provider networks. Because of that, Content Providers are commonly known as Over the Top Providers.

Providers which were introduced in this chapter have business relationships with each other. For example, Content Providers receive IP Transit service from Transit/Backbone Service Providers and have Settlement Free Interconnection with Local/Access Providers.

Content Providers might have their own CDN networks but they might be using third party CDN providers, thus they are paying to other CDN providers as well, to distribute their content closer to end user locations.

Different types of Fixed and Mobile, Wired and Wireless Service Provider businesses were introduced in this chapter. Hosting, Colocation, CDN, Edge Computing and many other non-Internet Service Provider networks and their interactions with each other was mentioned.

In the next chapter, we will have a look at what do Fixed, Mobile, Wired, Wireless services mean? What are the different broadband technologies and how they are provided to residential and the corporate customers by the Service Providers.

https://t.me/learningnets

Chapter 2

Chapter-2 Introduction to Service Providers Network and Services

Introduction

In the telecommunication world, the operators provide Internet, Voice, Video, Cable TV, Satellite TV and/or Internet, VPN, IPTV, Cloud, Hosting and many other services to their customers. Today, most of the operators provide more than one service to increase their revenues. Almost every operator provides an Internet service as of 2019.

These operators can be any of the following types:

- Landline Phone Companies
- Cable Companies
- Cellular Phone Companies
- Satellite Companies

Service Providers provide different types of services. Some Service Providers provide broadband services, while others provide mobile services, cloud services, edge computing, VPN, Internet or hosting services. Companies generally provide more than one type of broadband access to their customers to increase their revenue.

In this chapter, we will look at different types of fixed and mobile based broadband services such as XDSL, FTTX, Cable Broadband, Fixed and Mobile Satellite, Wireless Internet Service and Mobile Broadband LTE (Long Term Evolution).

https://t.me/learningnets

Broadband Services

Broadband signals compared to narrowband signals have much more band in the frequency spectrum. Thus, the term is referred as broadband. Higher band in the frequency spectrum allows faster data communication. Early Dial-up modems (over telephone lines) worked based on narrowband, thus they only provided voice communication and slow data speeds such as 56kpbs.

Broadband allows much higher data speeds such as 1Gbps or even more. DSL, FTTX, Cable Broadband, 3G, 4G Mobile Broadband are the examples of broadband technologies. These technologies will be covered in detail later in this chapter.

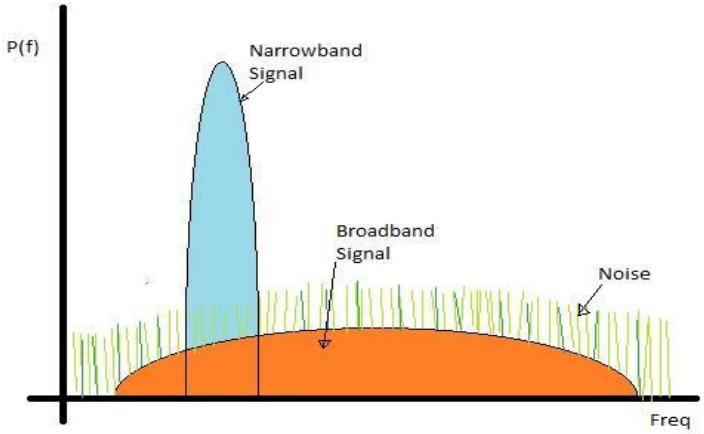


Figure 2-1 *Broadband*

Carrier and ISP are very commonly used terms in the industry. There is a difference between them.

- **Carrier** is the company that owns the phone lines and maintains them.
- **Service Provider** is the company that is responsible for making sure that the services (such as voice, VPN and Internet Service) are functioning properly.

Sometimes the Service Provider owns its hardware that provides the technical directions to the carrier, other times they manage the carrier's hardware. The difference between Carrier and Service Provider is similar to FedEx and EBay. One is bringing the service to you; while the other is selling those goods to you.

In this chapter, various Broadband Technologies will be explained in detail. Broadband Technologies can be categorized as Fixed or Mobile Broadband Technologies.

https://t.me/learningnets

The Fixed Services provided in this chapter are:

- XDSL
- FTTH
- Cable Broadband
- BPL
- Fixed Satellite
- Wireless (Fixed and Mobile Satellite Services will be explained under the same section)

The Mobile Services provided in this chapter are:

- 3G
- LTE

Satellite system can be Fixed or Mobile, as both will be explained later in this chapter under the same section. Let's first understand these services in more detail and then explain the design and architecture of a fictitious Service Provider (ATELCO), for these services in the following chapters.

Fixed Broadband Service Technologies

DSL

Dial-up modems were using analog transmission which were limiting the bandwidth around 56kbps. DSL uses digital transmission between telephony exchange (Central Office in U.S) and the customer modem. DSL provides broadband communication over ordinary copper telephone lines.

DSL services were initiated by the telephone companies which needed to provide higher bandwidth than 56kbps dialup, because their competitors such as Cable TV and Satellite companies begun providing 10Mbps and 50Mbps respectively.

The Upstream and Downstream speed can be different, which is then called Asymmetric DSL. DSL is also known as Digital Subscriber Loop. DSL was the most commonly deployed fixed broadband technology until recently.

https://t.me/learningnets

DSL vs. other fixed access method penetration in the world

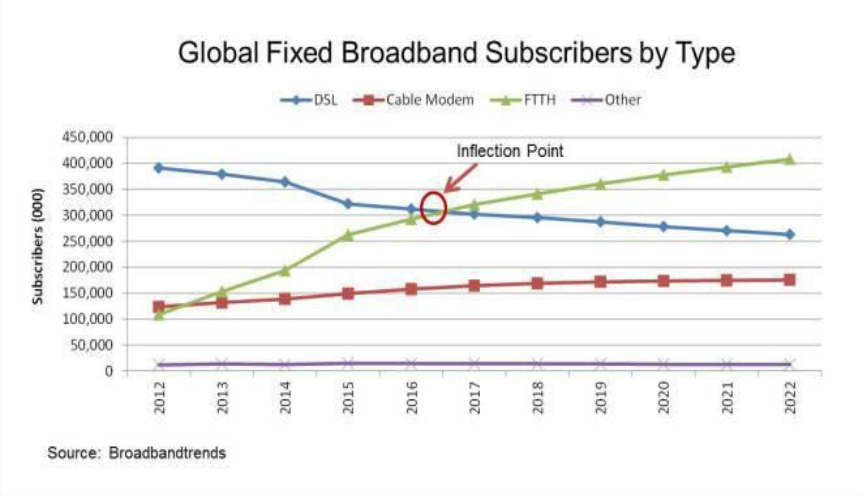


Figure 2-2 *DSL vs other fixed access method technologies penetration in the world*

The actual data rate depends on several factors such as the modulation method used, the number of sub channels used, distance between CPE and DSLAM and also the quality of the copper wire which is the main factor causing a noise on the connection.

Noise is a major bandwidth limiting factor for DSL connections. Bandwidth decreases because of copper signal attenuation and interference of surrounding signal (these two factors are simply called as noise). Noise is strongly dependent on the distance between CPE and DSLAM and also the copper quality.

Copper quality can be described as a combination of wire age, wire radius, number of junction points and isolation of cables. For an example; a 5-year-old 0,5mm radius copper wire with 2-3 junction points can carry higher speeds to longer distances than a 20-year-old 0,4mm radius copper wire with 5-6 junctions.

DSLAM is the physical DSL modem termination equipment and located in the telephony exchange or street cabinet of the Service Provider. The functionality of it is similar to CMTS in

https://t.me/learningnets

Cable broadband or eNodeB in Mobile broadband, as we will see later in this section. DSLAM which stands for Digital Subscriber Line Access Multiplexer is used to aggregate multiple DSL customers and send the traffic to the IP Backbone and vice versa.



Figure 2-3 *Old type of DSLAM used in 1990's and early 2000*

DSLAMs are placed very close to subscriber locations to provide faster speeds in recent years. Old DSLAMs were very large, since they were located in Telecom Exchange and terminated 1000s of subscribers, while new DSLAMs are pizza-box size and are located at the street cabinets.



Figure 2-4 *New DSLAM devices are pizza-box size*

DSL provides voice and Internet service over the same copper cable. DSL distances can go up to 5.5 km without a repeater. There are many different DSL implementations such as ADSL, ADSL2, ADSL2+, VDSL, VDSL2 and G.fast.

https://t.me/learningnets

DSL Technologies Speed Range

Family	ITU	Name	Ratified	Max.Speed
ADSL	G.992.1	G.dmt	1999	7 Mbps down, 800 kbps up
ADSL2	G.992.3	G.dmt.bis	2002	8 Mb/s down, 1 Mbps up
ADSL2plus	G.992.5	ADSL2plus	2003	24 Mbps down, 1 Mbps up
ADSL2-RE	G.992.3	Reach Extended	2003	8 Mbps down 1 Mbps up
SHDSL	G.991.2	G.SHDSL	2001	5.6 Mbps up/down
VDSL	G.993.1	Very-high-data-rate DSL	2004	55 Mbps down, 15 Mbps up
VDSL2	G.993.2	Very-high-data-rate DSL 2	2005	100 Mbps up/down

Figure 2-5 *DSL technologies speed range*

Higher capacity, lower distance!

Newer technologies use enhanced modulation techniques and also use higher frequency, thus they can provide higher data speeds (bandwidth). Higher frequency signals travel shorter distances compared to lower frequency signals, thus VDSL2 compared to ADSL2+ provides more bandwidth but requires much less copper distance from the modem to the DSLAM.

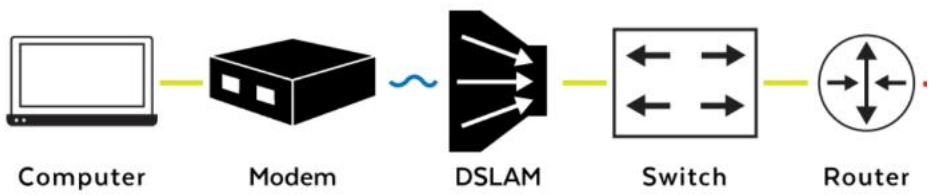


Figure 2-6 *Modem is connected to the DSLAM*

VDSL can be used to provide speeds up to 100Mbps over short distances but the main problem is interoperability due to lack of standardization.
VDSL2 is a standard based mechanism which provides higher capacity compared to VDSL, as VDSL2 uses pair bonding and vectoring.

https://t.me/learningnets

Port bonding is to use multiple copper pairs at the same time to get higher capacity, but when this is done, signal interference is created. Vectoring is used to avoid this interference and for noise cancellation.

As of 2019, DSL speeds can go up to 10Gbps in theory with the XG.fast technology which works over existing copper telephone lines. Nokia has achieved a connection speed of 5Gbps—about 625MB/sec—over 70 meters of conventional twisted-pair copper telephone wire, and 8Gbps over 30 meters. The trial used a relatively new digital subscriber line (DSL) protocol called XG.fast (aka G.fast2)

FTTX

Fiber to the X (FTTx) is a generic term for any broadband network architecture using optical fiber to provide all or part of the local loop used for last mile telecommunications. FTTN (Fiber to the Node), FTTC (Fiber to the Cabinet or Curb), FTTB (Fiber to the Building) and FTTH (Fiber to the Home or Premise) are the FTTX deployment models.

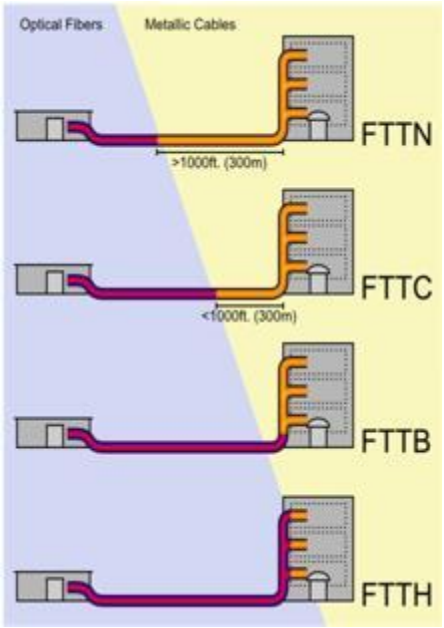


Figure 2-7 *FTTX*

With FTTN and FTTC, fiber is laid up to the cabinet/node and copper wire connection between the street cabinet to the destination will complete the connection.

With FTTB, FTTP and FTTH, fiber is laid up to the building, premise or home. Premise can be a home, apartment, condos, small businesses etc. Single mode fiber is used in all FTTx architectures.

https://t.me/learningnets

From the Operator's central office (Telephone Exchange) to the destination (home users, other ISPs or enterprises), fiber is distributed over ODN (Optical Distribution Network). ODN (Optical Distribution Network) is the physical path for optical transmission between OLT (Optical Line Terminal) and ONU (Optical Network Unit) equipment.

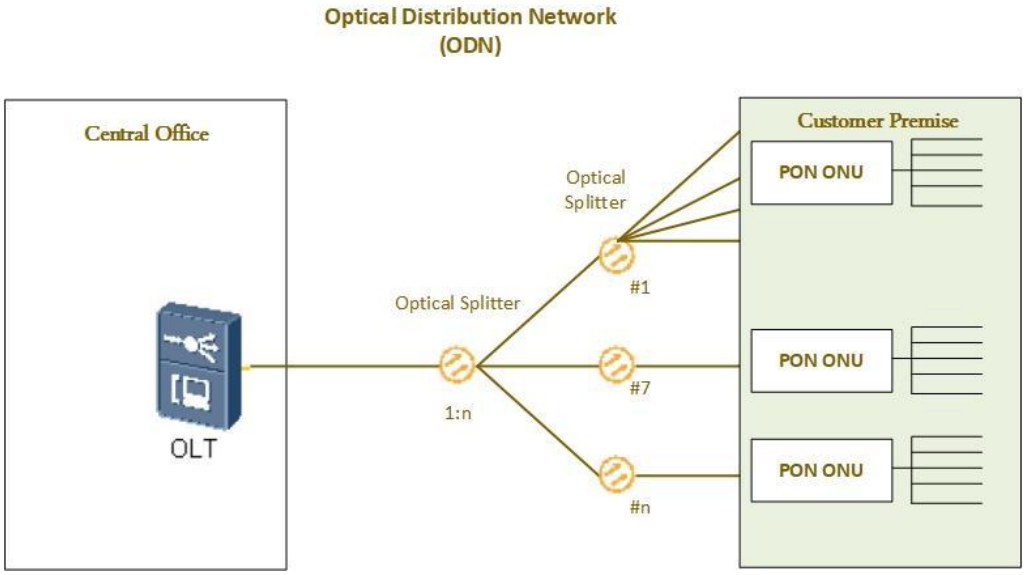


Figure 2-8 *Optical Distribution Network Component*

What is ONU in PON (Passive Optical Network)

Optical Network Unit (ONU) is a generic term denoting a device that terminates any one of the distributed (leaf) endpoints of an Optical Distribution Network (ODN), which runs a PON protocol. In the case of a multi-dwelling unit (MDU) or multi-tenant unit (MTU), a multi-subscriber ONU typically resides in the basement or a wiring closet (FTTB case) and has FE/GE/Ethernet over native Ethernet link or over xDSL (typically VDSL) connectivity with each CPE at the subscriber premises.

In the case where fiber is terminated outside the premises (neighborhood or curb side) on an ONT/ONU, the last-leg-premises connections could be via existing or new copper, with xDSL as the physical layer (typically VDSL). In this case, the ONU effectively is a "PON-fed DSLAM".

https://t.me/learningnets

OLT – ONU and ONT

OLT is the device which terminates the subscribers fibers, located in the Service Provider network. OLT has two directions: upstream (getting different types of data and voice traffic from users) and downstream, which is getting data, voice and video traffic from metro network or from a long-haul network and sending it to all ONT modules on the ODN.

ONT is the end user device which terminates the fiber, located at the user side (Apartments, Building, and Street Cabinet). ONT is also known as ONU. ONT is an ITU term (International Telecommunication Union).

ONU (Optical Network Unit) is an IEEE term; there is a slight difference though. In Europe, ONT term is used mostly, not ONU. Depends on the geographical location, these two terms might be used interchangeably.

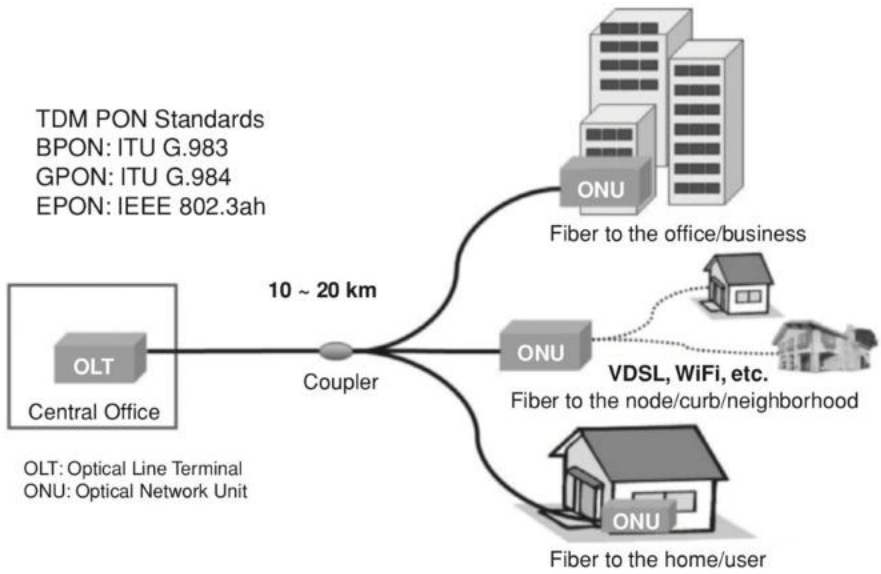


Figure 2-9 *OLT (Optical Line Terminal) and ONU (Optical Network Unit)*

https://t.me/learningnets

P2P and P2MP ODN (Optical Distribution Network)

Optical Distribution Network (ODN) can be setup as P2P (Point to Point) or P2MP (Point to Multipoint). P2P ODN has single fiber per customer, between OLT and ONT. P2MP ODN shares the fiber cable between many subscribers. P2MP ODN can be setup as AON (Active Optical Network) or PON (Passive Optical Network). The most popular P2MP architecture is PON (Passive Optical Network).

ATELCO (the scenario which will be explained later in the book) provides FTTH service. PON (Passive Optical Network) is the most common used technology for providing FTTH service. Other than Passive Optical Networking, Fiber connectivity can be provided via Active Optical Networking.

Active and Passive Optical Networking for FTTx

P2P ODN (Optical Distribution Network) can be provided to large Enterprises or other Service Providers. It is a dedicated fiber per customer.

Shared fiber ODN infrastructure with AON (Active Optical Networks) consists of equipment's which require electricity (Switches, Routers, amplifiers, repeaters). AON can reach much higher distance compared to PON, but almost all residential FTTx deployments are done with PON (Passive Optical Network).

In ATELCO's network, as it will be explained in the following chapters, there are some Active Optical deployments. They deployed it before PON getting mature and popular. At the very close place to the multi-floor (the multi-tenant buildings), ATELCO deployed Ethernet switches and used copper cable from these switches to the residential apartments.

ATELCO considered migrating from Active Optical networking to PON, but they found this migration too expensive. That's why; they don't have any plan to migrate AON residential deployments to PON.

https://t.me/learningnets

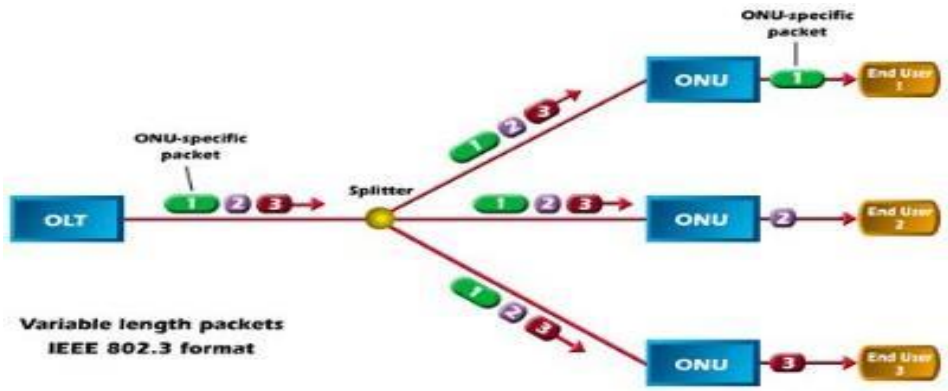


Figure 2-10 *GPON - Gigabit Passive Optical Network*

Optical Splitter and GPON

Optical Splitter is a passive equipment (Doesn't require electrical power). It broadcasts all the packets to the end users. Split ratio on the splitter can be up to 1/128.



Figure 2-11 *PON splitters*

GPON is one of the most popular PON standards. OLT sends the packets with ONU-ID, thus although the packets are broadcasted by splitter, the correct ONU accepts its own packets (Similar to MAC learning in Ethernet Switches).

Splitters can be connected to the two different OLTs for the redundancy. Although this provides higher availability, it increases the cost of PON deployment. The most common splitters deployed in a PON system are 1: N or 2: N splitter ratio, where N is the number of output ports. Generally, the 1: N splitters are deployed in Star/Hub and Spoke topologies, while 2: N splitters are deployed in ring topologies to provide physical network redundancy.

https://t.me/learningnets

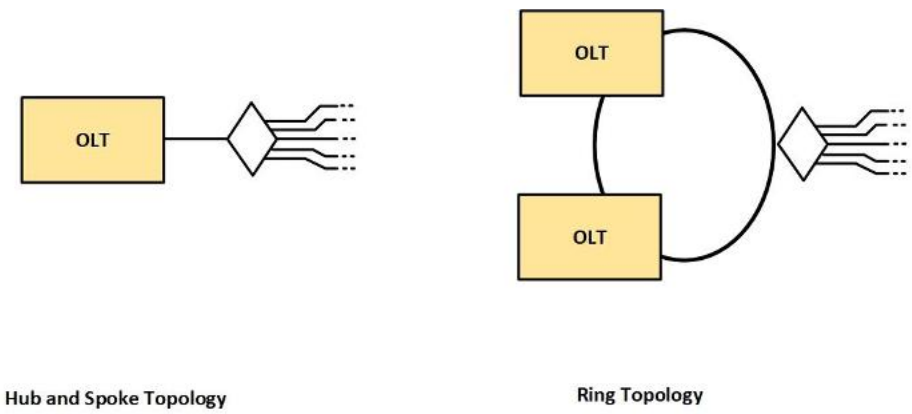


Figure 2-12 *SPLITTER to OLT connectivity in Ring and Star (Hub and Spoke) Topologies*

Using optical splitters in PON allows the service provider to conserve fibers in the backbone, essentially using one fiber to feed as many as 128 end users. A typical split ratio in a PON application is 1:32, which means one incoming fiber splits into 32 outputs. And the qualified fiber optic signal can be transmitted over 20 km.

If the distance between the OLT and ONT is small (in 5 km), you can consider about 1:64 or even 1:128. With higher split ratios, the PON network has both advantages and disadvantages.

Fiber optic splitters with higher split ratios can share the OLT optics and electronic costs as well as the feeder fiber costs and potential new installation costs. In addition, larger split ratio allows more flexibility and fiber management at the Headend. At the same time, higher split ratio splitters reduce bandwidth per ONU (optical network unit).

GPON vs. Traditional Ethernet Architecture

GPON (Gigabit Passive Optical Network) is used to reduce the number of active switching nodes in the network design. Network Design Best practice in Campus networks and many Datacenter networks (Not Massively Scale Datacenters), is to use Three-Tier; Access, Distribution and Core network design. Although the design decision depends on the scalability requirements in the Campus and DC, two layer; Access and Collapsed Distribution/Core design can be used. Figure 2-13 depicts common three tier Access, Distribution and Core design.

https://t.me/learningnets

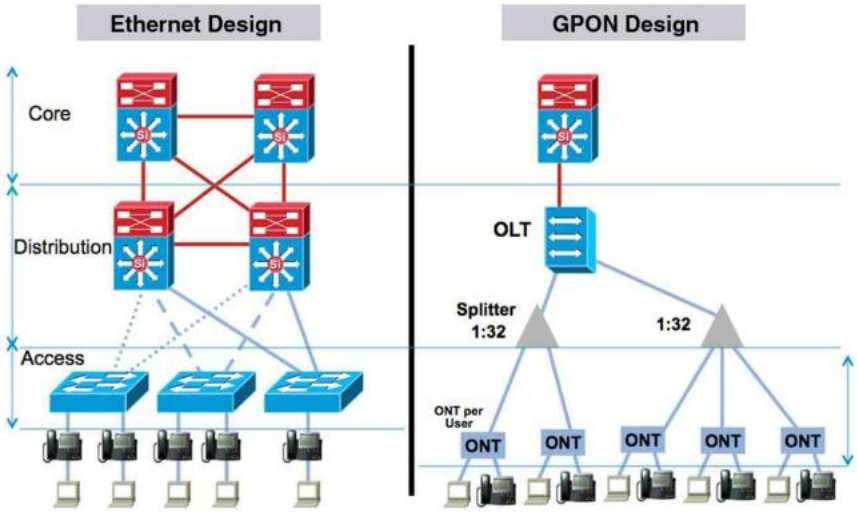


Figure 2-13 *GPON vs. Traditional Ethernet Architecture, Source: cisco.com*

In Three-tier Traditional campus networks, there are active Ethernet devices used in each tier. Active means, nodes require electricity. Active Ethernet switches forward traffic based on forwarding rules. If it's a Layer 2 network, traffic is forwarded based on Layer 2 information, if it is a Layer 3 design, traffic is forwarded based on routing protocol information.

GPON in the campus network replaces traditional three-tier design with two-tier optical network, by removing the Active access and distribution layer Ethernet switches with the ONT, Splitter and OLT devices. Although ONT requires power, the power requirement of ONT, compared to Active Ethernet switch is much less and the Splitter doesn't require power at all. Splitter is purely a non-Active device.

Analysis show that using GPON in the Campus network, instead of Active devices, reduces power requirement significantly.

Many capabilities which are provided by Active Ethernet switches, such as Vlan awareness, Security features, Quality of service, Multicast, Redundancy, loop prevention etc. are provided with GPON design as well.

In design, there is no best solution. Every solution has its advantages and the disadvantages. This is also true for the comparison between GPON vs. Active Ethernet.

So far, it has been given that GPON has many advantages. On the other side, GPON has its bandwidth limitation as a technology. 2.5Gbps Download, 1.25Gbps Upload limitation. Although with the newer generation of PON solution, it can provide more Download and Upload bandwidth.

https://t.me/learningnets

When we compare GPON with Traditional Active Ethernet, it is true that the Download bandwidth is 2.5Gbps and Upload bandwidth is 1.25Gbps, which is less than what traditional Ethernet can provide.

Depending on the Split ratio on the Splitter, 2.5Gbps bandwidth might be shared with 32, 64 or even with 128 different end points. Thus, it is true to say that when there is more bandwidth requirement per end point, Active Ethernet architecture can provide more bandwidth, probably with a better cost.

Cost analysis should be made carefully, as there might be different Fiber optic cable and transceiver requirements for each solution.

Cable Broadband

For providing a broadband Internet service over Cable TV network, the Cable Broadband/Internet requires cable modem at the customer premise and CMTS (Cable Modem Termination System) at the cable provider facility. These facilities are cable television headend (It is similar to Telco CO (Central Office)).

In DSL, bandwidth is dedicated on the access network per user.
In PON (FTTx), the bandwidth is shared on the access network between the end users.
In cable broadband the bandwidth is shared between the end users, similar to PON.

CMTS and cable modem is connected either via coaxial cable or HFC. Coaxial cable used by cable TV allows broadband communication by transmitting of several channels using distinct frequencies.



Figure 2-14 *Coaxial Cable is used in Cable Broadband*

https://t.me/learningnets

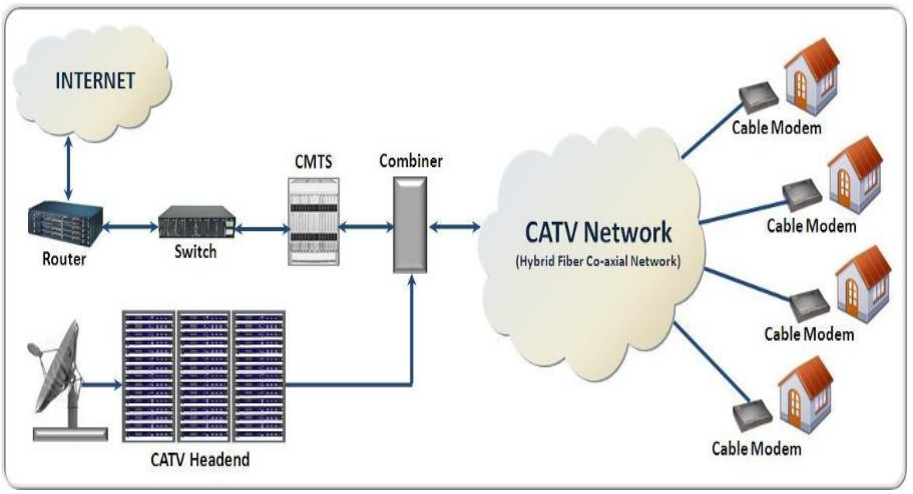


Figure 2-15 *CMTS to Cable Modem Connectivity*

HFC System (Hybrid Fiber Coax)

Most of the Cable Broadband systems use HFC which is Hybrid Fiber-Coaxial system.

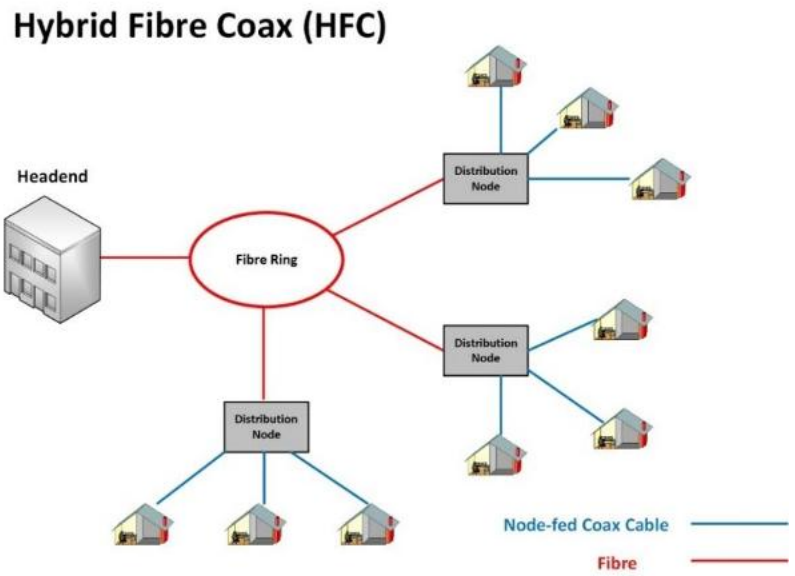


Figure 2-16 *Hybrid Fiber Coax - Fiber and Coaxial Cable infrastructure is used in Cable Broadband*

https://t.me/learningnets

HFC (Hybrid Fiber-Coaxial) and Speed Dependency

Since with Cable Broadband, users share the access network bandwidth, the actual transfer rate achieved using cable TV networks is related to the number of users connected to the optical node at the same time, as this system is based on the fact that all users will not be accessing the Internet at the same time (Statistical multiplexing). More users means lower bandwidth available to each individual user.

DOCSIS

The most common system used by cable TV companies to offer Internet access is called DOCSIS (Data Over Cable Service Interface Specification). It was founded by CableLabs (Non-Profit Organization). DOCSIS defines interface requirements for cable modems involved in high speed data delivery over HFC network.

DOCSIS and the other Broadband Standards

DSL systems use ADSL, VDSL etc. for CPE to communicate with network nodes (DSLAM/MSAN). FTTX networks use GPON, EPON, WDM-PON between CPE (ONU in this case) and OLT. Cable Broadband uses DOCSIS for all communication standards between Cable modem and CMTS (Cable Modem Termination System).

Cable Modem receives speed, IP address and time configuration parameters through CMTS from DHCP, TFTP and TOD (Time of Day) servers. The TFTP config file, sets the user broadband speeds, thus it is very crucial in cable broadband architecture to assign bandwidth and speed to the user (Some users try to upload their own tftp config file to change their broadband speed)

CMTS (Cable Modem Termination System)

CMTS is a provider edge system which connects the RF cable plant side to the provider IP core network. CMTS allows Cable Operator to offer broadband and other IP based services to the end subscribers connected to their cable network, including Voice and Video. CMTS functionality is similar to DSLAM in DSL and OLT in FTTx.

https://t.me/learningnets

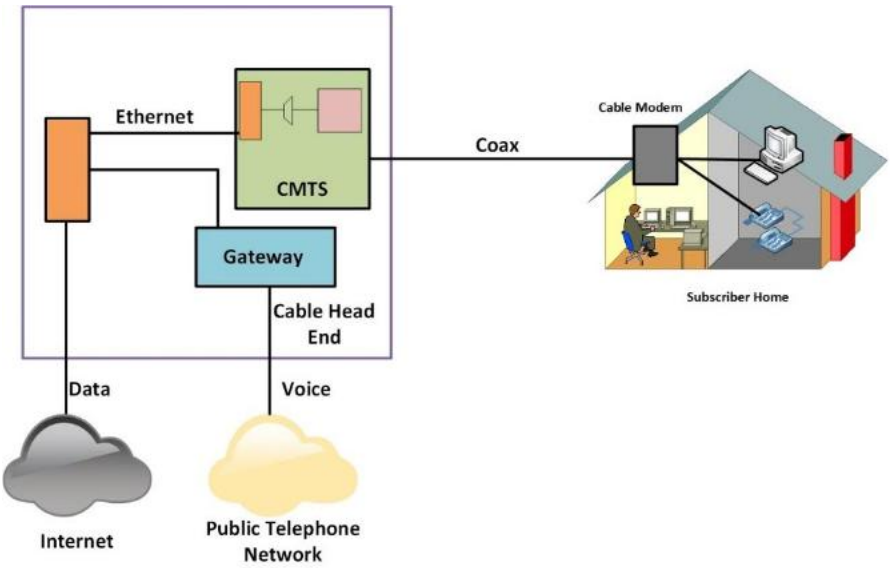


Figure 2-17 *CMTS (Cable Modem Termination System)*

Cable Broadband service is mostly used in United States. It is used in some European countries as well but more and more FTTX deployment is seen in Middle East, Europe and Africa.

Fixed Wireless Service

WISPs provide a fixed wireless Internet service. This means that it relies upon a direct, line-of-sight connection from the access point to the roof of your home. Access point in WISP environment is commonly known as Base Station (BS). WISP shouldn't be confused with mobile wireless technologies such as satellite and mobile broadband.

Satellite can provide Fixed or Mobile options and will be explained in the next section. WISP provides a fixed service which means end user devices are stationary (Fixed WISP service) or within very close diameter (Hot Spot WISP service). Fixed, Hot Spot and Hybrid WISP services will be explained later in this chapter.

https://t.me/learningnets



Figure 2-18 *WISP Network Components*

There are three main components in WISP networks, these are Base Stations (Access Point), Client Premises Equipment (CPE) (Reception device in figure 2-18) and Backhaul network.

Base Stations are the equipment used to distribute the wireless from a single point, mostly located on a roof, water tower or tall buildings in order to transmit signal over obstacles such as trees and buildings.

Base Stations allow for transmission of wireless signals at a range of anywhere between a couple of hundred meters up to distances of around 20km+ depending on the base station equipment used and environmental factors such as interference. On the receiving end, at the house or office that was mentioned earlier, you will need a device called a CPE. This is in order to receive and transmit data wirelessly to and from the Base Station. These come in many shapes and sizes but the type and size of CPE mainly depends on the distance, signal strength and overall performance required.



Figure 2-19 *Different types of WISP Customers*

https://t.me/learningnets

WISP provides a fixed wireless solution; so, the customer antenna or dish is fixed and the customer cannot use their broadband service outside of their home. WISP is different than a Mobile/Cellular Operator as the Mobile Operator provides an Internet service that can also be used outside of homes.

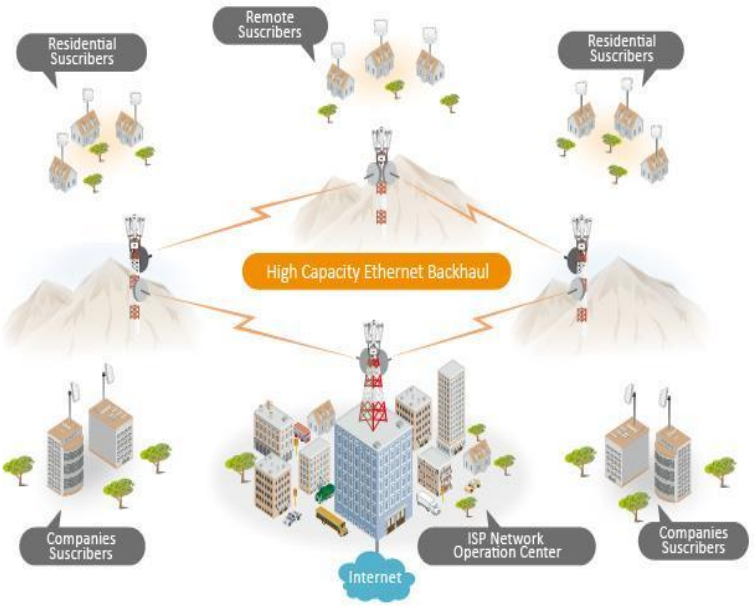


Figure 2-20 *WISP Setup and Backhaul Infrastructure*

In figure 2-20, Typical WISP setup and backhaul infrastructure are shown. Backhaul speed depends on bandwidth between the tower and the central hub location. WISP can be an only choice in rural areas for high speed broadband as bringing fiber or other technologies are economically not profitable in those areas. CPE (Customer Premise Equipment) is shown in figure 2-21. Generally, WISP CPE is located at the roof of the house.



Figure 2-21 *Typical CPE of a WISP Customer*

https://t.me/learningnets

In order to provide Internet access to end users, WISP needs a circuit to the Internet. This can be provided by terrestrial services such as fiber or microwave point-to-point circuits, connecting your WISP location to a telecom provider, which will deliver the user traffic to the Internet.

In many of the places that new WISPs are being established, these backhaul options are not available or are very expensive. Often, satellite broadband as we will see in the next section is the only reliable option to connect the community to the Internet backbone.

Different WISP Service Models

There are three types of WISP services in general. These are Hot-spot WISP, Standard WISP and Hybrid WISP. So far, we have been describing the Standard WISP setup.

Hot Spot WISP: A directional or Omni-directional access point is placed on a building or tower near a target location with subscribers such as a village, an entertainment venue, stadium, airport, etc. It delivers a powerful WiFi signal, so that those within the distance (typically up to 500 meters) can receive WiFi and if authorized, connect to the network.

This service option; leverages the standard WiFi built into most smartphones and laptops. By using repeaters or point-to-point bridges, additional towers can be installed, linked together, and more area can be covered with a WiFi signal. Subscribers of this service are typically the individual users.

Standard WISP: Similarly, an access point/base station is installed on the tower, but a different wireless technology, P2MP (Point to Multi-Point) is used, supporting much longer distances – several kilometers. In this case, small antennas or CPE's (Customer Premise Equipment) are installed at residences and offices.

CPE provides an Ethernet connection to which a switch or WiFi router can be connected to it, so that those within the building can connect using WiFi, or standard wired Ethernet. With Standard WISP service, subscribers are generally businesses or residences.

Hybrid WISP: A hybrid WISP design might use the standard WISP solution which includes CPE devices, to distribute the network to longer distances, where Hot Spot services might be used for users with their own WiFi devices. In this case, subscribers might be businesses and residents, as well as individual subscribers who use the Hot Spot services.

https://t.me/learningnets

WISP Billing Models

Following are the two basic WISP billing models:

1. **Unlimited services billing**: The subscriber typically signs up for a particular plan with a fixed MRC monthly recurring cost. This model is attractive to businesses that want a fixed price every month, and have no concerns about running out of capacity.

2. **Usage based billing:** These services are designed for residential clients and individual subscribers. A plan will typically include some data rate along with a data cap (quota). Data cap can be described as how much traffic the user is allowed to pass over the network within a time period, such as a month.

The WISP needs to be able to help the subscriber select a plan that will meet their needs.

Satellite Broadband

Satellite communication enables the connection for most of the rural areas and underserved areas (even in 2019) to the Internet. In this section, Satellite broadband system component and the challenges which Satellite broadband systems encounter, will be explained.

In addition to broadband communication, Satellite systems can provide weather forecasts, GPS, or satellite TV broadcasting. Other common satellite applications, such as extending cellular coverage, connecting ATM machines, and restoring communication infrastructure quickly among others, provides benefits to end users, governments and the companies.

Satellite communication networks are radio communication networks where a communication link is facilitated between two earth stations by using satellite, which serves as an in-space signal repeater for the transmitting earth station.

A satellite communication link can be made in one of two directions: earth-to-space, known as the uplink; and space-to-earth, known as the downlink.

Residential satellite broadband is generally provided through Geostationary (GEO) satellites. The Main problem with satellite communication is latency (delay), which is not acceptable for some applications such as online gaming or financial trading. Latency also decreases the overall throughput of the communication, so as an example, web browsing might seem very slow.

https://t.me/learningnets

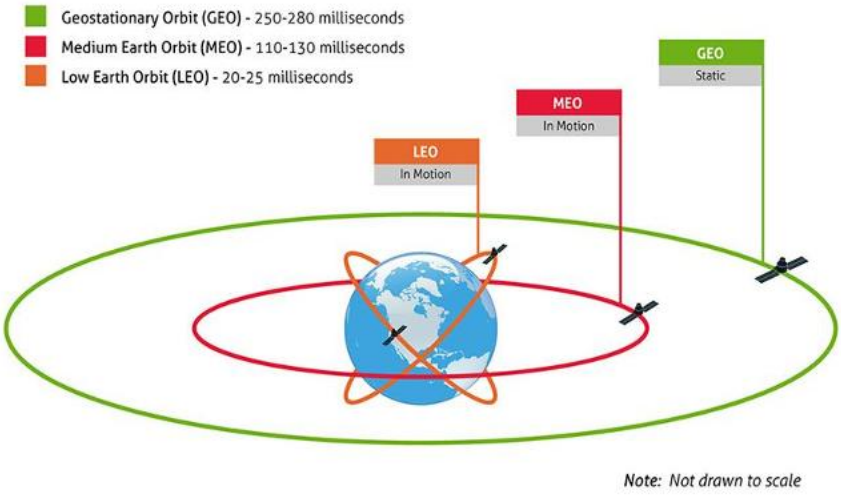


Figure 2-22 *Different types of Communication Satellites*

In figure 2-22, three different Orbits are shown. LEO, MEO and GEO based satellites are categorized based on their distance from earth. Based on the distance between Earth and the Satellite, the Satellite can be positioned in one of these three orbits.

LEO based systems have the lowest latency. On the other side, GEO based systems have the highest latency. The reason is that LEO (Low Earth Orbit) has the closest Orbit to the Earth, while GEO (Geostationary Earth Orbit) has the farthest Orbit to the Earth. Since latency is proportional to distance, closer orbit satellite system (LEO) has the lowest latency.

While broadband requirements have traditionally been met through technologies like fiber, copper, microwave and 2G/3G/4G, satellites can now deliver connectivity with similar performance, including multi-gigabit speeds and low latency.

Common thought by many network engineers and consumers about satellite is that it is slow. This is mainly because most of the satellite systems historically were deployed at the GEO orbit to cover the entire world with very few numbers of satellites.

But there are currently more MEO (Medium Earth Orbit) and LEO based communication satellite systems, operating in many different satellite bands, providing Broadband Internet access or VPN services to the end users and to the commercial business.

https://t.me/learningnets

Satellite System Components

VSAT stands for very-small-aperture terminal which is the customer's site antenna LNB. Indoor modem is a satellite modem which transfers data from the computer to the Antenna. Satellite systems (technologies) are either GEO, MEO or LEO based satellites.

Ground Stations are also called as Satellite Hub or Teleport and control all communications over the satellite link between the user and the Internet.



Figure 2-23 *Customer site VSAT Antenna (VSAT)*

VSAT dishes are commonly used in GEO based communication satellite systems. They are deployed at the residential or business side to receive satellite signals, focus the received energy and transmit the signals to the satellite modem which is located in the indoor location.

Figure 2-24 shows a satellite modem, or commonly referred as Satellite CPE. These modems are located indoor, receive signals and transmit the signals to the Wireless Access Points. These modems can also be manufactured with wireless capability.



Figure 2-24 *Satellite Modem with Coax and Ethernet Cables*

https://t.me/learningnets

Satellite Earth station or commonly called Ground Station as well, is shown in figure 2-25. Satellite signal is sent from Earth Station to the real satellite in the sky and vice versa. Earth station is connected to the Internet and user Internet request arrives to the Earth station.



Figure 2-25 *Satellite Ground (Earth) Station*

In figure 2-26, the actual satellite is shown. Some satellites are 10 meters high but some satellites are just couple centimeters long. Depending on the purpose, there are many different types of Satellite systems.



Figure 2-26 *Communication Satellite*

In this chapter, we are only talking about Communication Satellites which can provide us Broadband access connectivity. There is an O3B satellite system which is located in the MEO orbit and works in KA band (Higher frequency, high rain fade) just couple meters tall. In general, newer communication satellites are smaller compared to older satellite systems, such as ones who were launched 10 – 15 years ago.

After understanding the Satellite Internet components, let's have a look at how it works and what is the actual path when users request content from Internet. After that, we will also discuss the latency issue with Satellite Internet.

https://t.me/learningnets

How Satellite Internet works?

The orbiting satellite transmits and receives its information to a location on Earth called the Network Operations Center (NOC). NOC is connected to the Internet so all communications made from the customer location (satellite dish) to the orbiting satellite will flow through the NOC before its reached to the Internet and the return traffic from the Internet to the user will follow the same path.

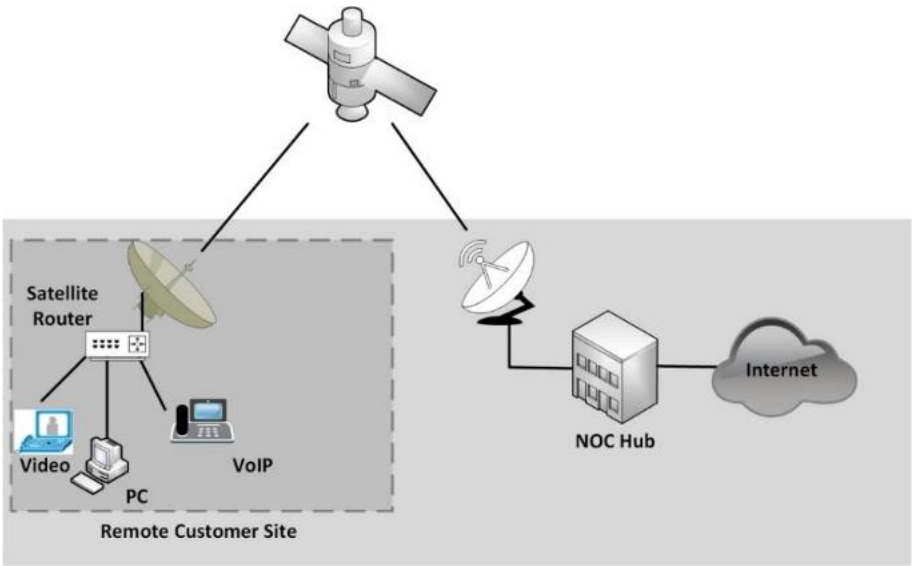


Figure 2-27 *End to end Satellite Internet communication*

Data over satellite travels at the speed of light and Light speed is 186,300 miles per second. The orbiting satellite is 22,300 miles above earth (This is true for the GEO based satellite).

The data must travel this distance 4 times:

1. Computer to satellite
2. Satellite to NOC/Internet
3. NOC/Internet to satellite
4. Satellite to computer

This adds a lot of time to the communication. This time is called "Latency or Delay" and it is almost 500 milliseconds. This may not be seen so much, but some applications like financial and real-time gaming don't like latency. Who wants to pull a trigger, and wait half a second for the gun to go off?
But, latency is related to which orbit the satellite is positioned. Let's have a look at different Satellite Orbits to understand the satellite latency and its effect to the communication.

https://t.me/learningnets

Geostationary (GEO) Satellites

Geostationary satellites are earth-orbiting about 22,300 miles (35,800 Kilometers) directly above the equator.

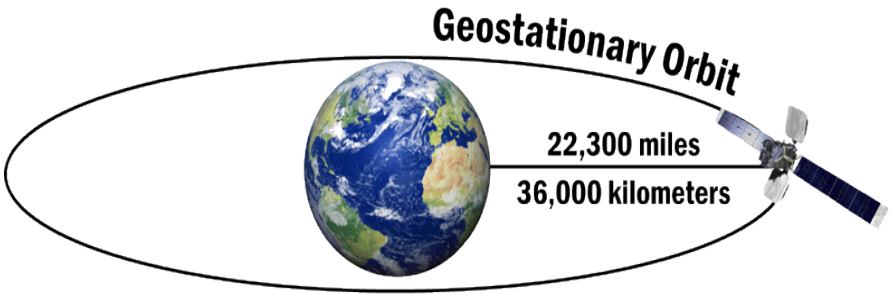


Figure 2-28 *Geostationary Earth Orbit*

They travel in the same direction as the rotation of the Earth. This gives the satellites the ability to stay in one stationary position relative to the Earth. Communication satellites and weather satellites are often given geostationary orbits, so that the satellite antennas that communicate with them do not have to move to track them, so it can be pointed permanently at the position in the sky where they stay.

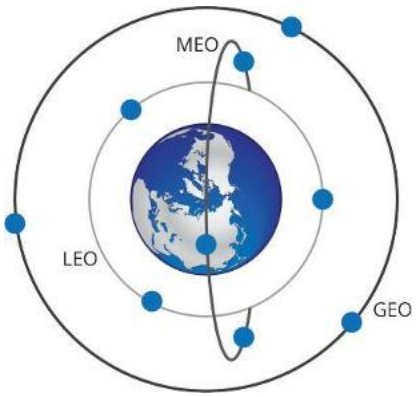


Figure 2-29 *GEO, MEO and LEO Orbits*

The latency in GEO Satellites is very high compared to MEO and LEO Satellites. The geostationary orbit is useful for communication applications, because ground based antennas, which must be directed toward the satellite, can operate effectively without the need for expensive equipment to track the satellite's motion.

https://t.me/learningnets

There are hundreds of GEO satellites in orbit today, delivering services ranging from weather and mapping data to distribution of digital video-on-demand, streaming, and satellite TV channels globally. The higher orbit of GEO based satellite means greater signal power loss during transmission, when compared to lower orbit.

Medium Earth Orbit Satellites

MEO is the region of space around the Earth above low Earth orbit and below geostationary orbit. Historically, MEO constellations have been used for GPS and navigation applications, but in the past five years, MEO satellites have been deployed to provide broadband connectivity to service providers, government agencies and enterprises.

Current applications include delivering 4G LTE and broadband to rural, remote, and underserved areas where laying fiber is either impossible or not cost effective – such as cruise or commercial ships, offshore drilling platforms, backhaul for cell towers, and military sites, among others. In addition, Service Providers are using managed data services from these MEO satellites to quickly restore connectivity in regions where the service has been lost due to undersea cable cuts or where major storms have occurred

MEO satellite constellations can cover the majority of Earth with about eight satellites. Because MEO satellites are not stationary, a constellation of satellites is required to provide continuous service. This means that antennas on the ground need to track the satellite across the sky, which requires ground infrastructure which is more complex compared to GEO based satellites.

Low Earth Orbit (LEO) Satellites

Unlike geostationary satellites, low and medium Earth orbit satellites do not stay in a fixed position in the sky. Consequently, ground based antennas cannot be easily locked into communication with any one specific satellite.

Low Earth orbit satellites, as their name implies, orbit much closer to earth. LEOs tend to be smaller in size compared to GEO satellites, but require more LEO satellites to orbit together at one time to be effective. Lower orbits tend to have lower latency for time-critical services because of the closer distance to earth.

It’s important to reiterate that many LEO satellites must work together to offer sufficient coverage to a given location. Although many LEOs are required, they require less power to operate because they are closer to earth. Choosing to go with more satellites in the LEO orbit on less power, or using fewer larger satellites in GEO, is the biggest decision to make here. Due to the high number of satellites required in LEO constellations, LEO satellites systems are expected to be high initial manufacturing and launch costs and more expensive ground hardware compared to GEO.

https://t.me/learningnets

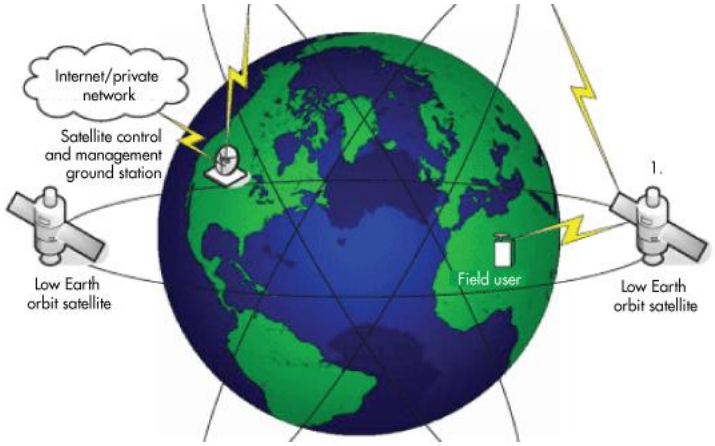


Figure 2-30 *LEO Satellite*

Satellite Operating Bands

Before sharing the different frequencies (bands) which are used in satellite communication, it is important to understand the phenomena called 'rain fade'. Rain fade is an interruption of wireless communication signals as a result of rain, snow, or ice, and losses which are especially prevalent at frequencies above 11 GHz. Satellite communications use microwave frequencies, which require direct line of sight between the receiving and transmission equipment.

The following frequency ranges are commonly used:

C Band (4-8GHz): These lower frequencies have longer wavelengths and require larger dishes (1.8-2.4m) for reception, but are not affected by "rain fade". Larger dish size requires more expensive equipment compared to Ku and Ka bands. The C-band frequency range has one significant problem. It is the frequency region assigned to terrestrial microwave radio communication systems.

There are an emerging number of these microwave systems located all over the world and they carry a large volume of commercial communications. Consequently, the VSAT locations are needed to be restricted in order to prevent interference with the terrestrial microwave communication systems.

C-band was the first band that was used for satellite communication systems. However, when the band became overloaded (due to the same frequency being used by terrestrial microwave links), satellites were built for the next available frequency band, the Ku-band.

https://t.me/learningnets

Ku band (12-18GHz): A shorter wavelength permits smaller dishes. The Ku-band frequency range is allocated to be exclusively used by satellite communication systems, thereby eliminating the problem of interference with microwave systems.

Due to higher power levels at new satellites, Ku-band allows for significantly smaller earth station antennas and RF units to be installed at the VSAT location. Ku-band is typically used for broadcasting and Internet communication.

Ka band (26.5-40GHz): Ka-Band is a relatively new frequency band for satellite broadband and will provide additional transmission capacity. Ka-band has several advantages, with perhaps the most significant as having the 2-3 GHz increase in bandwidth, which is double available bandwidth than in Ku band and five times more than C band.

Due to the smaller wavelength, Ka-band components are typically smaller, leading to smaller antennas on the same-sized platform. But its sensitivity to rain fade makes it difficult for rainy regions and will support the use of small antennas (<1.2m).

It is good to mention about recent ideas in communication satellite systems. Space-X, Google, Facebook, and many other Internet giants have low earth orbit or low altitude satellite communication system plans.

Elon Musk's company SpaceX, has been granted permission by the US Federal Communications Commission (FCC) to set up a vast network of thousands of low Earth orbit communication satellites. His Satellite project name is Starlink.

Starlink will mostly be targeting to high-frequency traders at big banks, who might be willing to fork out large sums for dedicated, faster connections. Initially, Starlink will consist of 4425 satellites orbiting between 1100 and 1300 kilometers up, a greater number of active satellites compared to what are currently in orbit.

When sending a network traffic via Starlink, a ground station will begin by using radio waves to talk to a satellite above it. Once in space, the message will be fired from satellite to satellite using lasers until it is above its destination.

From there, it will be beamed down to the right ground station using radio waves again. Between distant places, this will allow messages to be sent about twice as fast as through the optical fibers on Earth that currently connect the Internet, despite having to travel to space and back. This is because the speed of the signal in glass is slower than it is through space.

For most people, the regular Internet is fast enough already. But for certain applications such as high-frequency trading or online gaming, speed/latency is the most important aspect of communication.

https://t.me/learningnets

Mobile Service Technologies

Mobile broadband also referred to as WWAN (Wireless Wide Area Network), is a general term used to describe high-speed Internet access from mobile providers for portable devices. Just as wireless networking evolved from 802.11b to 802.11n with faster speeds and other improved performance and security features, mobile broadband performance continues to evolve, with having so many players in this growing field. Access network infrastructure links the backbone network to the customers.

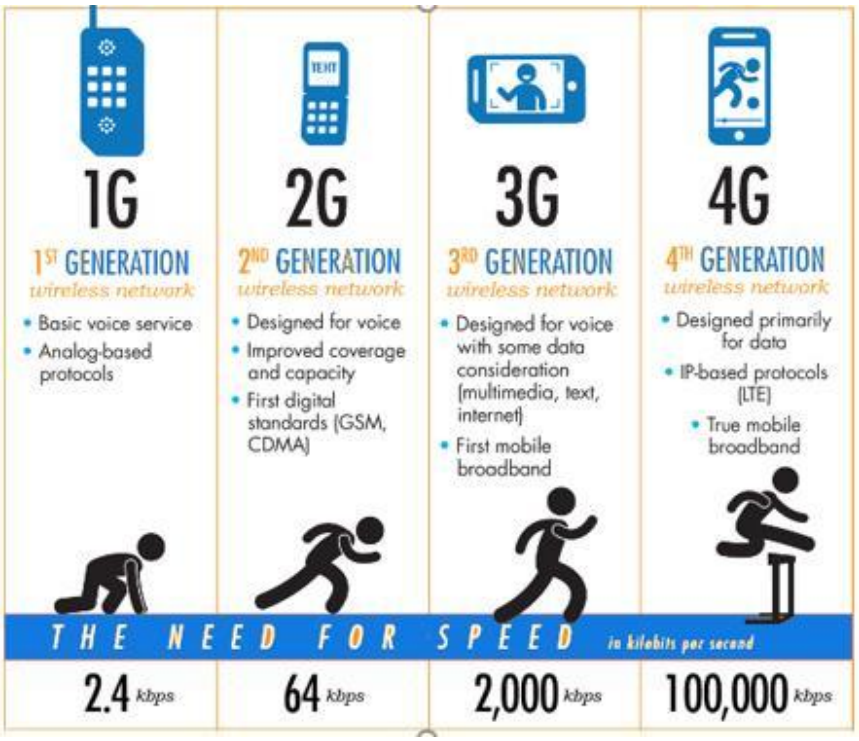


Figure 2-31 *Speed Evolution from 1G to 4G*

Started from 1G as 2.4kbps, Mobile broadband as of 2018 can provide more than 100Mbps.

https://t.me/learningnets

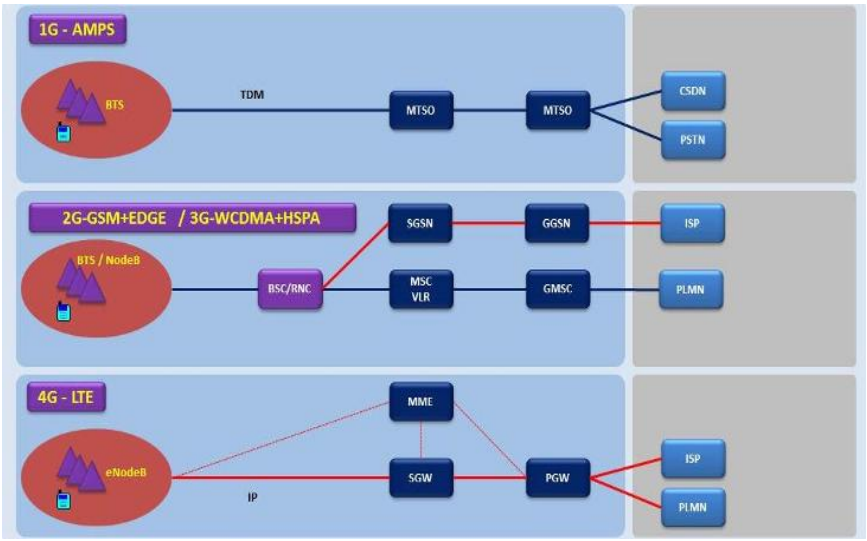


Figure 2-32 *Architectural Evolution from 1G to 4G*

LTE

LTE Architecture – Terms

UE is a User Equipment. SIM Card + Mobile device together is called UE.

eNodeB is Enhanced Node B, it is an enhanced base station because radio controller is located in the same place with the Base Station, this is different from 3G architecture.

MME is the mobility management entity. MME is the key control node for LTE access network. It is responsible for tracking and paging procedures including retransmissions, and also for idle mode of User Equipment (UE).

Serving GW is the gateway which terminates the interface towards E-UTRAN. For each UE associated with the EPS, at given point of time, there is a single Serving GW. SGW is responsible for handovers with neighboring eNodeB's, also for data transfer in terms of all packets across user plane which is known as data plane by many network engineers.

PGW is the gateway between the Radio Access Network and the IP network. It handles the User IP address allocation as well.

https://t.me/learningnets

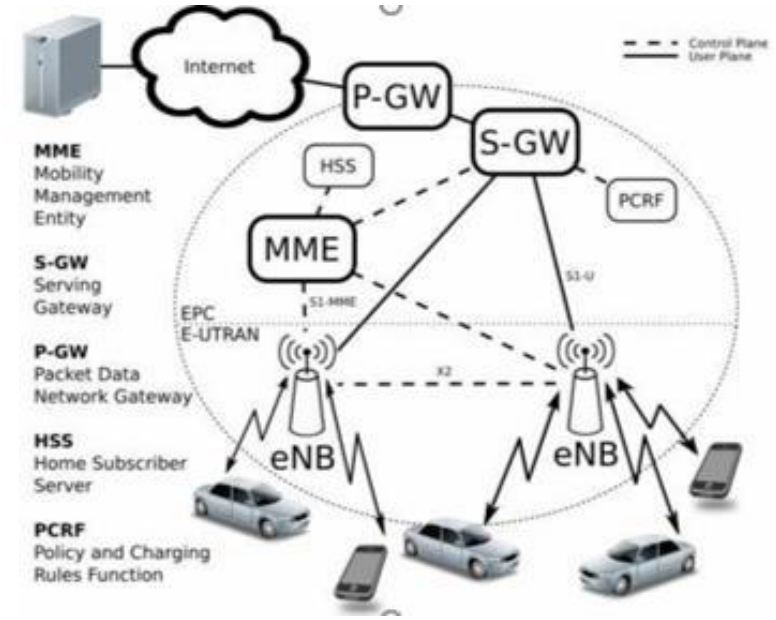


Figure 2-33 *eNB, MME, S-GW, P-GW*

From the classical/traditional network engineering point of view, PGW is the mobile networking node which is connected to the IP Core/backbone network. Mobile broadband comes with data cap. This limits the usability in home networks for using HD quality videos. Ultra HD video consumes so much bandwidth.

Although the cost of having frequency and deploying small cells to provide enough bandwidth puts pressure on mobile carrier, broadband penetration of mobile is increasing very rapidly.

4 Key Design Principles of Mobile Networks

Not all Mobile network systems are designed in Cellular based, but as of 2018, almost all Mobile network system designs are based on Cellular architecture.

Let me explain what would be the other deployment option for mobile networks, other than cell based systems and then we will highlight the 4 main characteristics of cell based mobile networks. With the previous cellular systems design, mobile network operators used to place their radio transmitters at the tallest buildings in the area which they wanted to provide coverage. Single, very high-power transmitters were used to cover very large geographic areas.

https://t.me/learningnets

Design principles of cell based mobile phone networks, are as follow:

1. Cellular Based Mobile Networks: With the cell based telephone systems, so many low-power, small coverage area transmitters are used instead of a single, powerful, monolithic transmitter to cover a wide area.

2. The second design principle which is frequency reuse, takes into account the fact that cellular telephony, like all radio-based services, has been allocated a limited number of frequencies by the regulators. Frequency reuse is commonly used by the wireless Internet Service Providers as they have serious frequency limitation issues. By using geographically small, low-power cells, frequencies can be reused by non-adjacent cells.

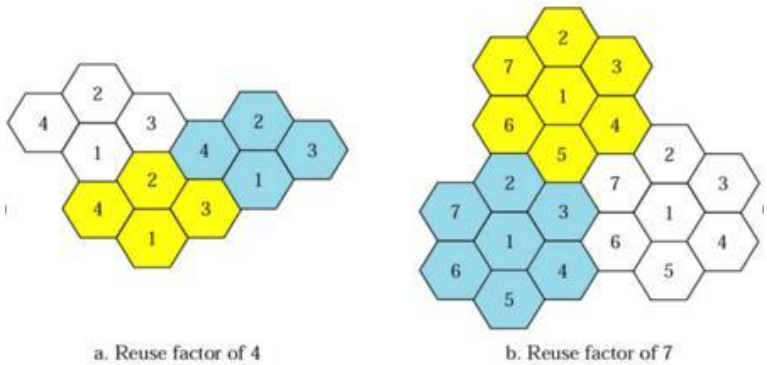


Figure 2-34 *Cellular networks frequency reuse*

3. Cell Splitting: When usage areas become saturated by excessive usage, cells can be split. This is the third design principle of cellular based mobile phone systems.

   When the network operator realizes that the number of callers are refused because of congestion reaching a critical level, they can split the original cell into two cells, by installing a second site within the same geographical area that uses a different set of non-adjacent frequencies. This has an extended impact; as the cells become smaller, the total number of calls that the aggregate can support increases dramatically.

https://t.me/learningnets

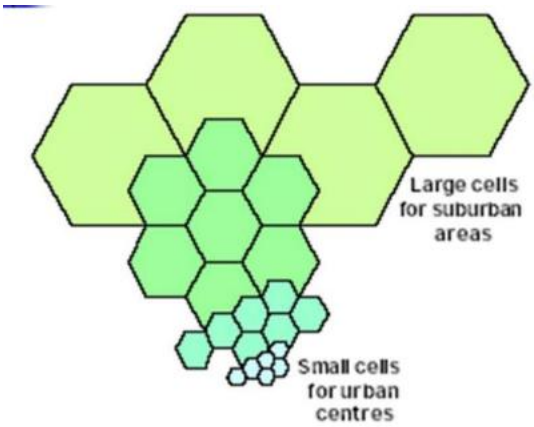


Figure 2-35 *Cell Splitting*

The smaller the cell, the larger the total number of simultaneous calls that can be accommodated. Of course, this also causes the cost of deployment to increase exponentially.

4. The Fourth and last design principle of cell based phone systems is cell-to-cell handover, which simply means that the cellular network has the ability to track a call as the user moves across a cell. When the signal strength detected by the current cell is perceived by the system to be weaker than the signal strength of the second cell, the call is handed over to the second cell.

   Ideally, the user doesn't sense the handover. Cell handover and other cellular system capabilities are under the central control of a mobile telephone switching office (MTSO, sometimes called as MSO – Mobile Switching Office or MSC – Mobile Switching Center)

https://t.me/learningnets

Summary

In this chapter, we provided a brief overview of the architecture of Service Provider Access types and showed various different types of Broadband Services.

Then various broadband access technologies were discussed and the major features of DSL, HFC, PON, Wi-Fi, WiMAX, Cellular and Satellite networks were presented in detail.

DSL can offer data rates over 10 Mb/s within a short distance. Efforts to develop next-generation DSL focus on increasing data rates and transmission distance.

With DSL, voice and data can be supported by a single phone line. In the past decade, DSL has become one of the dominant broadband access technologies worldwide. However, TV broadcasting and IP TV are still a technical challenge for DSL networks. With using VDSL and advanced data compressing techniques, transferring video over IP may be offered in the near future.

Traditionally, Cable TV companies offered analog TV broadcasting. With the development of many protocols and technologies such as DOCSIS, HFC, CMTS, Cable Modem, etc. Broadband Internet access is provided to subscribers.

Cable Access Providers have added VoIP and digital TV over HFC networks. Further development has extended the cable plant beyond 1 GHz, and higher data rates are the main focus of HFC networks.

As optical fibers can provide essentially unlimited bandwidth, PON (Passive Optical Network) is considered the most promising wired access technology. Most of the FTTx deployment is based on PON today.

In addition to wired broadband solutions, many wireless technologies have been developed to provide broadband Internet access, such as free-space optical communications, Wi-Fi, WiMAX, cellular and satellite networks.

Free-space optical communications can support gigabit per second data rates and a transmission distance of a few kilometers, but the atmospheric effects impose severe constraints for free-space optical communications.

Wi-Fi is widely used for wireless local area networks with transmission distances up to a few hundred meters. With a mesh topology, Wi-Fi networks can support extended reach and broadband Internet access. WiMAX can support wireless access over 10 km, but it requires higher transmitted power and the data rate is lower than in Wi-Fi networks.

https://t.me/learningnets

Cellular networks are used primarily for mobile voice communication. With digital encoding technologies, data transmission service is added to the Cellular networks. Currently 5G is developing in many countries and commercial deployment is expected in 2020.

Satellite networks are used primarily for direct video distribution, but data service over satellite is covering a large geographical area. The main disadvantage of satellite communication in the past was limited data rates. Many new satellite systems have been launched in recent years and it is possible to receive 10s of Mb/s of Broadband access over satellite with very low latency due to LEO (Low Earth Orbit) and MEO (Medium Earth Orbit) satellite systems.

The advantages of wireless technologies are mobility, scalability, low cost, and ease of deployment. Except for free-space optical communication, other wireless technologies use RF or microwave frequencies. Network capacity and reliability with Wireless are much lower compared to wired access networks. Emerging multimedia applications demand broadband access networks, and a number of wired and wireless broadband access technologies have been developed over the past decade. By solving the last mile bottleneck problem, it will be possible to use a variety of new applications.

https://t.me/learningnets

Chapter-3 Service Provider Physical Connectivity and Transport Network

Introduction

Service Providers use different types of physical connectivity as the transport to connect the customers to their network. Customers such as Residential Customers, Business/Corporate Customers, etc. (that have been explained in the previous chapter), all connect to the Service Provider network through many different types of physical connectivity layer, that will be explained in detail in this chapter.

Also in this chapter, transport network basics such as Fiber optic, Microwave, Comparison of Fiber and Microwave, SONET/SDH, WDM and Dark fiber will be covered. Terrestrial and Sub Marine/Undersea Cable Systems and the components of these systems will be shared.

Fiber Optic

Fiber-Optic cables carry information between two places using entirely optical (light-based) technology. A fiber-optic cable is made up of incredibly thin strands of glass or plastic known as optical fibers; one cable can have as few as two strands or as many as several hundreds. Each strand is less than a tenth as thick as a human hair and can carry something like 25,000 telephone calls, so an entire fiber-optic cable can easily carry several million calls.

https://t.me/learningnets

Total Internal Reflection

Fiber technology works based on total internal reflection. The basic functional structure of an optical fiber consists of an outer protective cladding and an inner core through which light pulses travel.

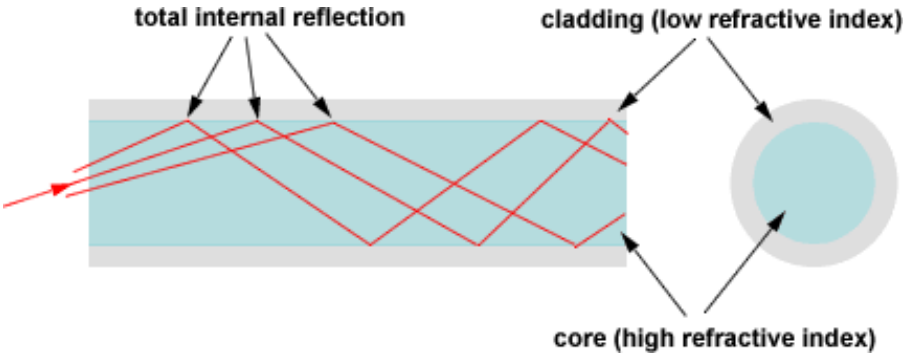


Figure 3-1 *Total Internal Reflection – How Fiber works*

The difference in refractive index of the cladding and the core allows **total internal reflection** in the same way as happens at an air-water boundary, as shown in figure 3-1. Thus, fiber optics use total internal reflection.



Figure 3-2 *Total Internal Reflection can be seen at the air-water boundary*

https://t.me/learningnets

Fiber Optic Cable Installation

Fiber can be installed in three ways:

- Aerial fiber optic cable installation
- Direct burying
- Sub marine fiber optic cable installation



Figure 3-3 *Aerial Fiber Optic Cable Installation*



Figure 3-4 *Direct Buried Fiber Optic Cable Installation*

https://t.me/learningnets



Figure 3-5 *Submarine Fiber Optic Cable Installation*

Fiber Optic Cable Types

Single Mode Fiber

The simplest type of optical fiber is called **single-mode**. It has a very thin core about 5-10 microns (millionths of a meter) in diameter. In a single-mode fiber, all signals travel straight down the middle without bouncing off the edges. Cable TV, Internet, and telephone signals are generally carried by single-mode fibers.

Multimode Fiber

Each optical fiber in a multi-mode cable is about 10 times bigger than one in a single-mode cable. This means light beams can travel through the core by following a variety of different paths, in other words, in multiple different modes. Multi-mode cables can send information only over relatively short distances.

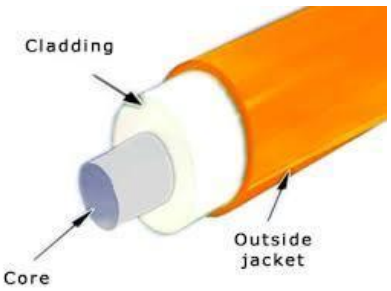


Figure 3-6 *Fiber Optic Cable*

https://t.me/learningnets

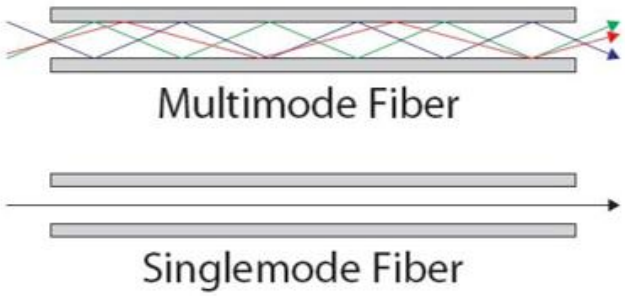


Figure 3-7 *Single vs. Multimode Fiber Light Propagation*

Microwave

Within the broader spectrum of radio frequency (RF) communications, point-to-point communications are usually carried out using microwave frequencies between 1 GHz and 100 GHz along line-of-sight paths called links. These frequencies and their propagation characteristics allow the transmission of vast amounts of data between remote communication sites without the need to lay cables between them.

The first practical application of microwaves in a communication system took place more than 80 years ago.

In the 1930s, an experimental microwave transmission system was used to connect United Kingdom with France—bridging the English Channel without cables.

Microwave is commonly used in the "Backhaul" part of the networks. When Microwave is used for backhaul purpose in Mobile Operator networks, it is called Microwave backhaul. Before we explain Microwave Backhaul further, it is important to share the definition of "Backhaul".

The term backhaul is often used in telecommunications and refers to transmitting a signal from a remote site of the network to another site, usually a central one. Backhaul usually implies a high-capacity line, meaning high-speed lines capable of transmitting high bandwidth at very fast speeds. Carrying the data, voice or video traffic from Cell Site (Access Network in the Mobile Operators) to Core network is known as Backhauling the traffic.

Many Mobile Operators transitioned their networks from 2G/3G radio access to 4G access technologies. These networks needed to deliver far more bandwidth to support richer applications while demand for backhaul capacity has been growing since the 2G era.
Microwave systems have emerged as a good way to deliver capacity with a better pricing compared to fiber option. The mobile backhaul infrastructure carries voice and data communication traffic from the Access network (cell sites) to the Core network.

https://t.me/learningnets

Most cell sites are connected to the Core network via 100Mbps to 1Gbps as of 2019 in the Mobile Operator networks.

Microwave has been used by many Mobile Operators for years, but when it comes to speed comparison, which one is faster? Question is not asking about the bandwidth capacity, it is the speed comparison. When it comes to bandwidth, Fiber is an obvious answer. But let's have a look at speed.

Microwave or Fiber, which one is faster?

As it is mentioned above, most of the Mobile (Cellular) networks use microwave to connect their cell towers to backhaul networks. For mobile operators, a reason using microwave is not the speed which microwave provides. It is used to connect their remote sites (Rural areas), because microwave is a faster and cheaper deployment option compared to fiber. (Reliability and security is a different aspect which is out of the scope of this book).

When more capacity (bandwidth) is required, fiber becomes more economical. The actual cost of fiber deployment is the laboring cost. For example, digging a trench and laying the fiber, getting the required permissions from the land owners and from the municipalities. You have to dig a trench that's hundreds (or thousands) of kilometers long, or lease access to ducts that have already been laid by infrastructure companies.

Geography of the land is very important for fiber deployments. For example, when faced with a mountain or river, do you go straight across at great expense, or do you make a diversion to the nearest bridge or tunnel?

Combine all of these factors and you'll understand why most of the world's terrestrial fiber networks are deployed alongside existing roads and railways. Let's go back to our initial discussion. Microwave or Fiber, which one is faster?



Figure 3-8 *Microwave Tower with many Radios*

In other words, which one has lower latency? (More bandwidth means more people can send a data over the link but faster means lower latency).

https://t.me/learningnets

A radio signal travels through air at just under the speed of light (299,700km per second).

Through a glass or plastic fiber, where light has to bounce along the refractive index rather than travel in a straight line, the speed of light is reduced to around 200,000km/s. This fact is known by transport network engineers but not by most of the network engineers.

Why is this important? Let's say we have two buildings with 200km distance between them. With microwave (radio signal through air), latency between the two buildings would be 1,064 milliseconds or 1064 microsecond. With fiber, the latency between the buildings would be 1.594 millisecond or 1594 microsecond. But really, is 500 microsecond difference critical? Yes, in fact for some business, it is a lot.

Have you ever heard HFT (High Frequency Trading) networks? In HFT business, end to end latency is called tick-to-trade latency. Tick is market data and trade is a buy/sell or any other type of order (Tick data flows from the stock exchanges to the HFT Company, Trade data flows from the HFT to stock exchanges).

The whole business relies on having a faster network than any other company in the market. Having an information from the exchanges, carrying it through the network to your servers, processing it and taking an action based on "algo" (Algorithm but in HFT environment there are tons of jargon) all happens under less than 100 microseconds.

In fact, some HFT vendors claim that they provide tick-to-trade latency as less than 50 microseconds. Transport medium such as microwave and fiber contribute to propagation and serialization delays.

Propagation delay is what I explained above, how long it would take to carry data between point A and B. Microwave is the winner in our comparison.
Serialization delay means, how long it would take to place a data into a wire from the networking device. It is a known fact that 10G fiber has much lower serialization delay than 1G fiber.

To be more precise, for 128- byte data, 1G (Gigabit Ethernet) serialization delay is 1 microsecond, 10G serialization delay is 0.1 microsecond. It seems that this is not a large difference, but you should consider that the delay would be accumulated at every port. In general, companies that desire to deploy low-latency infrastructures would implement 10 Gigabit Ethernet wherever possible. Especially in the HFT (High Frequency Trading) business.

https://t.me/learningnets

SDH/SONET

Synchronous Optical Networking (SONET) and Synchronous Digital Hierarchy (SDH) are standardized protocols that transfer multiple digital bit streams synchronously over optical fiber. SONET and SDH are essentially the same, SONET is used mostly in U.S and Canada, SDH is used in rest of the world. They were designed to support real time applications, thus it was made for carrying circuit mode transport such as E1, T1, DS3. Today, both SONET and SDH can transport ATM, Ethernet and IP as their client protocol.

Synchronous Digital Hierarchy (SDH) Level	SONET	Speed (Mbps)
STM-0	STS-1 (OC-1)	51.84
STM-1	STS-3 (OC-3)	155.52
STM-2	STS-6 (OC-6)	311.04
STM-3	STS-9 (OC-9)	466.56
STM-4	STS-12 (OC-12)	622.08
STM-8	STS-24 (OC-24)	1244.16
STM-16	STS-48 (OC-48)	2488.32
STM-32	STS-96 (OC-96)	4976.64
STM-64	STS-192 (OC-192)	9953.28
STM-256	STS-768 (OC-768)	39,813.12

Figure 3-9 *SONET and SDH Speeds*

SONET/SDH is Mostly Obsolete Now!

With SONET/SDH, the maximum speed is 40Gbps; the reason is that the development of these technologies has stopped. Both SONET/SDH are now obsolete technologies. There are multiple reasons, but mainly these technologies don't carry multiple wavelengths over single optical fiber and cost of the equipment is too much compared to Ethernet.

Today, Ethernet is ubiquitous and used in LAN, MAN and WAN, thus mass production of the Ethernet equipment made Ethernet products so cheap. With SDH you can only transmit a single wavelength per fiber and that is typically either 1310nm (short-haul) or 1550nm (long-haul). This is also referred to as black & white optics.

SDH does not scale very well and begins with the basic block of an STM-1 and scales up by a factor of 4 (i.e. STM-4, STM-16, STM-64, and STM-256). SDH was deployed in metro areas and can have a maximum reach of approximately 80km using 1550nm optics. SONET/SDH can be carried over WDM but not vice versa.

https://t.me/learningnets

WDM

WDM stands for Wavelength division multiplexing. There are two different WDM mechanisms, DWDM and CDWM. WDM is considered Layer 0 by some transmission people and SONET/SDH as Layer 1. Some transmission people consider WDM as Layer 1 and SONET/SDH as Layer 1.5, but keep in mind that SONET/SDH can be a client of WDM but not vice versa.

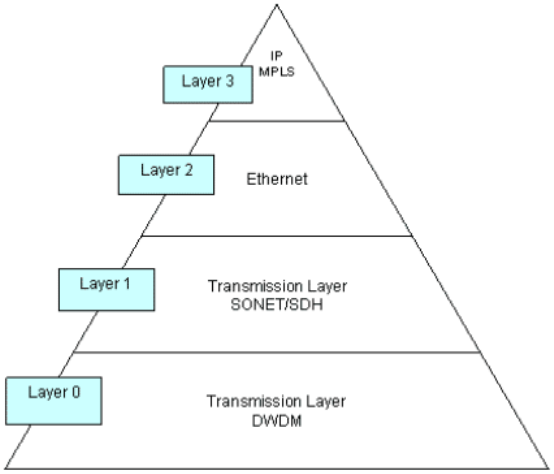


Figure 3-10 *DWDM is considered as Layer 0*

DWDM allows you to transmit multiple wavelengths on a single fiber and as the name suggests, it utilizes wave division multiplexing. DWDM can be deployed in ultra-long haul, long-haul, regional, and metro area's. DWDM also uses amplification to reach long distances and can be an expensive technology to deploy.

On the other hand, CWDM which is a metro technology, multiplexes a maximum of 8 wavelengths per fiber and does not require amplification (i.e. much cheaper than DWDM).

WDM is used to transport SDH/Ethernet/IP between regions or cities or to aggregate traffic where large bandwidth is required.

https://t.me/learningnets

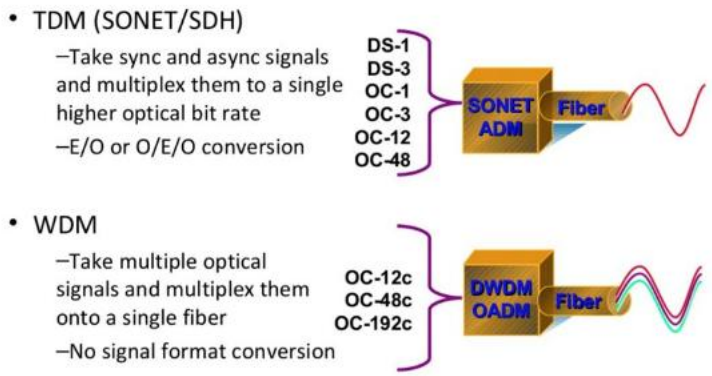


Figure 3-11 *SONET/SDH vs. DWDM*

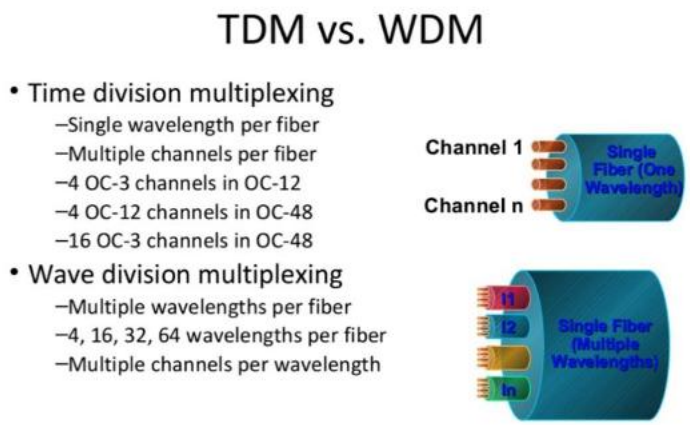


Figure 3-12 *Single vs. Multiple Wavelength in SONET/SDH vs. WDM*

https://t.me/learningnets

DWDM

Dense Wavelength Division Multiplexing (DWDM) is an optical multiplexing technology used to increase bandwidth over existing fiber networks. DWDM works by combining and transmitting multiple signals simultaneously at different wavelengths on the same fiber.

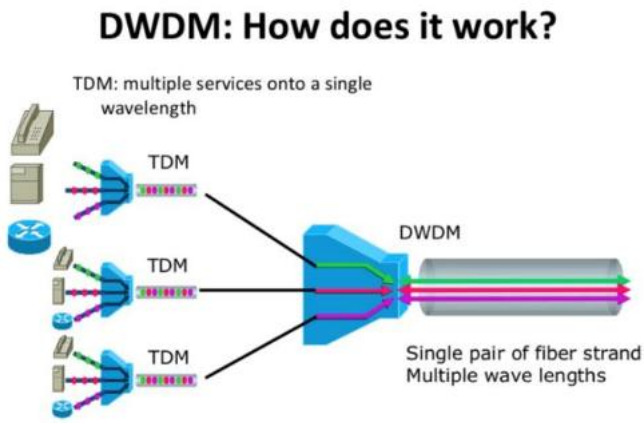


Figure 3-13 *How DWDM works*

IP Transport Evolution on Wide Area Network

The evolution started over ATM, then with POS, and today IPoDWDM (IP over DWDM) is very common. Less number of layers provides lower OPEX and CAPEX, also provides better management and less complexity.

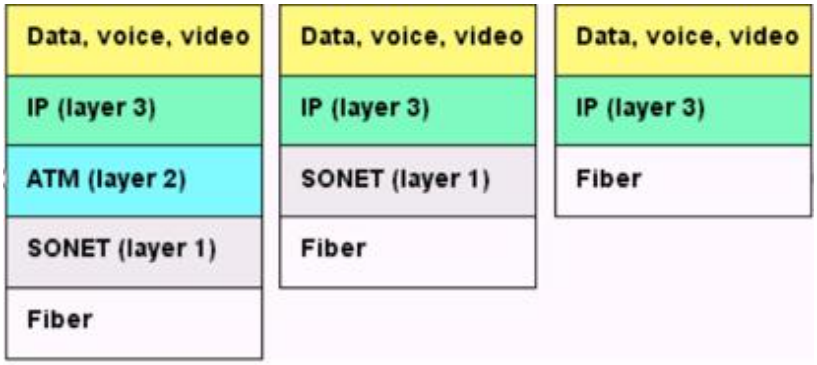


Figure 3-14 *IP Transport Evolution on WAN*

https://t.me/learningnets

Dark Fiber

Dark fiber is the fiber optic infrastructure that's not yet in use by a service carrier or a provider. It is an installation of cables currently lying dormant, ready to be connected at some time in the future. The name "dark" from dark fiber comes from the technology that the literally lacks the lasers that send light through the cables.

Purchasing and Leasing Capacity on Fiber Links

Service Providers can build their own fiber network or buy it from other providers which are generally called Carriers. Buying a fiber can be in two ways, leasing or IRU (Indefeasible rights of use) based.

Indefeasible right of use (IRU)

IRU is a permanent contractual agreement that cannot be undone, between the owners of a cable and a customer of that cable system. Indefeasible means "not capable of being voided or undone". The customer purchases the right to use a certain amount of capacity of the fiber system for a specified number of years.

Service Providers use the transport network of other carriers. This is very common; in fact, even the biggest networks use other carrier's transport/transmission infrastructure, especially outside of their main location. For example, an operator might provide services mainly in U.S, but wants to extend its network to Europe. Instead of setting up a fully-fledged telecom environment in Europe to provide a service, (for example to Business and Residential customers), one option is to use local carrier networks in Europe.

IRU is a permanent contractual agreement that cannot be undone between the owners of a cable and a customer of that cable system. Cable is mostly a fiber cable as fiber can carry more data than any other type of media. Customers who purchase IRU, can lease the capacity to other companies.

Let me give you an analogy.
Think of it in this way, if you are renting an apartment, you sign a contract with the Landlord as a tenant. You cannot rent that apartment to someone else. This is similar to leasing. But if you are the landlord, you can rent it to anyone you want. This is an example for Indefeasible Right of Use based agreement.

https://t.me/learningnets

IRU vs. Leasing a Fiber

IRU contracts are almost always long term, such as 10 to 20 years (Cable lifetime is generally considered as 25 years). Leased fiber doesn't have to be a long term contract. IRU period can start from 10 years. The currently most common version of IRU contract is 15 years. The most common leased service is IPLC which is International Private Leased Circuit.

IPLC can be half circuit or full circuit. IPLC unlike IRU doesn't dictate the buyers to pay the cost of fiber upfront. IPLC is not a prepaid service. Leasing is very flexible but IRU can be very cost effective. IRU gives the purchaser the right to use some capacity on telecommunications cable system, including the right to lease that capacity to someone else.

Should smaller companies purchase an IRU based fiber?

Smaller companies that need a leased line between 2 different countries, for example between London and New York do not buy an IRU. They receive capacity from a telecommunications company. Those telecom companies may also have received a larger amount of capacity from another company (and so on), until at the end of the chain of contracts there is a company that has an IRU, or wholly owns a cable system.

Carrying Network Traffic between Countries

How Internet Traffic is carried between the Countries? Internet traffic is carried between the countries mostly in two ways; terrestrial and submarine cables. Satellite can be used as backup, but limited capacity and latency makes it unattractive and as a last resort, for example U.S FCC (Federal Communication Commission) indicates that satellites account for just 0.37 percent of all U.S. international capacity. Terrestrial cables are used between the landlocked countries in general.

Terrestrial Fiber Optic Cables

There are 49 landlocked countries (According to Wikipedia) in the World, largest is the Kazakhstan and the smallest landlocked country is Vatican. Terrestrial cables are only used on land and a cross border agreement is signed between the countries.

Some terrestrial cables pass several countries. Terrestrial cables carry much higher capacity compared to submarine cables. Terrestrial cables are not only used for inter-country communication but commonly used inside the country for Metro Access and inter-city connections as well.

Terrestrial fiber cables can connect countries even when they have access to submarine cables. For example, Bangladesh operators are using newly deployed Bangladesh to India route, in addition to SEA-ME-WE 4 and SEA-ME-WE-5 Submarine cable systems

https://t.me/learningnets

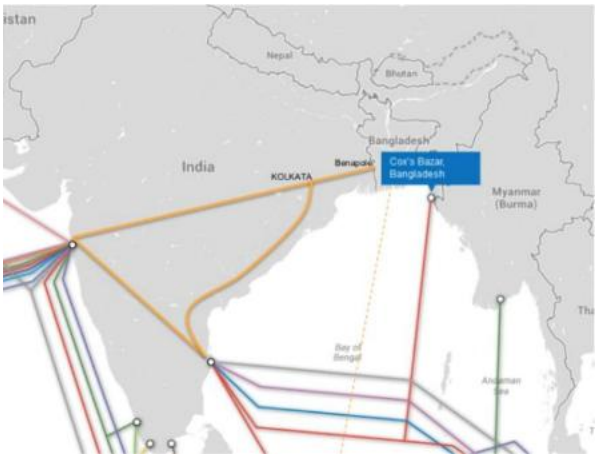


Figure 3-15 *Bangladesh to India Terrestrial Cable*

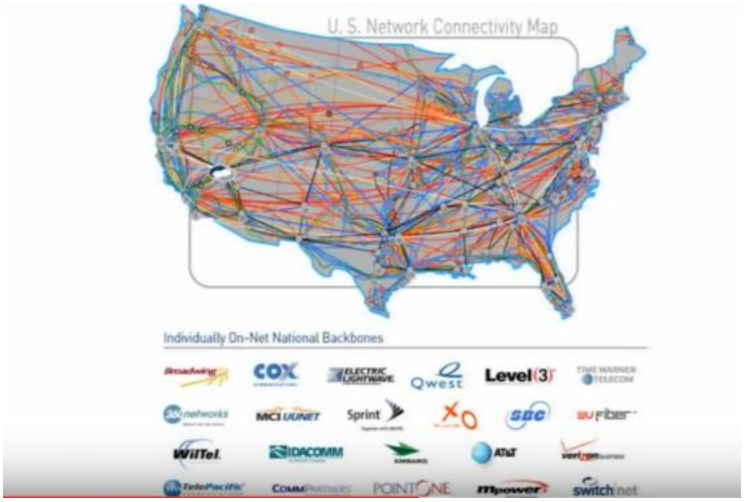


Figure 3-16 *U.S Terrestrial Backbone Network*

Submarine Fiber Optic Cable Systems

It is also called as subsea or undersea cable as well. Sub marine cables use fiber optic infrastructure. There are approximately 450 submarine cables in service as of 2018. Cables are typically as wide as garden hose but of course actual components which carries the light is a diameter of a human hair.

https://t.me/learningnets

Cables are laid in deep sea directly on the ocean floor, but nearer to the shore. Cables are buried under the seabed for protection, that's why people don't see the cables when they go to the beach. As of 2018, more than 1.2 million kilometers of sub marine cables are laid in the world. Some submarine cables are quite short, like the 131 kilometer CeltixConnect cable, while other submarine cables are quite long such as 20.000 km Asia America Gateway Cable.

Major route concept in sub marine fiber optic cable

Undersea cables are built between locations that have something "important to communicate." Europe, Asia, and Latin America all have large amounts of data to send and receive from North America. This includes Internet backbone operators ensuring that emails and phone calls are connected and also content providers who need to link their massive data centers to each other. This explains why you see so many cables along these major routes.

For example, there's just not much data that needs to go directly between Australia and South America. If that situation were to change, you can be sure someone would build a new cable between those two locations.

Who builds sub marine fiber cables?

Cables were traditionally owned by telecom carriers who would form a consortium of all parties interested in using the cable. In the late 1990s, an influx of entrepreneurial companies built lots of private cables and sold off the capacity to users. Both the consortium and private cable models still exist today, but one of the biggest changes in the past few years is the type of companies involved in building cables.

Content Providers such as Google, Facebook, Microsoft, and Amazon are major investors in new cables. The amount of capacity deployed by private network operators – like these content providers – has outpaced Internet backbone operators in recent years. Faced with the prospect of ongoing massive bandwidth growth, owning new submarine cables makes sense for these companies.



Figure 3-17 *Microsoft and Facebook completed Submarine Cable Between U.S and Spain , 160Tbps*

https://t.me/learningnets

Who uses submarine cables?

Users of submarine cable capacity include a wide range of types. Telecom carriers, mobile operators, multinational corporations, governments, content providers, and research institutions all rely on submarine cables to send data around the world. Ultimately, anyone accessing the Internet, regardless of the device they are using, has the potential to use submarine cables.

Submarine Cable Types

There are different Sub Marine Fiber Cables.

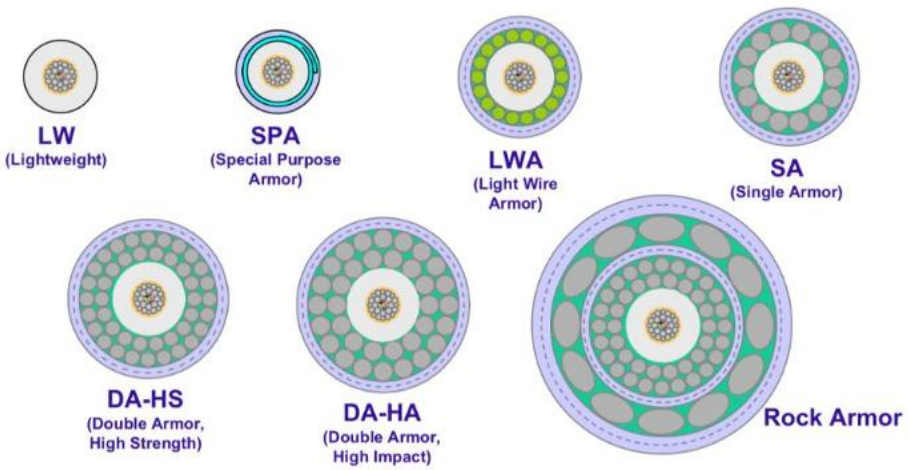


Figure 3-18 *Submarine/Undersea Cable Types (source: www.Tycom.com)*

Table 3-1 Submarine Cable Types

Submarine Cable	**Typical Application**
Double Armored (DA)	Beach Manhole to 400 meter depth
Single Armored (SA)	400 meter 900 meter depth
Light Weight (LW)	After 2000 meter depth
Special Application (SPA)	Between 900 meter to 2000 meter depth
Light-Wire Armored (LWA)	Buried

https://t.me/learningnets

What are the capacity of the Sub Marine Cables?

Cable capacity varies a lot and depends on the owners and the applications of the cable owners. Some submarine cables are capable of carrying 160Tbps. There are two principal ways of measuring a cable's capacity; Potential capacity and Lit capacity.

- Potential capacity is the total amount of capacity that would be possible if the cable's owner installed all available equipment at the ends of the cable. This is the metric most often cited in the press.

- Lit capacity is the amount of capacity that is actually running over a cable. Cable owners rarely purchase and install the transmission equipment to fully realize a cable's potential from day one. Because these equipment's are expensive, owners instead prefer to upgrade their cable gradually, as customer demand dictates.

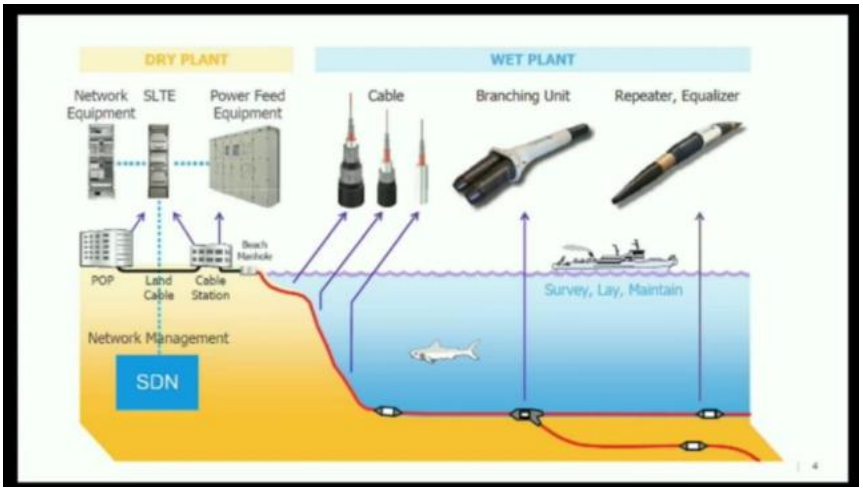


Figure 3-19 *End to End Sub Marine and Terrestrial Cable Architecture*

Cable Landing Point

It is a cable station for submarine cable system, also called landing station. Cable Landing Station is one important component of a submarine cable system which comprises of Wet Plant and Dry Plant.

https://t.me/learningnets

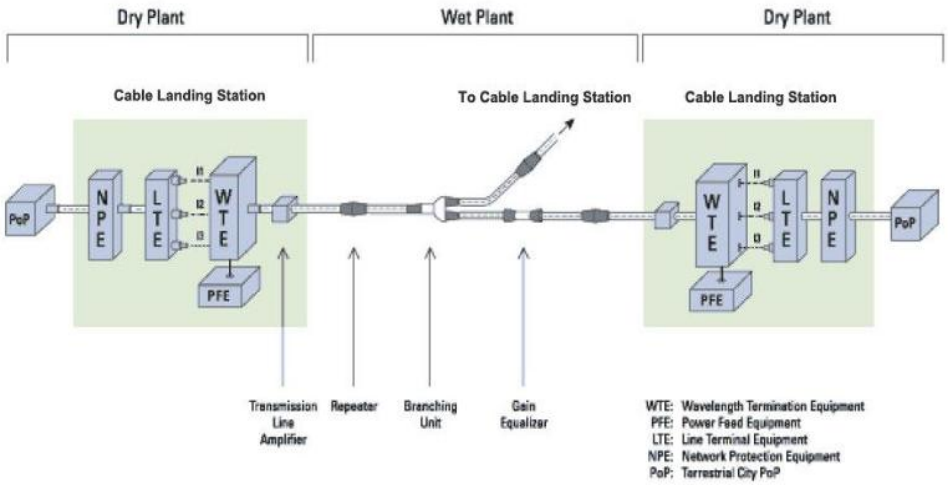


Figure 3-20 *Components of Submarine Cable Systems*

The PFE and the SLTE of a submarine cable system are installed at the cable landing station. The cable landing station is usually carefully chosen to be in the following areas:

1. Areas that have little marine traffic to minimize the risk of cables being damaged by ship anchors and trawler operations.
2. Sandy sea-floors so that the cable can be buried to minimize the chance of damage.

Cable Landing Station typically terminates undersea signals and connects the submarine cable with the domestic network. They are most often very close to beach. Their construction & features are very similar to other telecom offices, such as Central Offices. Large terminal stations cost $10-$15M. Very small stations can be less than $5M.

Some Cable Landing Station requirements can be listed as follows:

- Raised floor
- DC -48v power, with battery backup for at least 1 hour. Diesel Engine emergency backup
- HVAC to maintain room temperature between 22 - 24C
- Fire/smoke detection, with connection to emergency/control center 24-hour access for maintenance and repair
- Transmission Equipment
- Colocation for backhaul

https://t.me/learningnets

Beach manhole

The precise point of connection is known as “the beach manhole,” a small underground place where the undersea cable is connected to dry land.



Figure 3-21 *Beach Manhole of Submarine Cable System*



Figure 3-22 *Another view of Beach Manhole*

Cable is laid between Beach Manhole and the Landing Stations



Figure 3-23 *Cable Landing Station to Beach Manhole Fiber connectivity*

Worldwide Submarine/Undersea cable map might be beneficial for some readers, you can find it on https://www.submarinecablemap.com/

https://t.me/learningnets

Summary

In this chapter, many transport network components which are used in the Service Provider networks were explained.

Fiber and Microwave, as the important building blocks of almost all transmission systems were introduced. Different Transport network systems such as SONET/SDH, CWDM and DWDM were mentioned.

Traffic between the countries and the continents have been carried over Sub Marine/Undersea Fiber Optic Cable systems for decades. Detail architecture of these systems such as Dry and Wet Plans and the nodes, also the components of each Sub Marine system were explained.

In this chapter, different transport capacity purchasing options were explained as well. IRU, Dark Fiber and Leasing options for having transport network capacity were shared.

In the next chapter, the specific physical locations such as POP, Colocation Centers, Meet-me Room, IDC (Internet Datacenter) and the related concepts will be covered.

https://t.me/learningnets

Chapter 4

Chapter-4 Service Provider Physical Locations

Introduction

In Service Provider networks, there are physical locations which provide many different functions. These locations can be CO (Central Office), POP, Datacenter, Colocation Centers, Meet-me room etc. Understanding what kind of services these locations can provide, how they are connected to the rest of the network, and what are the IT functions and services that these places provide, will be the topics of this chapter.

The terminology of these locations can change based on the geography. In different parts of the world, same functional places can have different names. For example, in U.S, telecom facility which provides the last mile access to the customer is called CO (Central Office), but in Europe it is called as Exchange or Telephony Exchange.

Naming convention is not only changed based on the country, as different providers might call these places differently. For example, for the Core POP location, different provider network engineers might call it, Backbone POP, Main POP, or Tier 1 POP.

https://t.me/learningnets

CO (Central Office)/Telephony Exchange

CO is a U.S based term. In the rest of the world, it is commonly used as Telephony Exchange or Public Exchange. (It shouldn't be confused with Public Exchange which we will cover in the IXP topic later in this chapter).

Central Office is a telecom facility where subscriber's local loop (last mile connection) is terminated. Broadband termination equipment such as CMTS, DSLAM, OLT, PSTN and other Voice switching equipment are placed in the Central Office. Many Service Providers use the Access POP term rather than CO (Central Office).



Figure 4-1 *Subscriber Copper Lines at the Central Office*

POP – Point of Presence

POP is a place where communication services are available to subscribers. Internet Service Providers have multiple POPs in different geographic locations, so subscribers can connect to a location closest to them. POPs can be co-located at the Service Provider's central office (CO). Central Office term is mostly used in U.S and Europe. Access POP is the most commonly used term for the Central Office.

Generally, base stations, modems, switches, routers, servers, security and voice appliances are located in the POP sites. POP sites are the demarcation point between customer and Service Provider. POPs can be located at the Internet Exchange Points (IXP will be covered in the Interconnection chapter) as well as the Colocation Centers. POPs can include a meet-me-room which will be explained in this chapter.

Service Providers may classify their POP locations based on speed, hierarchy, technology used in the POP and so on. Earlier it was a trend to classify the POPs with their speeds such as Gigabit POPs, Terabit POPs and so on, but this is not used anymore. POPs are classified mostly based on their hierarchy such as Access POP, Distribution POP/Aggregation POP and Core/Backbone POP.

https://t.me/learningnets

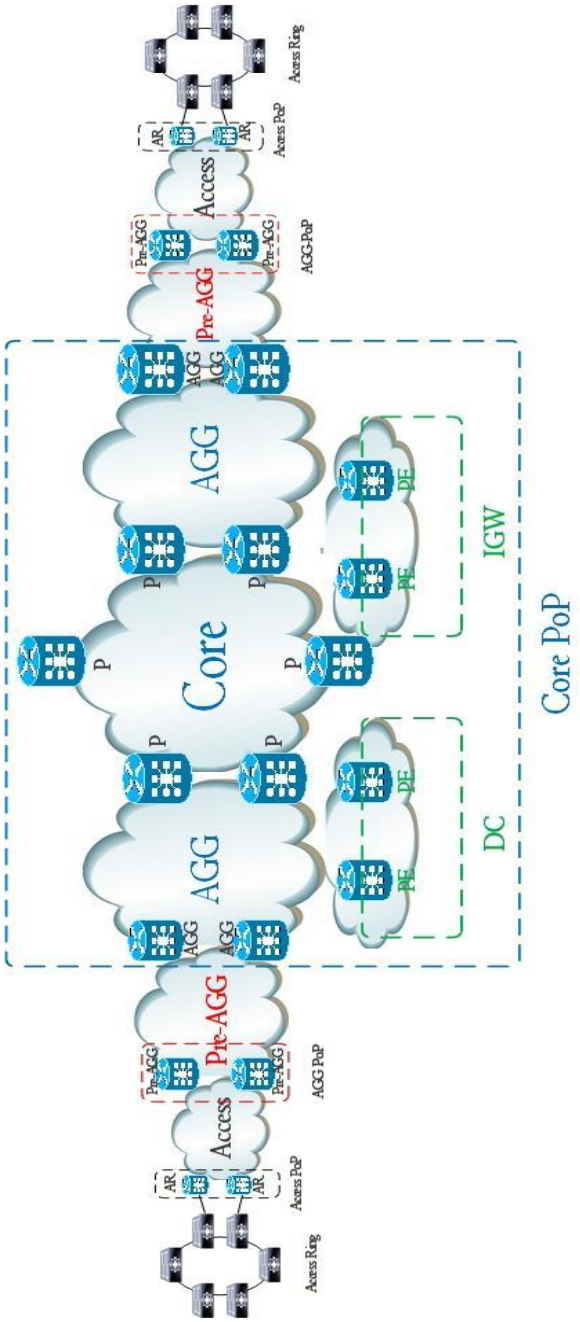


Figure 4-2 *Access, Aggregation and Core/Backbone POP*

https://t.me/learningnets

In each level of hierarchy, users can be terminated but generally the Backbone POP tends not to terminate residential, business or any other types of users. Also in Access, Distribution and Core/Backbone hierarchy, the Access POP link speed is less than Distribution and the Distribution POP speeds are less than Backbone POPs. In each level of hierarchy, inside the POPs, routers are connected to each other.

Access POP's have smaller number of link speeds and also smaller number of routers. Backbone POPs have much higher link speeds and usually the number of routers are more. Availability and Security of Backbone POPs are also critical.

For the Access and Pre-Aggregation POPs that are connected to different Aggregations POP's, traffic is transferred through Core POPs. If the traffic is transferred between two Access POPs which are behind the same Aggregation POP, traffic doesn't go through the Core POP site.

Sometimes Service Providers classify their POPs based on the area of coverage, such as Metro POP, Core POP etc. Figure 4-3 shows the XO (Large operator) network which was acquired by Verizon

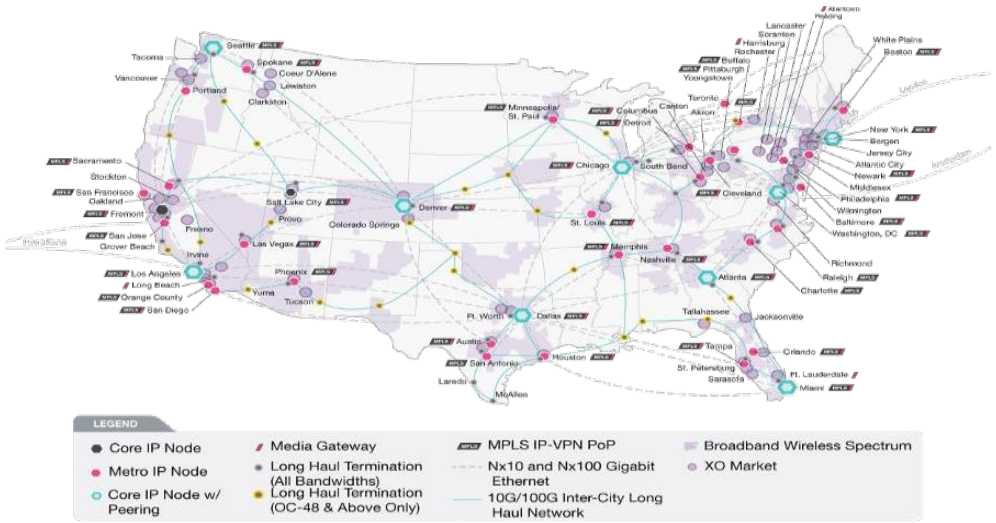


Figure 4-3 *POP Classification based on area of coverage (Metro POP – Core POP)*

https://t.me/learningnets

POP Interconnections

POPs are interconnected via redundant links when high availability is important. This is especially true between the Aggregation and Core POP locations. Access POP's may not have redundant connectivity, though most of the Access Network is redundant through the Ring topology. Link speed varies based on the services that are provided, number of users terminated and the type of the POP.

Optical fiber transport connects the POP locations to each other. In the Access, Aggregation and Core layer (Three Layer) POP design, Access POPs are connected to Aggregation and Aggregation POPs are connected to Backbone POPs. In the Two-layer POP hierarchy, Access POPs are connected to Backbone POPs.

Backbone POPs can be connected in a ring, partial mesh or full mesh topology, which depends on several factors such as the capacity requirements between the POPs, fiber distance and availability requirements. But mostly, in real life Service Provider network designs, partial mesh and full mesh is seen more. The number of Layer's depends on the scalability requirements of the company.

Colocation Centers

It is a type of datacenter where equipment, space and bandwidth are available for rental to retail customers. Customers can be Internet Service Providers, Enterprises or Small – Medium Business Owners. Colocation Centers provide space, power, cooling, and physical security for the server, storage and networking equipment for other companies.

These Colocation Centers connect their customers to a variety of telecommunications and network Service Providers. Internet Service Providers commonly provide Colocation Services inside their Internet Datacenter. This point will be mentioned In the Internet Datacenter Module in the next chapter.



Figure 4-4 *Typical Colocation Center*

https://t.me/learningnets

Some businesses provide Colocation to other companies. Their only business might be providing Colocation to other ISPs, Enterprises, Telco's, Content Providers, IXPs, CDNs etc.

Why companies place their systems inside a Colocation which is actually someone else's datacenter? Why they don't build their own datacenter?

There might be many reasons for these questions.

- Colocation is an environment where their equipment is secure and that ensures no down time. Colocation data centers have fully redundant network connections ensuring that customer's business critical applications always run uninterrupted. Colocation data centers offer power redundancy through a combination of multiple power grids, diesel power generators, double battery backup systems and excellent maintenance practices.

- Cost savings of not building and running your own data centers.

- Audit-ready data centers that meet all the compliance standards relevant to their industry. Data centers have top-notch network security, including the latest firewalls / IDS systems to detect and prevent unauthorized access to their customers systems.

- Secure connectivity to the many leading public clouds. If your company is looking at cloud computing in the near future, colocation provides a smooth transition by allowing you to move your equipment to an offsite facility with increased capacity and performance.

- Bringing many different Telco and Carrier networks to the colocation facility which they have provided. Beyond the cost, one of the most difficult aspects of running your own data center is convincing multiple carriers (necessary for redundancy) to extend their networks into your data center, which only serves your business. Colocation providers, with large facilities filled with multiple tenants, can more easily convince carriers to extend that "last mile" of fiber to their facilities.

- Colocation allows office flexibility by offloading servers to a remote location, so your business is freed from a physical location. This can help when it is necessary to move office space quickly due to growth or unexpected damage. Colocation allows businesses to scale their infrastructure without office space becoming a problem.

Colocation facilities generally have generators that start automatically when utility power fails, usually running on diesel fuel. These generators may have varying levels of redundancy, depending on how the facility is built as it will be covered in the next chapter.

Generators do not start instantaneously, so colocation facilities usually have battery backup systems. In many facilities, the operator of the facility provides large inverters to provide AC power from the batteries.

https://t.me/learningnets

In other cases, the customers may install smaller UPS in their racks. It should be obvious that Colocation services is an outsourced arrangement providing the physical building, cooling process, bandwidth and physical security while customers own and manage their hardware.



Figure 4-5 *Lockable Cabinet inside Datacenter/Colocation*

Space is an area that provides the raised floor, lockable cabinets, racks / cage, biometric and card reader on cage, power and physical connectivity inside the Colocation Center to other companies. This physical connectivity is commonly known as Cross Connect.



Figure 4-6 *Cage inside the Datacenter/Colocation*

A Cross Connect is a physical/Layer 1 connection between the companies inside the Datacenter/Colocation facilities. Some Colocation or Datacenter Providers offer a Remote Hands service. Remote Hands is a service offered by Colocation Providers that allows customers to delegate IT management and maintenance tasks within a Colocation facility to technicians hired by the provider. This allows the Colocation customer to focus on internal tasks within their own organization, and prevents them from sending their own technical staff to a Colocation facility.

https://t.me/learningnets

Remote Hands services generally covers a wide range of tasks, including:

- Physical Cabling
- Server refreshes and reboots
- Hardware and software replacement or installation
- Power cycling
- Inventory management and labeling
- Providing access to the engineers who can connect to the systems remotely
- Last but not least, remote hands provides reporting on equipment performance

Carrier Hotel

Carrier Hotel is a Company that owns large buildings and rents out redundant power and floor space. And of course, attracts many Telco's and Carrier networks to the building. Carrier Hotel often leases off large chunks of space to Service Providers or Enterprises. These companies operate the space as a datacenter or as an Office space. Carrier Hotel doesn't provide IP, VPN, MPLS services.

The carrier hotel, like Telehouse, does not operate an IP network and does not offer any services usually available from the "Datacenter". In this case, the datacenter would be the party that at least will offer IP, Cloud, IXP and additional datacenter services such as management, security, equipment lease, colocation, remote hand etc.

A carrier hotel doesn't deal with few rack spaces or small rack units. They generally do not have services like remote hands, IP, Cloud, or other typical datacenter services. These services will be explained in the next chapter.

Companies that classify themselves as a Carrier Hotel usually lease space out in big chunks, often floor by floor. On the other hand, the people that offer datacenter services usually sell the space by the Rack Unit or by the entire Rack/s. IP services is almost always a part of the offering with the datacenter business as well.

A Carrier Hotel is generally run by large management companies and have more close relationships with carriers and providers for connectivity. They are usually one step higher on the food chain.
They provide large spaces to the Service Providers or Enterprises (Content, OTTs, CDNs and many other Enterprise types) and in turn their tenant builds out their POP, Datacenter or Office location inside the Carrier Hotel.

In general, a Carrier Hotel company is in the real estate business but related to some degree into the I.T. business. On the other side, Colocation or Datacenter business tends to be an I.T. business related to some degree into the real estate business. In the Colocation or ISP Datacenter, they may have limited number of Carrier/Telecom Operator choices.

https://t.me/learningnets

There is also Carrier Neutral Datacenter (which will be explained in the next chapter), which many different carrier networks are hosted. As an example, Equinix is a Carrier Neutral Datacenter, which has 200 Datacenters in 24 countries on 5 continents.

As an analogy, a Carrier Hotel is more like a Carrier Neutral Datacenter. Generally there are so many Carrier choices, so ISPs, Enterprises and any tenant of the building (Carrier Hotel) can receive the service from any Provider.

Carrier Hotels are the huge connectivity hubs in major cities that bring together dozens (and sometimes hundreds) of networks and providers. They house large amounts of data center and telecom facilities, but typically also have office tenants.



Figure 4-7 *111 8th Avenue Carrier Hotel – Google Office and many other Carrier and ISPs Located*

Example of a Carrier Hotel is 111 8th Avenue, NYC. This building is the third-largest building in New York City with 260 thousand square meters of space. Google's New York office is located here along with many other large companies including MCI, Sprint, Level 3, Qwest.

Meet-me Room

It is a place within a Colocation Center where telecommunication companies can physically connect to one another and exchange data without paying local loop fees. Services provided across connections in a Meet-me-Room may be voice, data circuits or Internet traffic. Meet-me-rooms are placed in Carrier Hotel or in traditional Datacenters. For doing interconnection, companies order a patch from their cabinet to the meet-me-room and then arrange for the organization running the facility to connect them together. An IXP can be setup within a Meet-me-Room.

https://t.me/learningnets

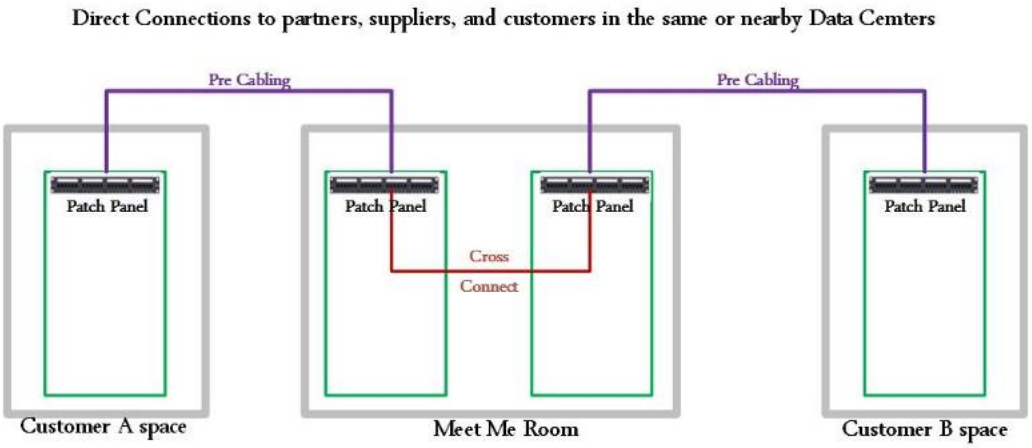


Figure 4-8 *Meet-me Room*

Summary

In this chapter, physical locations and terminologies such as Central Office, Telephony Exchange, POP (Point of Presence), Meet-me Room, Carrier Hotel & Colocation Centers and the Datacenters in Service Provider businesses were explained.

Datacenter, Colocation and Carrier Hotels as can be seen are very interrelated concepts but they are different businesses. You can imagine it as the Colocation is provided by Datacenter and Datacenter is built within Carrier Hotels. Colocation Providers and Datacenter Providers in this case usually have the same owner but Carrier Hotel owner is generally a different company. Throughout the book, these definitions will be beneficial for understanding the topics. Also, it is important to know that, some companies might only provide Colocation Services to their customers, as it was explained in the chapter.

In the next chapter, modules (commonly known as layers in the Service Provider networks) will be introduced. Access, Core/Backbone layers will be explained. Also in the next chapter, different services such as XDSL, FTTX, Fixed and Mobile Broadband network architecture will be introduced, but this time, you will be able to understand the whole picture, not just as an individual technology.

https://t.me/learningnets

Chapter-5 Service Provider Modules - The Big Picture

Introduction

Service Provider's provide many different types of services. Their services are categorized in many different ways.

One way of categorization is based on their customer profile:

Residential customers are apartments, multi-floor buildings which generally don't require higher speed connections.

Business/Corporate customers are smaller Service Providers, small medium Businesses and Enterprises. These customers require higher bandwidth, between 10Mbps to 10s of Gigabit per second.

Another way of categorization is based on the service as well. If end user devices are stationary, the service is called Fixed or wired. If end user devices are not stationary, then the service is called Mobile service. In this chapter, both service categorization (Fixed and Mobile services) and customer profile categorization (Residential and Business Customer service) will be explained. In Chapter 7 and 8 we will discuss a Fictitious Service Provider Company. In order to understand and create a Service Provider network from scratch, it is important and necessary to understand the Service Provider network environment. Thus, this chapter will cover many different parts of the Service Provider networks. Also, before starting the logical level, protocol and technical discussions about the Service Provider networks, it is necessary to understand each module and the physical layout of the Service Provider's network.

https://t.me/learningnets

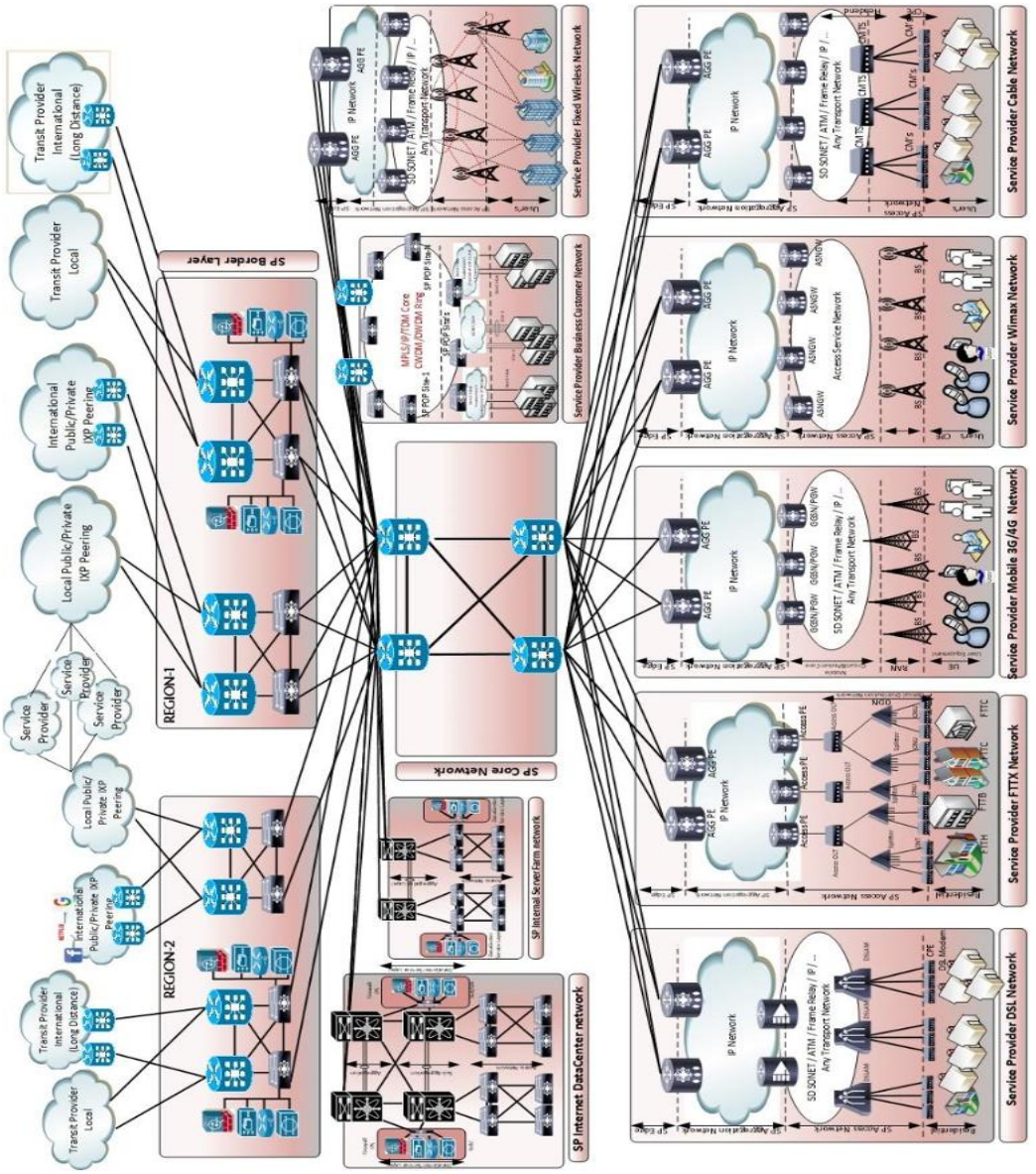


Figure 5-1 *High Level Overview of a Service Provider Network*

High quality of this image can be found at www.orhanergun.net/spbook/spnet.jpg

https://t.me/learningnets

In figure 5-1 (Service Provider Services) , the connectivity , names of the nodes in the specific service, common topologies in each of the services and briefly the Service Provider networks with sufficient detail is shared. Some operators provide different types of Access connectivity such as Public WIFI, BPL (Broadband Over Power Line), Satellite and also other types of connectivity, but in this book we tried to cover the most commonly deployed Access technologies used in real Service Provider networks.

In figure 5-1, there are 12 modules, which are listed as bellow. Each of these modules will be explained in this chapter.

1. Core/Backbone Network
2. Datacenter and Server Farm Modules
3. Border/IGW Layer
4. XDSL Access
5. FTTX Access
6. Cable Access
7. Mobile Broadband
8. Fixed Broadband Wireless
9. WiMAX
10. National Peering
11. International Peering and Transit
12. Business/Corporate Customers

Core Layer Module

The Core layer is the center module of the network, which has the responsibility of connecting all the modules of the Service Provider network together. The traffic between all modules are passed through the SP Core layer. Core/Backbone layer should have high capacity and there are not much protocol, technology and control plane policy found in this layer. It should be designed simple and with high capacity and redundancy.

The Core Layer provides connectivity nearly between each of the Service Provider Modules which are shown in the above list and provides connectivity between different regions of the Service Provider.

https://t.me/learningnets

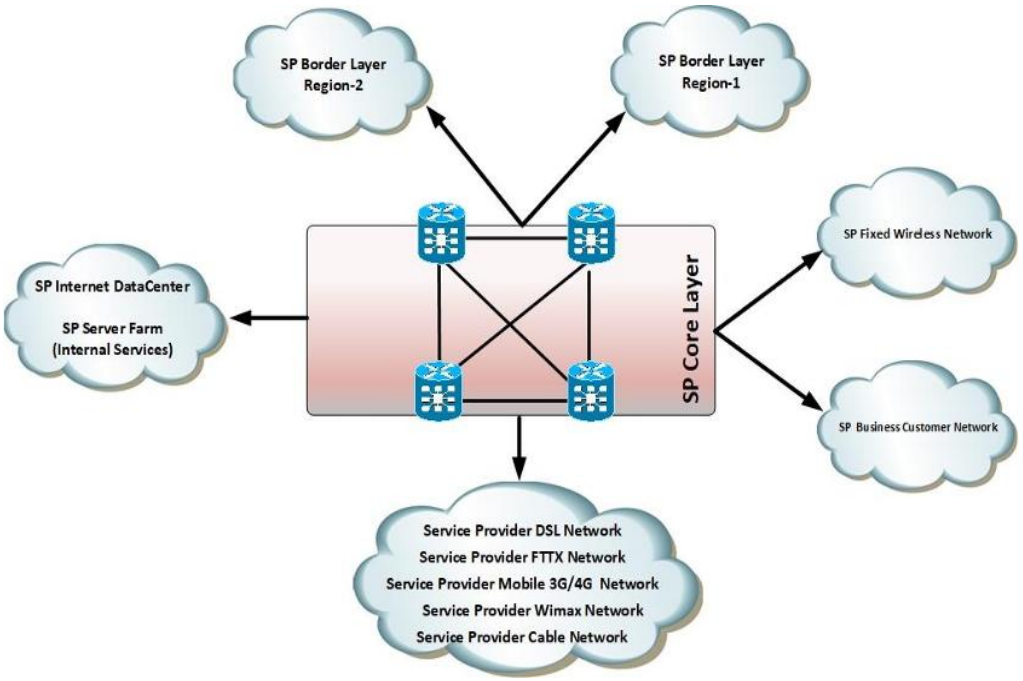


Figure 5-2 *Service Provider Network – Core Layer*

The key Characteristics of the Core Module are:

- High Speed Connectivity.
- Bringing the Internet Gateway, Access, Aggregation and Datacenter networks together.
- Redundancy and High Availability is so important. Redundant physical circuits and devices are very common.
- Failure impact is so high in this module, compared to other modules
- Full Mesh or Partial Mesh deployments are seen
- Commonly known in the Operator community as Backbone or 'P' Layer

Redundancy in this module is very important. Most of the Core network deployments in ISP networks are based on Full Mesh or Partial Mesh. The reason of having full mesh physical connectivity in the Core network is that full mesh connectivity provides the most optimal network traffic and the shortest path between the two locations.

But not every network can have full mesh architecture, because it is the most expensive design option. Instead, many operators connect their Core/Backbone locations in a partial mesh model.

https://t.me/learningnets

In partial mesh physical connectivity model, all of the Core locations are not connected to each other, instead only the Core POP locations which have high network traffic demand are connected together.

Core/Backbone provides scalability to the Service Provider network. Without this layer, many Aggregation layers are required to be connected to each other to provide end to end connectivity. This would be too costly and so many physical links are required to be provisioned. The Core Layer reduces the number of circuit requirements between different Aggregation networks.

Datacenter and Server Farm Modules

Internet Service Providers have their own Datacenter's which might provide Colocation, IXP, Hosting, Meet-me Room etc. for providing services to their customers. Internet Service Provider Datacenter is commonly referred to as IDC (Internet Datacenter). Datacenters commonly have the following components:

- Building, Physical Location
- Security of the Location
- HVAC (Heating, Venting, Air Conditioning)
- Power Infrastructure
- Network Operation Center (NOC) Personal

Let's have a look at these components a bit closer.

Building, Physical Location

Physical location of the building has huge impact on the cost of the data center, thus increases the cost of the services. Probably the most important data center selection criteria is fiber proximity. If the data center location is far from Fiber connections, bringing it to the data center can be a huge cost. If more fiber providers are found at the location, the data center will be more valuable.

Size and expandability of the datacenter location is important as well. The location must be large enough to accommodate cabinets and all other necessary physical equipment. Utility power and the unit price for power is important for the datacenter selection as data center consumes so much Power.

Datacenter should be close to the customers and also to the staff who will operate it. As it will be explained later, there are many Operators which have many data centers in a single country; main reason is being closer to the Customers.

https://t.me/learningnets

Security Systems of the Datacenters

Security systems such as Perimeter security and In-Building access control mechanisms are the key aspect of any data center.



Figure 5-3 *Perimeter Security to the Datacenter*

Monitoring systems, Surveillance cameras, Biometric based security such as Retina scanning, facial recognition, hand and finger geometry and Fingerprint readers are the common security mechanisms to protect data center assets.



Figure 5-4 *Fingerprint Scan*

Retina recognition/scanning technology captures and analyzes the patterns of blood vessels on the thin nerve on the back of the eyeball that processes light entering through the pupil.
Retinal patterns are highly distinctive traits. Every eye has its own totally unique pattern of blood vessels; even the eyes of identical twins are distinct. Although each pattern normally remains stable over a person's lifetime, it can be affected by disease such as diabetes and high blood pressure.

https://t.me/learningnets



Figure 5-5 *Retina Scanning (Source: https://arstechnica.com/)*

HVAC (Heating, Venting, Air Conditioning)

The recommended operational temperature range for the data center is 64° to 81°F (18° to 27°C) and the humidity range is between 20% and %65. Too little humidity generates static electricity and too much humidity can cause system failure.



Figure 5-6 *Cooling Plant inside Google's Finland Datacenter*

https://t.me/learningnets

Datacenter Power Infrastructure

Datacenter's Power has four main components. Main Power (generally comes from the city Power Grid), Generator, UPS and Monitoring System.

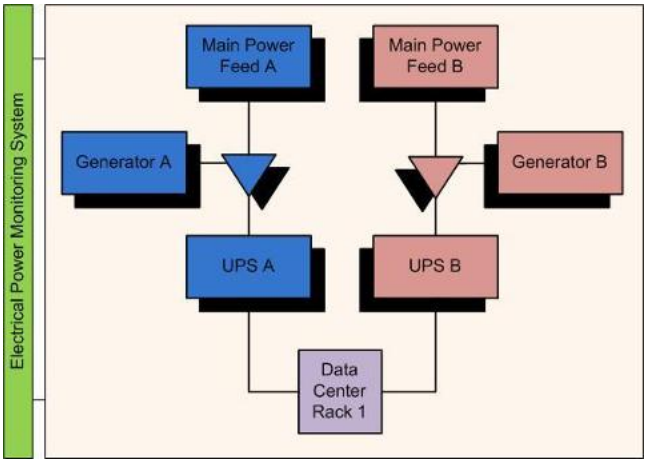


Figure 5-7 *Typical Power System in the Datacenter*

The following functions are performed by each component in figure 5-7.

- Main Power Feed which normally comes from regional or local electrical system. Main Power Feed which is also referred as Utility Grid provides the ultimate source of power for the data center. Some facilities are connected to two separate utility grids for redundancy if one goes down.

- Redundant AC Generators to provide enough power to run the complete facility. These are used as backup generators. Backup generators are diesel powered electrical generators that produce electricity in the event the Utility Grid goes offline.



Figure 5-8 *Diesel Backup Generators*

https://t.me/learningnets

- Redundant UPS to provide uninterrupted supply to racks and other equipment.



Figure 5-9 *UPS (Toshiba)*

- Robust fail-over systems which takes care of any failure in the system. For the failover between the Main Power (Utility Grid) and the Backup Generators in the Datacenter, there is an Automatic Transfer Switches (ATS). ATS is able to switch the source of power from the Utility Grid to the Backup Generators without interruption as it can be seen from figure 5-10.

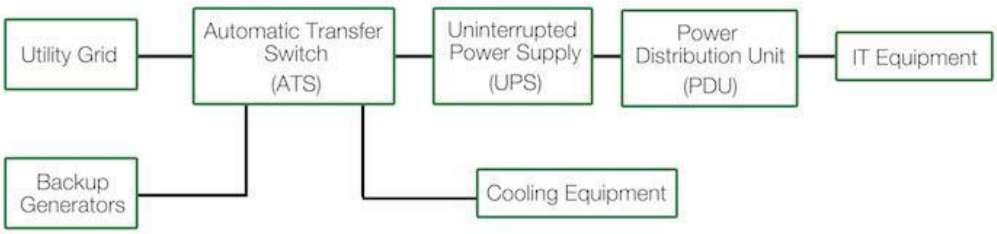


Figure 5-10 *End to End Electrical Diagram of the Datacenter*

- Robust monitoring system to monitor and measure electrical power drawn by the units.

Last but not least for the Power Systems in the Datacenter, not only the redundancy of power but also the price of power is important for selecting the data center location. In some cities in the country and in some countries in the world, power unit price is totally different.

https://t.me/learningnets

Network Operation Center (NOC)

Data centers require Operation staff. They monitor the power, electricity, heating, humidity, manage the network cables and manage the fiber and power connectivity inside the data center. They notify the customers when there is planned activity in the data center. When things break, they repair the issue and communicate it with the customers.



Figure 5-11 *Network Operation Center*

Internet Service Providers generally have more than one data center inside a single country. Almost all medium to large scale Service Providers which the Author has a consultancy agreement with them, have more than one data center and some of them have 10, 20 or even 30 data centers inside the country.

The main reason of having this amount of data center is proximity, which means they want to be closer to their users, thus they deploy their DC's in major cities which they have many business/corporate customers. Also, with this type of design, they could have higher layer of redundancy for their services. But the main reason is to be closer to users, so cost of reaching to the data center for customers are reduced.

The Service Provider's internal operational servers such as NOC systems, OSS, BSS Systems, and Packet Core Network (for Mobile Network operators) are placed in separate servers at the data centers.

These servers are either placed in different zones in their existing Internet Datacenters which they provide services to their customers, or are placed in a totally separate data center for security and availability reasons.

https://t.me/learningnets

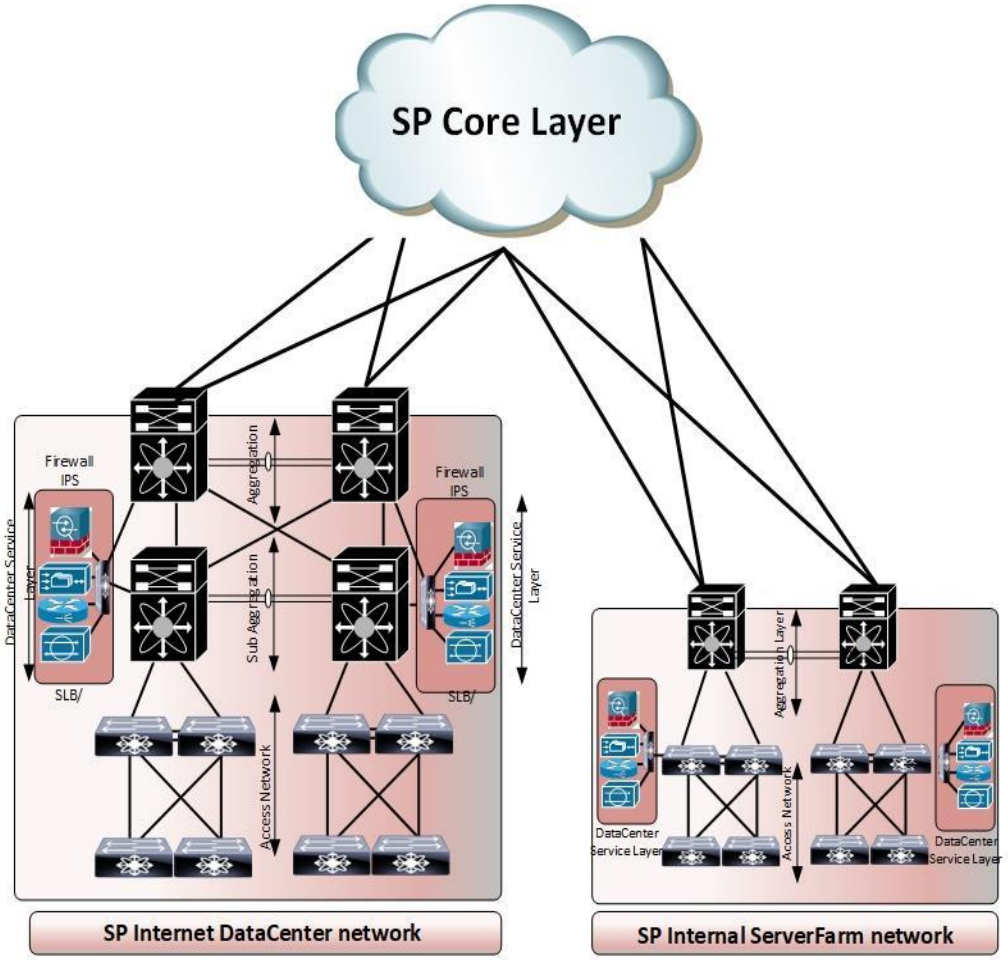


Figure 5-12 *DC and SP Server Farm connectivity to the rest of the network*

https://t.me/learningnets

Border/IGW Module

The Border Module is also called the Internet Gateway (IGW) Module, which hosts services that will be used before reaching out the Internet. It also provides Internet Access and connects the Core/Backbone network to the Internet content.

Services in this module are generally Firewalls, CGNAT, CDN Caches, DNS, Load Balancer, Parental Control boxes and so on. This layer generally runs IBGP with the internal network, and EBGP with other Autonomous Systems.

If Residential or Corporate customers want to reach the Internet, if content is cached inside the Service Provider, it is served directly from the IGW/Border layer. If the content is not cached in the SP network, then National Peering is checked first.

Content might be found inside the country through the National Peering connectivity. Inside the country there might be IXP (Internet Exchange Point), which will be explained in detail later in this chapter. In this case, the content is delivered from the IXP in the country through the National peering router that resides in the Border/IGW layer.

https://t.me/learningnets

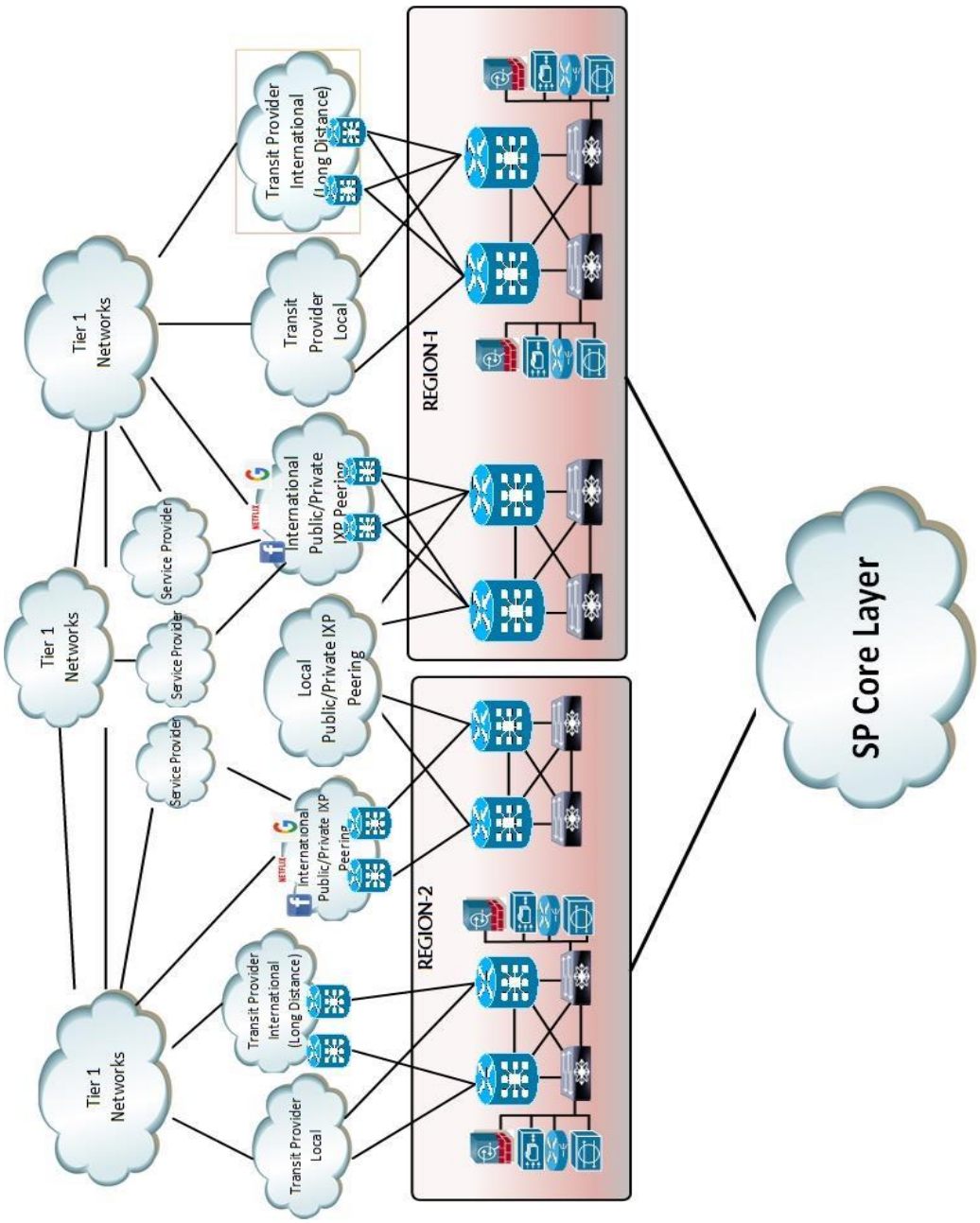


Figure 5-13 *Service Provider Border/IGW Module*

https://t.me/learningnets

XDSL Service Module

DSL provides broadband over ordinary copper telephone lines. DSL was initiated by the telephone companies by the need to provide higher bandwidth than 56kbps dialup. Their competitor such as Cable TV and Satellite companies began providing 10Mbps and 50Mbps respectively. In the XDSL technology, the upstream and downstream speeds can be different, which is known as Asymmetric DSL.

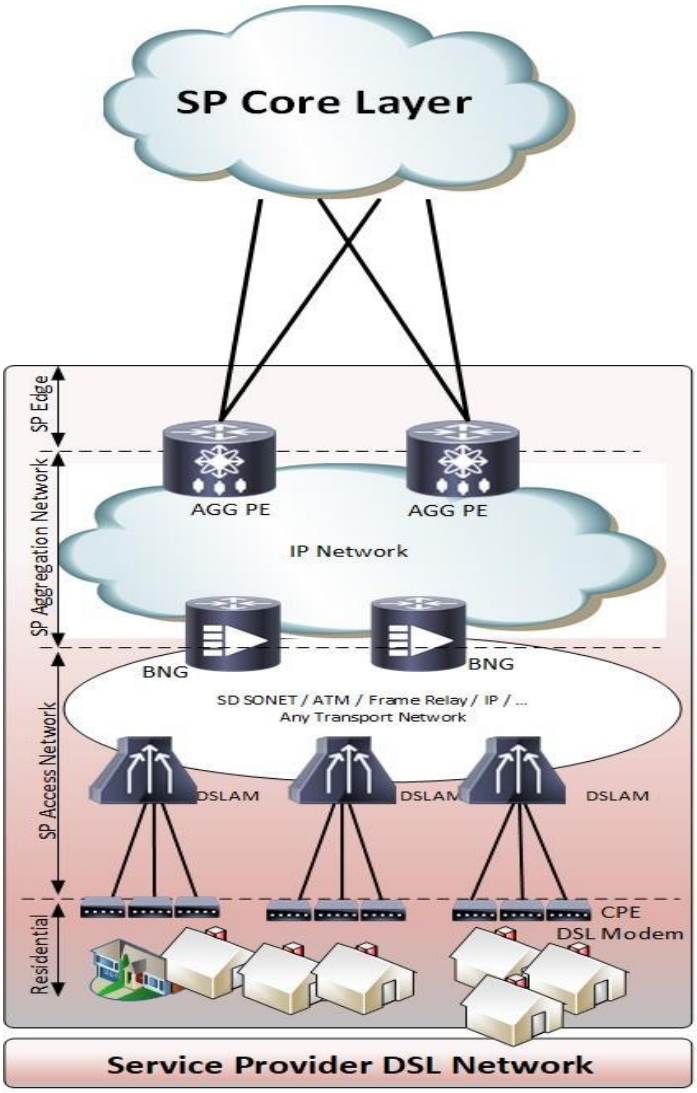


Figure 5-14 *Service Provider DSL module*

https://t.me/learningnets

DSL is also known as Digital Subscriber Loop. It was the most commonly deployed fixed broadband technology until the recent years.

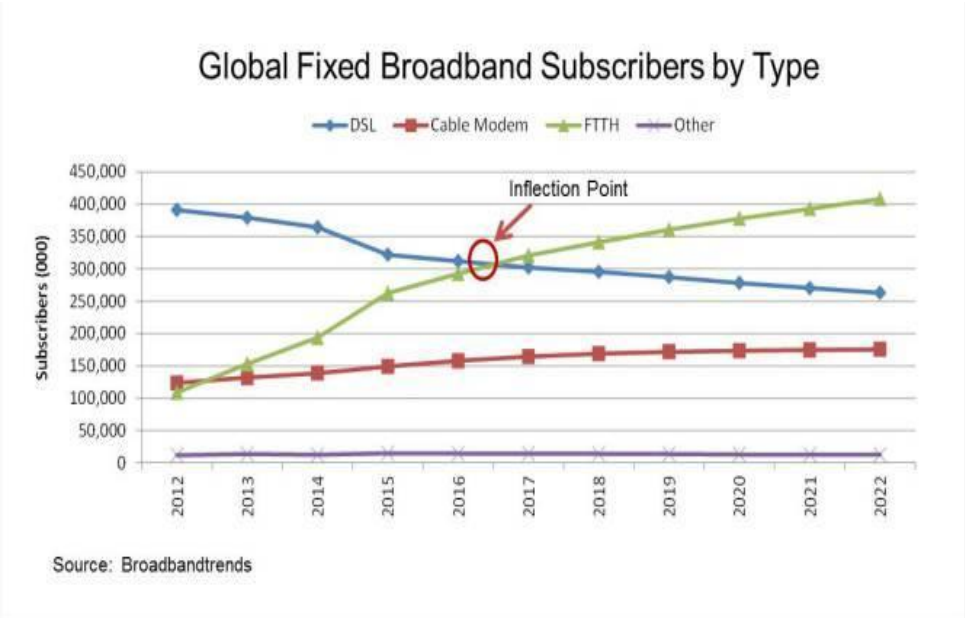


Figure 5-15 *Global Fixed Broadband Subscriber by Type*

In xDSL technology, bandwidth decreases with distance, because of the copper signal attenuation. Bandwidth doesn't decrease by distance in fiber technology. DSLAM which stands for Digital Subscriber Line Access Multiplexer is used to aggregate multiple DSL customers and send the traffic to the IP Backbone and vice versa.

BNG /BRAS

The Broadband Network Gateway (BNG) which is also commonly known as BRAS (Broadband Network Access Server) allows subscribers to connect to a broadband network through an access network. It creates and manages subscriber sessions, aggregates traffic from various subscriber sessions from an access network, and routes it to the network of the Internet Service Provider.

The role of BNG is to pass traffic from the subscriber to the ISP. The manner in which BNG connects to the ISP depends on the model of the network in which it is present.

https://t.me/learningnets

There are two types of network models:

- Network Service Provider
- Access Network Provider

LAC and LNS

In the Network Service Provider model, the ISP (also called the retailer) directly provides the broadband connection to the subscriber. BNG is deployed at the network edge and it is connected to the core of the network through uplinks.

In the Access Network Provider model, a network carrier (also called the wholesaler) owns the edge part of the network infrastructure and provides the broadband connection to the subscriber. However, the network carrier does not own the broadband network. Instead, the network carrier connects to one of the ISPs that manages the broadband network.

BNG is implemented by the network carrier and its role is to hand off the subscriber traffic to one of the several ISPs. The hand-off task, from the carrier to the ISP, is implemented by Layer 2 Tunneling Protocol (L2TP) or Layer 3 Virtual Private Networking (VPN).

L2TP requires two distinct network components:

- L2TP Access Concentrator (LAC)—The LAC is provided by the BNG.
- L2TP Network Server (LNS)—The LNS is provided by the ISP.

DSLAM is the physical DSL modem termination equipment and it's located in the telephony exchange or street cabinet of the Service Provider. It is similar to CMTS in Cable broadband or eNodeB in Mobile broadband.

Telephony exchanges are connected in a ring topology as it provides redundancy at the cheapest cost. Thus, in a large scale XDSL Service Provider network, there are many access rings in the network. BNG, LAC and LNS have critical functionalities in XDSL networks.

https://t.me/learningnets

FTTX Service Module

From the Operators Central Office (Telephone Exchange), to the destination (home users, other ISPs or enterprises), fiber is distributed over ODN (Optical Distribution Network). BNG/BRAS is the subscribers default gateway. IP address is provided by BNG/BRAS to the end customer with FTTX service, same as how the IP address is provided to the customers with the XDSL service

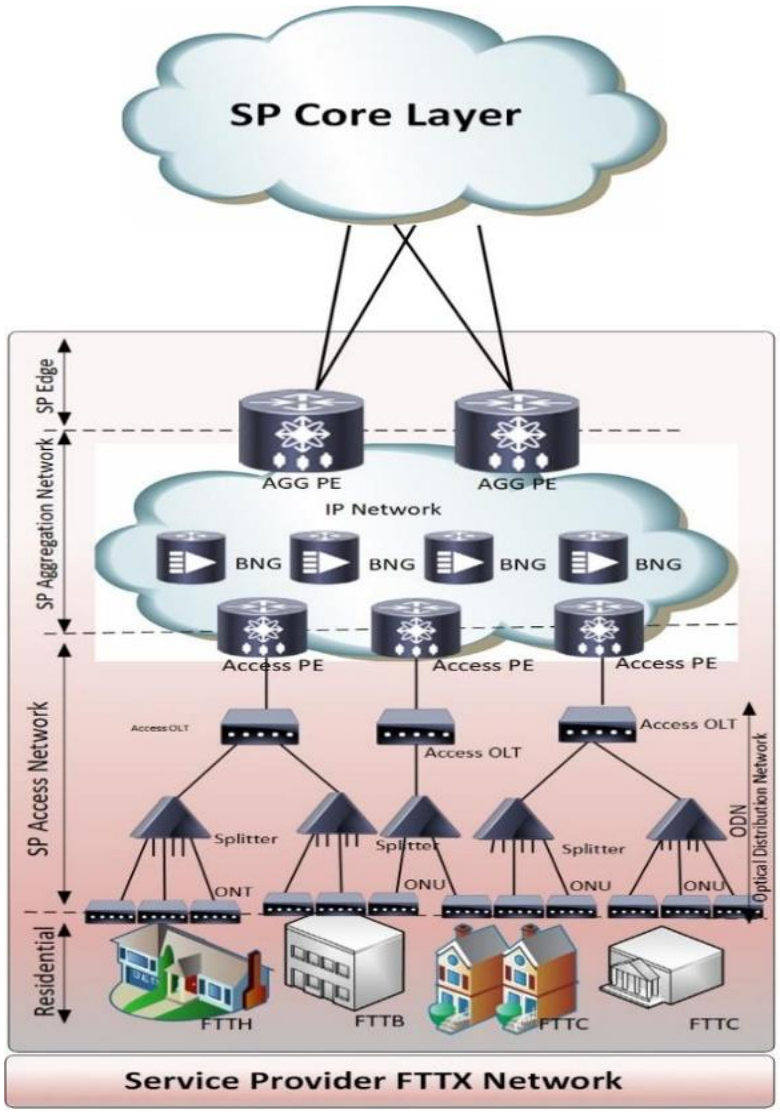


Figure 5-16 *Service Provider FTTX module*

https://t.me/learningnets

Passive Optical Network (PON) is primarily deployed for FTTX service in the Access network of the Service Provider. ONT provides physical termination to the end user/subscriber.

ODN (Optical Distribution Network) is the physical path for optical transmission between OLT (Optical Line Termination) and ONU (Optical Network Unit) equipment.

OLT is the physical termination of the Splitters. OLTs are aggregated at the Access PE routers. OLTs in general are connected to each other physically in a ring topology. A reason using Ring topology is to connect OLTs to each other to provide redundancy with a cheaper cost compared to other topologies such as Hub and Spoke/Star or Full Mesh topologies.

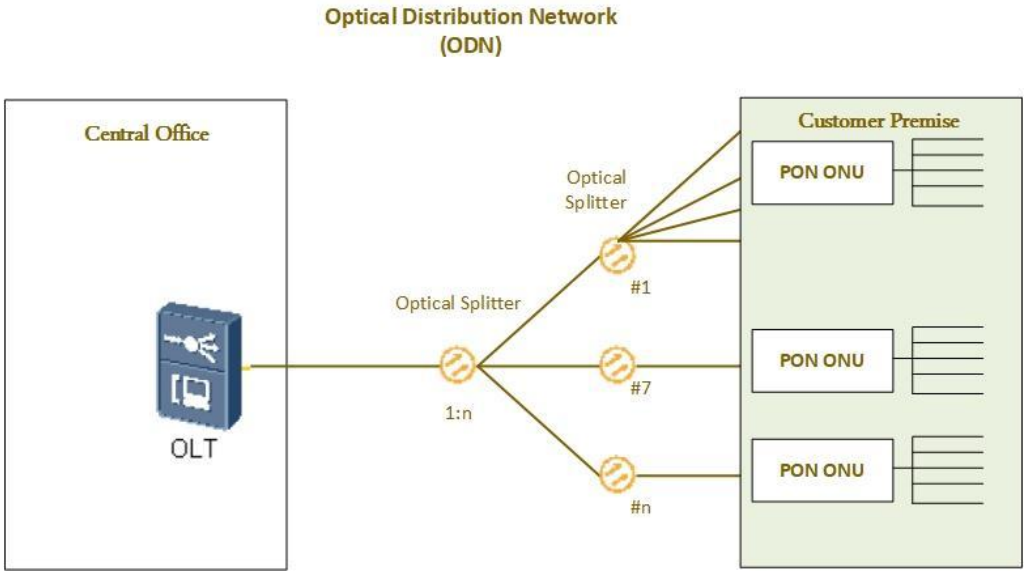


Figure 5-17 *Optical Distribution Network Component*

https://t.me/learningnets

Cable Broadband Service Module

The cable network comprises a number of elements:

- The headend
- The fiber and coaxial cable (HFC) to the customer's premises
- The individual customer's terminal equipment

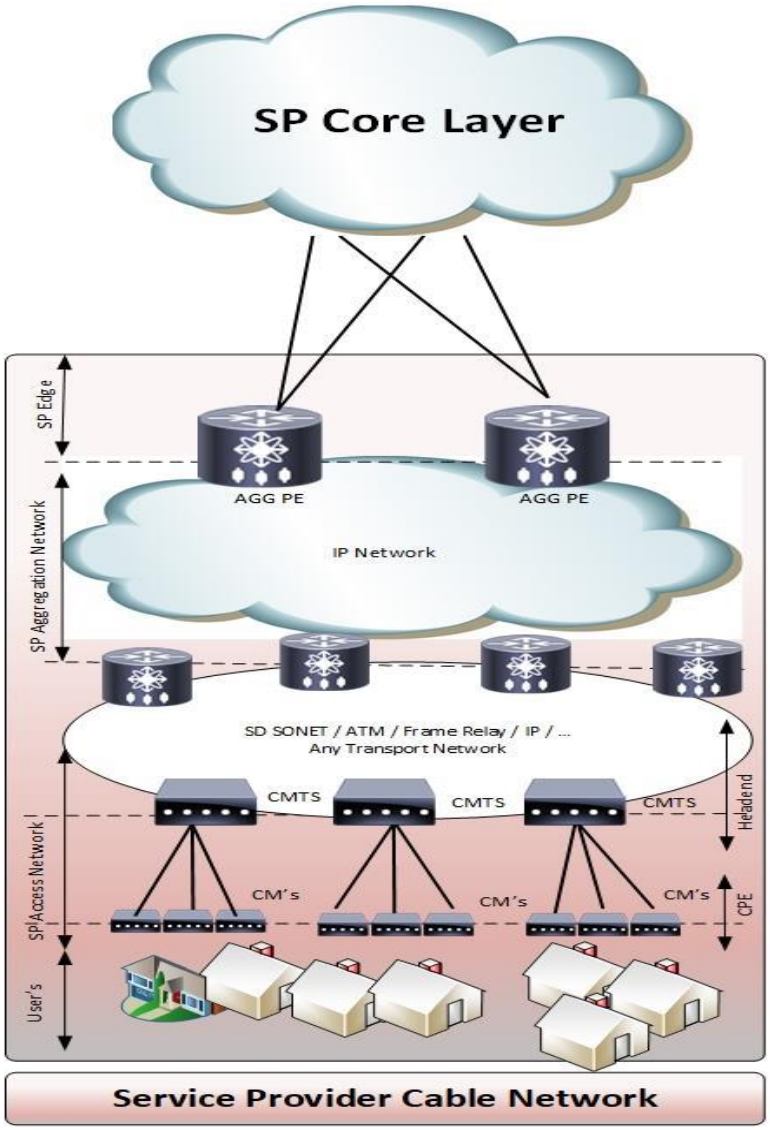


Figure 5-18 *Service Provider Cable Network*

https://t.me/learningnets

For providing a broadband Internet service over Cable TV network, the Cable broadband/Internet requires cable modem at the customer premise and CMTS (Cable Modem Termination System) at the cable provider facility. These facilities are cable television headend (It is similar to Telco CO (Central Office)).

In figure 5-18, the headend is where the broadcast content is received, either from a satellite or a local TV antenna or sometimes via a direct fiber link from a studio. The headend processes and assembles the content for onward delivery to the customer. It also connects with other networks and the Service Providers.

In DSL, bandwidth is dedicated on the Access Network per user. In PON (FTTx), the bandwidth is shared on the Access Network between the end users. In Cable Broadband, the bandwidth is shared between the end users similar to PON.

CMTS in Cable Broadband network (similar to OLTs in FTTX networks) are connected to each other to provide redundancy with the cheapest possible cost.

Mobile Broadband Service Module

Mobile broadband is the marketing term for wireless Internet access through a mobile phone, portable modem, USB wireless modem, tablet or other mobile devices.

The first wireless Internet access became available with 2G mobile technology but this was not considered as mobile broadband. The first mobile broadband was available via 3G. LTE-Advanced is the most current technology as of 2019 and the industry is working on 5G technology as well.

The Mobile broadband module is connected to the Core network through the Aggregation network via GGSN or PGW (Packet Gateway) node, depending on the Mobile network generation.

This module is found in Mobile Operator companies. If a Service Provider only delivers fixed based services, then Mobile network elements which are shown in this module are not deployed in their network. In the RAN (Radio Access Network), there are many BS (Base Stations) that are connected physically in a Ring topology. The main reasons for using ring topology are redundancy and less cost compared to other physical topologies. This is true for Mobile Broadband Access as well as FTTX, DSL or Cable Access networks. In large scale Mobile Operator networks, there are 10s of thousands of Base Stations.

https://t.me/learningnets

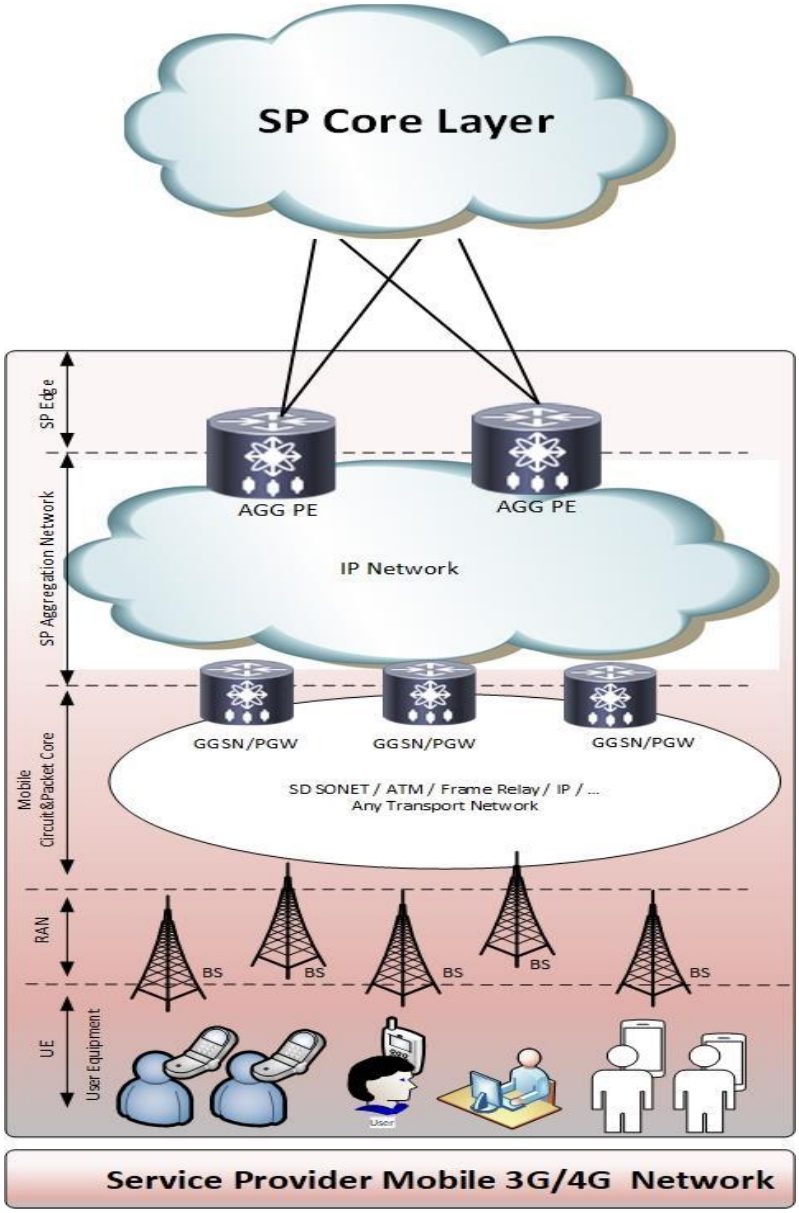


Figure 5-19 *Service Provider Mobile 3G/4G Network*

https://t.me/learningnets

Fixed Wireless Service Module

Fixed wireless Internet is an accomplished method of business connectivity that can provide broadband Internet access to the location, without the use of phone or cable lines. Fixed wireless is a microwave-based technology that allows you to send and receive high-speed data between two fixed sites or locations. It is not mobile technology, nor Wi-Fi where bandwidth is shared on a "one to many" basis. There are many compelling advantages for fixed wireless as compared to wired services such as fast installation, free from trenching and construction. Fixed wireless relies on microwave signals that are used to connect customers to a point of presence (POP). These provide high-performance bandwidth speeds that range from 1.5Mbps to 1Gbps. Unlike satellite systems, these signals have much lower latency and are less affected by inclement weather, which results in higher availability compared to satellite.

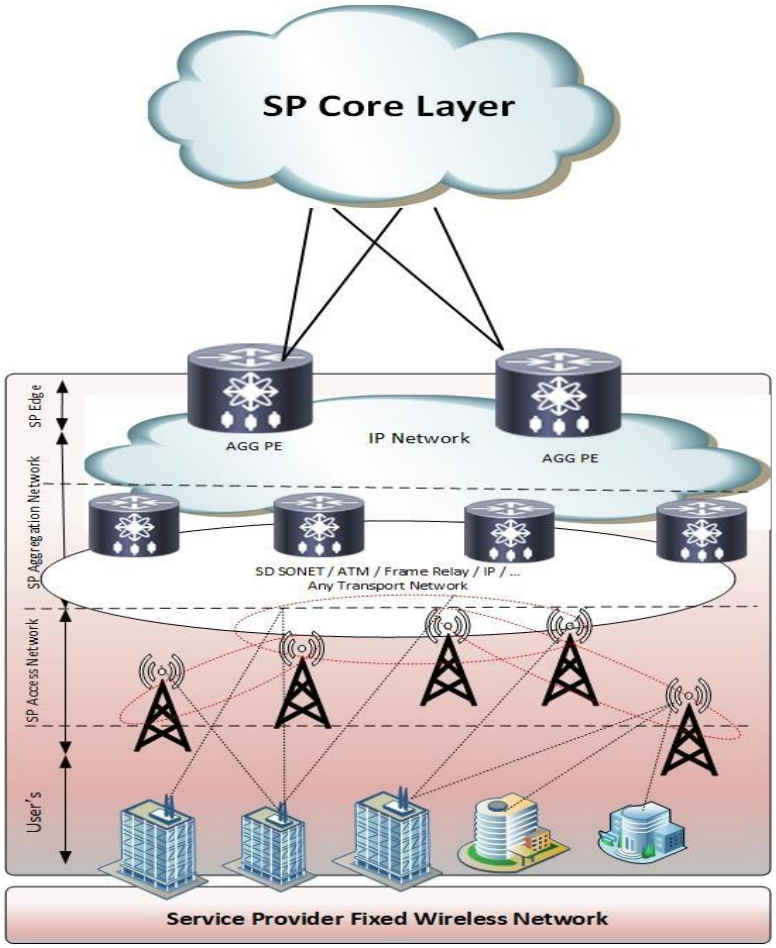


Figure 5-20 *Service Provider Fixed Wireless Network*

https://t.me/learningnets

WIMAX Service Module

Based on IEEE 802.16 standards, WiMAX access networks can provide wireless broadband Internet access at a relatively low cost. A single base station in a WiMAX network can support data rates up to 75 Mb/s for residential or business users. However, since multiple users are served by a single base station, actual bandwidth delivered to end users is likely to be 1Mb/s for residential subscribers and a few Mb/s for business clients.

Compared to the transmission distance of a few hundred meters supported by Wi-Fi (802.11a/b/g/n), WiMAX promises wireless access range up to 50 km. Therefore, WiMAX can provide citywide coverage and QoS capabilities, supporting multimedia applications from non-real-time data to real-time voice and video. Furthermore, as an IP-based wireless technology, WiMAX can be integrated seamlessly with other types of wireless or wire line networks.

The original IEEE 802.16 standard defines backhaul point-to-point connections with bit rates up to 134 Mb/s using frequencies in the range of 10 to 66 GHz. IEEE 802.16d/e specifies point-to-multipoint wireless access at bit rates up to 75 Mb/s. The newest standard, IEEE 802.16m, supports data rates up to 1Gbps but with a much shorter transmission range.

In WiMAX networks, WiMAX base stations are connected to the wire line networks (usually, optical metro networks) using optical fiber, cable, and microwave high-speed point- to-point links. Theoretically, a base station can cover up to 50-km radius, but in practice it is usually limited to 10 km. The base station serves a number of subscriber stations (deployed at the locations of residential or business users) using point-to-multipoint links.

A WiMAX network can be configured with a Hub and Spoke topology or a mesh topology; each has advantages and disadvantages. Hub and Spoke topology can support higher data rates; mesh topology provides a longer reach and faster deployment.

https://t.me/learningnets

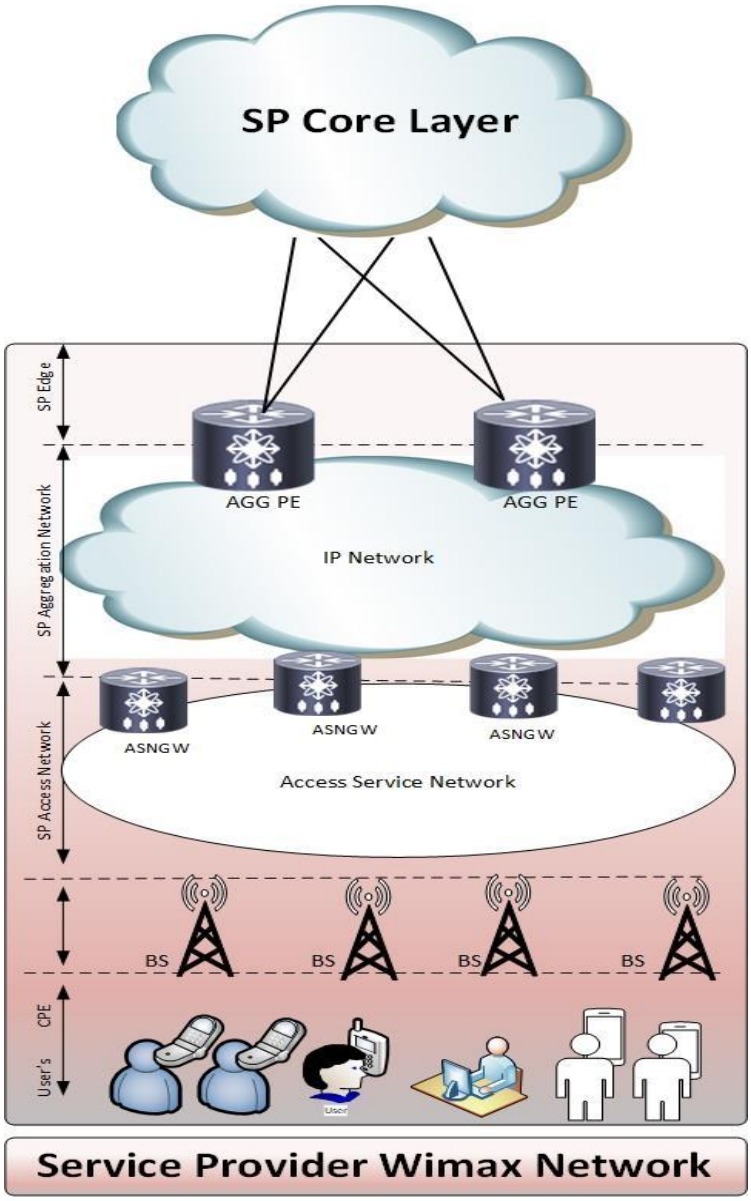


Figure 5-21 *Service Provider WiMAX Network*

https://t.me/learningnets

As of 2018, many people ask whether WIMAX is still alive. Yes, indeed there are hundreds of WIMAX installations worldwide. Many wireless experts believe that WIMAX is an excellent wireless technology but currently living in the shadow of LTE.

National Peering Module

National Peering Routers are located in the IGW Layer. Internet Service Providers commonly have more than one Internet Gateway POP location. National Peering means connecting to other networks inside the same country and having a Peering (Settlement Free) with each other.

ISPs are connected to each other. Mobile Operators are connected to each other. The Enterprises, Government Organizations, Application and Content Providers all peer with each other. In all of these cases, if the traffic will stay inside the country, it is defined as National Peering.

For example, when two Providers (any type of provider, ISP-ISP, Content Provider – ISP etc.) peer with each other inside their own country through National Peering, their customers enjoy the low latency. Operators save money because they don't use upstream networks to reach each other's customers. When they save money, they can provide cheaper service to their customers. Customers can be happy because they get better price from their operators and the voice quality can get better due to low latency as the peering is done within the country.

Service Providers are connected to each other and to other networks such as Enterprises to exchange network traffic. They announce to each other their customers' prefixes. For example, two Mobile Operators connect their network to each other to exchange traffic directly instead of using upstream operator networks.

Mobile Operators in the past used GRX network for exchanging traffic globally. By the time this book is published, IPX is the most common approach by the Mobile Operators to exchange the traffic.

GRX (GPRS Roaming Exchange)

General Packet Radio Service (GPRS) technology is a mobile packet delivery service in GSM era. GRX network was established in 2000, to cater GPRS roaming. Only Mobile Network Operators (MNO) are allowed to connect to this network. Later, other services like Universal Mobile Telecommunication System Roaming (UMTS), MMS interworking and WLAN data roaming also were added in this model.

GRX network was built on a best effort method in the transport network. So the quality of service or security is not guaranteed in the transport.

https://t.me/learningnets

IPX (IP Exchange)

Don't confuse IP Exchange with Novel's networking protocol IPX.
IPX is also an interconnect method; however, it's considered as evolved GRX. Although, the concept is the same as GRX, it has a lot of advantages. Here, not only mobile operators, any operator can connect to it. For example, Fixed Network Operators, Internet Service Providers (ISP), and Application Service Providers (ASP) etc. can connect to this network. This model ensures end to end service levels on Security and Quality of Service.

This is an open model to interconnect and all technical interconnect mechanisms are feasible between operators or peering partners. There are four service models or class of services defined in IPX; they are Conversational, Streaming, Interactive, and Background. MPLS backbone plays a major role in this model of connectivity. Here, we get fully managed transport layer (MPLS) with all quality of service guaranteed end to end for each operator or partner core. Each interconnect partner or carrier will follow the same.

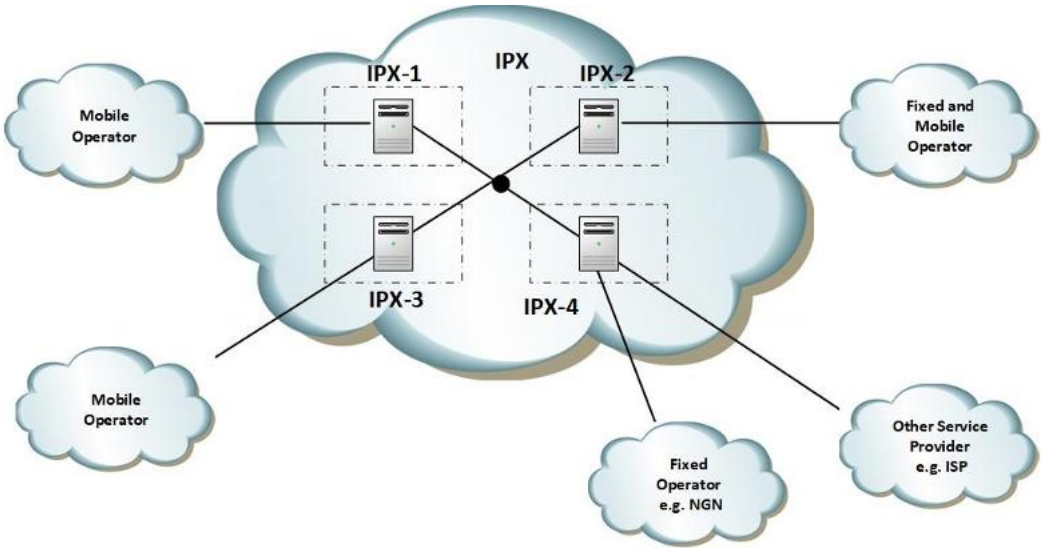


Figure 5-22 *IPX Network – LTE Roaming*

On the other side, IP Traffic for the Internet service is exchanged at the IXP (Internet Exchange Point) or directly between two companies. If it is only done by connecting two companies, in this case it is called Private Peering. If it is done at the IXP which is a shared network, in this case it is called Public Peering.

https://t.me/learningnets

In Public Peering arrangement, more than two companies exchange their customer prefixes with each other for Internet service.
Using national peering instead of International peering is more cost effective as it reduces the International bandwidth usage. Generally country wide public Internet exchange is called a Local IXP (Internet Exchange Point).

A big reason to setup an IXP is to keep the local traffic as local. Imagine two ISPs in the same country, one of them has a server and another one has a client who wants to reach that server, and their only connection is through another ISP which is located at another country. Without having a Local IXP, the client traffic would go to the other country and come back to reach the server.

All IXPs start as Local IXP and some of them by the time evolve to a Regional IXP. In the same country, there might be several Local IXPs. Figure 5-23 shows the Local IXPs in Canada.



Figure 5-23 *Local IXPs in Canada*

There are many IXPs and some companies are connected to more than one IXP in the same country. Connecting to more than one IXP provides higher availability and accessibility to different prefixes.

https://t.me/learningnets

International Peering and Transit Module

Peering can be done outside of the country as well. In fact, many Service Providers have peering agreement with other companies, both inside and outside the country. If content is delivered directly from the IXP inside the country, then International bandwidth usage will be reduced. But in some countries, such as very small African countries or Islands, there is no IXP and the only option to reach some content from the Content Providers such as Netflix, Facebook etc. is receiving the "IP Transit service".

With the "IP Transit service" the enterprise or ISP customers receive the full Internet routing table from another Internet Service Provider. This service can be received within the country or from the International Internet Exchange Points.

For example, Local ISPs in Turkey can receive IP Transit Service (Full Internet Routing Table) from another Local ISP in Turkey, or from an International ISP which has a POP location in Turkey or even from an International Transit Internet Service Provider at Amsterdam, London, Germany etc. Receiving an IP Transit Service from different Internet Service Providers will have different impacts.

For example, a Tier 2 ISP in Turkey might be selling IP Transit service, but the overall latency and the quality of routes to the destination may not be the same as when a customer in Turkey is receiving the IP Transit from a Tier 1 Operator such as Level 3, Tata etc.

There are some IXPs in some countries which have more than 1000 networks doing peering with each other. When Service Providers and other networks such as Content Providers and CDN's peer with each other in the IXP or if they do Peering outside of the IXP, they might have cost savings. Instead of paying to Upstream Internet Service Providers for IP Transit Service, Settlement Free Peering arrangement can be done at these Large IXPs, or through the Private Peering, so they can reduce the Internet traffic cost.

https://t.me/learningnets

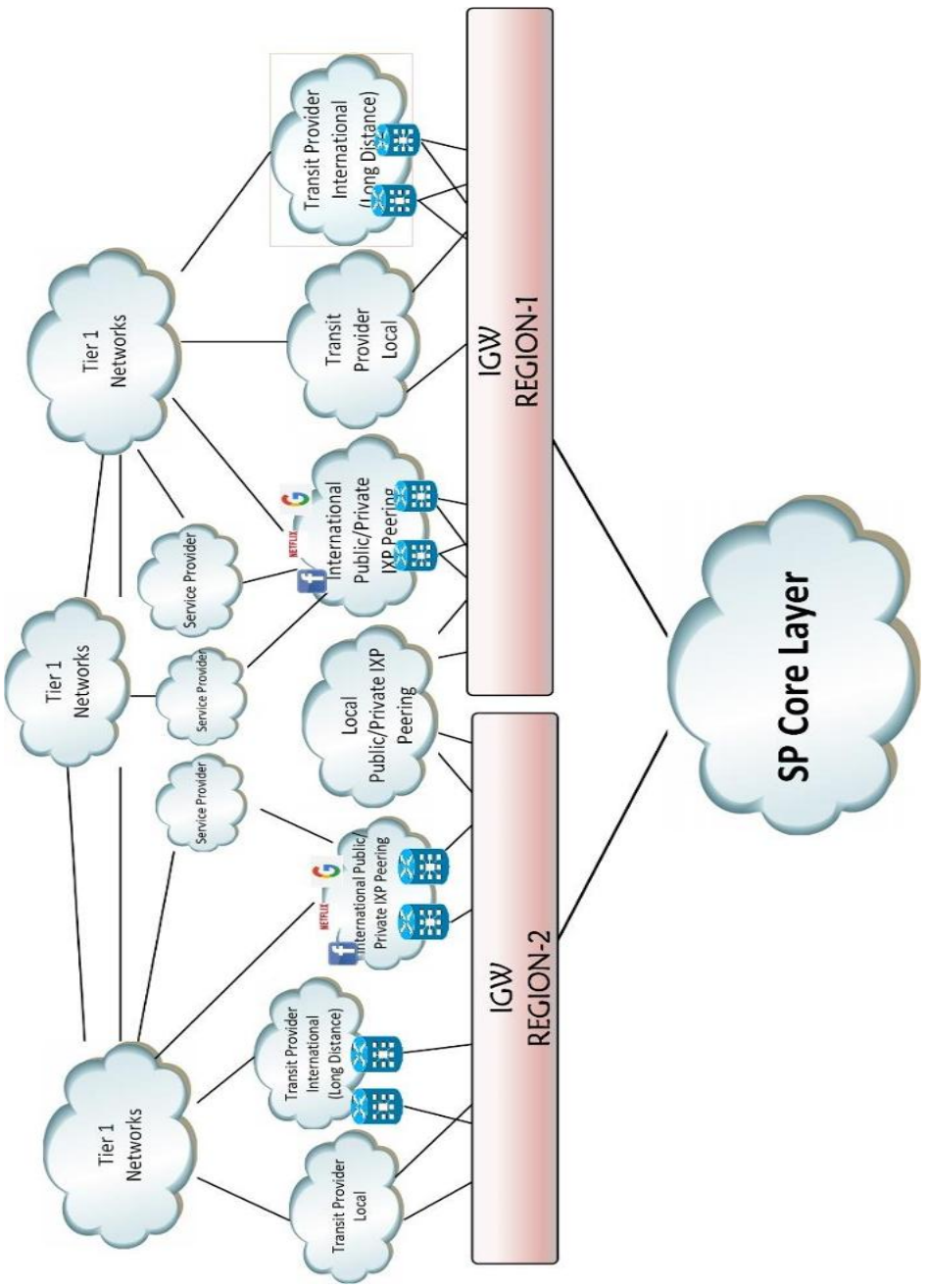


Figure 5-24 *International Peering Module*

https://t.me/learningnets

The IXPs that have the most number of networks and unique prefixes are considered as Internet Hubs, as the most of the Global Internet traffic goes through these IXPs. In the late 90's, content were centralized in the U.S, but in the early 2000 with the advent of CDNs and Content Provider Networks, the content are distributed around the world. New Internet Hubs are the biggest IXPs in Europe.

In the U.S, these are the Global Internet Hubs as the most of the content as of 2019 are located at these points. Some Asian countries such as Singapore and Hong Kong have large number of networks connected, so they are considered as Regional Hubs. There are currently more than 500 IXPs in the World. Many European countries and the U.S have multiple IXPs in the country.



Figure 5-25 *IXPs around the World*

Business/Corporate Customer Module

The main difference between corporate and residential customers is the type of Internet service. Usually, corporate customers prefer symmetrical Internet service. However, almost all residential services which ISPs offer are asymmetrical Internet service (which means the upload bandwidth is significantly lower than the download bandwidth). The download bandwidth for residential service is usually 4 or even 5 times higher than the upload bandwidth.

Residential customers who require symmetrical service (Same upload/download bandwidth) or require higher upload speed than download bandwidth, are redirected to corporate services by the Service Providers. A reason SPs redirect them to the corporate service is that they cannot change the service models based on small number of customer requests. They arrange residential service packages according to millions of users.

https://t.me/learningnets

Some Corporate customers might require to receive higher upload bandwidth than download bandwidth. These are generally Hosting Providers/Colocation Providers which provide server space, or allow other companies to place their servers to their facilities. This kind of content heavy companies require higher upload bandwidth than download bandwidth.

For corporate customers, performance and price will also be the key considerations in selecting a broadband service, as well as support for value-added services such as VPNs, Quality of Service, and Security etc.

Corporate customers require higher bandwidth and uptime compared to Residential customers. Corporate customers require SLA (Service Level Agreement) to set out the following items:

- The terms and conditions of the ISPs service
- Their commitments regarding the quality and availability of service
- Their responsibilities as agreed with the customer

For an effective SLA, Businesses are continuing to request for more bandwidth as they deal with greater volumes of voice, data, and video traffic. Still the best service that can serve better KPI values in terms of delay and packet loss is TDM circuits such as STM. However, TDM circuit costs are expensive and not technically flexible.

Because of optimum cost with superior SLA commitments, Ethernet is the best choice for corporate customers, which can seamlessly connect the offices of a business whether they are 10, 1000 or even 10,000 kilometers apart.

Ethernet access is the fastest growing access technology available today. With a range of bandwidth and configuration options that reach Gigabit speeds, Ethernet services are adapting to meet virtually any challenge. Most business services today are based on Ethernet.

https://t.me/learningnets

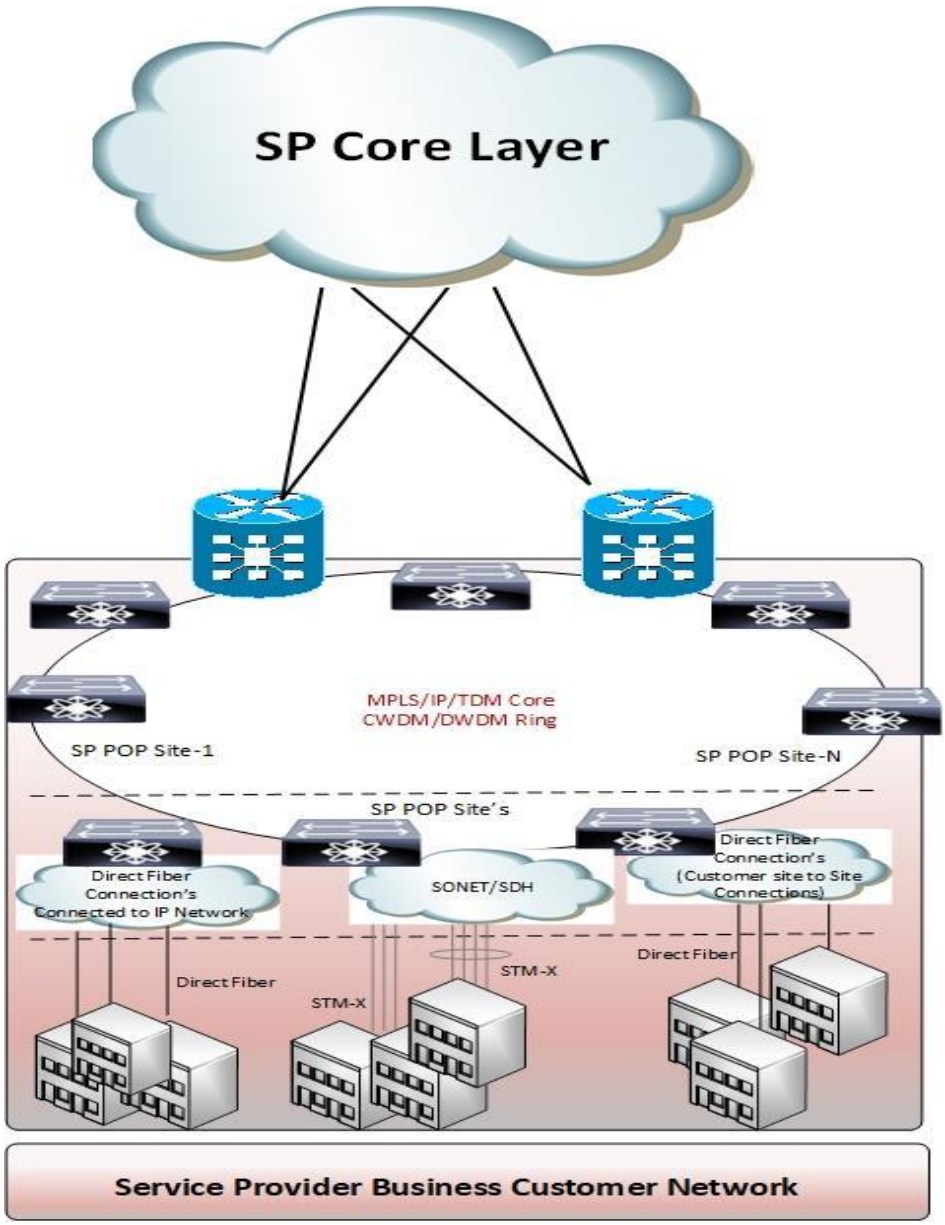


Figure 5-26 *Service Provider Business Customer Network*

https://t.me/learningnets

Metro Ethernet is an Ethernet transport network that provides point-to-point or multipoint connectivity services over a metropolitan area network (MAN). Ethernet originated as a LAN technology, and became a replacement for WAN technologies.

Service Providers use Metro Ethernet architecture to provide services to customers. These services can be such as E-Line (Ethernet-Line) similar to Leased Line, E-LAN (Ethernet LAN) which is multi point Ethernet Service, L2 MPLS VPN, L3 MPLS VPN and Internet Access. Metro Ethernet can use many different technologies as its transport.

For example, between the Service Provider and the Business Customer, there can be 802.1q, QinQ, VPWS, VPLS technologies deployed.

Although the Customer might receive Direct Fiber, and not a shared fiber, the technologies between the Operator and the customer don't change. Most of the current business connectivity is provided via Metro Ethernet. The physical topology in the Access Network (Customer to SP connectivity) is based on Ring topology.

There are still SONET/SDH based TDM systems and Business/Corporate Customers might request these technologies because of very high reliability and OAM support of SONET/SDH, but eventually the enhancement with the Ethernet will replace all TDM technologies.

In figure 5-26, the Business Access Module connectivity and the technologies are shared. You can see there are still SONET/SDH customers, Direct Fiber, Shared Fiber customers which might request Internet Access, E-Line, E-LAN, MPLS Layer 2 VPN, MPLS Layer 3 VPN etc. Due to higher bandwidth requirement of Business/Corporate Customers, it is often a case that ISPs connect Business/Corporate Customers to their Pre-Aggregation or Aggregation network, rather than Access Network, as the Pre-Aggregation and Aggregation network has much higher capacity compared to the Access network. This is shown in the figure 5-27.

https://t.me/learningnets

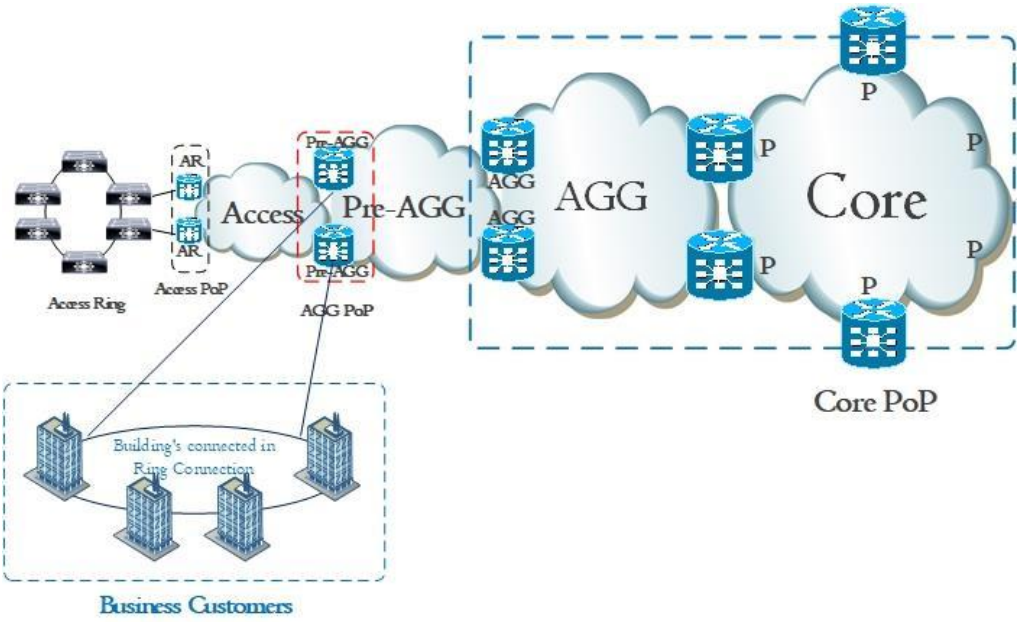


Figure 5-27 *Business/Corporate Customers are connected to the Pre-Aggregation Layer*

In figure 5-27, which will be explained in much detail later in the chapters that are covering the "ATELCO Scenario", Business/Corporate Customers are connected to Pre-Aggregation Layer, rather than Access Layer as Pre-AGG provides more bandwidth compared to Access Layer connections. This may not be the case for every ISP, because the bandwidth requirement of the Business/Corporate Customers will be playing a critical role when the ISP selects a location for the Business Customer termination.

If both Corporate customers and Residential customers would be connected to the Access network and a bottleneck occurs in the Access network, then the corporate customers would be affected. Most of the time corporate customers sign a contract with ISPs and they receive an SLA. If the ISP cannot fulfill the SLA terms, they have to pay penalty.

https://t.me/learningnets

Summary

In this chapter, the high level physical infrastructure of the Service Provider, physical connectivity of most commonly deployed services such as DSL, Cable, FTTX and Mobile Services, also Internet Datacenter, SP Server Farm, Internet Gateway, Global Network Connectivity and the Backbone network of the Service Providers were explained.

Some details might be different based on the networks, but these are the common parts and connectivity models in the SP networks.

After demonstrating the full picture, each module with sufficient detail was covered. It should be enough to help readers to understand and visualize the information, services and end to end packet flow of each service in the Service Provider network without too much complexity.

In the next chapter, Interconnection which is the heart of the Internet, will be explained in detail. There are some engineers who only deal with the Interconnection, Settlement Free Peering and the services related with it.

https://t.me/learningnets

Chapter-6 Service Provider Interconnections and Peering

Introduction

Peering is a BGP session between two Routers. When different companies have Peering with each other, they exchange network traffic over the peering session. There are three reasons to have BGP peering on Internet:

- Company wants to receive an Internet service
- Company wants to sell an Internet service
- Two companies exchange their customer prefixes and exchange network traffic but don't pay to each other, which is called Settlement Free Peering.

BGP is important, but this is not the topic of this chapter. It is important to understand the business models between the companies on the Internet.

- Who peers with whom?
- Which company pays to who?
- Why they peer or receive a service?

Service Providers sell Internet connectivity. Mostly they purchase Internet service from each other. Enterprises peer with Service Providers as well. From this chapter, you will be able to understand Settlement Free Peering, IP Transit, Paid Peering, Remote Peering and any "Peering" related topics. When you finish this chapter, you will not use the term "Peering" only as BGP neighborship, instead this term will remind you about Settlement Free Peering or Settlement Free Interconnection.

https://t.me/learningnets

Settlement Free Peering

Settlement Free Peering is the main topic of this chapter but different variations of Settlement Free Peering such as Paid Peering and Remote Peering will be explained as well.

Settlement Free Peering is also referred as Settlement Free Interconnection and here onwards, to make it short, SFI term will be used. SFI is an agreement between different Service Providers. It is an EBGP neighborship between different Service Providers to send BGP traffic between them without paying to the upstream Service Provider. Settlement Free Interconnection - SFI is a way of life for some Internet networkers. It is how the business of the Internet works.

SFI is both a business and technical relationship between two networks. It is an agreement where the networks concede to trade traffic between each other's customers, without payment or without settlement. There are multiple steps for creating an SFI relationship. There are also many ways to walk down those steps. Majority of traffic on the Internet flows over PNIs (Private Peering). The Private Peering concept will be explained later in this chapter.

Without private connection based SFI, the Internet would be centralized across intermediate platforms. Risk would be higher. Performance could be lower. Costs would be substantial. We generally reduce the risks and costs of the interconnection.

SFI relationships require more than just knowledge of how BGP works. Business realities must be understood and mixed skillfully into the recipe. This includes technical aspects to ensure the proper outcome. Simply throwing fiber over a cage wall worked 25 years ago.

Today, lawyers, accountants, project managers and performance testers are stakeholders. They are part of the conversation. They add collective value to the business decision. As a collective, they ensure that an agreement is as comprehensive and practical as possible. Not all SFI agreements are created equal. Many are based on handshakes or email.

There are three primary relationships between the companies on Internet. The Customer, Peer and Provider as it was briefly mentioned above.

Business relationships between the networks:

1. **Provider:** Typically, someone who is paid and has the responsibility of routing packets to/from the entire Internet.

2. **Customer:** Typically, someone who pays a provider with the expectation that their packets will be routed to/from the entire Internet.

3. **Peers:** Two networks that get together and agree to exchange traffic between each other’s networks, typically for free. There are generally two types of peering: public and private. Both will be explained in this session.

https://t.me/learningnets

Let's have a look at figure 6-1 to better understand the relationships between the Customer, Provider and Peer.

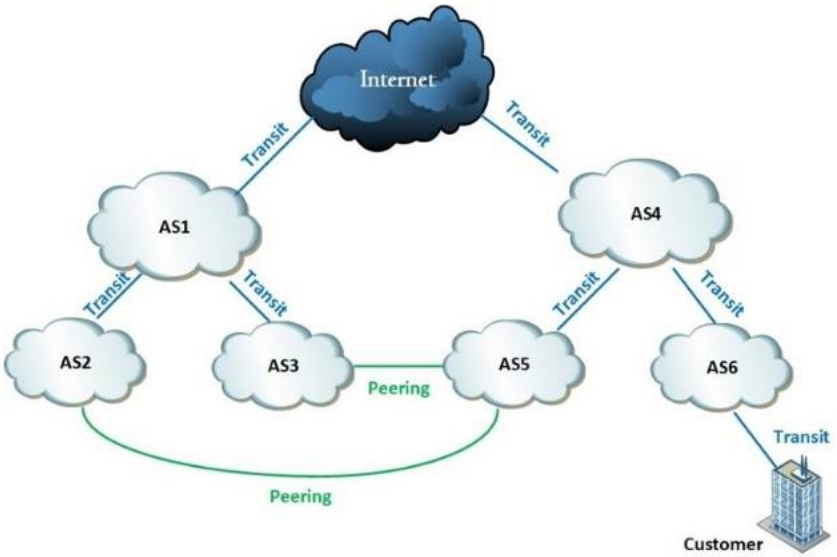


Figure 6-1 *Business relationship between the networks.*

In figure 6-1, AS2 and AS5 are peers with each other. They send their customer prefixes to each other. They don't exchange money for exchanging traffic between their customers. AS1 is the Provider of AS2 and AS3. Both AS2 and AS3 pay AS1 to reach the entire Internet.
As it can be seen from the figure, AS3 and AS5 are Peers as well, but this doesn't mean that AS2 and AS3 also need to be Peers. In fact, as it can be seen from the figure, AS2 and AS3 are not Peers. Peers don't send their Peers traffic to each other. Only their own prefixes and their customer prefixes are exchanged with their Peers. Let's have a look at different categories of Settlement Free Peering.

Private BGP peering

Private Peering is a direct interconnection between two networks, using a dedicated transport service or fiber. It is also known as bilateral peering in the industry. It may also be called a Private Network Interconnect, or PNI. Inside a data center, this is usually a dark-fiber cross-connect. It may also be a Telco-delivered circuit as well. If there is big amount of traffic between two networks, Private Peering makes more sense than Public Peering. Private peering can be setup inside Internet Exchange Points as well. Larger companies generally use Private Peering rather than Public Peering, since they want to select who they are going to be peer with. Also based on the large amount of traffic that is exchanged between them, they don't want to exchange traffic with everyone by joining the Public Peering.

https://t.me/learningnets

Public BGP Peering

Typically, Public Peering is done at the Internet Exchange Point. BGP Route Servers are used in Public Peering to improve scalability.

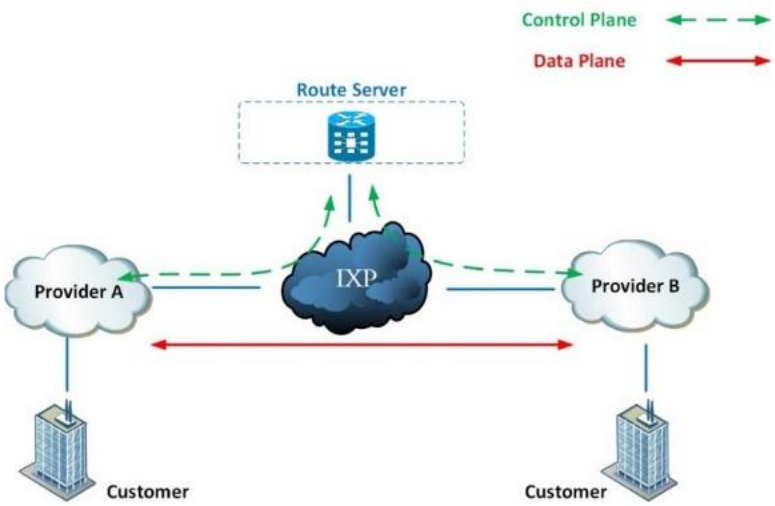


Figure 6-2 *BGP Route Server*

BGP Route Server is used at the Internet Exchange Point to simplify the BGP Peering process. Instead of managing and maintaining hundreds of Peering sessions in large Internet Exchange Points, BGP Route Servers are used. Every BGP speaking router has a BGP session with the BGP Route Server. Route Servers don't change the BGP Attributes, although the type of BGP Peering session is EBGP.

The BGP Route Server also doesn't change the next-hop to itself, thus it is used only as a Control Plane device, not a Data Plane. This means that the actual traffic is passed between the companies that participate to the Public Peering Internet Exchange Point.

BGP Route Servers are very similar to BGP Router Reflectors which are used in IBGP topologies. The difference is that BGP Route Server is used in EBGP, but Route Reflectors are used in IBGP.

Internet Exchange Point is commonly referred as IX. It is the BGP prefix exchange point between the Service Providers. The IXP concept will be explained later in the chapter with great detail. Internet Exchange Point networks are basically Layer 2 LAN. Each IXP participant is assigned an IP address from the common LAN and BGP neighborship is setup between the participants and the Route Servers of IX.

https://t.me/learningnets

Bilateral Peering

Bilateral Peering is when two networks negotiate with each other and establish a direct peering session. This is generally done when there are large amount of traffic between two networks. Tier 1 Operators just do Bilateral Peering as they don't want to peer with anyone other than Tier 1 Operators. The rest of the companies are their potential customers, not their peers.

Multilateral Peering

As mentioned above, Bilateral Peering offers the most control, but some networks with very open peering policies may wish to simplify the process, and simply "connect with everyone". To help facilitate this, many Exchange Points offer "Multilateral Peering Exchanges", or "MLPE".

- An MLPE is typically an exchange point that offers "Route-Server", allowing a member to establish a single BGP session and receive routes from every other member connected to the MLPE.
- Effectively, connecting to the MLPE is same as agreeing to automatically peer with everyone else connected to the MLPE, without requiring the configuration of a BGP session for every peer.

Basically, Public Peering and MLPE is almost the same thing and used mostly interchangeably. Objectives for an interconnection agreement to consider are:

- Provides cost savings and performance improvements
- Ensure the exchange of traffic is secure, stable, and resilient
- Establish timely cooperation for security and network incidents
- Usually includes a non-disclosure agreement

The team of engineers are responsible for the health and welfare of the network. They negotiate utilization, capacity and management parameters. Legal terms are negotiated by lawyers. While plain language is always best, legal language is what makes it an enforceable agreement.

Networks should cover all the necessary business, technical and legal points in the scope, such as term, jurisdiction and venue. Include all necessary parties in the conversation. Including a wide audience in the conversation helps to set realistic business goals.

Looking Glass

It is a server commonly deployed by an IXP to provide a view to the prefixes in the specific IXP. It gives publicly available information so any network owner can check the available prefixes in the IXP before they decide to join to that particular IXP.

There are many public available Looking Glasses in the world. They are configured as read-only, so anyone who wants to check the particular looking glass cannot change the BGP routing information. From http://www.bgplookingglass.com you can see the Looking Glass database.

https://t.me/learningnets

Benefits of Settlement Free Peering

Let's discuss the benefits of Settlement Free Peering in this section. Let's explain the important benefits.

1. Benefits for the Companies/Participants

   - Reduced operating cost: A transit provider is not being paid to deliver some portion of your traffic. Peering traffic is free! Through Settlement Free Peering, Transit Cost is reduced.

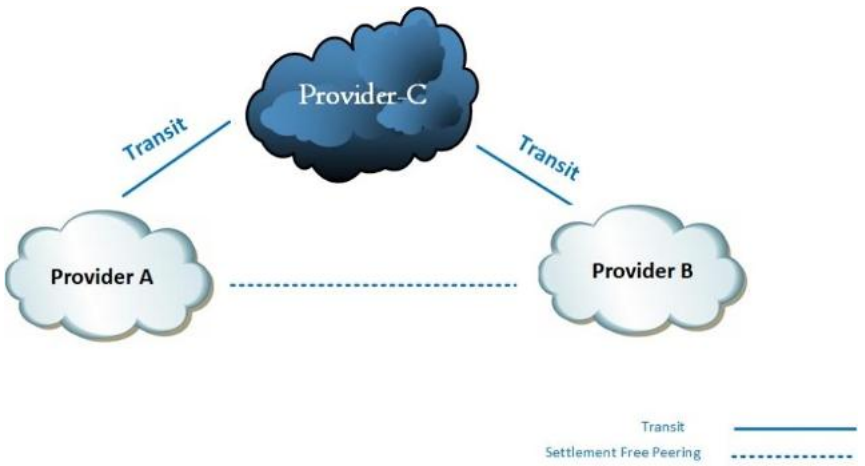


Figure 6-3 *Cost reduction through Settlement Free Peering & Transit Links*

   - Improved routing: By directly connecting with another network which you exchange traffic, you are eliminating a man-in-the-middle and a potential failure point.

   - Distribution of traffic: By distributing traffic over interconnections with many different networks, the scalability, reliability and redundancy are improved.

     In figure 6-4, Two Service Providers create a BGP Peering with each other and only send their customer and aggregate subnets. They don't need to use their upstream Transit Service Provider to reach each other's customers. Almost every country has Internet Exchange Points where Service Providers, Content Networks, CDNs, Enterprises, Mobile Operators, Carriers, TLD (Top Level Domains) and Root DNS Servers can meet.

https://t.me/learningnets

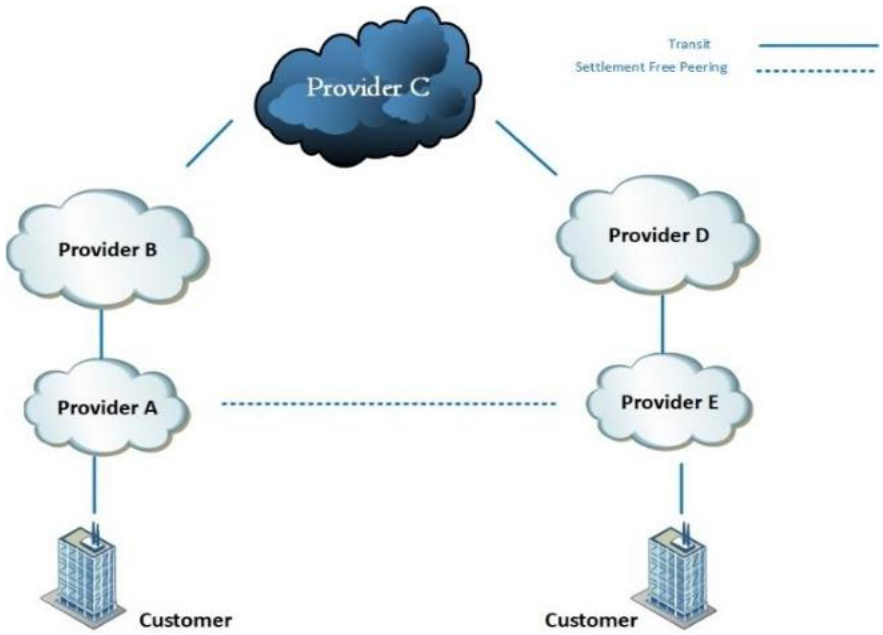


Figure 6-4 *Improved Routing Through SFI*

2. Benefits for Users

   Having Settlement Free Peering between the companies provides benefits to end users as well. Lower latency, higher reliability and better performance are some of these benefits.

   As the destination will be closer with direct peering session (compared to sending traffic to the Transit), latency is reduced and throughput can increase for end users. Also, even though companies peer with each other, they still keep their Transit Service as a backup, thus peering is also considered to provide reliability.

   More peering locations provides more reliability in general, though in network design more redundancy does not always mean more reliability.

3. Benefits for Operators and Other Companies

   Settlement Free Peering is useful for many different companies. Motivations for different companies might be different. Internet Service Providers might want to have Settlement Free Interconnections to reduce their Transit cost, while Content Providers want to have SFI to have predictable routing.

https://t.me/learningnets

4. Lower costs

   As the Transit bandwidth requirements are high for most of the Internet Service Providers, they get the benefits from SFI, as SFI provides cost advantages compared to IP Transit service for them.

5. Higher reliability

   For the companies having SFI, in addition to their Transit Service, they have redundancy which provides reliability.

   - More predictable routing.
   - Better performance for customers.
   - Better security (DDOS Mitigation for example).

Peering Requirements by IXP

Common requirements by Internet Exchange Points:

- Public AS (Autonomous System number).
  Internet Exchange Points request the participants to have Public AS numbers. 2 byte or 4 byte ASN space is accepted. The IXP itself has its own AS number which is used on the BGP Route Server to have BGP peering with the IXP participants.
- Public IP Address block
- Many IXP's require to be presented on the peeringdb (www.peeringdb.com) where all the AS related information is listed for the operators

Peering Requirements by Participants

- **Backbone capacity and Diversity:** Peer can ask other Peers about the specific amount of backbone capacity and redundant paths for reliability.

- **POP Locations:** It is important to have POP locations where there is maximum amount of traffic or closer to eyeballs. Number of POP's can be a peering requirement as well.

- **Traffic Requirement:** Is an important factor for peering, which economically makes sense for providers to peer with someone that can exchange a good amount of traffic with them, otherwise buying an IP Transit service might be cheaper.

- **Operational Requirement:** Is also an important requirement. As an example the peers should have a 24/7 NOC team which can respond when failure or attack happens.

Generally, companies don't want to peer with their existing customers so simultaneous peering and customer relationship doesn't happen. Though a network becomes larger in factors of backbone size, numbers of customers and so on, it can become a peer of their provider by time.

https://t.me/learningnets

Peering Policies

There are generally three types of Peering Policies. Companies announce their peering policies on *peeringdb*. Peering policy of a company reflects their willingness to peer with others. The Peering policy can be **Open, Selective or Restrictive.**

1. **Open** Peering Policy means the network is generally willing to peer with anyone without imposing specific tests or conditions. Content Providers and Search Engines generally have an Open Peering Policy.

2. **Selective** Peering Policy means the network is generally willing to peer with those who meet a specific set of requirements, but may not peer with everyone. Generally, Tier 2 Service Providers have Selective Peering Policy.

3. **Restrictive** Peering Policy means the network is generally inclined not to add any new peers. Restrictive networks may list a set of specific requirements, but with the bar so high that the intention is for no one to reach it. Generally, Tier 1 networks have Restrictive Peering Policy so they only have peering with other Tier 1 ISPs.

Peering Rules

- **Consistent Routing** is one of the most important peering rules (Same prefix length, same BGP attributes should be announced on all peering points)
-

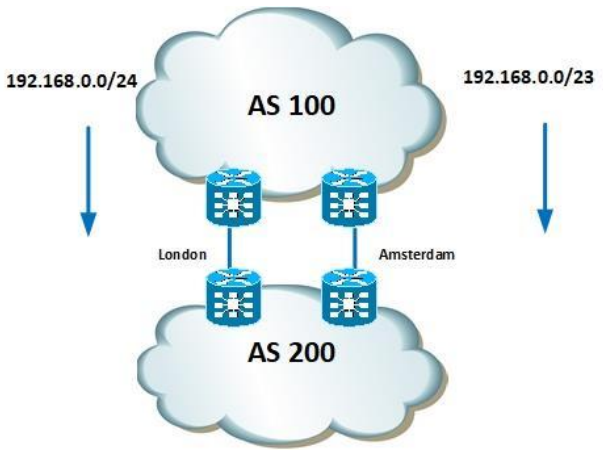


Figure 6-5 *Consistent Routing*

In figure 6-5, two Autonomous Systems peer at two different locations. One in London, another in Amsterdam. Based on the peering rule, 192.168.0.0 network should be advertised with the same subnet mask from both of the locations. In this topology, AS 100 is advertising 192.168.0.0/24 from London and 192.168.0.0/23 from Amsterdam.

https://t.me/learningnets

Based on the longest prefix length rule, AS200 chooses London as the exit point for 192.168.0.0/24 subnet. Traffic starting from Amsterdam to 192.168.0.0.0/24 subnet needs to be carried to London, inside AS 200 network. By advertising prefixes with different subnet masks, AS 100 is forcing AS 200 to perform "Cold Potato Routing".

- Don't point static route or default route to peer, this means that your peer can use your network as a Transit. Also, don't create a GRE tunnel to use your peer as transit, this can also be considered as this rule.
- In peering relationship, only customer routes and your peer routes are announced, not the Default Free Zone, otherwise this would be the same effect of having default route

What is IXP (Internet Exchange Point)?

IXP is a layer 2 network where multiple network entities meet, for the purposes of interconnection and exchanging traffic with one another. Internet Exchange Points start with a single Layer 2 switch at one location. Networks, peer with each other in this facility. When the number of participants grows, more switches are added at that location and more locations are added to the IXP itself.

For example, AMS-IX in Netherlands has many places inside many data centers and for each data center, they have more than one switch for Settlement Free Interconnection.

- IXP is often referred to as an Internet Exchange (IX), or "Public Peering".
- Today most Exchange Points are Ethernet based LANs, where all members are sharing a common broadcast domain, and each member is given a single IP per router out of a common IP block (such as a /24).
- IXP's are typically done at the carrier neutral data centers and in many cases; Datacenter owners provide racks for the peering fabric for free.

https://t.me/learningnets

IXP Best Practices

- The IXP provides a layer 2 shared switch fabric for peering networks to interconnect.
- There must be at least 3 networks (ASNs) connected for a facility to be considered as an IXP.
- The IXP should have neutral ownership and or management.
- The IXP should be at a secure location.
- The IXP should have a website with basic information's such as: Contact Information, Peering member's information, Peering statistics, Membership & Joining and Peering and technical policy.
- The IXP design is a Layer 2 fabric with most possible reliability, security and stability features (Many IXP's are considering to move their Layer 2 fabric to VPLS).
- Access to the IXP by members and fiber providers should be easy.
- IXP location should be closer to the participant as much as possible.
- IXPs are not generally involved in the peering agreements between connected ISPs, though some IXPs encourage their users to sell or buy Transit and involve peering negotiations.
- IXPs do not provide services that compete with their members such as Commercial Hosting Services and Transit Services

Why Networks Peer at the IXP?

An Exchange Point acts as a common gathering point, where networks who want to peer can find each other. A network new to peering will typically go to an exchange point as their first step, to also be able to connect to many other like-minded networks interested in peering with them. The more members an exchange point has, the more attractive it becomes to new members that are looking to interconnect with the most other networks. This is called as "critical mass".

Where are the Internet Exchange Points?

Most of the IXPs in the World are in Europe. There are many IXP's in North America as well. IXPs in Europe mainly work based on Membership model, while IXPs in U.S work based on Commercial model. There are exceptions in each case though. Most European IXPs grew from non-commercial ventures, such as research organizations. Most African IXPs were established by ISP Associations and Universities.

https://t.me/learningnets

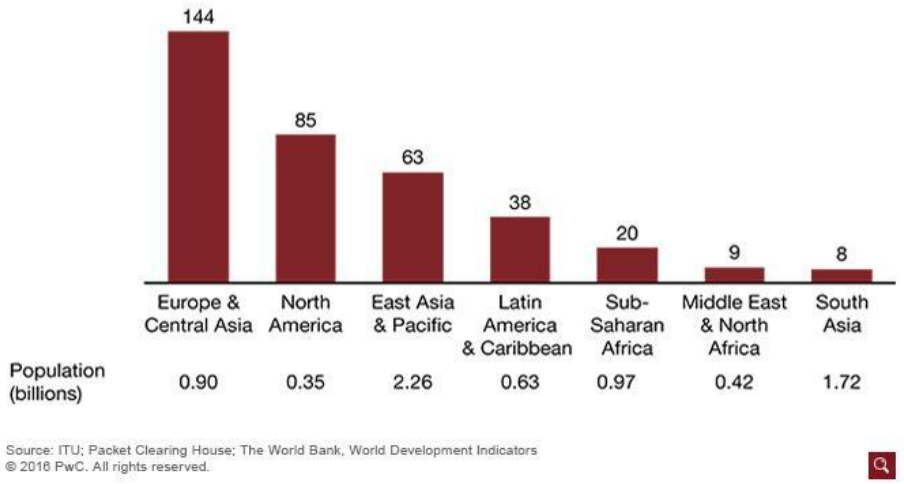


Figure 6-6 *Number of IXPs in the World (source: www.PCH.com)*

Packet Clearing House (PCH) provides an excellent information about the Internet Exchange Points, number of participants in each IXP, their country and the established cities, whether IPv6 is enabled and many other useful information. It can be found on https://www.pch.net/ixp/dir

IXP Membership vs. Commercial Models

Table 6-1 Commercial vs. Membership Based IXP Models

	Commercial IXP	**Membership Based IXP**
Services	More	Less
Carrier Neutrality	Not a Problem	It is an issue
Location	Mostly in U.S	Mostly in Europe
Independency	Dependent	Independent
Managed by	IXP Owner	Participants

In the comparison chart shown in Table 6-1, the IXP Membership model has been compared with the Commercial Model. One of the comparison criteria is ‘Carrier Neutrality’

https://t.me/learningnets

What is Carrier Neutrality

According to Internet Society:
Whether commercial or not, virtually all IXPs are owned and managed neutrally with respect to carriers, ISPs and Co-location Providers.

Many ISPs have express strong feelings about the importance of neutrality of IXPs and believe that the success of the IXP is related to their neutrality. If the IXP is owned and operated by a Carrier, other ISPs and the Carrier networks don't want to become a part of that IXP. One reason is that companies don't want to help their competitor's to success. In the future, if the IXP grows, their competitor will be the HUB for the Interconnections and this would be bad for their business interest. Thus, companies would like to join to an IXP which is not owned by the Carriers/Other ISPs.

European IXPs

Most of the IXPs around the World are running in Europe.
London, Amsterdam, Germany, France and Russia are the main locations and locate the biggest IXPs in the World.

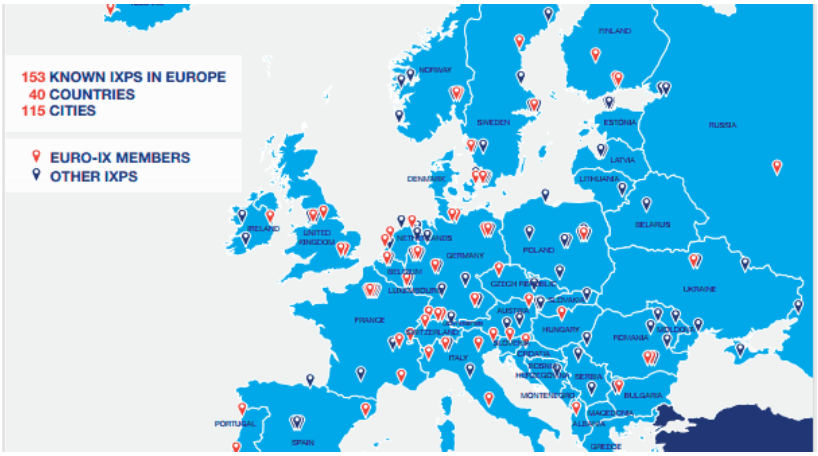


Figure 6-7 *European IXPs*

https://t.me/learningnets

IXPs around the World

Figure 6-8 shows the IXPs located around the world.



Figure 6-8 *IXPs around the World*

What is Local IXP

Generally, country wide Public Internet Exchange is called a Local IXP (Internet Exchange Point). A Big reason to setup an IXP is to keep local traffic as local. Imagine two ISPs in the same country, one of ISPs has a server and the other ISP has a client which wants to reach that server, and their only connection is through another ISP which is located at another country.

Without having Local IXPs, traffic would go to the other country and come back.
All IXPs start as Local IXPs and some of them by the time evolve to a Regional IXP. In the same country, there might be several Local IXPs. In U.K, Germany, Russia and U.S there are many IXPs for example.

What is Regional IXP

There are also IXPs which serve not only to their cities or inside their country, but they provide services in many countries in their region. Regional ISPs peer with each other and show up at several of these Regional IXPs.

There are many regional IXPs such as AMS-IX, LINK, DE-CIX. These IXPs have hundreds of ISPs, Enterprises and other network operators outside of their local country. Some of these IXPSs even have Global presence. For example DE-CIX and AMS-IX operate in Europe as well as in U.S and ASIA.

https://t.me/learningnets



Figure 6-9 *DE-CIX Regional IXP*

DE-CIX Regional IXP is shown in figure 6-9. DE-CIX Regional IXP can be considered as a Global IXP, but layer 2 connection is not extended between the countries.

Who are the Internet Hubs in the World?

The IXPs which have the most number of networks and the unique prefixes are considered as Internet Hubs, since most of the Global Internet traffic goes through these IXPs. Content was centralized in the U.S in late 90s and early 2000, but with the advent of CDNs and Content Provider Networks, content is distributed around the world. This phenomenon is called “Flat Internet”.

New Internet Hubs are the largest IXPs in Europe and largest IXPs in U.S. These are the Global Internet Hubs as most of the content as of 2019 are located at these points. Some Asian countries such as Singapore and Hong Kong have large number of connected networks, so they can also be considered as Regional Hubs.

What are Tier 1, Tier 2 and Tier 3 Internet Service Providers

A Tier 1 Service Provider is a network which does not purchase transit from any other network. It only peers with other Tier 1 networks to maintain global reachability. A Tier 2 Service Provider is a network which has Internet service customers and some Settlement Free Peerings, but still buys full transit to reach some portion of the Internet.

https://t.me/learningnets

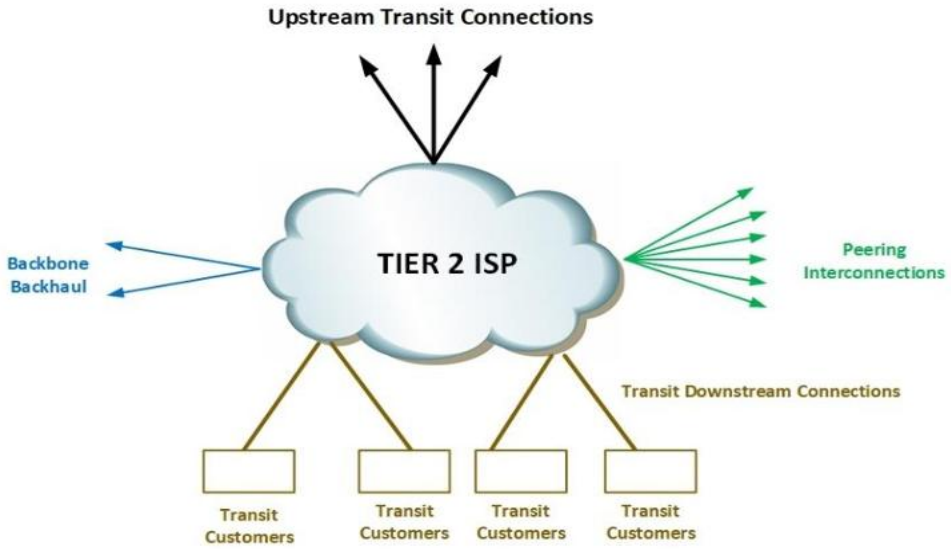


Figure 6-10 *Tier 2 ISP and its Connections*

A Tier 3 Service Provider is considered as a stub network. They are generally considered as Local/Access ISPs. They don't sell any IP Transit service to anyone. Sometimes, Tier 3 ISPs definition is used to describe Enterprise, SMB or End Users.

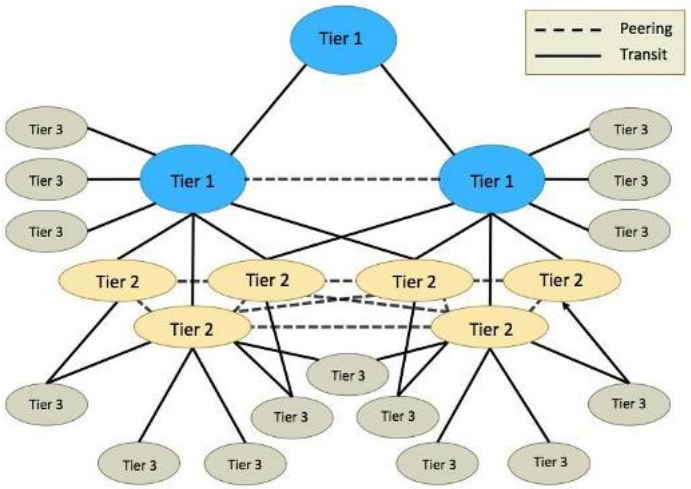


Figure 6-11 *Tier 1, Tier 2 and Tier 3 ISPs*

https://t.me/learningnets

Tier 1 Service Providers

A Tier 1 ISP is an ISP that has access to the entire Internet Region solely via its settlement free peering relationship. Tier 1 ISPs only peer with other Tier 1 ISPs and sometimes with CDN and the Search Engines. They don't have any Transit ISP but they are the top tier ISP.

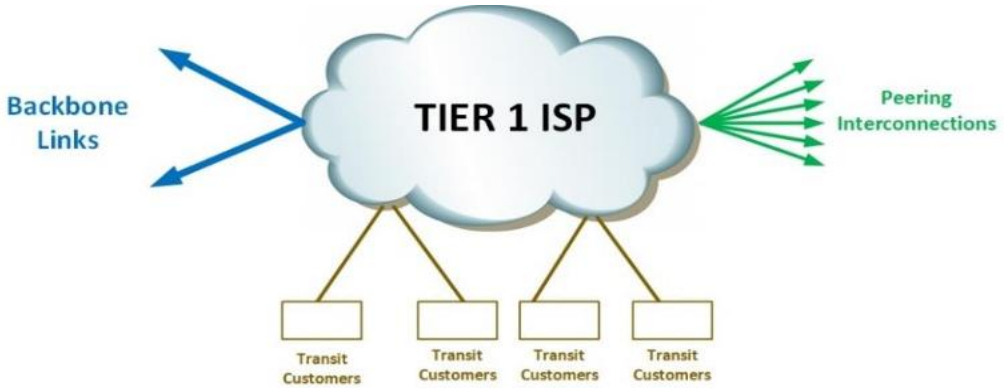


Figure 6-12 *Tier 1 ISP and its Connections*

As of 2016, there are 13 Tier 1 ISPs which don't have any transit provider. Baker's Dozen is considered as the Tier 1 ISP List and every year the list is updated with the ISP ranking. The list is provided by measuring the Transit IP Space of each ISP.

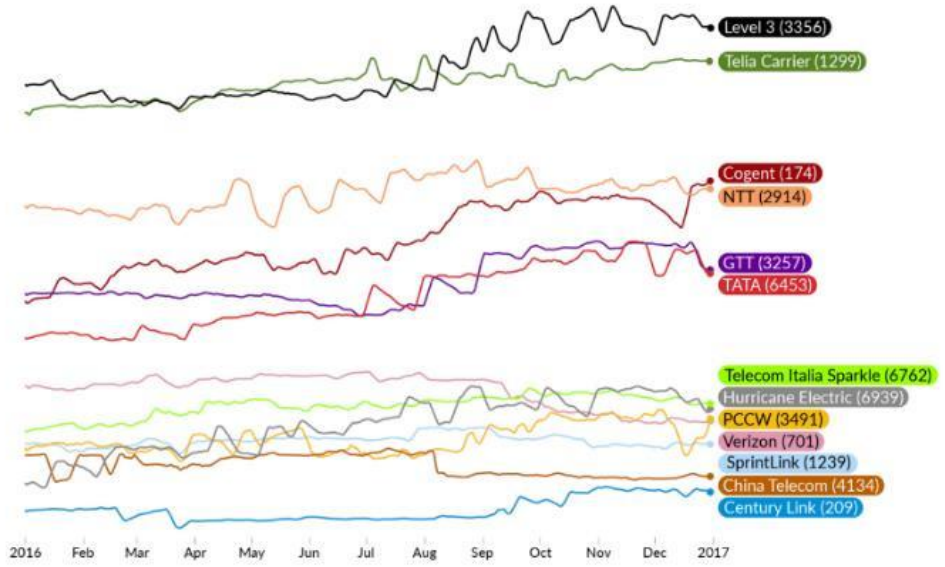


Figure 6-13 *2016 Baker's Dozen Tier 1 ISP Rankings*

https://t.me/learningnets

Is there a Global Tier 1 Internet Service Provider in the World?

All of the Service Providers in the Baker's Dozen list shown in figure 6-13 are considered Global Tier 1 ISPs. None of them pay for IP Transit Service, but they may lease or get an IRU for fiber access, so it doesn't mean they don't pay any cost for their communications. In fact, since they carry huge amount of traffic between countries and continents, they pay the cost for optical networks, colocations, IXP port costs and so on, even more than any of the other Providers.

On the other hand, for example, Turkish Telecom is a National Tier 1 ISP (Since they don't pay to any other providers to reach the IP prefixes in their country). However, they buy an IP Transit Service to reach many other destinations in the world, thus Turkish Telecom is not a global Tier 1 ISP (of course they have peering in the IXPs outside of Turkey, especially in Europe, to reduce their IP Transit costs).

Flat Internet vs. Hierarchical Internet

Content placement has changed a lot in the recent years. User traffic is not carried to the content anymore, in fact content is carried closer to the user as it can be seen from figure 6-14. Around 10 years ago, most of the content was only accessible from North America, but today, Internet is considered as "Flat Internet" due to Content Distribution Network and Settlement Free Peering.

When Facebook, Amazon, Google. Apple. Microsoft and other large Content Providers and CDN companies have new (mostly popular) content, they send it to their worldwide locations, so users reach the content via the closest location to them. Akamai, Amazon, Cloudflare and Limelight are the biggest CDN providers as of 2017. They have thousands of content caches around the world. You can see the network map of Akamai CDN network at the following link:

https://www.akamai.com/uk/en/solutions/intelligent-platform/visualizing-akamai/media-delivery-map.jsp

During the last decade, most of the major content providers (such as Google, Facebook, Netflix, Apple, etc.) have built their own CDN network. They prefer to distribute their content via their CDNs and use peering connections as less as possible. And as a new trend, some Tier-1 operators have built their own CDNs and became popular with using their network power all around the world. Level-3, TATA and Telia are good examples of this trend.

https://t.me/learningnets

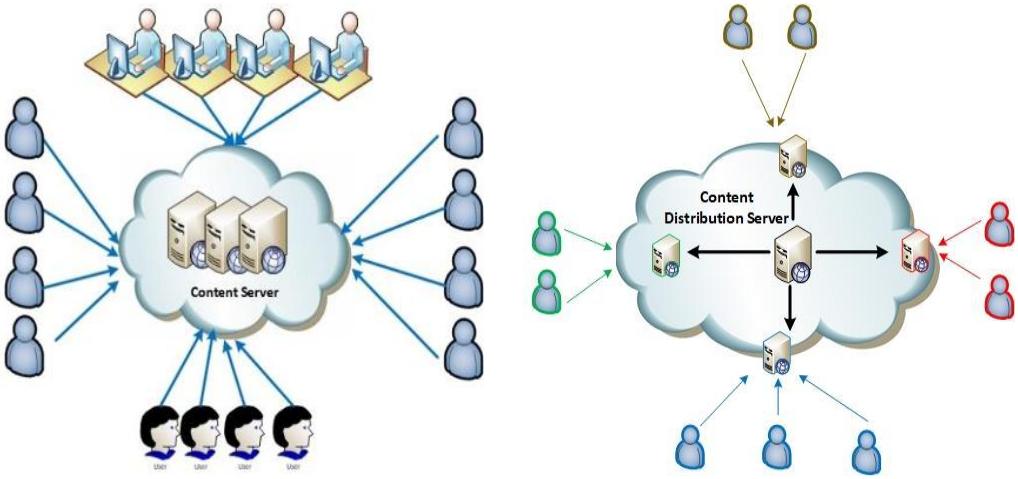


Figure 6-14 *Distributing content to the edge of the networks (Closer to the End users/Eyeballs)*

When you try to access a content, you most likely use the networks of one of these companies. When you have a cacheable content, which will be accessed all over the world, placing them on a CDN reduces the latency for the users who want to access it. CDN providers prefer to locate their servers (also called as clusters) in networks (which is closer to eyeball users) as wide as possible. Mostly, they choose the main operators in a country/region and deal with them about placing clusters in their networks.

CDN clusters are physically placed in Service Providers data centers and are connected to the operator's core network via BGP, such as a SFI connection. This allows the CDN provider to reach eyeball users with high quality and low latency. The operator benefits from this connection with reducing transit or SFI costs and also provides better quality for delivering the content to its customers.

CDN providers locate some clusters inside the Service Providers data center network as a public cluster with an Internet access. The objective behind public clusters is serving closely to several small ISP customers in specific regions. Thus, placing a content on CDN reduces remote user's latency, increases throughput of the sessions and provides better quality of experience for the content consumers.

When Latency decreases, throughput increases!

https://t.me/learningnets

Once, CDN providers deploy a cluster to a network, they fill it with cacheable content and start serving this content from cluster. Cacheable content is the common and mostly non-dynamic content such as media files or software packages. To identify cacheable content, CDN providers and content providers use enhanced complex custom algorithms to achieve maximum optimization. As an example, when a new IOS release announced for iPhone, Apple immediately pushed this software package to clusters all around the world, so all Apple customers could reach the new IOS via the closest CDN cluster to them (which is most probably in their ISPs network) instead of going somewhere else.

Traffic Ratio Problem between Content Providers and Eyeball Networks

First let's explain Hot Potato Routing as the following problem happens because the ISPs default Policy is based on Hot Potato Routing.

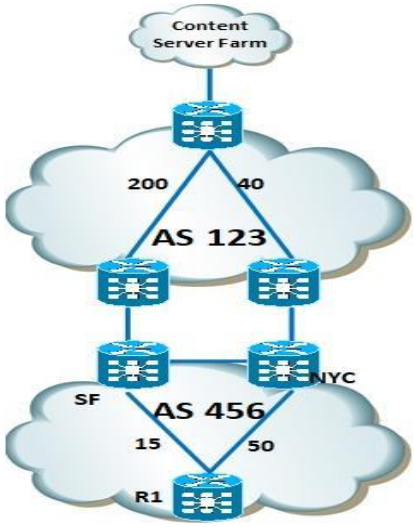


Figure 6-15 *Hot Potato Routing*

In figure 6-15, there are two ISPs. AS456 is the Customer and AS123 is the Transit ISP. Both ISPs by default send the network traffic to the other network as quickly as possible. In networking term, sending as quickly as possible is achieved through IGP metric. If there are multiple paths to the BGP next-hop, cumulative IGP smallest cost path is considered as the quickest path.

In the topology shown in figure 6-15, AS456 has two paths to the Content Farm which is located in AS123. Path 1 through SF, Path 2 through NYC. IGP metric to BGP next-hop is shown in the topology. To the SF, the IGP cost is 15 and to NYC the IGP cost is 50. Since IGP cost to SF is smaller than NYC from R1 (R1's point of view), SF is the quickest way to send the network traffic to AS 123. This is called as Hot Potato Routing.

https://t.me/learningnets

Dispute between the Peers can happen due to Traffic Ratio. Content heavy Providers have an advantage from the Hot Potato Routing point of view.

- Hot Potato Routing is the ability of a network to hand off traffic to other networks at the earliest possible moment. It is also known as “closest exit routing”.
- Cold Potato Routing is a situation where a network retains traffic on its network for as long as possible. It is also known as “best exit routing”.

Eyeball networks need to carry the traffic longer on their own network than content networks, so eyeball networks are considered as pull heavy (Downstream traffic), while content networks are push heavy (Upstream traffic).

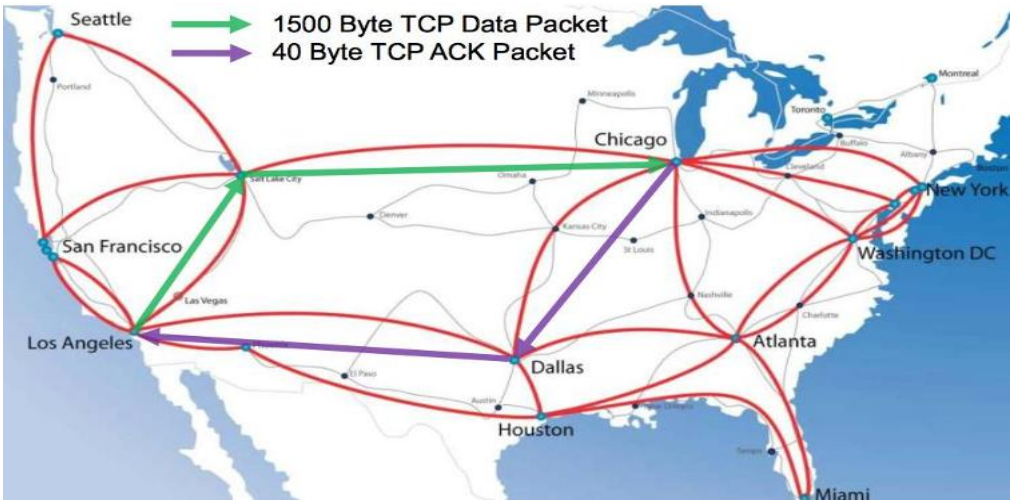


Figure 6-16 *CDN Traffic*

As an example, Content network in Los Angeles sends 1500 Byte Packet vs. Eyeball in Chicago that sends only 40 Byte. The Eyeball network might complain about the Traffic Ratio. Eyeball heavy networks, such as Access ISPs, might say that from our customer to the content server, only a single packet is sent but in reply 10, 20 or even 30 times bigger traffic is sent to our network. This traffic needs to be backhauled by our transport circuits and costs us money.

Access ISPs in this case might think that they don’t get mutual benefit from the Peering agreement and they may want to DE-PEER the connectivity. CDNs are Content Delivery Networks and they generally create Traffic Ratio Problems; even traffic ratio problem doesn’t prevent other networks to have peering with CDNs.

https://t.me/learningnets

What happens when two Tier 1 Providers dispute?

When two Tier 1 providers have a dispute, both of their customers get affected. A Customer who has a single upstream to one of these Tier 1 providers has no choice, so they suffer from the outages. When two Tier 1 providers have a dispute, generally they may not immediately "de-peer" each other. They just don't upgrade the bandwidth on their peering links which creates poor performance. Customers in this case may not have any other option than finding a different Transit Provider, so both providers may lose revenue.

Example: How Netflix distributes its content?

Netflix uses a Transit ISP in U.S as well as having their own CDNs in U.S and throughout the world, as it can be seen from figure 6-17. They started to deploy their own caches in the Service Providers and in the IXPs (Internet Exchange Points) where there are Netflix users. As a video content company, being as close as possible to their customer is very important since the latency and throughput is crucial for live shows and other real time videos (Though some videos can be buffered and many providers use Adaptive Bit Streaming, such as YouTube).

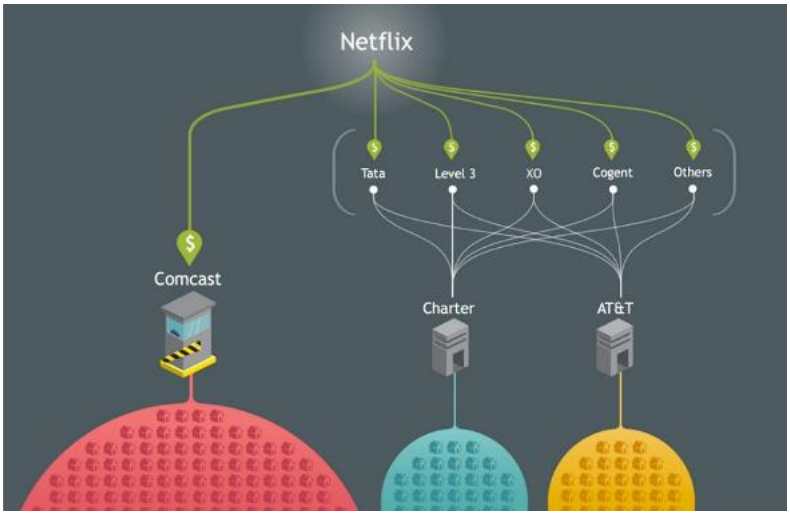


Figure 6-17 *Netflix Interconnection Network*

https://t.me/learningnets

What is Remote Peering

- Remote Peering is peering without a physical presence at the peering point. Only the remote peering provider nodes are placed at the IXPs
- Remote peering providers are located at the many IXPs and per IXP, vlans are extended up to customer location.
- Trunk is setup between Remote Peering Provider Router and the Customer
- It reduces the total peering cost of the customers.

Remote Peering Model

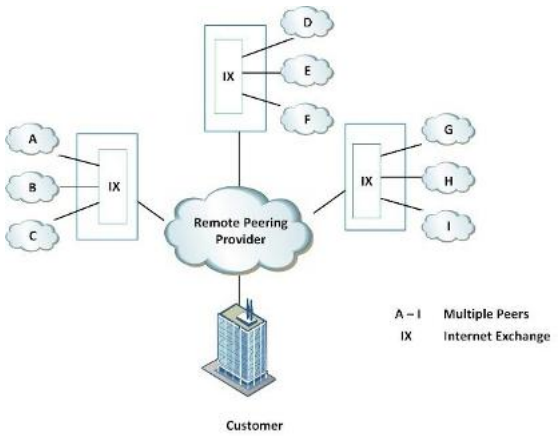


Figure 6-18 *Remote Peering Model*

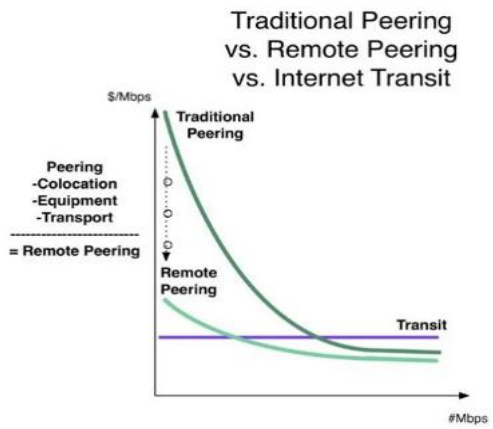


Figure 6-19 *Traditional Peering vs. Remote Peering vs. IP Transit*

https://t.me/learningnets

Pros and Cons of Remote Peering

With traditional peering, the real goal is to keep local traffic as local but with remote peering, customers from different countries can receive the content from different countries which increases the latency. Thus, remote peering may not make sense for the networks outside of the region. Remote Peering can reduce the cost of peering significantly, especially when the peering traffic is low.

Remote Peering and IP Transit total cost of ownership should be compared by the small network operators.

Remote Peering Providers

There are couples of Remote Peering Providers which have a connection to many IXPs Globally. Console Connect, Remote IX, IX Reach and so on. They have the connections to the largest IXPs such as LINK, DE-CIX, AMS-IX, MSK-IX. Not all Remote Peering Providers have a presence in all these big IX, so if you are considering to buy remote peering service, check their service locations, pricing and flexibility in terms of bandwidth which you can receive. Most of these Remote Peering Provider companies provide flexible bandwidth. For example, IXP may not allow you to receive 100Mbps service but Remote Peering Providers do.

IP Transit Service

IP Transit is the service of allowing traffic from a customer network to cross or "Transit" the provider's network, usually used to connect a smaller Internet Service Provider to the rest of the Internet. It's also known as Internet Transit. ISPs simply connect their network to their Transit Provider and pay the Transit Provider, which will do the rest.

Selling an IP Transit Service in the IXP

Selling an IP Transit Service in the IXP is common. Many big Service Providers such as Tier 1 or Regional Tier 2 Providers join an IXP as they see that an IXP is not only a peering point but a location where they can sell their IP Transit Services. Although some IXPs don't allow selling or buying an IP Transit, there is no real control mechanism which can prevent this situation. When companies have a peering, they still receive an IP Transit Service (Except Tier 1s) and they use IP Transit as backup connection.

Why should you place Peering Router at the IXP, Not at your Location?

In traditional peering (Not remote peering), the Peering router is located at the IXP, not at the customer site. The benefit of this is that unwanted traffic can be filtered directly at the IXP router and is not carried over the backhaul link to the SP network. This is important especially when the providers peer is in a different country, so they don't consume their expensive International bandwidth capacity.

https://t.me/learningnets

Summary

Are you still using the term “Peering” only when you talk about traditional BGP neighborship or you think that you will use “Peering” when you discuss Settlement Free Peering which is exchanging customer prefixes with each other by two companies.

In this chapter, the Internet Interconnection model was explained. Settlement Free Peering (SFI – Settlement Free Interconnection) is the backbone of the Internet. Without SFI, today we wouldn’t have Internet. We didn’t talk about the Internet History in this book as it has been covered in numerous books earlier, but as you can remember it started between couple government entities in U.S and then evolved with the NAP (Network Access Point) model. Today Internet Exchange Points are the extensions of early Internet NAPs.

In this chapter, Tier 1, Tier 2 and Tier 3 ISP models were introduced. There are 13 Tier 1 ISPs in the world, this number can change anytime with the merger and acquisition but the idea with the Tier 1 ISP is that they don’t have any Upstream Transit ISP.

Tier 1 ISPs reach the whole Internet just via their Settlement Free Peering agreements with each other. Tier 1 ISPs don’t peer with Tier 2 ISPs as Tier 2 ISPs are their customers. However, they can peer with Content Providers and CDNs.

Most of the Peering connectivity in the world is now done between Content Providers (such as Google, Facebook, Microsoft, etc.) with Tier 1 and Tier 2 ISPs and also between Tier 1 ISPs with each other and Tier2 ISPs with each other. Also, there are significant number of Peering connections between Tier 2 ISPs.

Tier 1 ISPs reach the whole Internet just via their Settlement Free Peering agreements with each other. Tier 1 ISPs don’t peer with Tier 2 ISPs as Tier 2 ISPs are their customers. Internet Transit or IP Transit model was introduced in this chapter and the relationship between the Customer and Provider was mentioned. The traffic ratio problem was mentioned and the power of Access Providers was demonstrated. The idea is that all interconnection is made through “handshake” is a myth. While that approach may work for creating BGP sessions on an IXP, it does not work to protect critical business interests. Operating a network at scale is hard. Removing as much risk as possible is important.

Settlement Free Peering agreements might be complex, but with the simplest explanation, if two companies are doing peering, they both get benefits from the peering. For decades, both Private Peering and Public Peering have been done by many companies in the World. As of 2019, there are almost 500 Internet Exchange Points (IXP) in the World and the number is growing.

Not only the number of IXPs but also the number of members in the IXPs is growing as well. Bandwidth is also growing as well. Today in some large IXPs there are 800+ participants. They became a critical Hub for data acquisition. If you are in a Service Provider or Large Enterprise Business, knowing the topics in this chapter is must for you. If you have skipped reading this chapter, it is highly recommended.

https://t.me/learningnets

Chapter 7

Chapter-7 ATELCO National Internet Service Provider Design

Introduction

ATELCO is a leading Telecommunication Service Company in Middle East. They have nationwide backbone infrastructure in fictitious country Greenland. They have around 11 million customers in the country; most of them are residential customers.

They have Business/Corporate customers from different fields such as other Internet Service Providers, Hosting Providers, Airports, Banks, Hospitals, Holding Companies, Newspapers, Application Providers, Content Providers, Universities, Hypermarkets, Government Companies etc.

https://t.me/learningnets

Regional Connectivities of ATELCO Network

ATELCO's nationwide network consists of 4 regions in the country. They have their POP locations across 60 cities in the country. ATELCO has an international connectivity from each of their 4 regions. Figure 7-1 shows ATELCO's Network Consisting of 4 Regions.

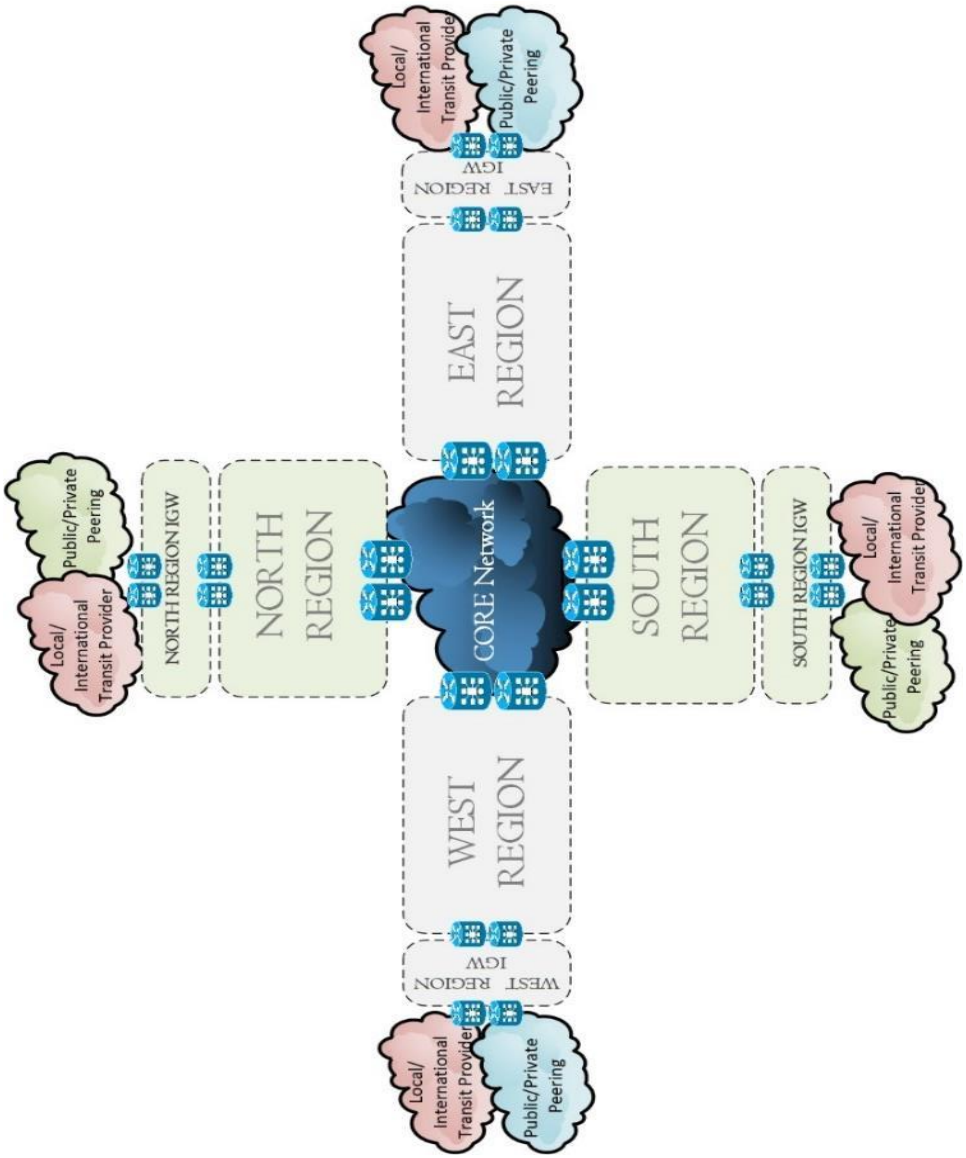


Figure 7-1 *ATELCO's Network Consisting of 4 Regions*

https://t.me/learningnets

1. Northern Region
   This region collects all customer traffic from the North part of the country and consists of around 100 Access POP locations, more than 10 Aggregation POP locations and 2 Core POP locations. Also, ATELCO has 1 Internet Gateway POP location, placed in one of the two CORE POP locations in a major city. At the Internet Gateway location, they have shared services routers, Internet Settlement Free Peering connections, IP Transit Provider termination and many cache servers of the content providers.

   The Northern region also serves as a backup for all the other regions, so in case of any Internet outage in other regions, it will provide Internet connectivity to customers in those regions.

2. Southern region
   This Region collects all customer traffic from the South part of the country and consists of around 100 Access POP locations, more than 10 Aggregation POP locations and 2 Core POP locations. Also, ATELCO has one Internet Gateway POP in the Southern region, placed in one of the two CORE POP locations in a major city. At the Internet Gateway locations, they have shared services routers, Internet Settlement Free Peering connections, IP Transit Provider termination and many cache servers of the content providers.

3. Eastern region
   This region collects all customer traffic from the East part of the country and consists of around 100 Access POP locations, more than 10 Aggregation POP locations and 2 Core POP locations. ATELCO in the Eastern region has one Gateway POP location, residing in one of the two CORE POP locations where they have shared services routers, Internet Settlement Free Peering connections, IP Transit Provider termination and many cache servers of the content providers.

4. Western region
   This region collects all customer traffic from the west part of the country and consists of around 100 Access POP locations, more than 10 Aggregation POP locations and 2 Core POP locations. ATELCO has one Gateway POP location in the Western region, residing in one of the two CORE POP sites where their shared services routers, Internet Settlement Free Peering connections, IP Transit Provider termination and many cache servers of the content providers exist.

As a summary of ATELCO's POP locations in the 4 regions, ATELCO only has one Internet Gateway POP location in each region, but various numbers of Access, Aggregation and Core POP locations. In figure 7-2, ATELCO's POP interconnections are shown.

https://t.me/learningnets

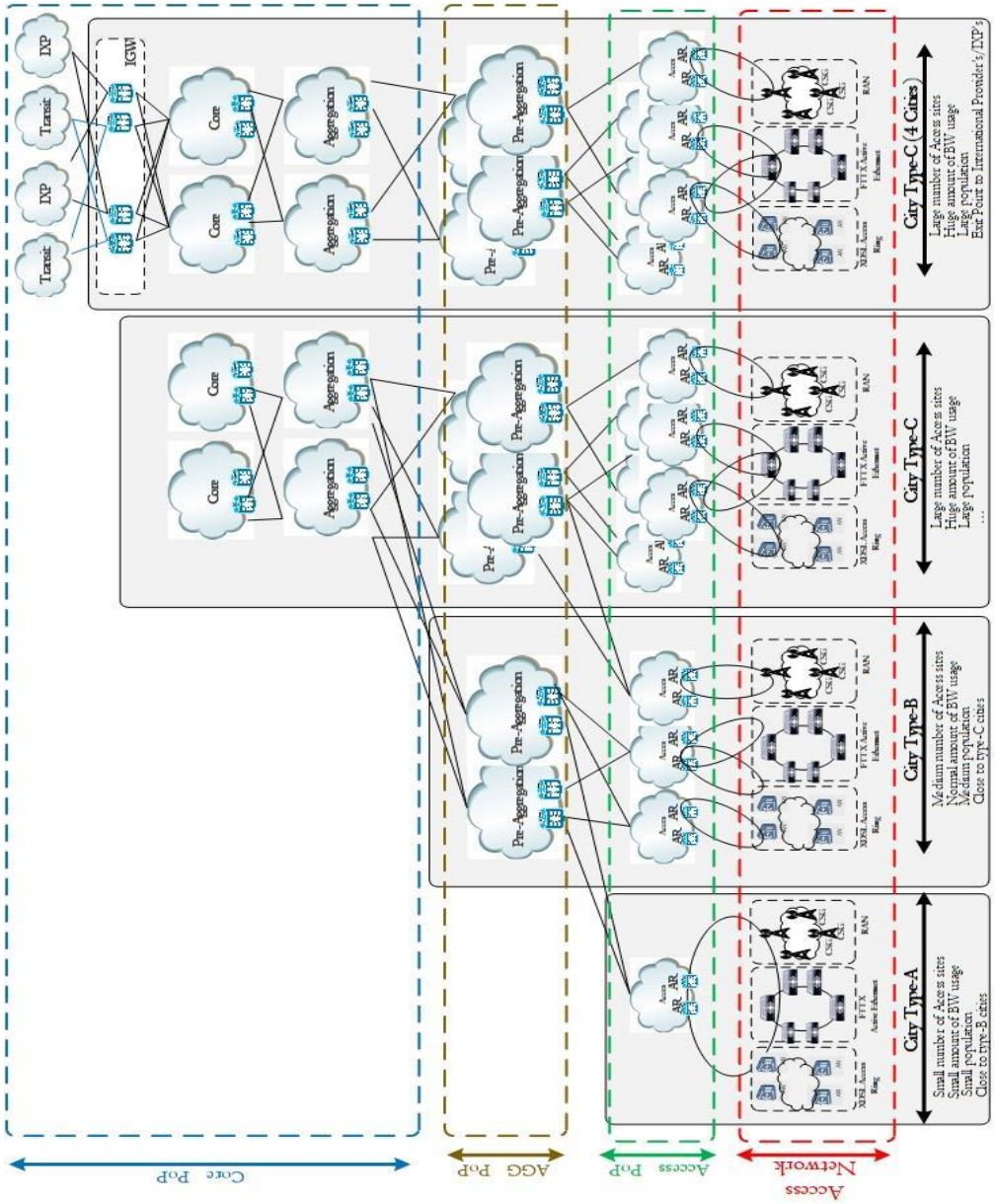


Figure 7-2 *ATELCO's POP Interconnections – Metro network connections*

For a high quality view of this Figure, visit www.orhanergun.net/SPBook/CityTypes.jpg

https://t.me/learningnets

There are three type of cities in each region which are named as Type-A, Type-B and Type-C cities. Cities are categorized by ATELCO based on the following criteria:

- Population of the city
- Number of service subscribers
- Aggregated bandwidth for the users

Based on this type of categorization, each city type is explained bellow.

Type-A Cities

This type of city has only Access POP locations. The Access POPs are connected to each other via Ring topology. For connecting the Access POPs to the backbone of the network or connecting to the Internet, these cities are connected to the nearest Type B-City, where there is an Aggregation POP location.

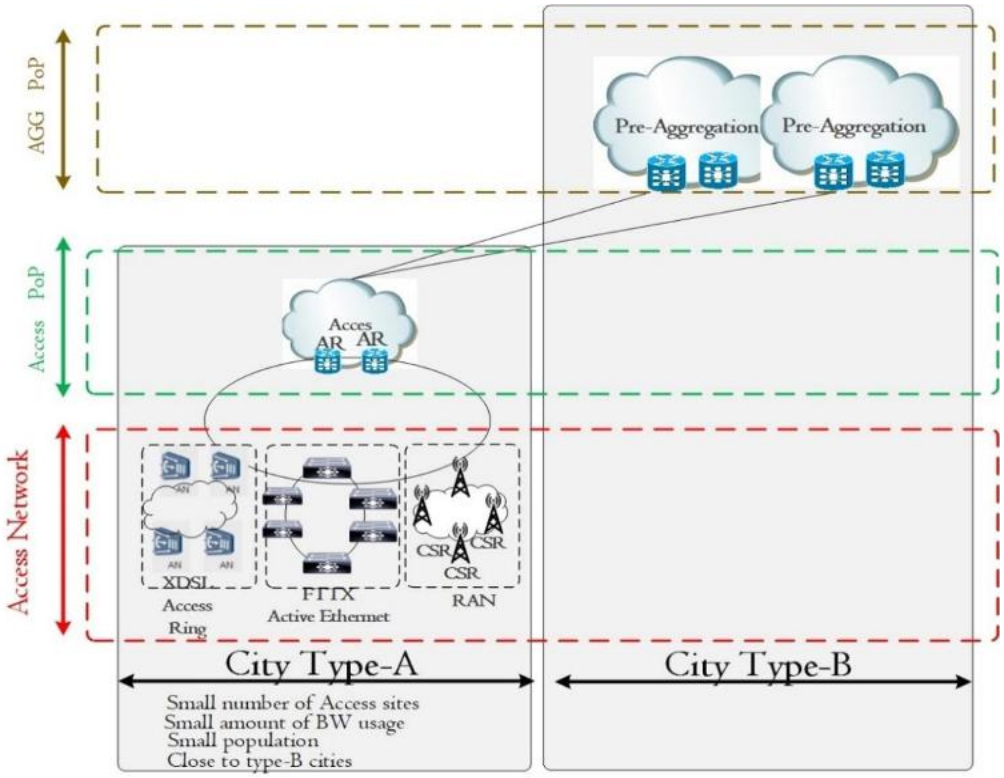


Figure 7-2 *(A)* *ATELCO's POP Interconnections – Type A Cities*

https://t.me/learningnets

Type-B Cities

This type of city has both Access and Aggregation POP locations. Access and Aggregation POPs are connected in a Ring topology and for connecting to the backbone of the network or connecting to the Internet, these cities are connected to the Core POP sites of the nearest Type-C city.

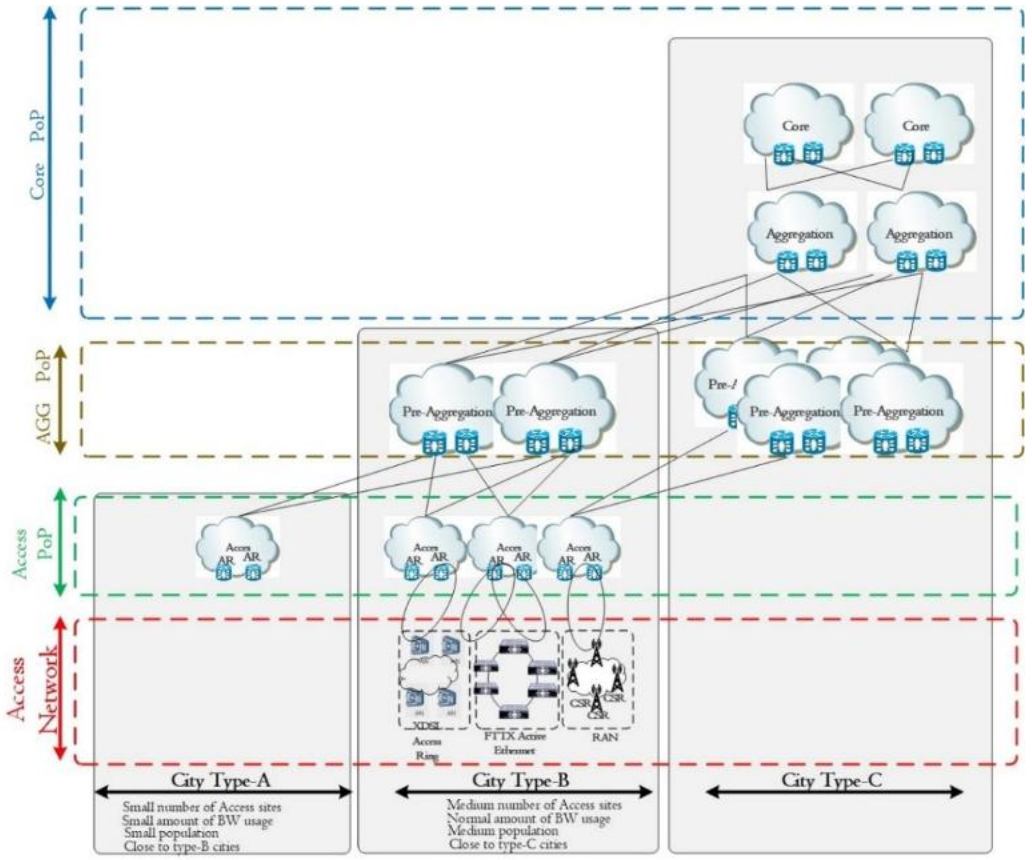


Figure 7-*2 (B)* *ATELCO's POP Interconnections – Type B Cities*

https://t.me/learningnets

Type-C Cities

This type of city has all Access, Aggregation and Core POP locations. The 4 Core site locations around the country are Type-C cities. Although these 4 Core sites have all the specifications of the Type-C cities, they also have the Gateway POP locations connecting to transit providers and peering with other service providers.

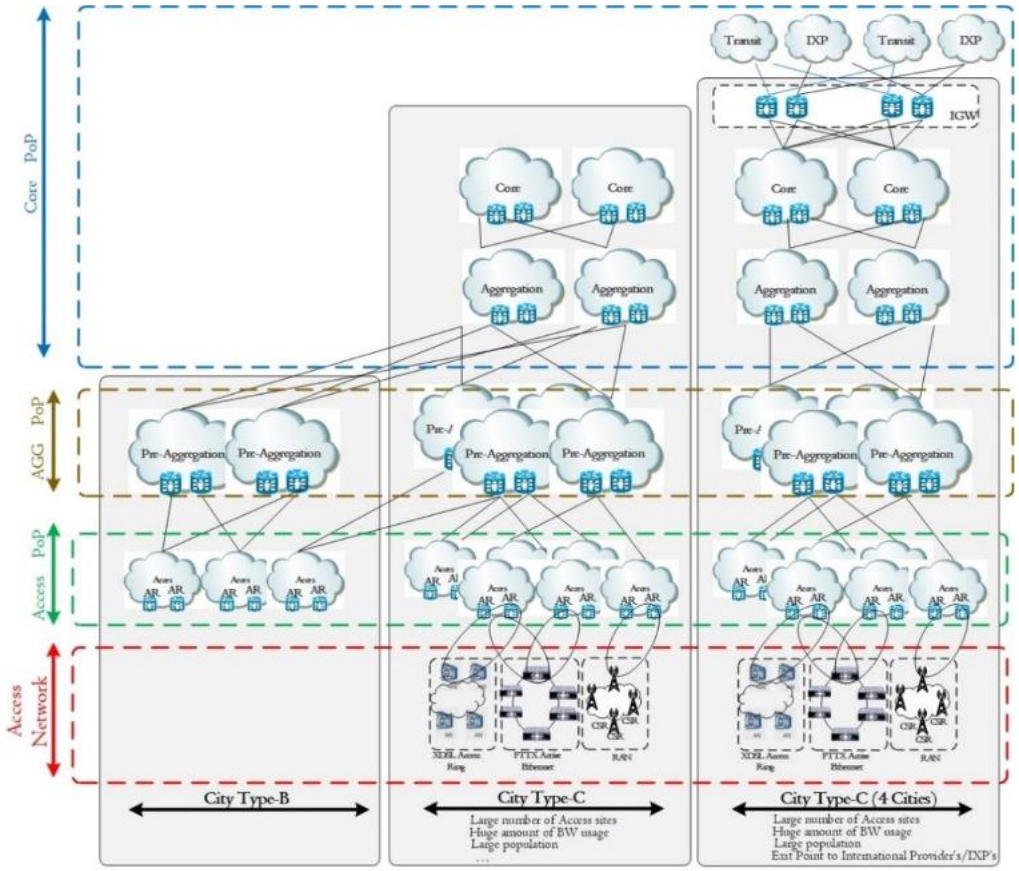


Figure 7-*2 (C)* *ATELCO's POP Interconnections – Type C Cities*

There are totally 60 cities in ATELCO's network. As it was mentioned above, there are 8 Core POP locations in 4 regions, meaning 2 Core POP sites in each region. ATELCO believes that the Core of the network should be highly available, that's why they deployed all their Core POP locations in different Type-C cities to achieve the most availability as possible.

https://t.me/learningnets

ATELCO has a total number of 4 Type-C cities around the country that are located in 4 regions. Each region has one Type-C city and each Type-C city has two Core POP sites for redundancy. Core POP locations are connected in a full mesh way. Figure 7-3 shows the Core POP locations in Type-C cities of each region.

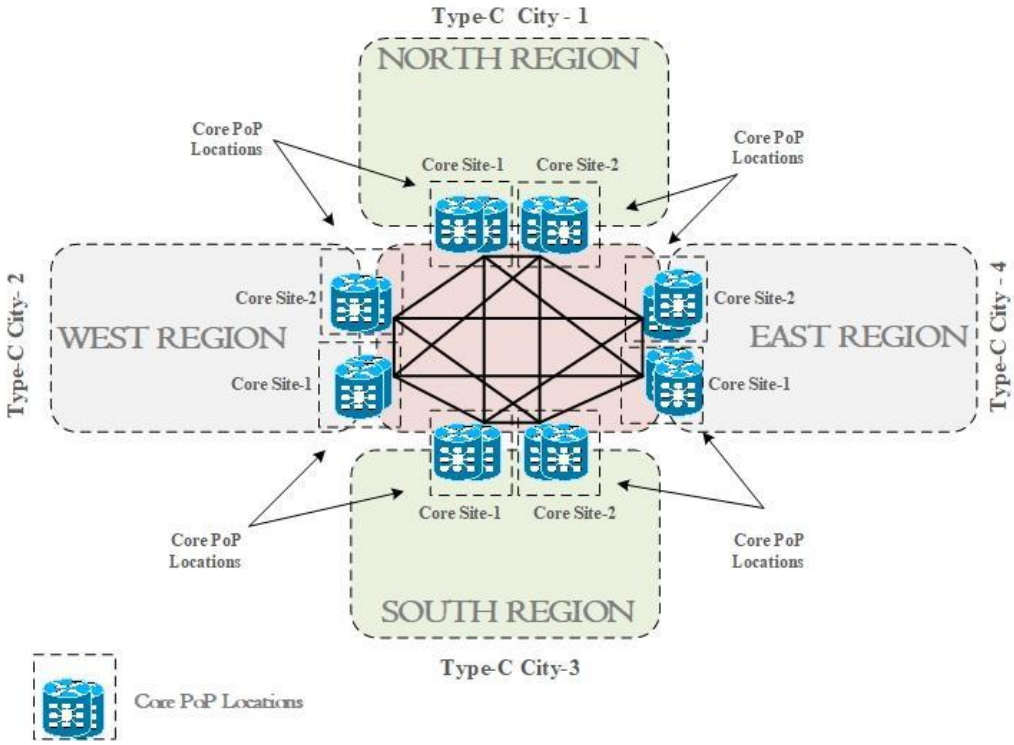


Figure 7-3 *Two Core POP sites in each region, full-mesh connectivity between Core POP sites*

ATELCO's Logical Network (End to End)

ATELCO PSN, (Packet Switched Network) consist of three logical layers. After demonstrating and explaining each layer, the end to end full picture will be presented for complete understanding.

1. IP/MPLS Multi Service Network
2. WCL (Worldwide Connectivity Layer)
3. Internet Gateway/Shared Services Layer

https://t.me/learningnets

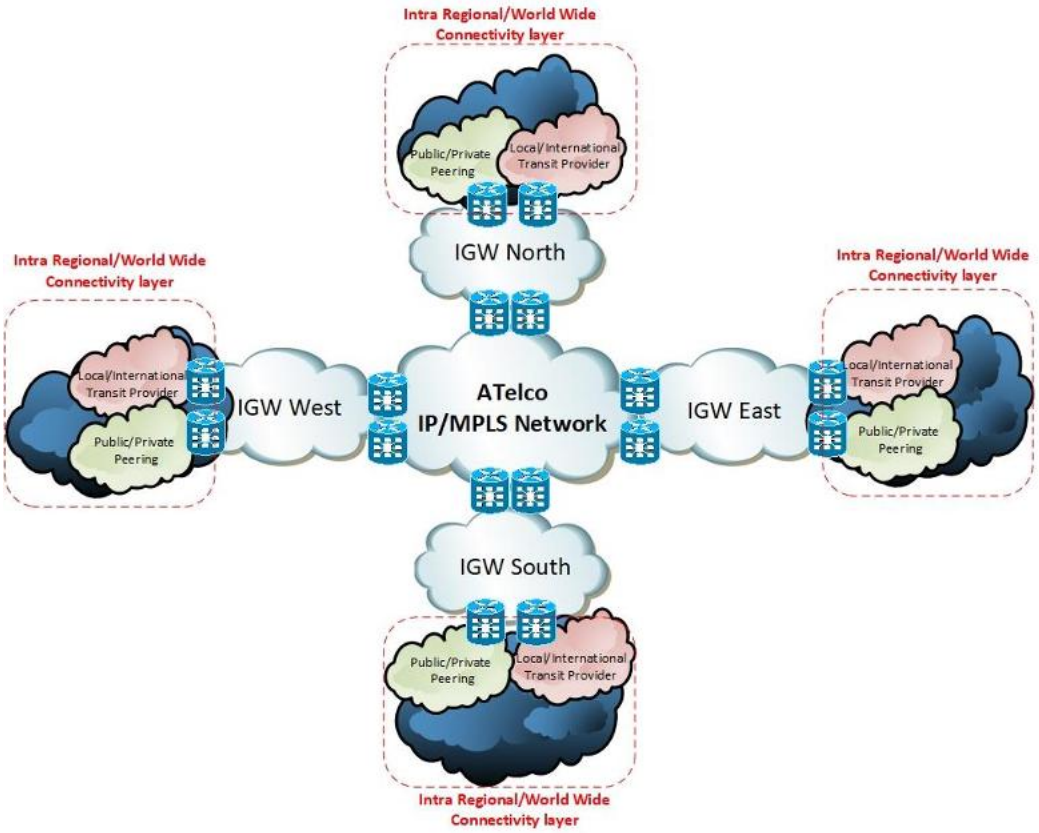


Figure 7-4 *IP/MPLS, IGW and World Wide Network Connectivity*

IP/MPLS Multi Service Network

IP/MPLS Multi Service Network consist of Core, Aggregation, Pre-Aggregation and Access networks. ATELCO has many different types of access network connectivity options, providing fixed and mobile broadband connections to their Residential and Business/Corporate customers.

https://t.me/learningnets

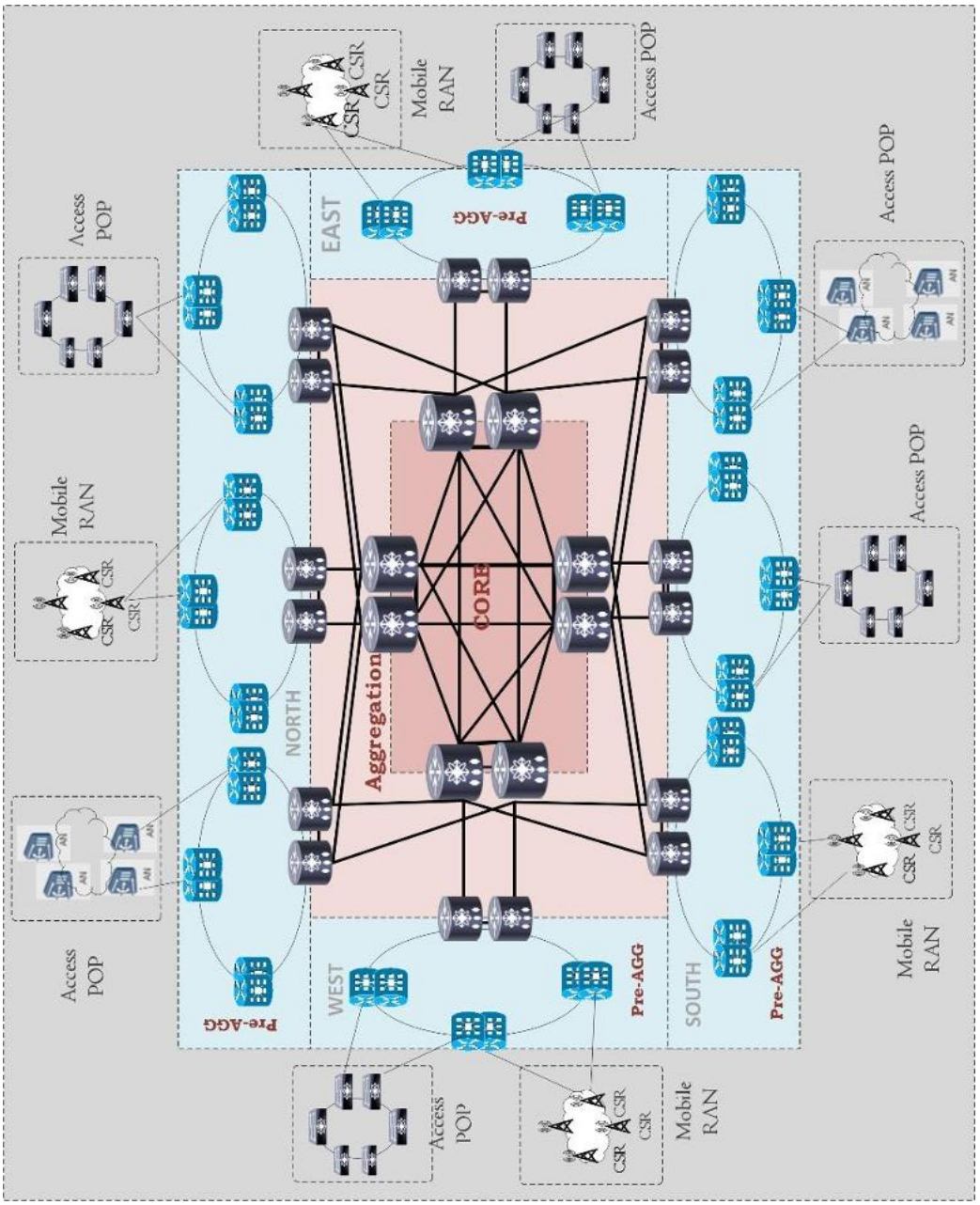


Figure7-5 *IP/MPLS Multi Service Network – Deployed on Four Regions*

https://t.me/learningnets

ATELCO's Multi Service IP/MPLS network has four layers.

1. Core
2. Aggregation
3. Pre-Aggregation
4. Access Layers

Network devices that form each layer are explained as follow:

- P routers form the Core Layer
- AGG PE routers form the Aggregation Layer
- Pre-AGG PE routers form the Pre-Aggregation Layer and provide connectivity between AGG and Access Layers. The Pre-Aggregation Layer is based on Ring Topology.
- AR (Access Router) PE/non-PE devices provide Access Layer connectivity for the different access network types. The access layer physical connectivity for both Mobile and Fixed services is based on ring topology.

Access Layer devices reside in the Access POP locations, Pre-Aggregation Layer routers reside in the Aggregation POP locations and Aggregation and Core Layer devices reside in the Core POP locations.

https://t.me/learningnets

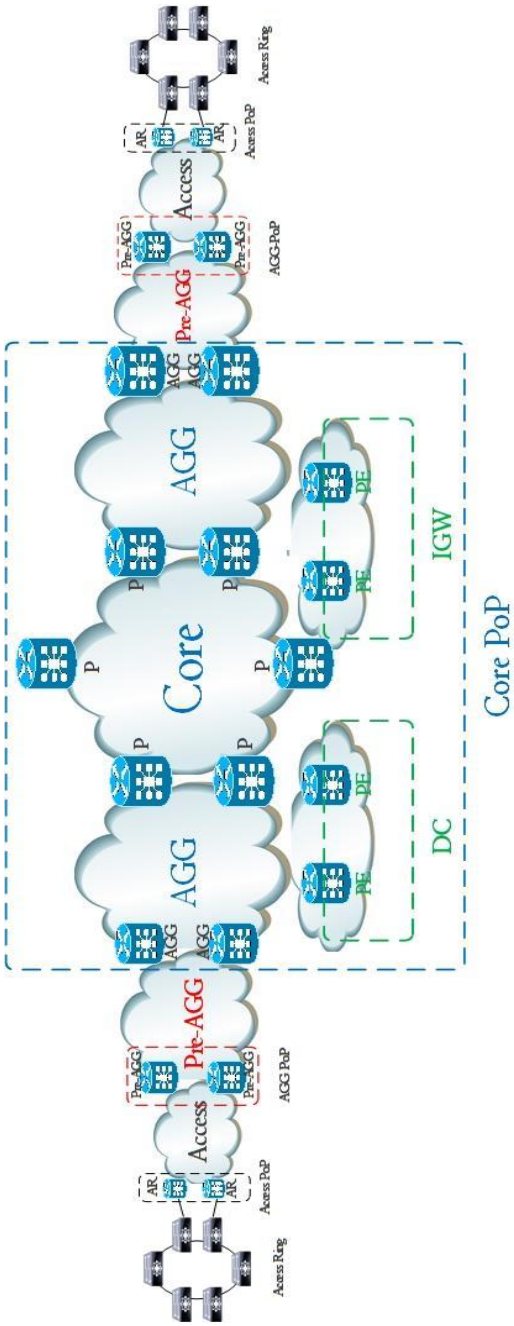


Figure 7-6 *Physical POP locations and the layers*

https://t.me/learningnets

WCL (Worldwide Connectivity Layer)

ATELCO's WCL (Worldwide Connectivity Layer) has Peering and IP Transit Connectivity to Europe, Dubai, USA and Hong Kong Internet Exchange Points for Global Internet Connectivity. ATELCO has many Private Settlement Free Peering agreements. They have also joined to the Public Peering in the IX through Route Servers in those locations. They have Selective Peering policy and their PeeringDB records are updated.

WCL provides the Internet connectivity for ATELCO's Residential and IP Transit Customers. ATELCO's entire network is not limited to the single country. Although they sell their services in the single country, they have many business agreements with other ISPs and Content Providers in the World.

In figure 7-7, ATELCO's Global network is shown. They internally refer this network as WCL (Worldwide Connectivity Layer).

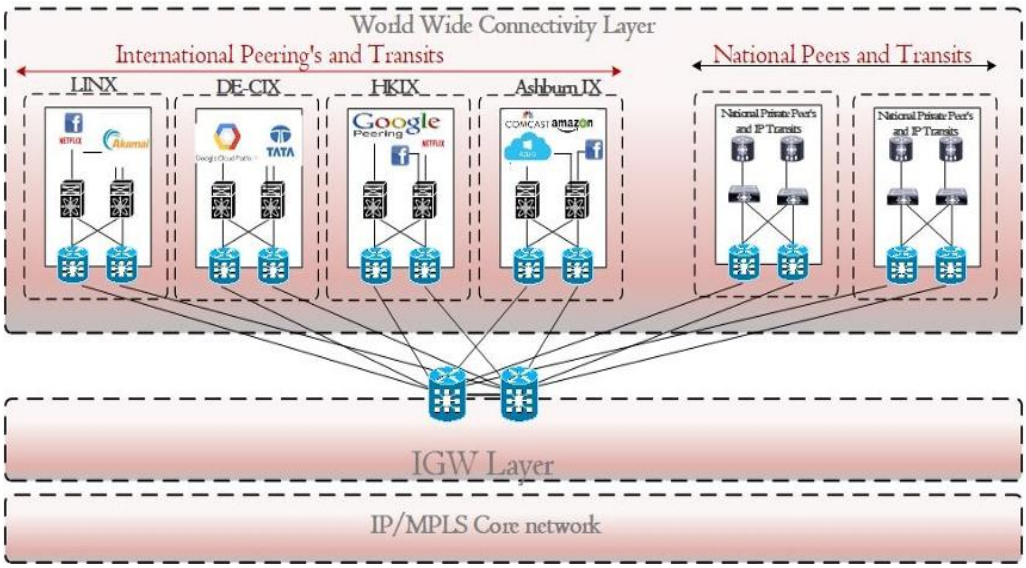


Figure 7-7 *World Wide Connectivity/International Locations of ATELCO*

Currently they have presences at London, Frankfurt, Ashburn (U.S.A) and Hong-Kong. In the country where ATELCO has their operation, there is no IXP (Internet Exchange Point). They have Settlement Free Peering agreements in LINX in London, DE-CIX in Frankfurt, U.S.A and HK-IX in Hong Kong. ATELCO's peering policy is Selective Peering Policy.

https://t.me/learningnets

Based on their Peering Policy, their peers should satisfy the following conditions:

1. **Traffic Requirement:** Peer should have at least 10Gbps traffic with ATELCO
2. **Presence Requirement**: Peer should have a presence at least in two Exchange Points
3. **Operational Requirement**: Peer should have a 24/7 NOC team which can response when failure or attack happens.

ATELCO is using BGP AS number 65000 in the World-Wide Connectivity Layer devices as well as the IGW Layer and also inside the IP/MPLS Multi Service Network. Their International Leased and IRU based fiber to International sites are done via all **4 Regional IGW Gateways**. ATELCO has two routers in each IXP, which they use for both Settlement Free Peering and IP Transit purposes.

ATELCO has the following connections to the International Peering's and Transits:

- 10G Leased circuit to Ashburn
- 10G Leased circuit to Hong-Kong
- 10G Leased circuit to London and Frankfurt connectivity

When the first Submarine Cable system was deployed to ATELCO's country, ATELCO had signed an IRU agreement with the Submarine Cable operator. The bandwidth on the cable is 80Gbps.

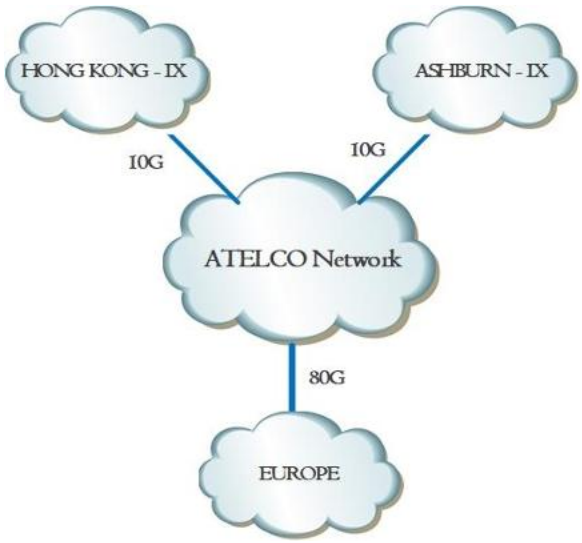


Figure 7-8 *ATELCO's International Private Leased Capacity & Locations*

https://t.me/learningnets

The submarine cable lifetime which connects the company to Europe is 25 years and 10 years has already passed. The total International bandwidth of the company is 300Gbps, which 200Gbps is received by an International Tier 1 Operator in the country. As shown in the figure 7-8, ATELCO's total amount of International Private Leased Circuit is around 100Gbps.

ATELCO has Facebook and Akamai caches inside their network, located in the country. They also have Settlement Free Peering with these companies and many other Content Providers which are located outside of the country in the IXPs. The reason is that that Content Providers don't keep all the content in each country.

The content which is not cached in Greenland where ATELCO's network resides, can be found in Europe which can be higher level Caches, or at the Origin, let's assume it in North America. ATELCO runs EBGP with all other companies in its World Wide Connectivity Layer and the Traffic Engineering is done via BGP.

Currently when there is a problem on one of their International circuits, or when there is congestion either inbound or outbound to and from their network, BGP configuration is done by their expert engineers which are ATELCO's WCL Team (Worldwide Connectivity Layer).

Although engineers are very experienced on BGP, the company has many ISP customers and many International agreements making their BGP Policy so complex to operate, so they are looking to automate their Interconnections as soon as possible.

As shown in figure 7-9, inside the IXPs in LINX and DE-CIX IXPs, they have 40Gbps connectivity to the IXP Switch Fabric on each link from each of the two routers. The two routers are connected back to back via 2 x 100Gbps circuits.

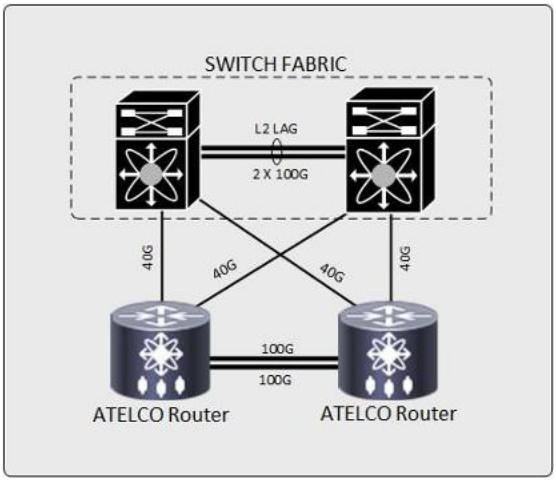


Figure 7-9 *ATELCO's Internal LINX and DE-CIX IXP Connectivity*

https://t.me/learningnets

As shown in figure 7-10, Inside Ashburn and HK-IX, they have 10Gbps ports connected to the IXP LAN from each router on each link and the routers are connected to each other via 2X 10Gbps circuits.

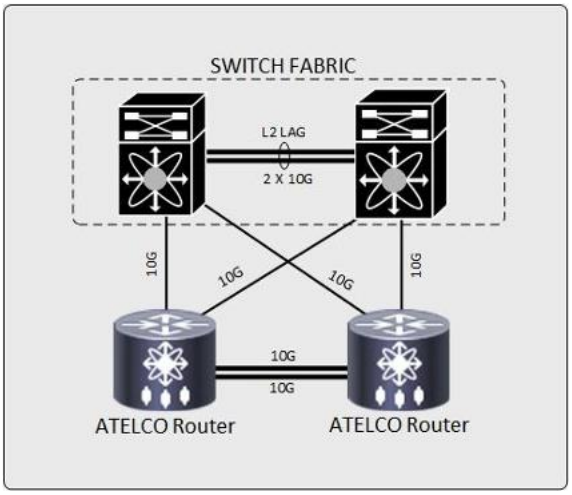


Figure 7-10 *ATELCO's Internal HKIX and Ashburn IXP Connectivity*

They filter unwanted traffic on these routers to prevent carrying traffic from these locations to their country, so they can avoid unnecessary bandwidth usage.

Internet Gateway (IGW) and Shared Services Layer

Internet Gateway (IGW) and Shared Services Layer provides Internet Access and connects the IP/MPLS network to the WCL network. All the service nodes, such as Firewalls, CGNAT, CDN Caches, DNS, Load Balancer, and Parental Control boxes and so on reside in this layer.

As shown in figure 7-11, this layer runs IBGP with ATELCO's internal network (IP/MPLS Service Layer) and runs EBGP with the Worldwide Connectivity Layer (WCL)

https://t.me/learningnets

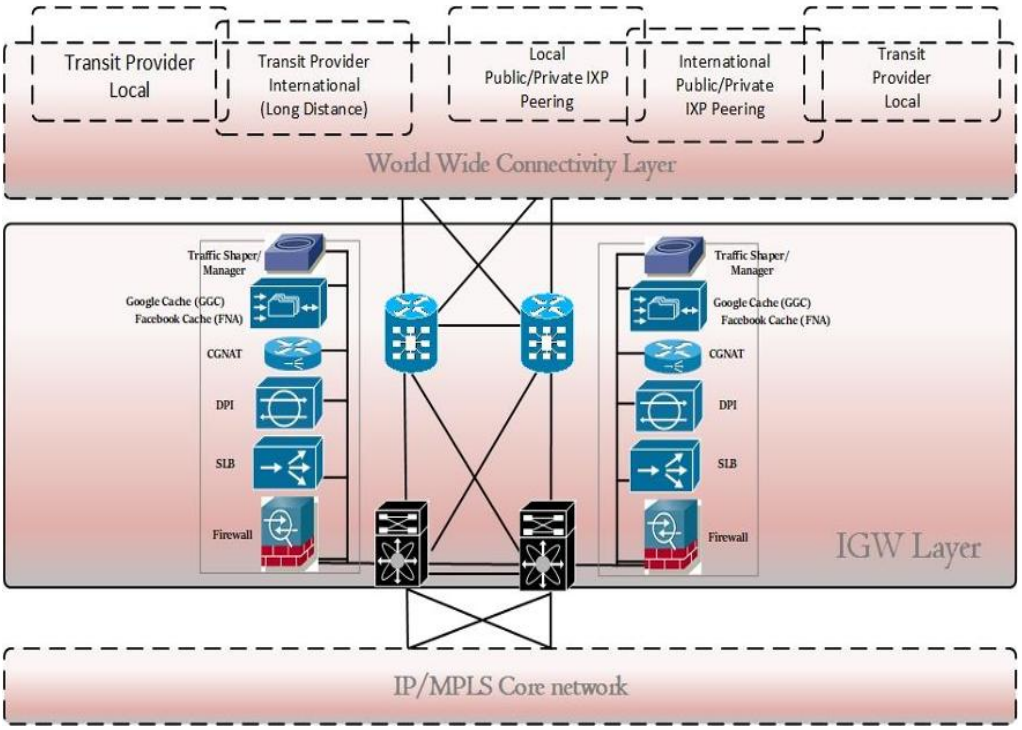


Figure 7-11 *ATELCO's IGW Layer*

For many years, the IGW team were taking care of the Internet service activities inside the country and the WCL team were taking care of all the Internet service activities for outside of the country. These two layers (IGW and WCL) were operated by different teams, but recently they formed a team called AWN (ATELCO Worldwide Network) which will be responsible for ATELCO's worldwide connections.

ATELCO's newly formed AWN is taking care of the following activities:

- Connection to the Internet Exchange Points
- Connectivity with other Operators inside and outside the country
- Settlement Free Peering
- IP Transit
- IRU agreements
- WCL Capacity Planning
- Service Layer Functions (Parental Control , Firewalls , DNS , Load Balancers , CGNAT deployments)
- Service Chaining in the AWN
- All BGP (EBGP and IBGP) related activities.

https://t.me/learningnets

The forming AWN team removes the silos, thus reduces the OPEX and provides efficient operation.
As it was explained in the beginning of the scenario, ATELCO has four regions and they have Gateway POP locations in each region. These IGW sites are located at the Datacenters.

ATELCO currently has IP Transit Service from three International Tier 1 Operators inside the country. Two of these Tier 1 Operators provide IP Transit service in each region. As the North region is backup of the other regions, all the Tier 1 Operators (3 operators) are connected to the North Region IGW Routers.

Figure 7-12 shows 4 Regions of ATELCO's network connecting to the Transit Providers inside the Country (Greenland).

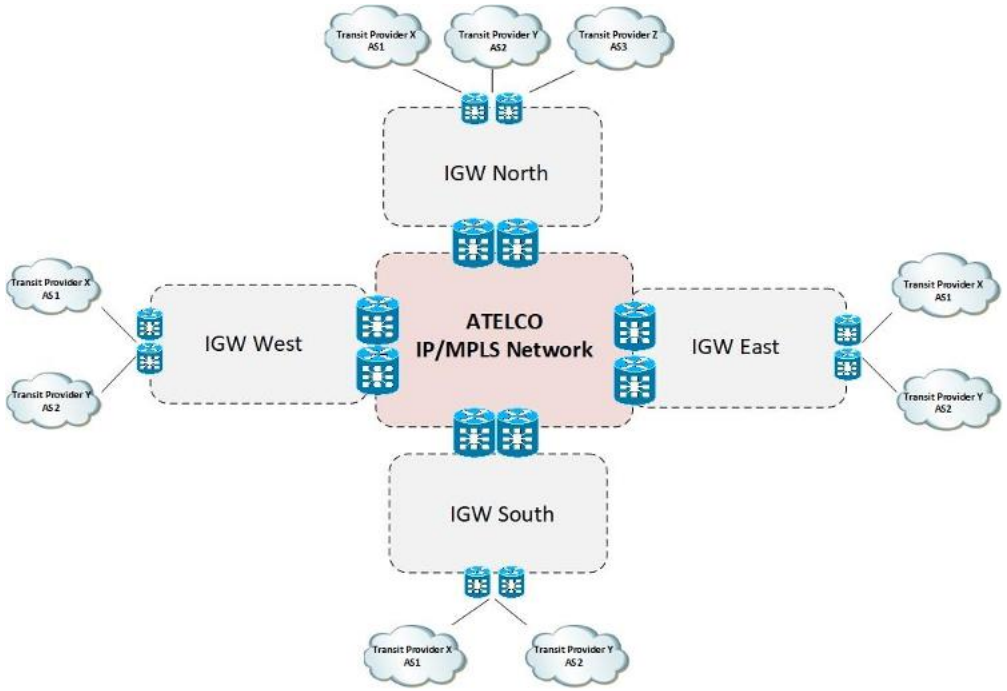


Figure 7-12 *Three Transit Providers of ATELCO inside its Country*

ATELCO is using all four Regions to have a connection with other ISPs inside the country. ATELCO has four datacenters in the country. As it was mentioned before, they have one IGW POP location in each Region; these IGW POP locations reside in the Datacenters where their IP/MPLS Core POPs are located.

ATELCO has two Core POP Locations in each Region as it was explained above. One of the two Core POP Locations reside in the Datacenter, another Core POP is not placed in the Datacenter.

https://t.me/learningnets

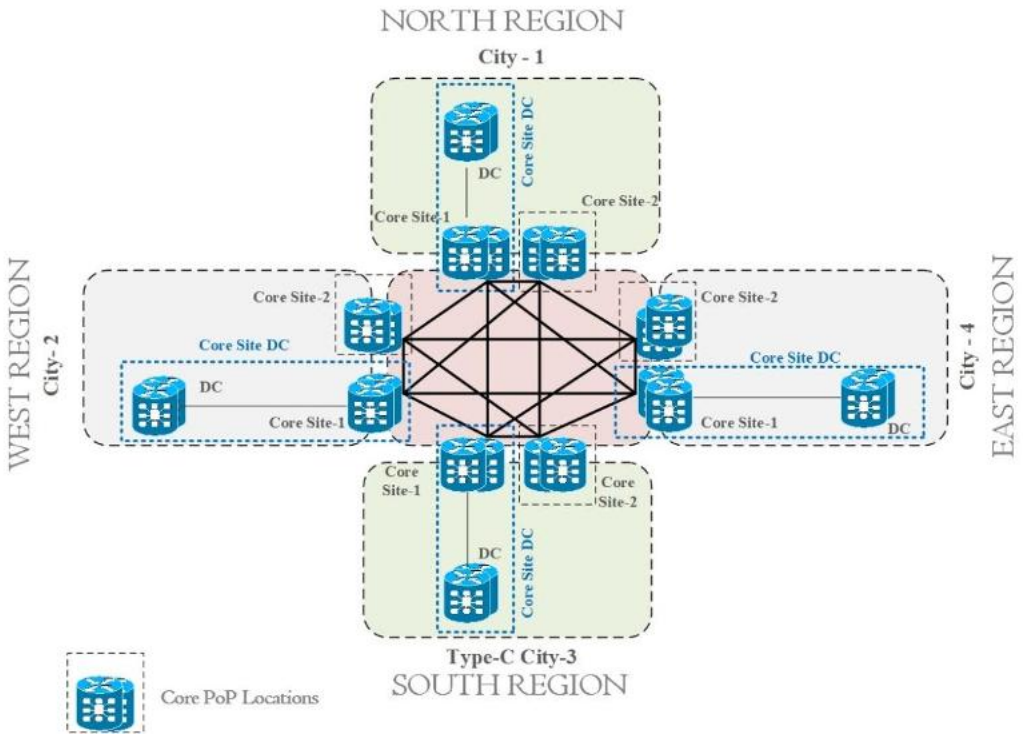


Figure 7-13 *Each Region with 2 Core POP's, with Single DC and IGW POP*

These datacenters are used for ATELCO's internal operation such as hosting all the Internet Gateway Layer devices, service nodes, cache engines, business customer services and also provides many Telecommunication services.

ATELCO also provides datacenter services to their business/corporate customers, such as:

- Datacenter Colocation
- Infrastructure as a Service Cloud Offering
- Managed Hosting
- Meet-me room
- Cross Connect
- Internet Access (inside the datacenter)
- MPLS VPNs (inside the data center)

Their Internet Gateway Layer (IGW) routers reside in these Datacenters. Internet Gateway Routers provide connectivity between the IP/MPLS network and the World-Wide Connectivity Layer.

https://t.me/learningnets

In this layer, there are many service nodes such as Firewall, IDS/IPS, Load Balancer, LSN/CGNAT (Carrier Grade NAT) devices and also many cache engines from many content providers.

ATELCO has BGP in the IP/MPLS, IGW and WCL, so it's all three logical networks. In this section, we will explain BGP design for their Internet service, as in the IGW layer, ATELCO only has Internet service.

In the IP/MPLS section, BGP is used for both Transport and Service. There will be Internet and VPN services, as it will be explained later in this chapter.

In the IGW network, ATELCO is currently using physical routers as BGP Route Reflectors, but they are evaluating to deploy Virtual BGP Route Reflectors.

In each IGW POP location which are located in the datacenters, ATELCO has two Internet Route Reflectors. These Route Reflectors are the BGP peers of Service Route Reflectors in the IP/MPLS network.

Figure7-14 shows the BGP Route Reflectors in the IGW POP locations and their connectivities.

https://t.me/learningnets

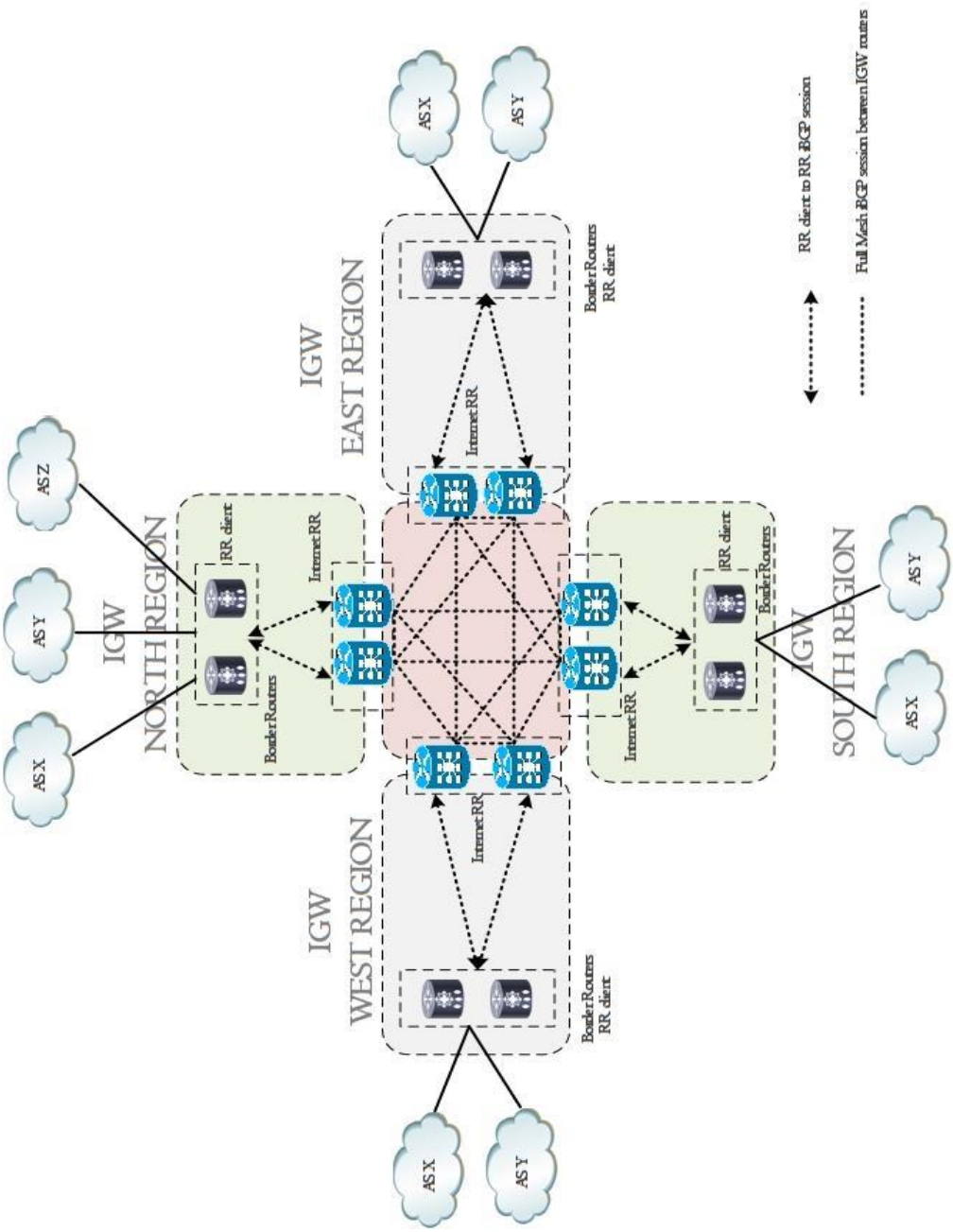


Figure7-14 *BGP Route Reflectors in the IGW POP's and their connectivity*

https://t.me/learningnets

Same cluster IDs are used in each datacenter but different datacenter Route Reflectors have different cluster ID's. On the IPv4 Route Reflector, Add-Path feature is enabled to send available paths to the Route Reflector clients. ATELCO is evaluating to deploy BGP ORR in addition to its BGP Add-Path feature to achieve optimal routing and path diversity in the network.

Transport RRs are inline RRs, thus MPLS and OSPF are enabled on these RRs. For the service layer RR's, MPLS is extended to the Route Reflectors, but OSPF is not enabled on them. ATELCO is considering to deploy BGP Egress Peer Engineering feature in this layer to have better capacity and resource utilization. BGP Egress Peer Engineering will be explained in great detail later in the book.

Content Provider/Over the Top Networks in ATELCO

Among the well-known content providers, ATELCO has GGC (Google Global Cache), Facebook (FNA), Akamai and Netflix Caches but also, they have regional content provider companies such as TV-channels and regional music and gaming providers. These nodes reduce International bandwidth requirement by 70%. All the service nodes and the cache engines are connected to chassis based switches via multiple 40G uplinks. Between the switches there are multiple times 100G connections.

Switches are providing port density and they don't do MPLS switching. As it was mentioned above, the current International bandwidth usage of ATELCO is around 300Gbps. Without those caches, all the content had to be delivered over International bandwidth capacity, because there is no IXP (Internet Exchange Point) currently in the country which ATELCO provides its services. Figure 7-15 shows the Content Provider Cache Servers in the IGW Layer of ATELCO.

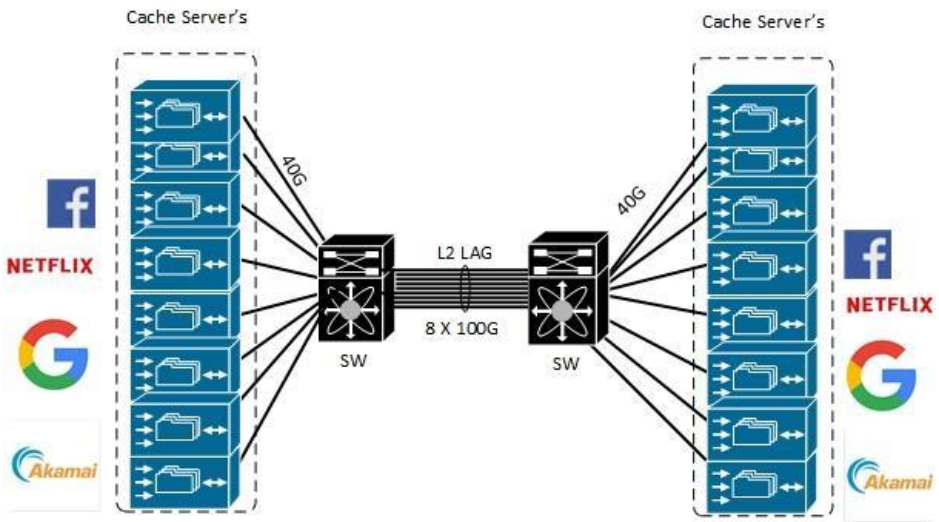


Figure 7-15 *Content Provider Cache Servers in the IGW Layer of ATELCO*

https://t.me/learningnets

Services Provided by ATELCO

There are several types of services provided by ATELCO.

- Residential Customer Fixed Services
- Residential Customers Mobile Services
- Business Customer Services

These services will be explained in very detail in this section.

Residential Fixed Services

Current Fixed services provided by ATELCO for their residential customers are FTTX and XDSL. Their FTTX design is based on GPON technology. The Current FTTX deployment of the company is purely based on FTTH technology as they think, FTTH is the ultimate goal and will fulfill the ever-increasing bandwidth demand.

As providing DSL services, ATELCO provides ADSL and VDSL services. For the ADSL services, they only provide ADSL v1 .For VDSL services they provide VDSL 1 and VDSL 2 service, which provides much more bandwidth compared to ADSL.

Currently ATELCO has more than 1.6 million FTTH customers coming to their network from all four regions in the country and 2.4 million xDSL customers, so in total 4 million fixed broadband customers which is approximately 20% of the population of the country.

ATELCO has around 400 Access POP locations around the country in four regions and they have their DSLAMs and MSANs for DSL termination devices and OLTs as FTTH gateways in these Access POP locations.

https://t.me/learningnets

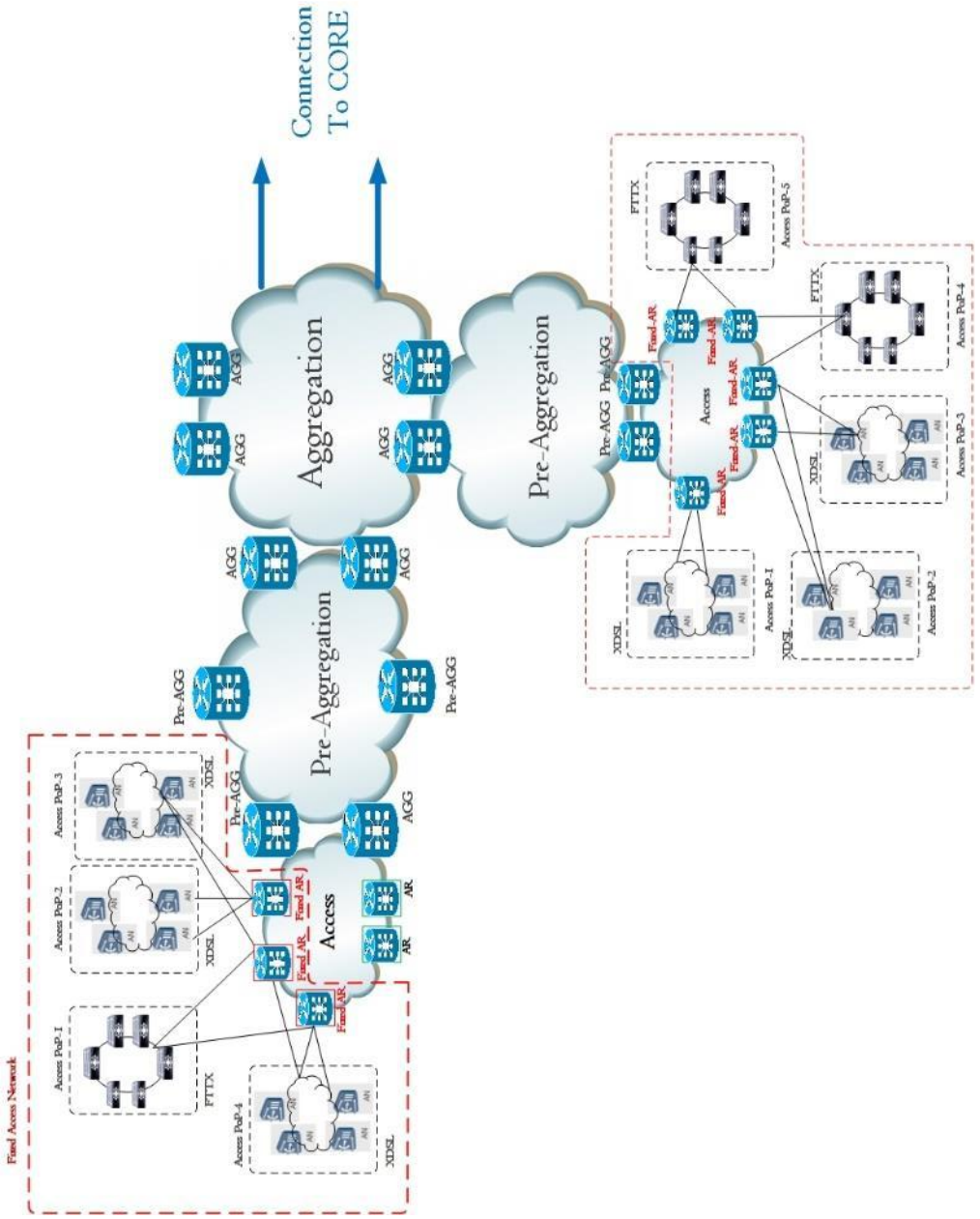


Figure 7-16 *ATELCO Fixed Residential Service, Access and Pre-AGG rings*
High Quality Image can be seen at www.orhanergun.net/spbook/F-R-Service.jpg

https://t.me/learningnets

For the fixed services, DSLAM and OLT devices are terminated at the Fixed Service Access PE (AR-Access Router) sites. As it can be seen from figure 7-16, Access POP locations are connected to each other in a ring topology as the ring topology is the cheapest topology which provides redundancy.

Access Rings are terminated at the Pre-AGG layer devices which are connected in another Ring based topology. Pre-AGG devices are connected to Aggregation Layer devices which are connected to IP/MPLS Core devices in a Hub and Spoke topology.

Due to their capacity and redundancy requirements between their four regions, the current IP/MPLS CORE/Backbone network of the company is designed and implemented as a full-Mesh topology.

Table 7-1 *Access, Pre-AGG, AGG and Core Network Connectivity*

Network Types	**Physical Connectivity**	**Connected to**
Access Network	Ring	Pre-Aggregation
Pre-Aggregation	Ring	Aggregation
Aggregation	Hub and Spoke	Core
Core Network	Full Mesh	Other Core POPs

ATELCO has only IP based DSLAM's and they don't have any ATM based DSLAM's in their network. The BRAS/BNG devices are connected to the Aggregation Layer Routers. Both XDSL and FTTX services use the same BNG for the subscriber access control and gateway functionality.

Figure 7-17 shows that the BNG's are connected to the Aggregation Layer which is located inside the Core POP locations. This shouldn't confuse you, as it was explained earlier in the chapter, Aggregation Layer Routers and the Core Layer Routers are placed in the same POP locations which is considered as Core or Backbone POP.

https://t.me/learningnets

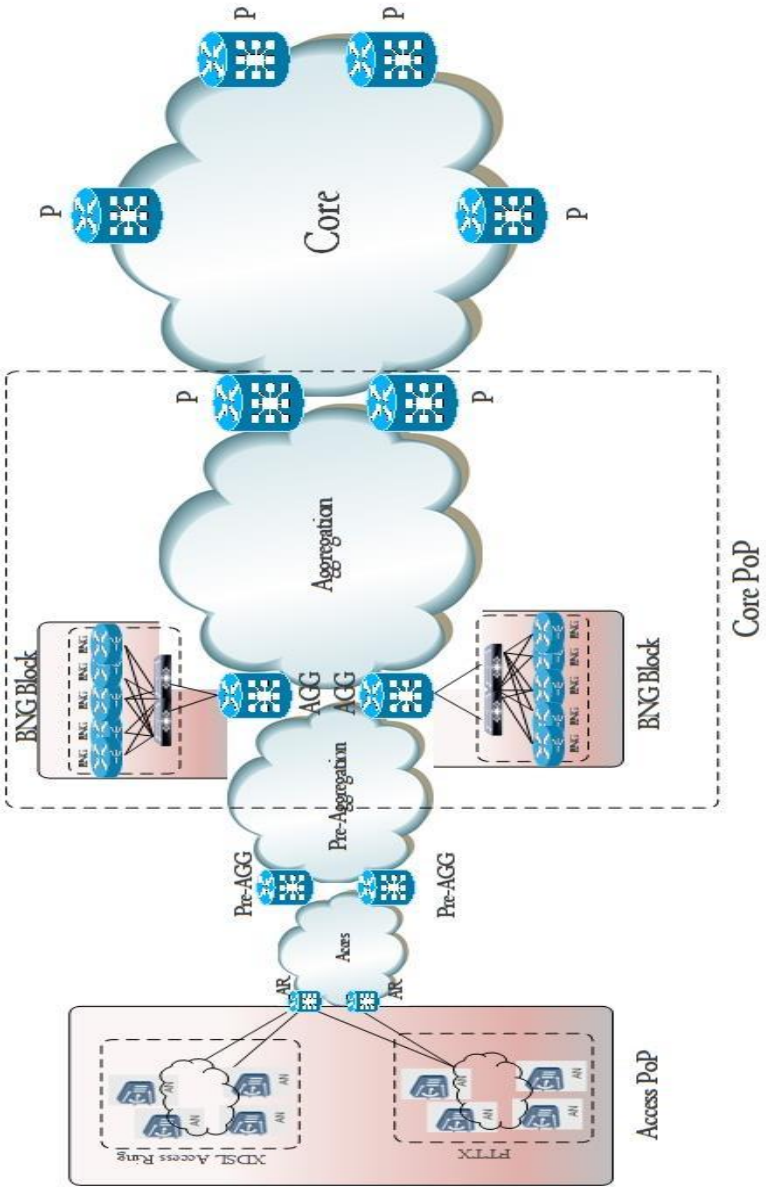


Figure 7-17 *BNG Connection in ATELCO network – Centralize BNG deployment model*

From the customer location, the customer's modem creates a PPPoE session which is terminated on the BNG. The IP Default Gateway of the customer modem (DSL and FTTX modem) is the BNG. Figure7-18 shows the PPPoE session from the Modem to the BNG.

https://t.me/learningnets

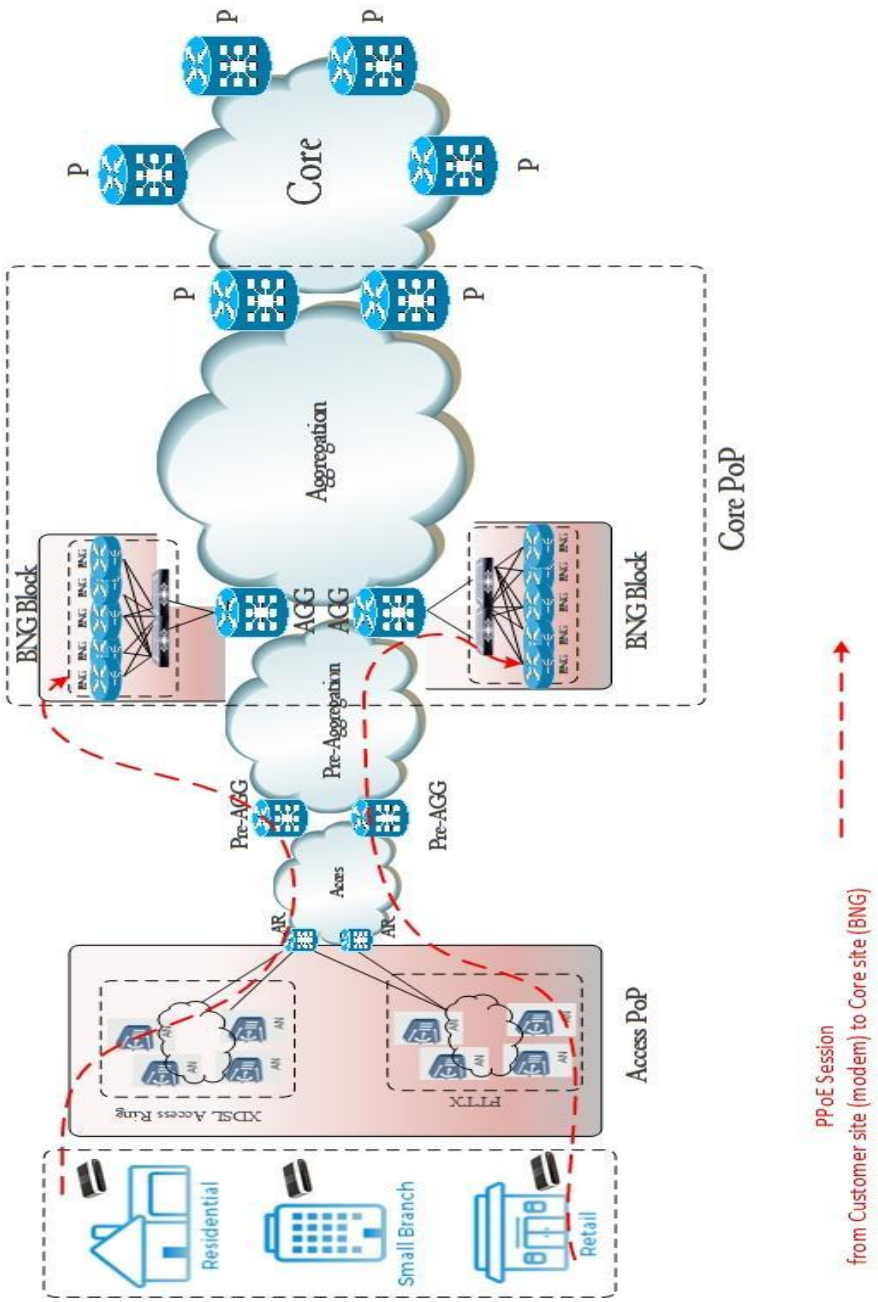


Figure 7-18 *PPPoE session from Modem to the BNG*

https://t.me/learningnets

ATELCO's current Fixed Access network connectivity options, upstream and downstream bandwidth support of each technology and the physical distance of these technologies for the advertised speeds for the residential customers, are shown in Table 7-2.

Table 7-2 Current Fixed Network Bandwidth and Supported Distance for each Service

Service	Downstream	Upstream	Max Reach
ADSL	8 Mb/s	1 Mb/s	5.5 km
VDSL 1	50 Mb/s	30 Mb/s	1.5km
VDSL 2	100 Mb/s	30 Mb/s	0.5 km
GPON	2.5 Gb/s	1.25 Gb/s	20 km

Residential Mobile Service

For the Mobile service, ATELCO provides 3G and LTE services and is planning to roll out 5G services within 2 years.

Currently ATELCO has more than 7 million Mobile Broadband customers coming to their network from all four regions in the country and which is approximately 25% of the population of the country.

As mentioned before, ATELCO has a total population of 20 million people in the country, which 11 million of the population are the customers of ATELCO. There are 4 million Fixed Broadband customers and 7 million Mobile Broadband customers. There are also thousands of corporate customers that are connected to ATELCO's network.

For the Mobile service, CSR (Cell Site Router) devices are terminated at the Mobile Service Access PE (Mobile AR – Access Router) nodes.

Figure 7-19 shows ATELCO's Mobile Service Access Connections, the Access and Pre-Aggregation Rings.

https://t.me/learningnets

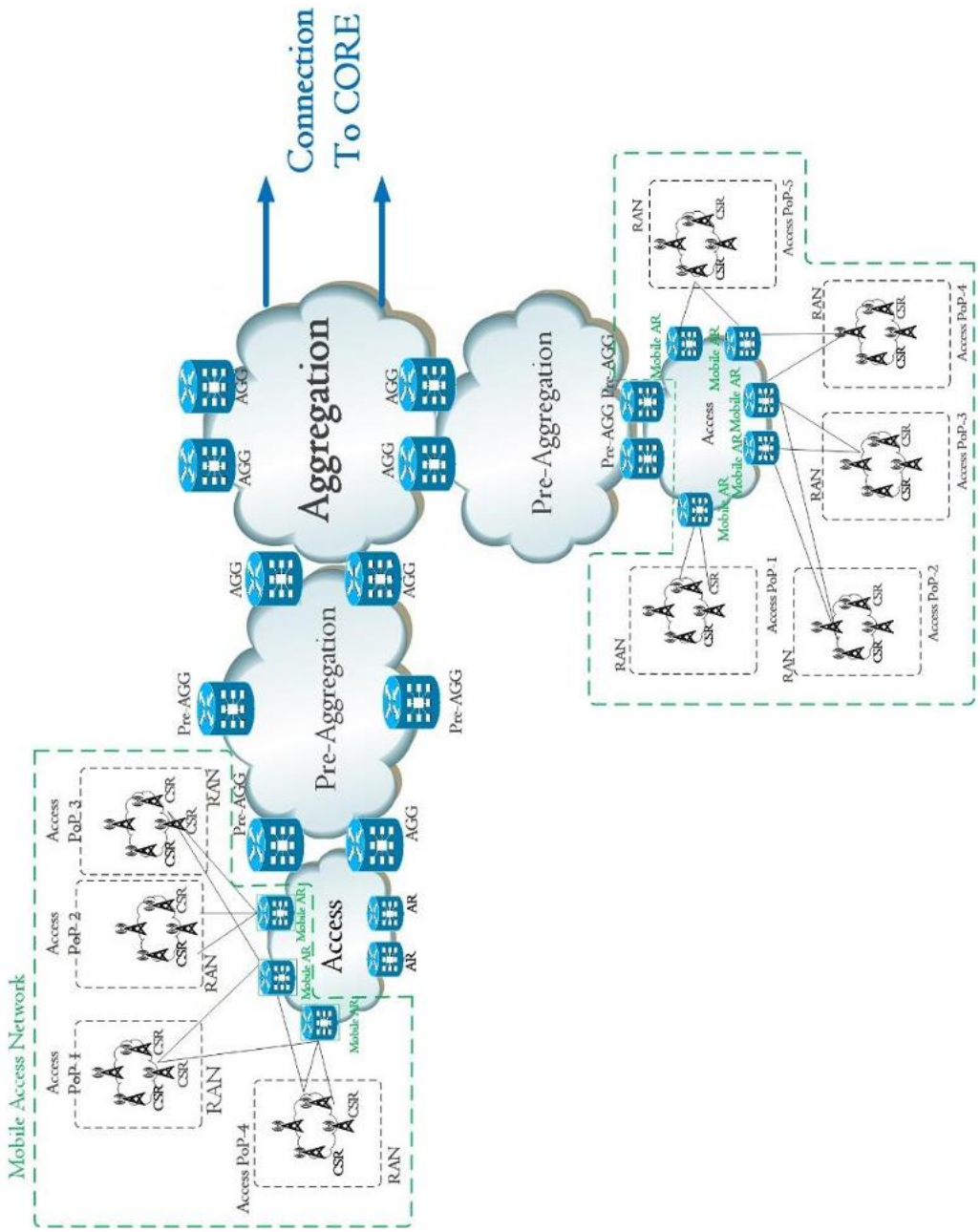


Figure 7-19 *ATELCO Mobile Service - Access and Pre-Aggregation Rings*

https://t.me/learningnets

Mobile penetration rate of the country is around 90% and still in some rural areas, satellite is the only connectivity option. ATELCO currently doesn't have any satellite service, but they are investigating O3B which is Medium Earth Orbit Satellite System, that provides decent bandwidth in Ka Band.

Business/Corporate Customers Access Connections

For the Business/Corporate customers, ATELCO has a Metro Ethernet Service which they deploy their switches at the customer location. If multiple companies share the same building, they place larger devices to accommodate more customers.

Buildings are connected to each other in a ring topology. If the capacity requirement is larger in some customer buildings, then the number of nodes in the rings is less. Typically, 6 nodes are used in each business customer ring in ATELCO's network. G.8032 (Ethernet Ring Protection) is used to provide fast convergence in the business customer rings.

These rings are terminated at the Pre-AGG layer in ATELCO's network. At the Pre-AGG layer, if the customer is looking for Layer 2 or Layer 3 VPN service, MPLS starts at the Pre-AGG Layer.

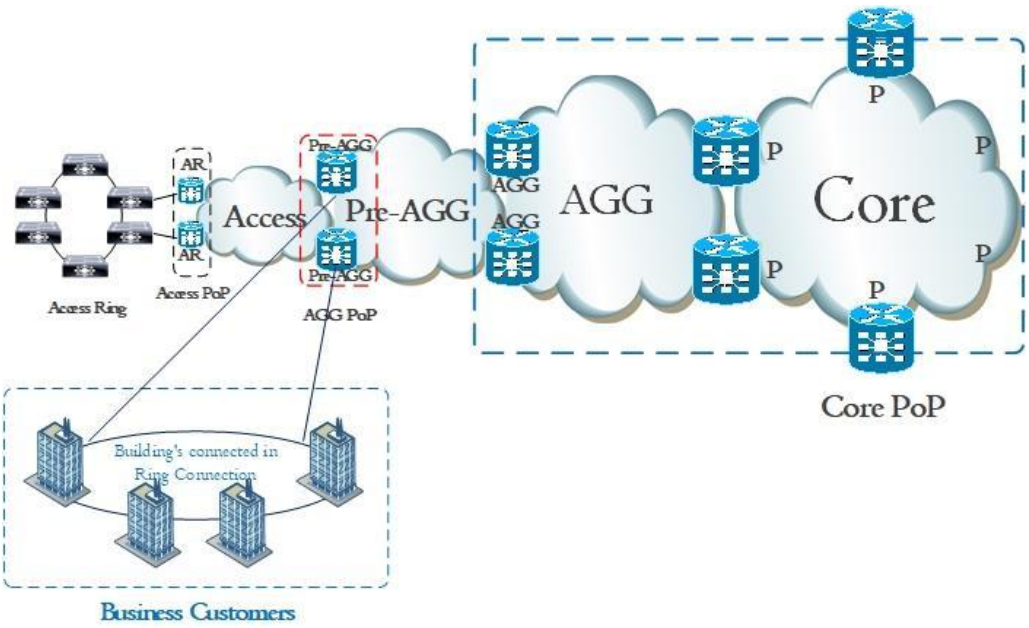


Figure 7-20 *Business/Corporate Customers connection to Pre-AGG Layer*

https://t.me/learningnets

There are some small cities in each region. Earlier in this chapter, different city types which were Type-A, Type-B and Type-C cities were mentioned. As explained in that part, in small cities, there are no CORE/Backbone POP locations. In these cases, Aggregation POP locations are connected to the closest CORE/Backbone POP locations. Ideally, the Aggregation POPs are connected to 2 Core POPs that are located in 2 different cities, but there are some exceptions in the network.

ATELCO places its own managed CPE devices at the Business customer sites. These CPE devices generally are the Ethernet switches. In Multi-tenant buildings, there might be more than one business customer sharing the same Ethernet Switch (CPE) and fiber connection, which connects it to ATELCO's POP location.

In ATELCO's network, Pre-AGG routers have more resources and have higher capacity than Access PE (AR) devices, thus business/corporate customers are connected to the Pre-Aggregation Routers which are located in the Aggregation POP locations. These POP locations are connected in a Hub and Spoke Topology and terminated at the Core POP Locations.
Martini based Layer 2 VPN is used for these customers that require Layer 2 VPN. If redundancy is a requirement for the customer, then two-way PW redundancy is deployed for them. Figure 7-21 shows Business/Corporate customers Layer 2 VPN PW between Pre-AGG routers.

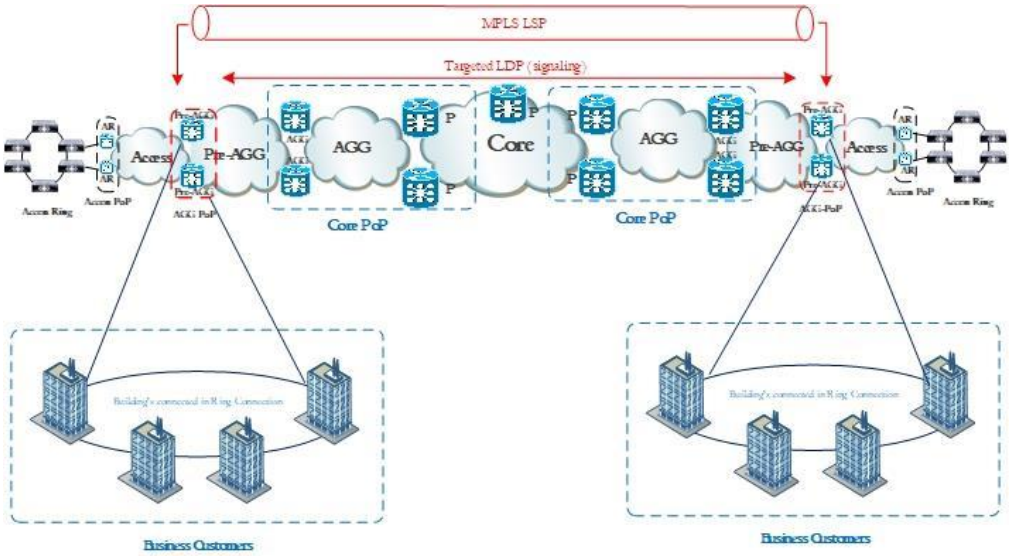


Figure 7-21 *Business/Corporate Customers Layer 2 VPN PW between Pre-AGG Routers*

Customers who require Layer 3 VPN, can run static routing or EBGP with ATELCO. ATELCO doesn't provide OSPF, IS-IS, RIP or EIGRP as the PE-CE MPLS Layer 3 VPN routing protocol.

https://t.me/learningnets

For the EBGP connection with the customers, ATELCO has BGP Route Limit Policies. According to their policies:

- 1500 prefix is the allowed limit
- If advertised prefixes reach to 70% of the allowed limit, NMS system is notified
- If allowed limit is exceeded, session will be tearing down and after 300 seconds, the router retries to bring the session up.

Protocols/Technologies used in the three layers of ATELCO Network

In the following section, the protocol and technologies that are used by ATLECO in the 3 layers are explained. These 3 layers consist of:

1. IP/MPLS Multi Service Network
2. WCL (Worldwide Connectivity Layer)
3. Internet Gateway/Shared Services Layer

In the WCL – World-Wide Connectivity Layer, ATELCO has EBGP connections with other companies. There are no other protocols in this layer. In the IGW Layer, two protocols are important for ATELCO. IGP and BGP.

ATELCO runs OSPF in the IGW Layer same as in the IP/MPLS Layer. OSPF is used to advertise Route Reflectors and Border Gateway addresses to the Core Network and also advertise the Service Layer Route Reflectors (which reside in the Core) to the IGW Layer.

IP/MPLS network is connected to the IGW network via CORE (P) routers as it is depicted in figure 7-22.

In the IGW Layer, Layer 2 Spanning Tree protocol is used on the Service Switches which terminate devices such as Firewall, DPI, Parental Control, Load Balancers and Cache Engines.

In the IGW Layer, most of the protocol operation is done with BGP. ATELCO is considering to deploy Segment Routing TI-LFA on their IGW network as it will be explained in the Chapter 8 in detail.

https://t.me/learningnets

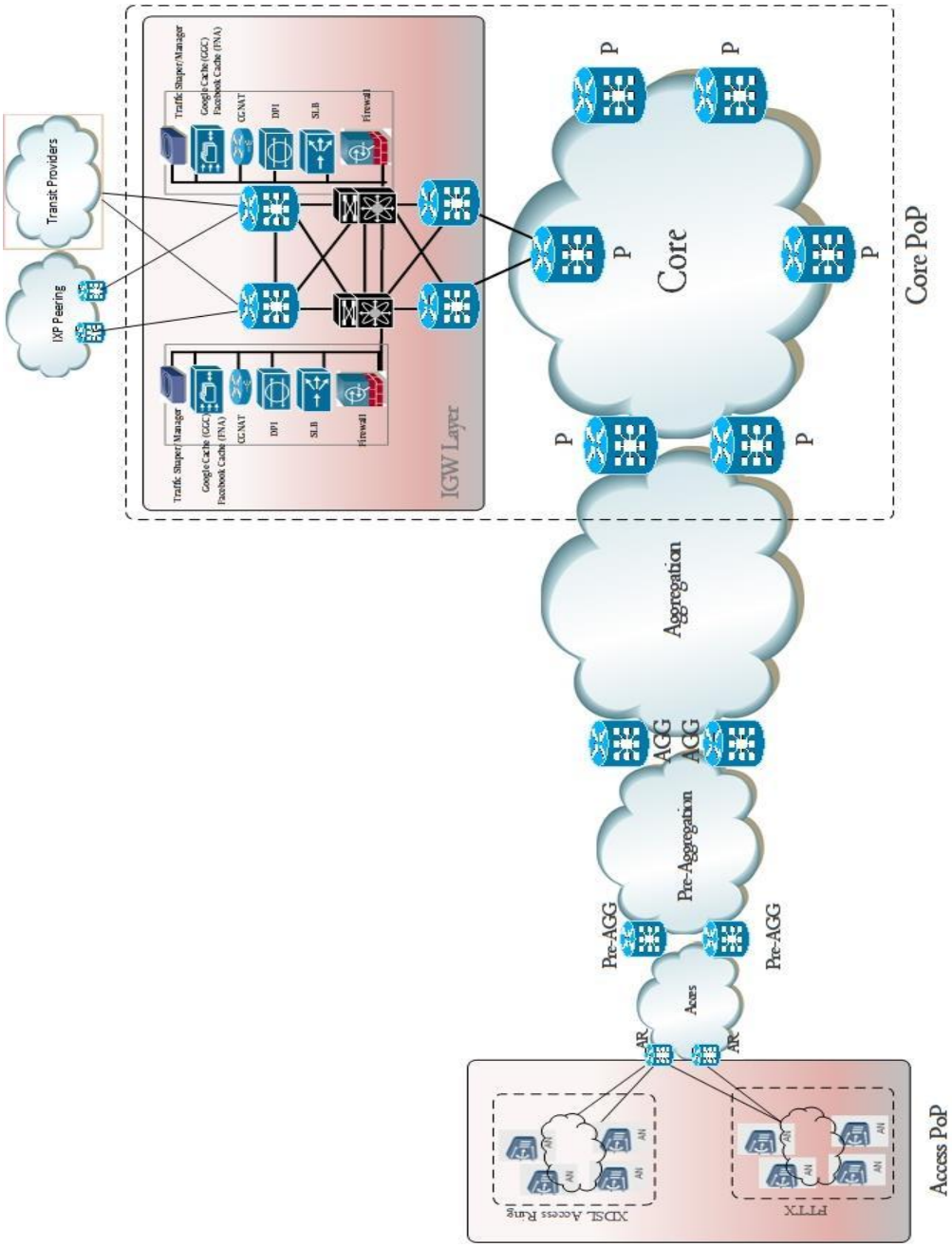


Figure 7-22 *ATELCO IP/MPLS to IGW Connection through Core Routers*

https://t.me/learningnets

IP/MPLS network is connected to the IGW/Shared Services Layer through Core Layer devices, as it can be seen from figure 7-22. OSPF is extended up to the IGW BR (Border Router) nodes.

Most of the Layer 2 and Layer 3 protocols run in the IP/MPLS network, and Services are provisioned within the IP/MPLS network without touching devices located in IGW or WCL networks. IGW and WCL network is mainly used for Internet Service. The technologies and the protocols in the IP/MPLS Multi Service Network will be explained next and it will be the longest section of this chapter.

IP/MPLS Multi Service Network

Through the IP/MPLS Multi Service Networks, the company carries FTTX, XDSL, 3G, LTE, MPLS Layer 2 VPN Point to Point, VPLS, MPLS Layer 3 VPN and Corporate customers Internet Traffic.

They use MPLS L2 VPN and MPLS L3 VPN for their internal purposes as well. Service deployments will be explained in this chapter.

Pre-Aggregation, Aggregation and the Core Layer of the company carry both Fixed and the Mobile services, so all the technologies and the protocols used in these layers are the same for both fixed and the mobile services.

One difference is at the access layer, which the fixed service is based on Layer2, while the mobile service access layer is based on Layer 3. This is done for scalability as the company has Unified/Seamless MPLS design in IP/MPLS Multi Service Network.

LDP is the signaling protocol that has been used for many years in ATELCO and they are investigating advantages and disadvantages of Segment Routing and the applicability of Segment Routing on their network.

LDP is used within each IGP domain and end to end reachability is achieved with BGP LSP.

ATELCO's IP/MPLS Network has IGP-LDP Synchronization feature enabled to avoid black holing in case of failure. They have LDP, extended up to the Access Network nodes which are CSR (Cell Site Router) devices for the Mobile service and to the Fixed AR (Access Router) nodes in the fixed services as it will be explained in more detail.
The fixed network Access part is purely based on Layer 2, thus there is no LDP, IGP or BGP in the Access network for the fixed network.

https://t.me/learningnets

ATELCO IP/MPLS Network IGP and BGP Design

BGP is used to carry all the device loopbacks in the network. BGP + Label address family (RFC 3107) is enabled on all BGP enabled devices in IP/MPLS Multi Service Network, except the Fixed Service Access nodes (DSLAM, OLT) as they don't run BGP. There is Hierarchical BGP RR (Route Reflector) deployment in ATELCO IP/MPLS Multi Service network.

There are full mesh IBGP sessions between the Core Routers (P Routers). Core P devices are deployed as Inline Route Reflectors in IP/MPLS Multi Service network and the Aggregation Layer Routers are the RR Clients of P routers. Aggregation Layer Routers are the Route Reflectors of the Pre-Aggregation Layer Routers.

There are two types of Access PE devices. Mobile Access PE devices and Fixed Service Access PE devices. Mobile Access PE devices terminate Cell Site Routers. Fixed Service Access PE devices terminate DSLAM and OLT devices.

There is a small difference between ATELCO's IGP design in IP/MPLS network for Fixed and the Mobile Service. For better understanding, the IGP and BGP design in IP/MPLS network will be explained for Mobile and Fixed Services separately.

ATELCO IP/MPLS IGP and BGP Design for Mobile Service

OSPF is used in the IP/MPLS network. Different OSPF Process ID's are used between the Access, Pre-AGG and the Core Block (AGG + Core layers). OSPF Router ID is set manually on all the OSPF enabled devices and Loopback 0 IP address will be used as OSPF Router ID on the devices. The CORE network OSPF Process ID is 10.

Fixed and Mobile network devices are only different at the Access Layer. Fixed services devices such as DSLAM, OLT, are terminated at the Fixed Service AR (Access Router), while the Mobile services devices such as CSR (Cell Site Router) are terminated at the Mobile AR (Access Router).All Pre-Aggregation, Aggregation and Core Layer devices in IP/MPLS network carry both Mobile and Fixed network service traffic. Core and Aggregation domains are in the same OSPF Process in ATELCO's network.

https://t.me/learningnets

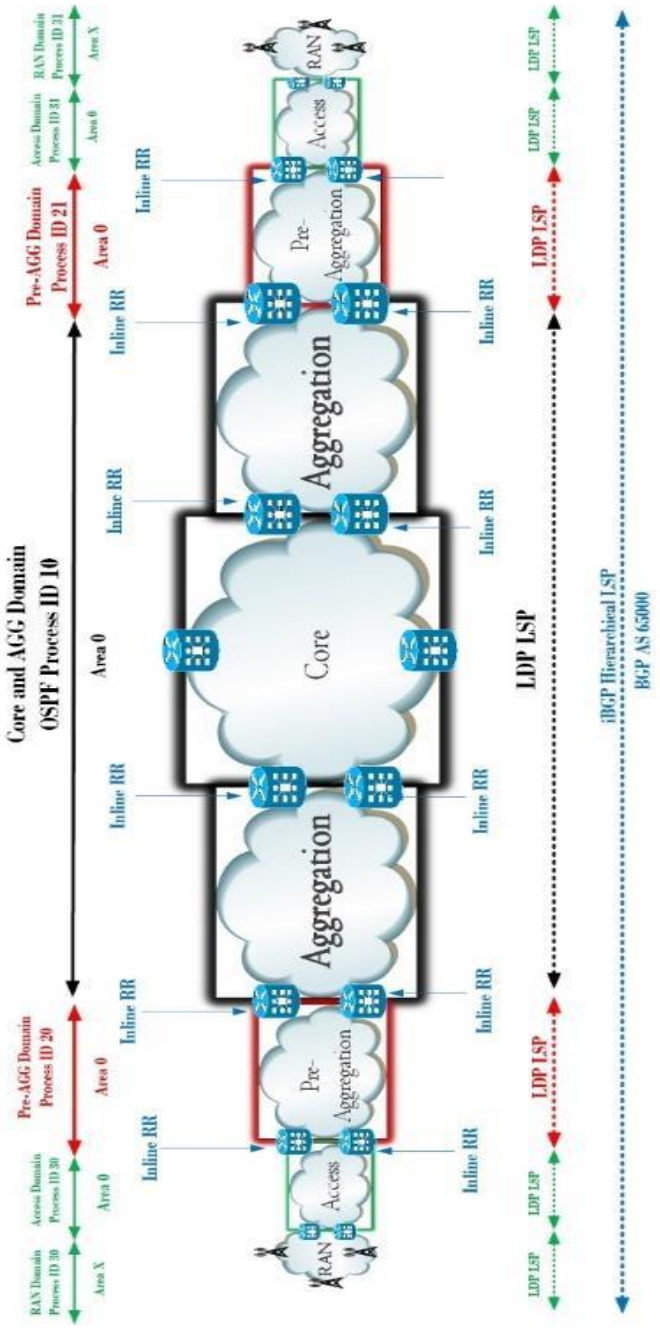


Figure 7-23 *OSPF/MPLS Transport and Inline BGP RR - Mobile Customers*

https://t.me/learningnets

Pre-Aggregation and Access Domains have different OSPF Process ID. RAN (Radio Access Network) has the same OSPF Process ID with the Access Network (Mobile Access Routers) but they are in different OSPF Areas.

Also, in order to provide more scalability in OSPF, OSPF Prefix-suppression feature and Incremental SPF is enabled.

BGP AS number 65000 is used in IP/MPLS Multi Service network. Access, Pre-AGG, AGG and Core Layer nodes run the same Autonomous System, that's why IBGP is used between the layers.

BGP is used for both Transport and MPLS Layer 3 VPN connectivity between Access CSR nodes and the Packet Core Nodes in the Mobile network as depicted in the figure 7-24.

In figure 7-24, Transport and Service Route Reflectors are shown.

- Pre-Aggregation Layer Routers are the Route Reflectors of Access routers.
- Aggregation Routers are the Route Reflectors of Pre-Aggregation routers.
- CORE Routers are the Router Reflectors of the Aggregation routers.

All of these routers are providing Transport connectivity as ATELCO has Unified MPLS design for Mobile Service. We have two Types of Route Reflectors in the network. Transport RR and Service RR.

All of the Pre-Aggregation, Aggregation and Core Routers are running BGP and the upper layer Routers act as Route Reflector to the lower layer routers, so it provides an end to end transport. This means it provides reachability between the devices as BGP is used to provide reachability between the device loopbacks of different layers in Unified MPLS.

ATELCO is providing LTE service to its customers. MPLS Layer 3 VPN is used for both X2 and S1 interfaces. For the MPLS Layer 3 VPN service, BGP Route Reflectors are used as well. Centralized Service Route Reflectors are placed in the Core network and CSR's have an IBGP session with the Service Route Reflectors.

https://t.me/learningnets

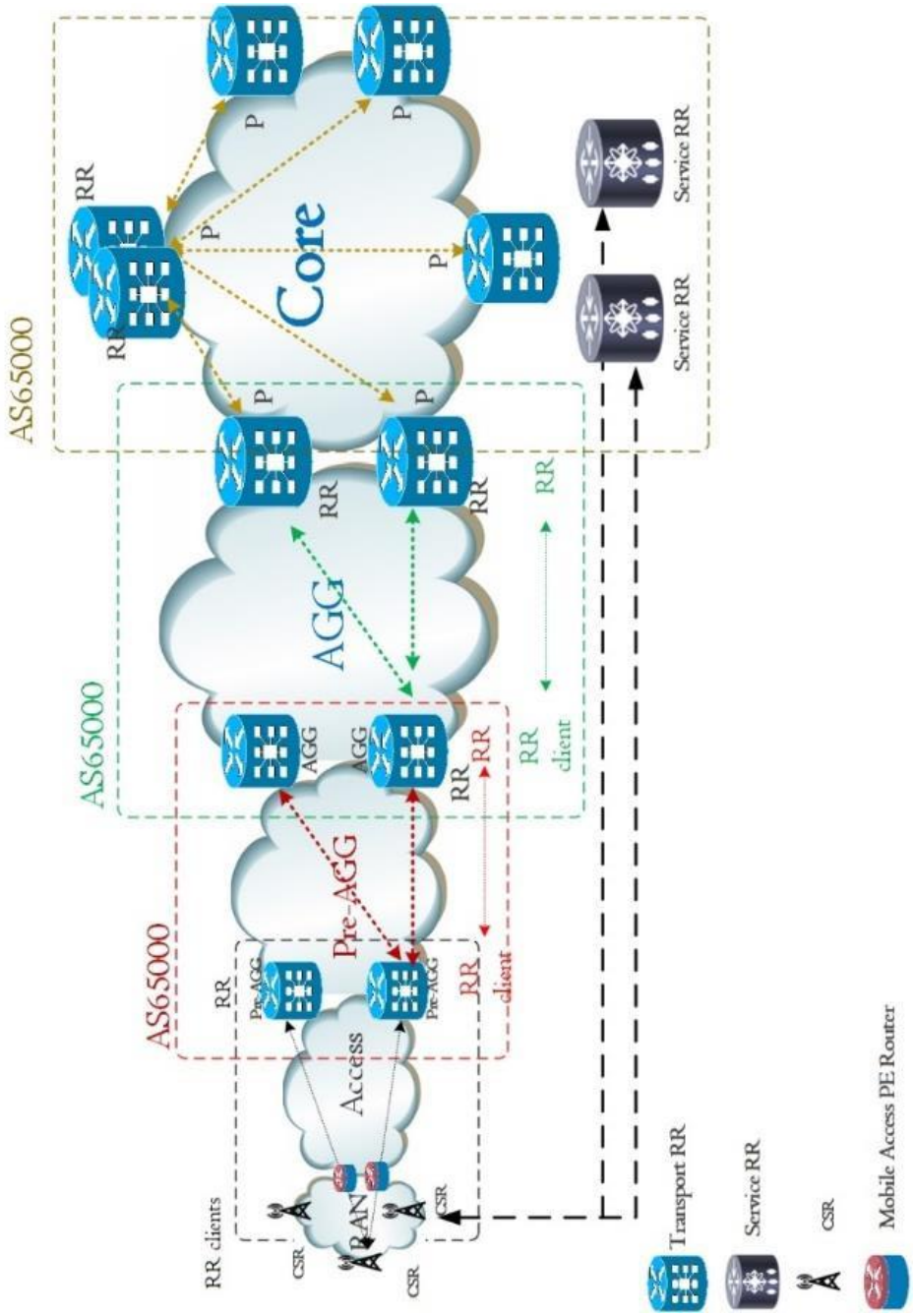


Figure 7-24 Mobile Service Transport and Service BGP Connectivity

https://t.me/learningnets

ATELCO IP/MPLS IGP and BGP Design for Fixed Service

The difference between Mobile and Fixed service from the IGP design point of view is that in the fixed network, there is no IGP in the Access network. For example, DSLAM' or OLT devices don't have OSPF, they work on Layer 2. Different service traffic, such as Video traffic and the Data traffic are carried via different VLAN's from the DSLAM and OLT to the AR (Access Router) devices.

The same Pre-Aggregation, Aggregation and Core Layer devices are used for both Fixed and Mobile Services, thus the IGP (OSPF in ATELCO network) design is the same.

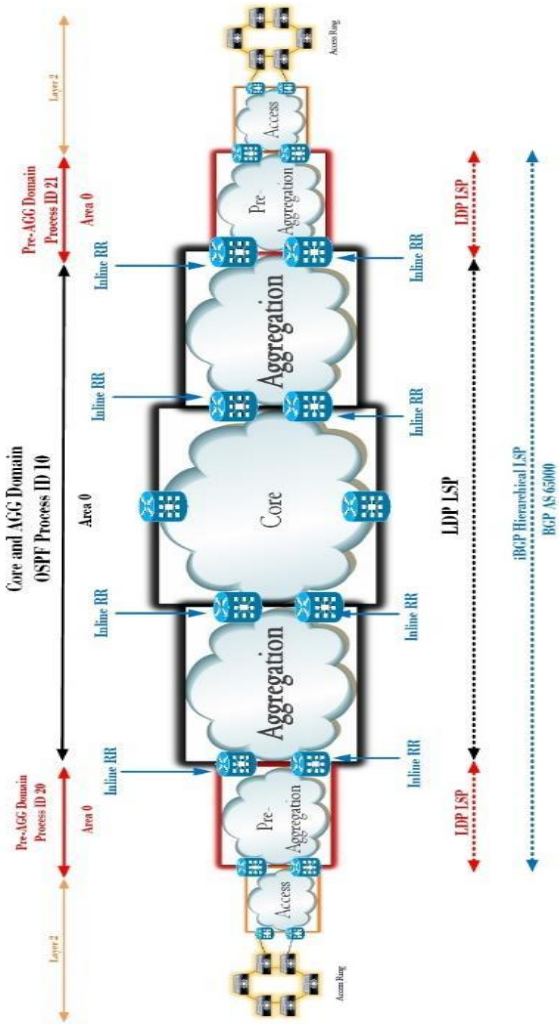


Figure 7-25 *OSPF/MPLS Transport & Inline BGP RR - Fixed Customers*

https://t.me/learningnets

In figure 7-26, Transport and Service Route Reflectors are shown. Pre-Access Layer runs just Layer 2, that's why there is no BGP in the Access Layer.

- Aggregation Routers are the Route Reflectors of Pre-Aggregation.
- CORE Routers are the Router Reflectors for the Aggregation.

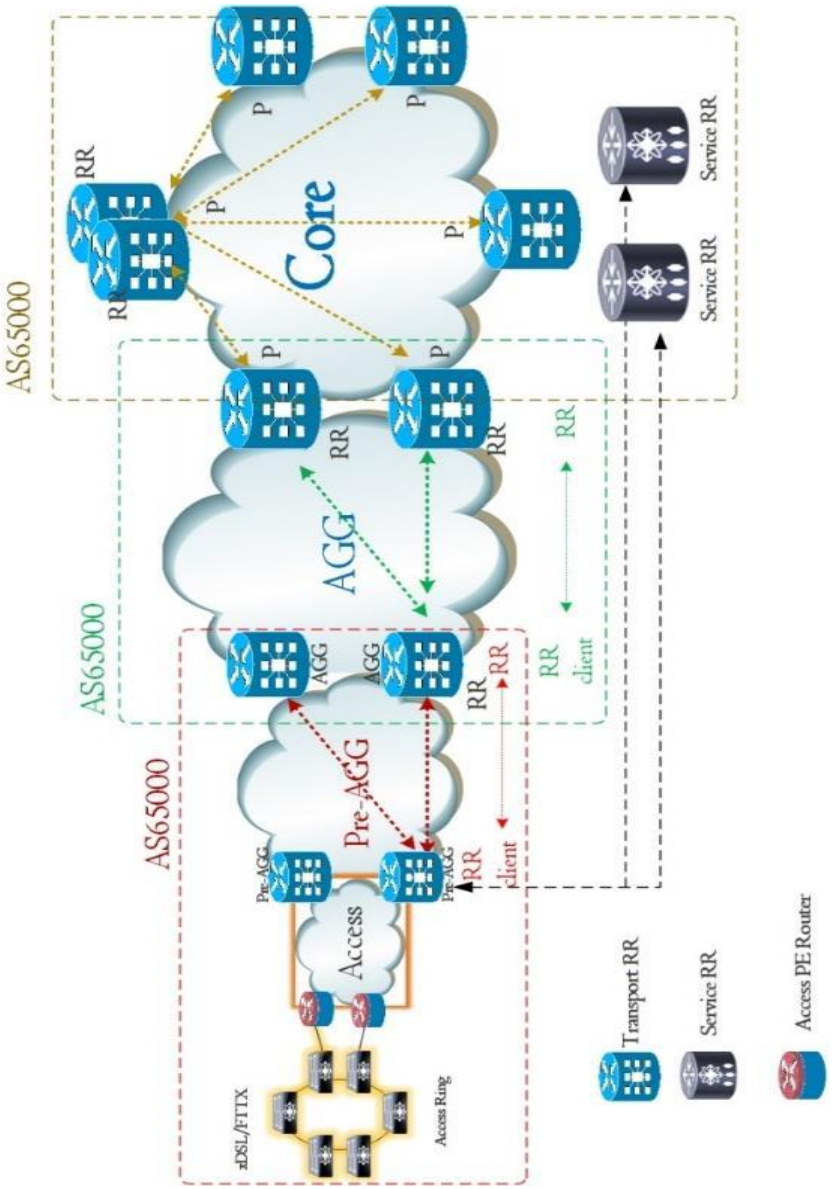


Figure7-26 *Fixed Residential Service BGP Design*

https://t.me/learningnets

All of these routers are providing Transport as ATELCO has Unified MPLS design for Mobile Service.

We have two Types of Route Reflectors in the network. Transport RR and Service RR. All of the Pre- Aggregation, Aggregation and Core Routers are running BGP and the upper layer Routers act as Route Reflector to the lower layer routers and provide an end to end transport. This means that they provide reachability between the devices as BGP is used to provide reachability between the device loopbacks of different layers in Unified MPLS. These explanations was for the Transport Route Reflectors which is used for the IP reachability between the routers inside the network.

For the Internet service, prefixes in the global routing table are learned through the Service RR. Also for VPN services, VPN prefixes will be advertised thought the Service Route Reflectors, which will be explained later in this chapter.

ATELCO is providing Fixed Access to both Residential and Corporate Customers. For the Residential Customers, the service is provisioned at the Access Routers. For the Corporate customers, the service is provisioned at the Pre-Aggregation Layer Routers. Thus, the Pre-Aggregation Layer Routers have an IBGP session with the Service Route Reflectors which are located at the Core POP locations.

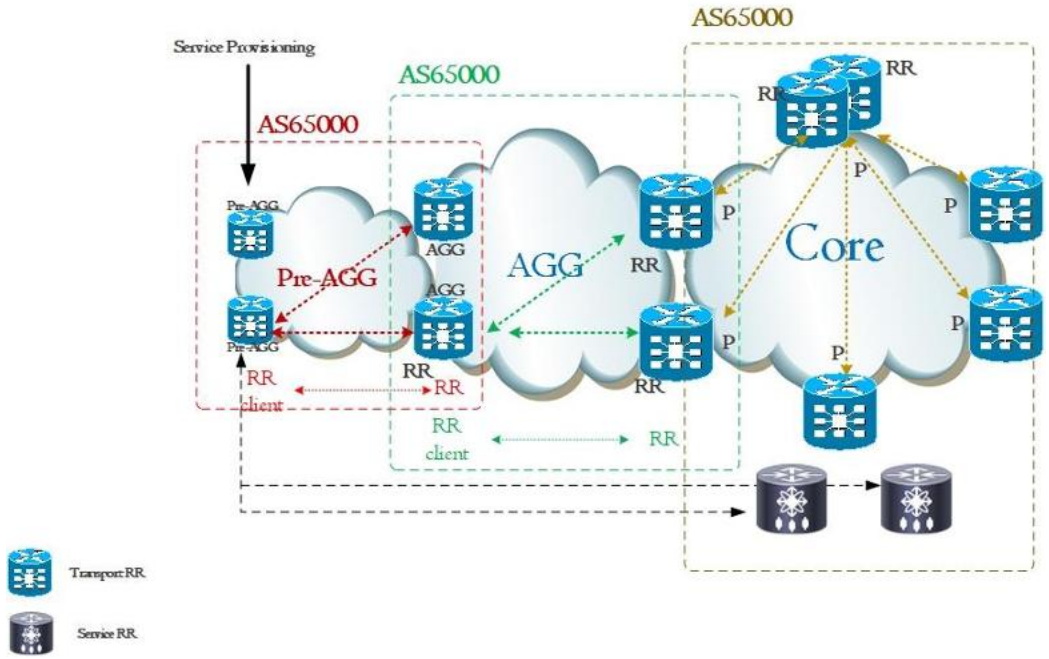


Figure 7-27 *Fixed Service Business/Corporate Customers BGP design*

https://t.me/learningnets

Convergence Mechanisms

BFD is deployed on the network for OSPF. BFD ‘min-rx-interval’ and ‘min-tx-interval’ is set to 50ms and the detection-multiplier is set to 4.

ATELCO recently deployed IP Fast Reroute Mechanism in Pre-AGG and Aggregation Layer and they had RSVP based Fast Reroute Link and Node Protection in the CORE network for many years. Before they deployed IP Fast Reroute, they had tuned OSPF timers.

Their OSPF timer configurations are:

- LSA Arrival time is 50ms
- LSA Throttling timers: 10 100 2000
- SPF Throttling timers: 10 100 2000
- LSA Retransmission timer is 60ms

They recently deployed IP FRR in the CORE network as well, and removed RSVP Fast Reroute configuration from the network. They kept OSPF tuned timers, so in case there is a destination which is not protected by IP Fast Reroute, they will still have good convergence time.

At the current stage, LFA, Remote LFA and BGP PIC is enabled in the IP/MPLS network. Based on the tests, ATELCO has the following convergence times:

- Single Link Failure < 50 msec.
- Single Node Failure < 50 msec.
- Edge Link and Node Failure < 200 msec.

All Fixed and Mobile Service customers will have the same benefit from IP Fast Reroute deployment as all the services use same Pre-Aggregation, Aggregation and Core IP/MPLS network. Services are only terminated on different Access nodes.

https://t.me/learningnets

IP Addressing in ATELCO Network

ATELCO is using /30 subnet for all the Point to Point interfaces in IP/MPLS Multi Service Network and OSPF network type is set to ‘Point to Point‘. This reduces the number of LSA’s in the link state database and provides easier troubleshooting and eliminating OSPF DR/BDR election process, to provide faster convergence.

All Loopback interface subnet masks are configured as /32. Loopback interfaces will be used for the device access and management. It will also be used in many protocols such as OSPF, BGP, LDP and LLDP in the network.

ATELCO IPv6 Design

Recently, Greenland regulatory body announced new regulation rules and has asked every Internet Service Provider in the country to deploy IPv6 service. The timeline for the deployment is 2 years.

IPv4 address exhaustion was a problem for ATELCO. They checked the IP Transfer option, but due to company security policies, they immediately gave up this idea. They requested couple times from their Regional Internet Registry – which is RIPE in their case, but RIR (Regional Internet Registry) responded as there is no IPv4 address space left and has put them in a waiting list. Finally they deployed their LSN (CGNAT) – Large Scale NAT devices three years ago.

Currently LSN is used only for ATELCO’s Mobile customers as they assign Public IPv4 address to their DSL and FTTx customers.

https://t.me/learningnets

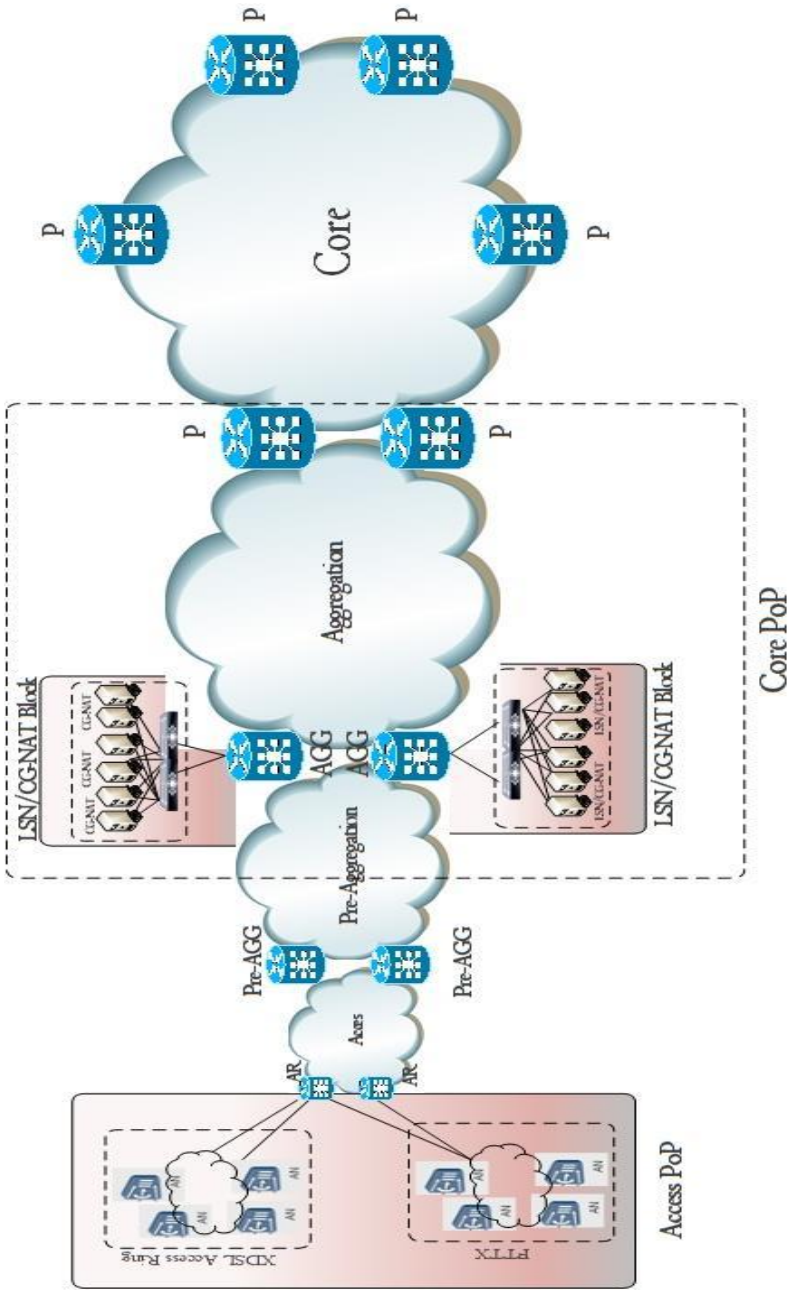


Figure 7-28 *Large Scale NAT Connection - ATELCO network – LSN/CGNAT*

https://t.me/learningnets

In figure 7-28, ATELCO's LSN deployment is shown. LSN boxes are connected to the Aggregation Layer routers. Earlier it was explained that BNG (Broadband Network Gateway) in ATELCO's network are connected to the Aggregation Layer Routers as well. LSN devices and BNG boxes effectively create an Edge Layer for ATELCO's network.

With deploying IPv6 service on their network, the load on the LSN devices are reduced. This means less number of NAT sessions are required, thus ATELCO doesn't need to buy new large LSN boxes. Also, keeping NAT logs and comparing them to logging and monitoring applications, requires large amount of storage space and processing on the server which needs more time to analyze user activities in case of criminal or hacking issues or government requests.

Because of regulatory reason's, ATELCO will deploy IPv6 on their internal network devices first. Although they have many DSLAM's which don't support IPv6, but all other Access Routers, Aggregation and Core devices support IPv6. They know that they need to enable IPv6 for their services, network operation center, routers, firewalls, load balancers, DNS, IP peering and transit connections, which are generally not the hard ones.

As in any other IPv6 deployment, the biggest concern is with the IPv6 transition mechanism.

ATELCO will choose and deploy IPV6 on their CPEs and Mobile devices. They are considering 464XLAT, and MAP-T solutions currently as there are many Mobile Operators around the world which have deployed these solutions.

MTU and Neighbor Discovery

As it is necessary and best practice to deploy higher size MTU numbers in ISP networks, ATELCO uses the following consistent MTU numbering scheme across all IP/MPLS network devices.

- Physical MTU is 9000
- IP MTU is 1600
- MPLS MTU is 9000

IETF Standard neighbor discovery protocol LLDP (Link layer Discovery Protocol) is enabled on all IP/MPLS Multi Service Network nodes.

Synchronization in Mobile Service

For synchronization in Mobile network, ATELCO uses both PTP and Synch-E in their network.

https://t.me/learningnets

Security Policy

Due to the corporate security policy in every layer except the Access layer, OSPF MD5 and LDP MD5 authentication will be enabled. For the BGP protocol, Flowspec rules are sent by the Border Routers to each Edge Router through Router Reflectors.

Also, ATELCO has been configuring AS-Path filters for their Business/Corporate customers to prevent some type of BGP based misconfigurations and attacks but they are researching other solutions to prevent possible future BGP based incidents.

Multicast in ATELCO Network

ATELCO has Rosen GRE SSM Multicast deployment for their IPTV service and currently IPTV is provided only to their FTTH customers. IPTV set top-box is provided and managed by ATELCO and all of them are IPv6 capable devices.

Summary

In this chapter, ATELCO'S fixed and mobile services for both residential and corporate customers, connectivity inside the country, worldwide connectivity, their network technical background and requirements were shared.

ATELCO's IXP connections, their transit provider connectivity and their bandwidth requirements for their inside country connections and also outside country connections were explained.

ATELCO's IGP, BGP, MPLS, Multicast, Convergence characteristics, Security policies for IGP, LDP and BGP, Timing and Synchronization, IP Addressing, MTU deployment, LSN (Large Scale NAT) –CGN deployment, IPv6 design and many other technology and protocol usage were highlighted.

ATELCO has fixed and mobile services, both services use same Pre-Aggregation, Aggregation and Core networks. Different services have separate Access networks. These information were shared in this chapter along with their future plans.

This chapter was considered as an overview by the Author of the book, next chapter will go into the detail of their fixed and mobile services for residential and corporate customers and will explain the alternative methods with the advantages and disadvantages. Also, ATELCO's IP/MPLS Unified MPLS design will be explained in great detail in the next chapter.

https://t.me/learningnets

Chapter 8

Chapter-8 ATELCO Network - Design Detail Explanations

Introduction

During the scenario, many information about ATELCO network was provided. In this section, detail explanation will be provided about their network design and also many real-life ISP network design examples will be shared with the readers.

Many information and the topologies which were shared in the previous chapter about ATELCO's network will be repeated here as a reminder. The reason of having this chapter as a separate chapter in the book is to keep the previous chapter short and understandable. So we will provide more detail in this chapter about the current design of ATELCO and discuss different alternative designs which other Internet Service Providers have in the World.

After this section, you will understand what kind of services other Internet Service Providers provide as well, what are the current trends, how they design and implement the technologies and the protocols on their network along with pros and cons of the available methods that are used.

In each section of this chapter, we will share the text which was given in the previous chapter during ATELCO's design review.
We will discuss the alternative design options, pros and cons of the current design of ATELCO and we will have a look at the future roadmap for ATELCO by keeping the Evolving Technologies in mind.

In the previous chapter, it was given that ATELCO is a leading telecommunication services company in Greenland (A fictitious country). They have nationwide backbone infrastructure in

https://t.me/learningnets

fictitious country Greenland. They have around 11 million customers in the country; most of them are Residential (fixed and mobile broadband) customers.

11 million customers for many countries is very normal as the subscriber numbers. In fact, there are many Operators that have more than 11 million customers in their country.

Out of 11 million customers, most of them are residential customers. ATELCO has Enterprise Customers as well. They haven't shared the number of Enterprise customers but it was given in the previous chapter that ATELCO terminates Enterprise Customers at the Pre-Aggregation Layer. We will discuss Service termination for the Residential and the Enterprise customers in detail later in this chapter.

We will cover this chapter in many parts. It will start with the Physical Infrastructure of ATELCO, secondly Logical Architecture will be shared. In the Logical Architecture section, IP/MPLS, IGW and Global Network Connectivity will be covered. After that, Services and last but not least Technologies and the Protocols in ATELCO will be explained in great detail.

ATELCO Physical Network

Regional Connectivities of ATELCO Network

We will start to introduce ATELCO's network. ATELCO has four regions in the country called Greenland. Internet Service Providers can be categorized as follow:

- Nationwide
- Regional (which is more local)
- International

Internet Service Providers which provide nationwide service, generally divide the country to different regions. Some operators use the term 'Regional Provider' for those who provide services to more than one country. So, the definition of 'Regional Provider' is either the Provider who serves a specific part of a country or provides a service to more than one country. Definitions might change based on the place in the world.

In U.S, regional provider mostly refers to one which provides the service in a part of the country, in Europe; Regional Provider refers to one which provides the services to more than one country.

As per Author's experience, in general, ISPs divide the country as three, four and even five regions (many cities inside each region) and manage the traffic flow, deploy the physical devices, start the services and connect the Internet Interconnections based on these regional arrangements.

Operators which divide the country to three regions have East, West and Central regions. They do it for optimal routing which provides better resource utilization, so leads to reduced costs. Each regional traffic reaches to the Internet from their local connectivity. So, the East Region users

https://t.me/learningnets

reach to the Internet from East Region Internet Gateways, West Region users from West Region Internet Gateways and Central Region users use Central Region Internet gateways.

Traffic doesn't always go to the Internet due to content cache engines. When a user wants to watch popular TV series, which the Service Provider has the local copy through Cache Engine's inside their network, this traffic doesn't leave the ISP network. Instead, it serves the content from the cache nodes inside the ISP network, which reduces latency, so the user gets better experience. When the geography is large, more regions in the ISP network is seen.

In the ISP networks, 80% of the traffic stays in the network (On-Net Traffic), while 20% leaves the network to reach to the destination (Off-Net Traffic). Many engineers call this the 80 to 20 rule. It is important to explain On-Net and Off-Net Traffic concept here.

Off-Net Traffic is the traffic that is handed off to another network at some point in its journey between source and the destination.

On-Net Traffic is the traffic under the control of the same network. On-Net Traffic remains on the same network between origination (source) and the termination (destination). ATELCO has four regions, North, South, East and West.

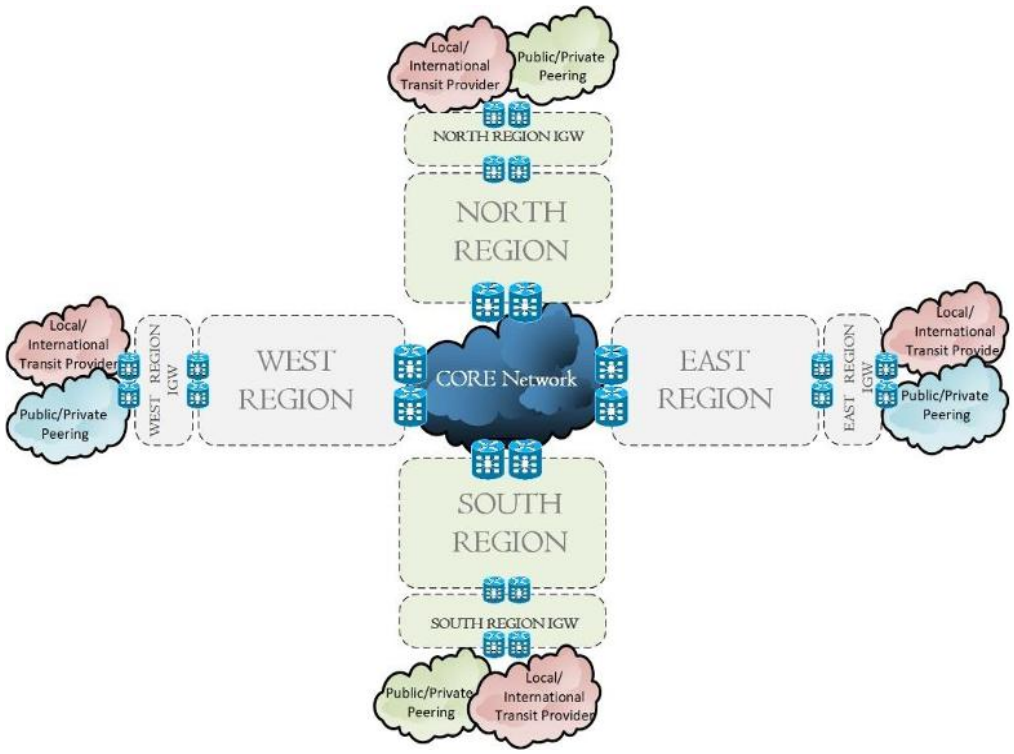


Figure 8-1 *ATELCO's Network Consists of Four Regions*

https://t.me/learningnets

As it was explained during the scenario, ATELCO has Internet exit points from all four regions. The North region serves as a backup if one Region completely loses its Internet connectivity. They have cache engines in each region.

As you can see from figure 8-2, Regions are connected to each other from the IP/MPLS Core-P routers. Both topologies shown in figure 8-1 and figure 8-2 were shared during the previous chapter.

https://t.me/learningnets

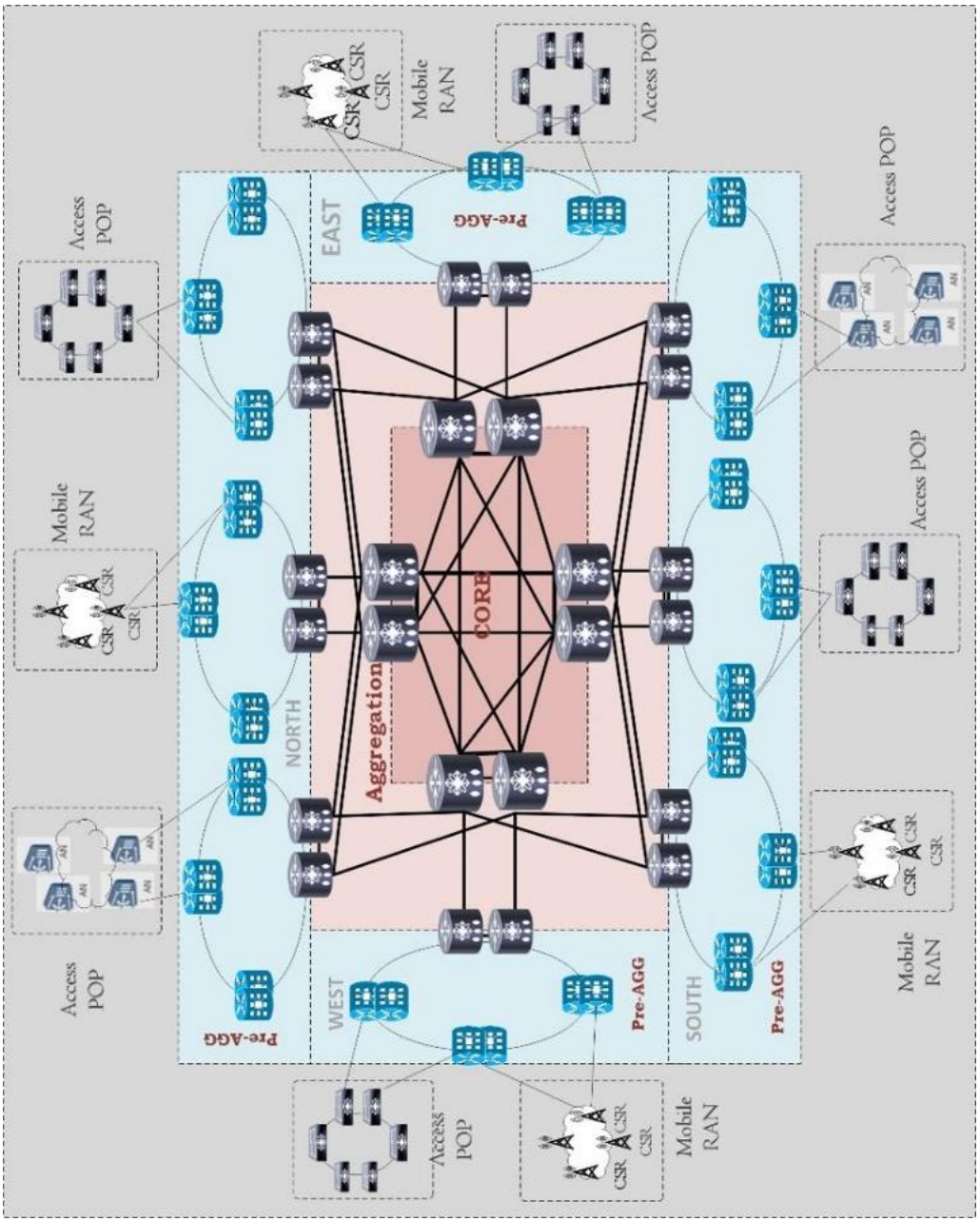


Figure 8-2 *4 Regions of ATELCO network connected via IP/MPLS Core Network*

https://t.me/learningnets

In figure 8-2, the four regions of ATELCO are shown. Access devices, such as DSLAM, OLT and CSR are connected to Access Routers. The Access Routers are terminated at the Pre-Aggregation Layer Router.
In turn, Pre-Aggregation Layer routers are connected to Aggregation Layer Routers and Aggregation Layer Routers are connected to Core Routers.

In Each region, ATELCO has two Core POP Locations as it was shared during the scenario in the previous chapter.
In some ISP networks, there might be less Layers. Pre-Aggregation Layer might not be seen. This is related with scalability. Maybe the country might be small or the number of customers may not be so much.

So far, ATELCO's design is not much different than classical Campus network design. We have multiple layers and different layers have different functionalities. Access Layer is the user termination, Core Layer is for high speed packet forwarding. Pre-Aggregation and Aggregation Layers connect different Access locations to each other, which provides extendibility to the network.

ATELCO has Access, Aggregation and Core POP locations. During the previous chapter, if it wasn't clear for you, ATELCO has Access, Pre-Aggregation, Aggregation and Core Layer, so totally four layers of hierarchy is in the network, but they have three level of POP hierarchy which is Access, Aggregation and Core POP locations.

The reason for this is that the Access Layer devices are located in the Access POP sites, Pre-Aggregation Layer devices are located in the Aggregation POP sites, and finally the Aggregation and Core Layer devices are located in the Core POP location sites.

There are different design options, but consolidating the different layers of devices in the same POP location is mainly done for OPEX reduction and it's a very common approach in ISP networks.

https://t.me/learningnets

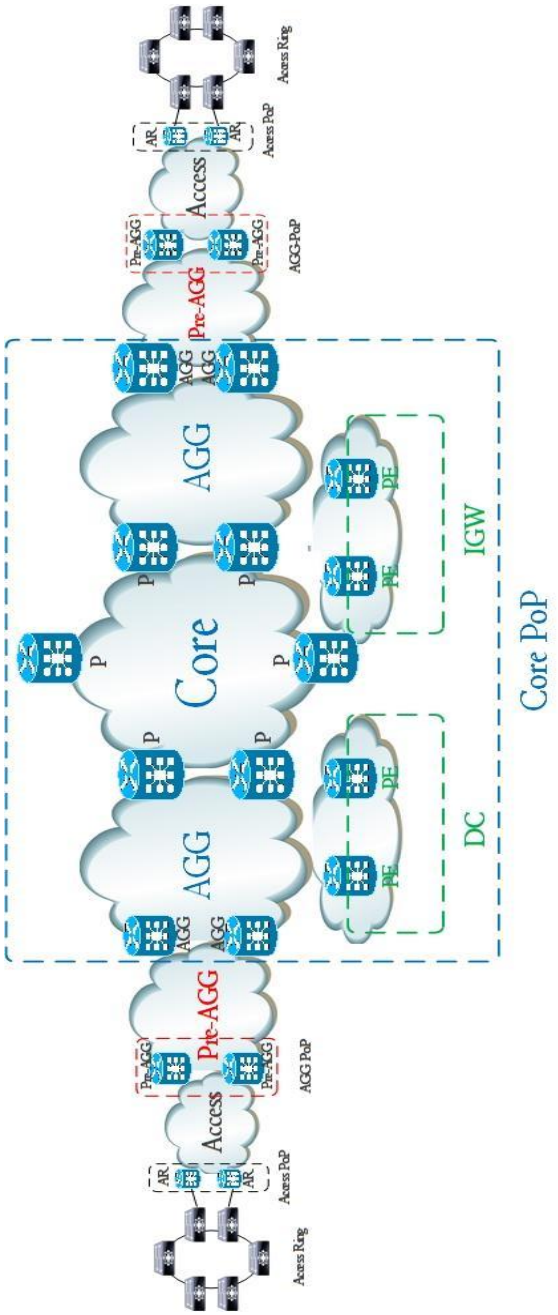


Figure 8-3 *Access, Aggregation, Core POP and the Layers in ATELCO*

https://t.me/learningnets

In Access POP locations, there are Access Layer connectivity devices such as DSLAM, MSAN, and OLT. Also, fiber connectivity nodes such as SONET/SDH and DWDM equipment are located there. In the Access POP's, these nodes are terminated at the AR (Access Routers). The reason for this is that instead of connecting many DSLAM or OLT devices to another POP location (which in ATELCO's network is the Pre-Aggregation Layer in different sites), multiple DSLAM or OLTs are terminated at the Access Routers in the Access POP location, so much less connections are carried between the POP locations.

If all DSLAM devices would be connected to another physical POP location, topology would be Hub and Spoke, which is more expensive than Ring topologies.

ATELCO has Ring topology in the Access and Pre-Aggregation network. In Aggregation POP, higher capacity devices are used. These are called Pre-Aggregation Layer or UPE layer devices in ATELCO's network. It is important to understand from the design point of view that why ISPs use Ring topologies, what are the pros and cons of having Ring vs. other common WAN topologies.

Table 8-1 *Different physical network topologies - comparison table*

Physical Topology	**Ring**	**Hub and Spoke**	**Partial Mesh**	**Full Mesh**
Redundancy	Low	Low	Medium	Good
Cost	Low	Medium	High	Too High
Complexity	Low	Medium	High	Too High
Optimal Routing	Bad	Medium	Good	Best

The Ring topology is the cheapest topology but still provides redundancy as every node is connected to two other nodes. Many Mobile Operators and Fixed Network Operators prefer Ring topology for their Access networks. Failure impact of Access node is very low compared to upper Layer nodes, such as the Core layer. This phenomenon is called 'Blast Radius 'as it was explained in the earlier chapters.

The DSLAM and OLT devices as well as Mobile devices such as Cell Site Routers are terminated at the Access Routers. Fixed AR (Access Router) is used to terminate DSLAM and OLT, Mobile AR is used to terminate CSR (Cell Site Router). Access Routers are terminated at the Pre-Aggregation routers. Access Routers are connected in a Ring topology and their traffic is carried to the Pre-Aggregation Layer which is located at the Aggregation POP as it can be seen from figure 8-4.

https://t.me/learningnets

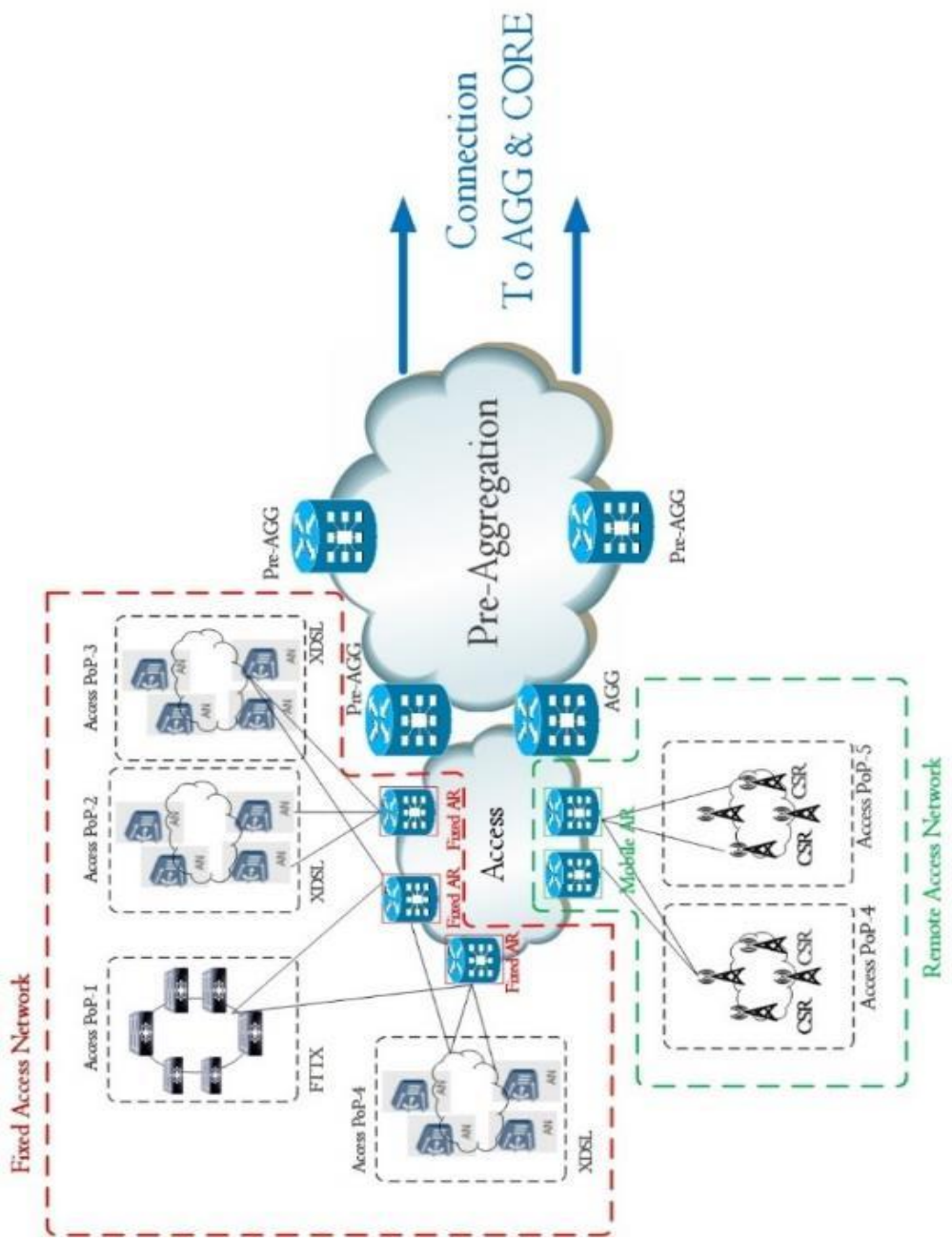


Figure 8-4 *Fixed and Mobile Access Routers connected to Pre-AGG Routers*

https://t.me/learningnets

Core POP sites provide connectivity between all the POP locations, as well as connecting all customers to the Internet and to other modules of the SP network.

Gateway POP locations provide connectivity to other networks and to the Internet, while they host cache engines of the content providers as well.

ATELCO Intra Region Physical Connections

During the previous chapter, it was mentioned that in each region, ATELCO has only one Internet Gateway POP location, but various numbers of Access, Aggregation, POP and Two Core POP Locations.

In each region the Internet Gateway POP is located in the Datacenter. From the 2 Core POPs in each region, one of the Core POPs are located in the Datacenter and the other Core POP is not placed in the Datacenter. Based on the importance of the Core POP locations that are aggregating all of the Aggregation POP sites in their region, and also collecting all the customer traffic from their region for sending it to the Internet or to other regions, the Core POP locations are equipped with the highest redundancy standards as possible (Security, Power, Cooling, etc.).

The buildings that the Core POP sites are located (either the Core POP that's located in the Datacenter or the Core POP that is not placed in the Datacenter) are protected buildings, located in safe areas and have the highest standards for high availability.

ATELCO has four Datacenters. Each one located in one of the most major cities of those four regions.

Let's understand how different cities are connected in ATELCO's network.
Figure 8-5 was provided in the previous chapter for introducing ATELCO's different City Types.

https://t.me/learningnets

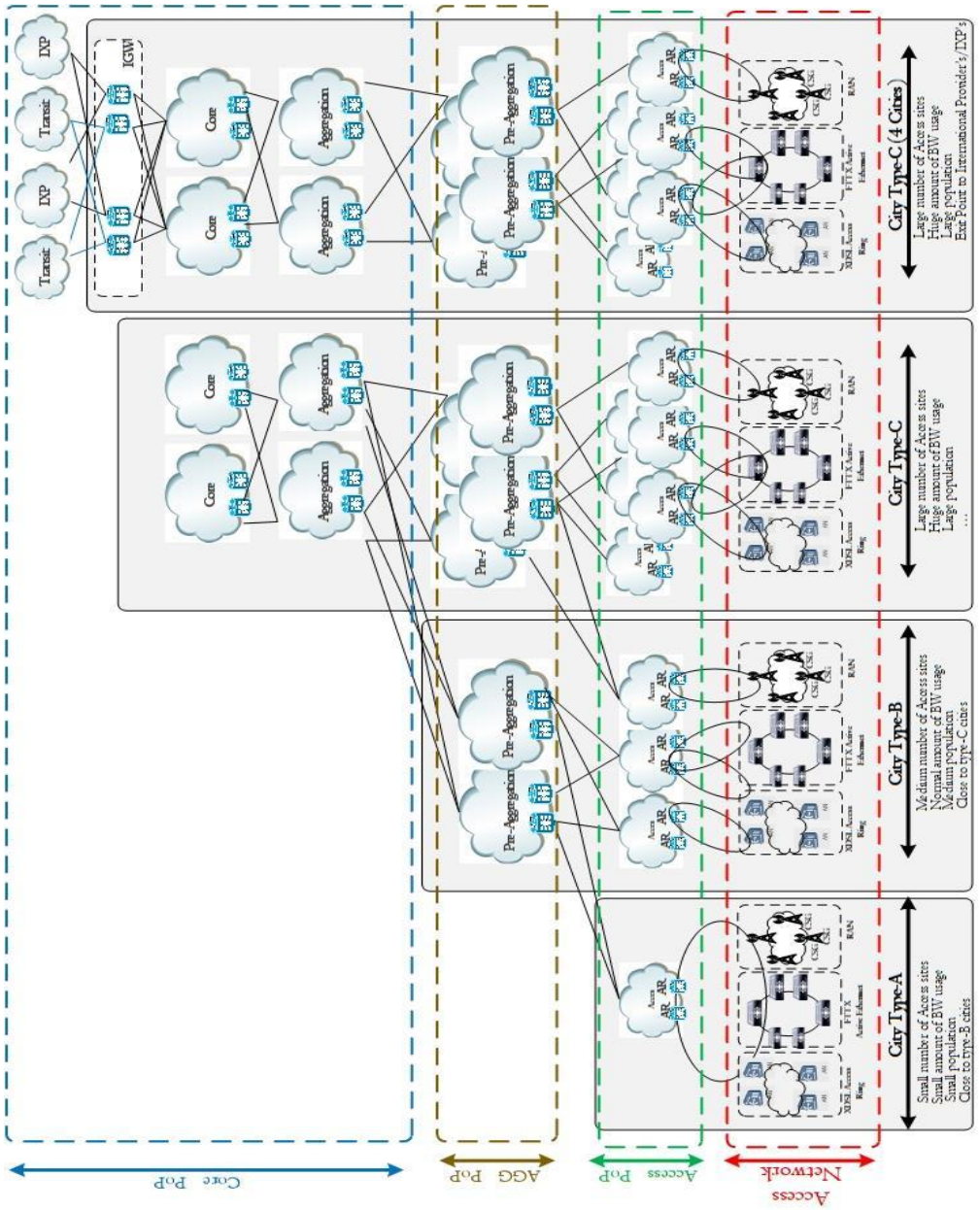


Figure 8-5 *POP Interconnections – Metro network connections*

For a better view of this Figure, visit www.orhanergun.net/SPBook/CityTypes.jpg

https://t.me/learningnets

ATELCO sells its services in a single country. It has connections to different countries. It does Settlement Free Peering at the IXPs in those countries, but it does not sell any services in those countries. The main reason ATELCO has a presence at those countries is to get the content and Transit service outside of its country.

There are three types of cities in each region which are named as Type-A, Type-B and Type-C cities. Cities are categorized by ATELCO based on the following criteria:

- Population of the city
- Number of service subscribers
- Aggregated bandwidth for the users

In figure 8-5, different city types are shown. Based on population, number of subscribers and aggregated bandwidth, ATELCO has classified the country and has deployed different kinds of POP locations in different cities in the country.

As it was mentioned in the previous chapter, ATELCO provides their services in 60 cities. Based on this type of categorization, each city type is explained as follow:

Type-A Cities: Has only Access POP locations and the Access POPs are connected to each other via Ring topology and for connecting to the backbone of the network or connecting to the Internet, Type-A cities are connected to the nearest Type-B city, where there is an Aggregation POP location.

Type-B Cities: Has both Access and Aggregation POP locations. Access and Aggregation POPs are connected in a Ring topology and for connecting to the backbone of the network or connecting to the Internet; Type-B cities are connected to the Core POP sites of the nearest Type-C city.

Type-C Cities: Has all Access, Aggregation and Core POP locations.

In figure 8-5, there are actually four types of cities. The topology shown at the right side of the picture, which has the IGW POP, is also classified as Type -C city as well. The only difference is that, from all the Type-C cities, four of those Type-C cities have IGW POPs, in addition to Core POPs.

POP location selection criteria in ISP networks are similar to city type categorization. Population of the city, number of subscribers and the number of bandwidth required for those subscribers are some of the reasons that ISPs select what type of POP locations are deployed in each city.

https://t.me/learningnets

Table 8-2 shows the important design characteristics of different types of POPs in Internet Service Providers.

Table 8-2 *Design characteristics of different types of POPs in ISPs*

Design Characteristics	**Access POP**	**Aggregation POP**	**Core/Backbone POP**
Physical Security	Low	Medium	High
Location Size	Small	Medium	Large
Contract with Location Owner	Short	Medium	Long Term or Ownership
Uplink Capacity	Low	Medium	High
Uplink Redundancy	Low/Ring Based	Medium Ring or Hub and Spoke	High, Partial Mesh or Full Mesh
Power Requirement	Low	Medium	High
Power Redundancy	Low	Medium	High
HVAC	Non-Critical	Important	Highly Critical

In the previous section we mentioned that there are two CORE/Backbone POP locations in each Region, which are located in the Type-C cities. The Regions are connected to each other via these CORE POP locations with a Full Mesh topology. This is shown in figure 8-6.

https://t.me/learningnets

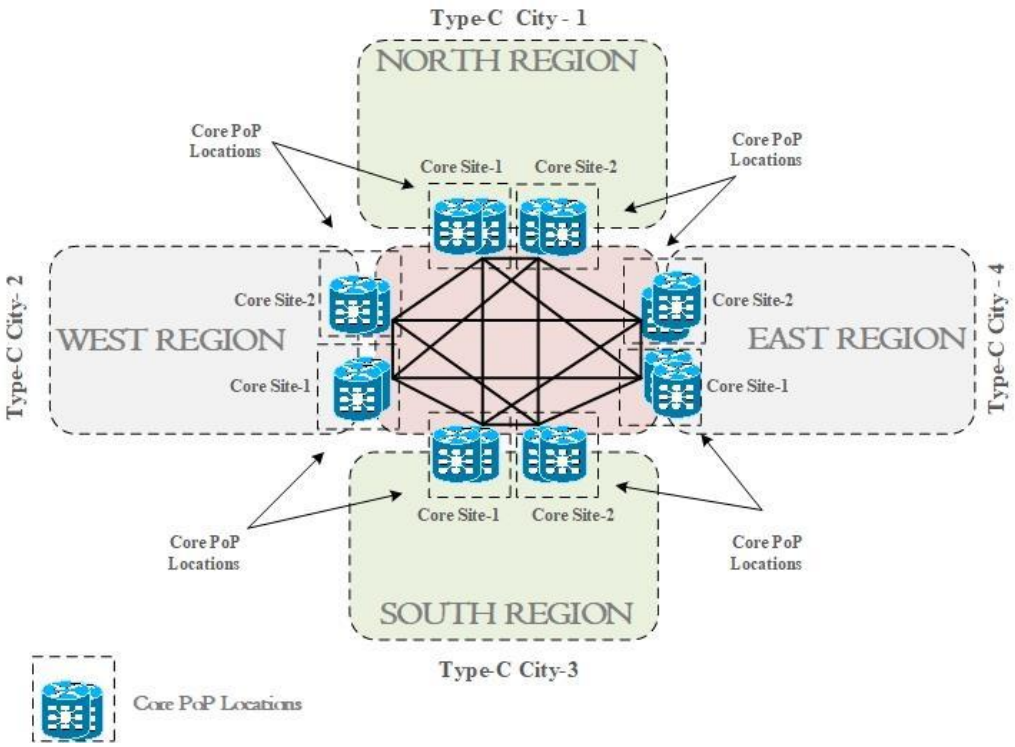


Figure 8-6 *Two CORE POP's in each region – Full Mesh Connectivity*

In ATELCOs network, there are two Core POP sites and one datacenter in each region. Thus, one of the two Core POP locations in each region is located inside the Datacenter and another Core POP is located outside of the datacenter. The CORE POP location of the Internet Service Provider has much more capacity, much more availability mechanisms and much more security (network and system level), even if there is no Datacenter at the location.

ATELCO Network – Logical Architecture

In the last section, ATELCO's physical network connectivity, different city types, POP, POP Interconnections and selection criteria were explained. For example, we mentioned that the IGW POP and a single Core POP in each region reside in the Datacenter. But we haven't explained how the Access, Aggregation and Core network is connected to the Global Internet and how VPN customers of ATELCO send the network traffic between their sites etc. In this section ATELCO's logical network architecture will be explained.

https://t.me/learningnets

It was explained in the previous chapter that ATELCO's network consists of three different segments.

- IP/MPLS network
- Internet Gateway Layer
- Global Network Layer (if there is Interconnections in other countries)

We will first cover all three network segments of ATELCO in detail and then we will look at Services and Protocols in each of these segments. Finally, we will finish this chapter by looking at the end to end picture of ATELCO and explaining the end to end traffic flow of some of their services.

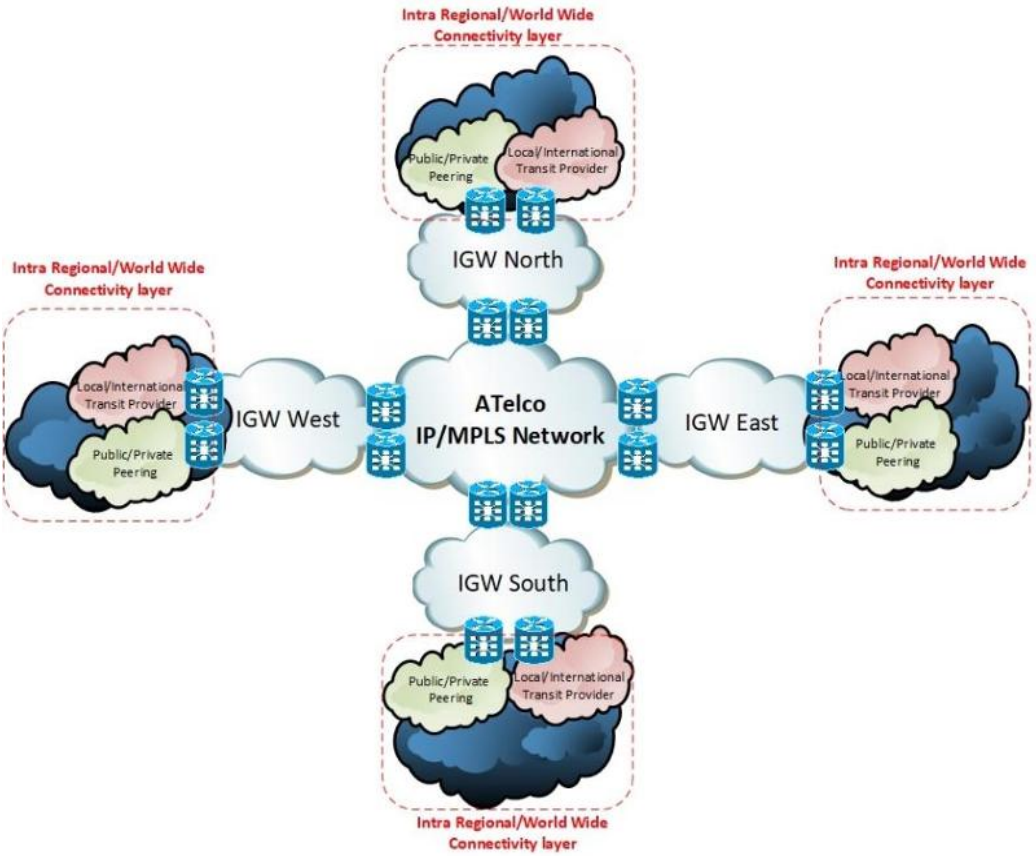


Figure 8-7 *IP/MPLS, IGW and World Wide Network Connectivity*

Figure 8-7 shows three different segments of ATELCO's network. IP/MPLS network is connected to the IGW (Internet Gateway) which is connected to the Worldwide Connectivity Layer to reach the Global Internet.

https://t.me/learningnets

IP/MPLS Multi Service Network

IP/MPLS Multi Service Network is extended up to the Access nodes for the Mobile service, but not for the fixed services. In ATELCO's network, there are different types of Access connections. Access connections are grouped in three different categories:

1. Fixed Residential Access
2. Mobile Access
3. Fixed Business Customer Access

ATELCO's Mobile access nodes such as Cell Site Router are MPLS enabled.

The fixed Service access network is based on Layer 2. Thus, FTTX, XDSL or Business Customer access services are based on Layer 2.

Services and the Protocols in the ATELCO network will be explained in detail later in this chapter. There are many Internet Service Providers which use Layer 3 based fixed access networks, which means their DSLAM and OLT are MPLS enabled.

When it is MPLS based, technologies such as LFA, Remote LFA and RSVP based TE Fast Reroute can be used, but when it is Layer 2 based, G.8032 and REP are the commonly used approaches for fast convergence. (At the time of this book written)

These technologies provide fast failover to the alternate/backup link or node in case of primary failure. Many Mobile Operators have deployed RSVP based Traffic Engineering in their network, including their Access network for Fast Reroute and Traffic Engineering purpose.

ATELCO's Multi Service IP/MPLS network has four layers.

1. Core
2 Aggregation
3 Pre-Aggregation
4 Access Layers.

Network devices that form each layer are explained as follow:

- P routers form the Core Layer
- AGG PE routers form the Aggregation Layer
- Pre-AGG PE routers form Pre-Aggregation Layer and provide connectivity between AGG and Access Layers. Pre-Aggregation Layer is based on Ring Topology as well.
- AR (Access Router) PE/non-PE devices provide Access Layer connectivity for different access network types. The access layer physical connectivity for both Mobile and Fixed services is based on ring topology.

https://t.me/learningnets

Different physical topologies have different design characteristics. For example, full mesh is the most expensive topology but provides the lowest latency and highest redundancy. Ring is the cheapest topology which provides minimum redundancy as every node is connected to two other nodes.

Access networks in most ISPs, if not all, are based on ring topology. The main reason for that is cost and failure impact. It is cheap and provides minimum redundancy. In case of failure, the impact on ATELCO's network is small, because small number of users will be affected. If the failure would be in the upper layers, such as CORE, as there are many Aggregation and below layers connected to the CORE, the failure impact on ATELCO's network would be much bigger, thus redundancy in the CORE is much more important compared to other places in the network.

That's why CORE networks are designed based on fully meshed or near fully meshed topologies. Not only redundancy, but also traffic demand dictates having full or near-full mesh connectivity in the CORE network. For traffic to pass between the regions, it should go through the Core network. In order to send the traffic optimally, direct connectivity might be required. Having direct connectivity between each CORE location turns the topology to Full-mesh topology.

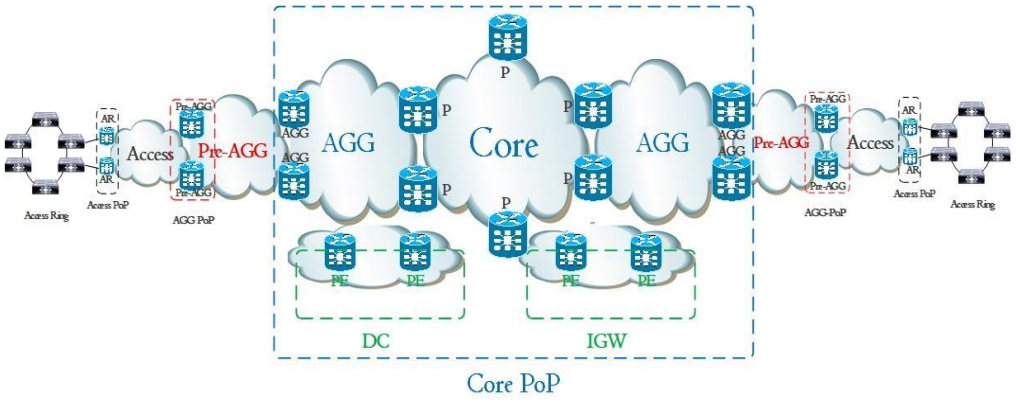


Figure 8-8 *POP Architecture and Multi-Layer Network of ATELCO ISP*

In figure 8-8, it is shown that Access Layer devices reside in the Access POP, Pre-Aggregation Layer routers reside in Aggregation POP locations and finally Aggregation and Core Layer devices reside in the Core POP locations.

This is how ATELCO's network devices are located in different categories of POP locations. In another ISP network, this design might be different. Different Layer devices can be located differently in the POP locations.

https://t.me/learningnets

For example, Access Layer devices can be located in the Access POP. There might be less layers, such as Aggregation and CORE, without having Pre-Aggregation. In this case, the Aggregation Layer and CORE can be collocated in the CORE POP location. Or the Aggregation Layer devices can be located in the Aggregation POP, and Core Layer devices be located in the CORE POP locations.

Aggregation Layer devices are the ones that require enough capacity to terminate Pre-Aggregation and Access capacity. CORE layer devices are the ones that require enough capacity to terminate Aggregation Layer and below layers. Device selection criteria is mainly based on the capacity (data plane) and the functions, such as MPLS LER, LSR, BNG, CG-NAT functions.

World Wide Connectivity Layer

World Wide Connectivity Layer is commonly seen in Tier 2 and Tier 1 ISPs. If ISPs don't have any Settlement Free Peering or Transit Connectivity outside of their country, they receive the Internet Service from other ISPs in their country. If this is the case, they don't have the World Wide Connectivity Layer.

Service Providers who have IP Transit or Settlement Free Peering connections outside of their country, generally still have Peering connections and still receive an IP Transit service within the country from other national providers or from the International providers. Receiving IP Transit service within the country as well as outside of the country is used as a backup connectivity to the Internet for Settlement Free Peering.

Level 3, TATA Communication, KPN, Seacom, Verizon, AT&T and Turkish Telecom International are the sample operators which provide International services outside of their main territory.

World Wide Connectivity Layer, Global Connectivity Layer and Global Network are the common terms in the industry to define company's presence at the other countries. Today, Europe and U.S have many IXP's (Internet Exchange Points) where hundreds of companies peer with each other and exchange their customer traffic.

ATELCO joins to the Exchange Points and peers with other companies as it was explained during the scenario. These locations are:

- London – U.K
- Frankfurt - Germany
- Hong Kong - China
- Ashburn - U.S

https://t.me/learningnets

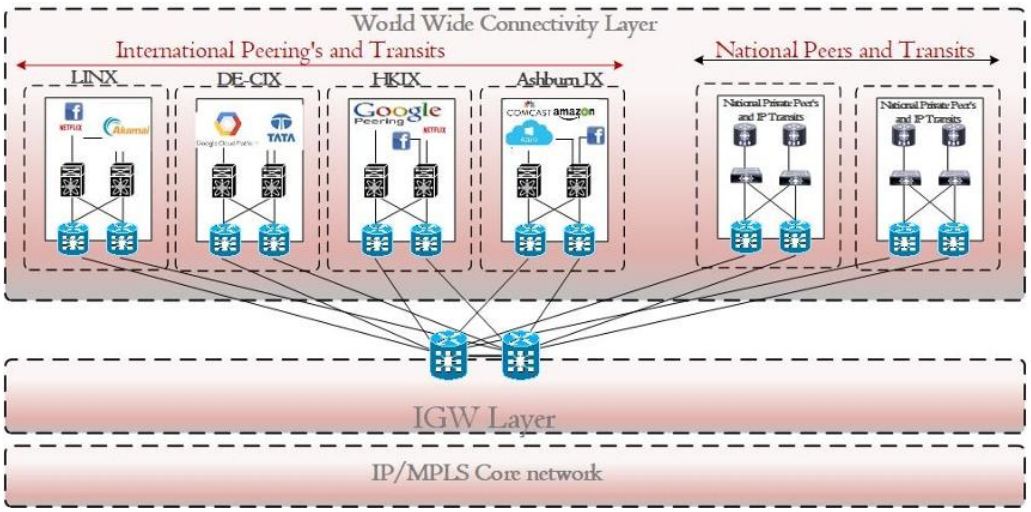


Figure 8-9 *World Wide Connectivity of ATELCO*

Internet services today are mostly provided by a small set of CDN providers, including Facebook, Amazon, Google, Akamai, Microsoft, Apple, Netflix, Cloudflare and perhaps just a couple of others. All other service providers essentially push their traffic into these large content networks. Some of these Content Providers bring their content into the ISP network but when there is no enough bandwidth meeting the expectations of the Content Providers, they don't bring their cache engines into the ISP network.

Bandwidth is a very important consideration for the Content Providers to place their caches on the networks. If enough bandwidth is provided by the IXP (Internet Exchange Point), then Content Providers such as Google, Facebook and Microsoft place their nodes there and Internet Service Providers access the content from an IXP. Generally, these content providers require 10s of Gbps of traffic to bring their cache engines to the network.

For larger content providers, per unit cost of cache server is much less compared to smaller content providers, due to economy of scale. For example, Google can build their cache engines much cheaper compared to smaller content providers.

ATELCO has some of these large content providers in their network. This means their users have less latency. Also, when ATELCO serves the content directly from their network, they don't use their International bandwidth capacity to access the content which can be reachable from these caches.

https://t.me/learningnets

As it was shared in the scenario, there is no IXP in the country where ATELCO is operating. But also, it is said that ATELCO receives IP Transit Service from the Tier 1 Operators inside the country.

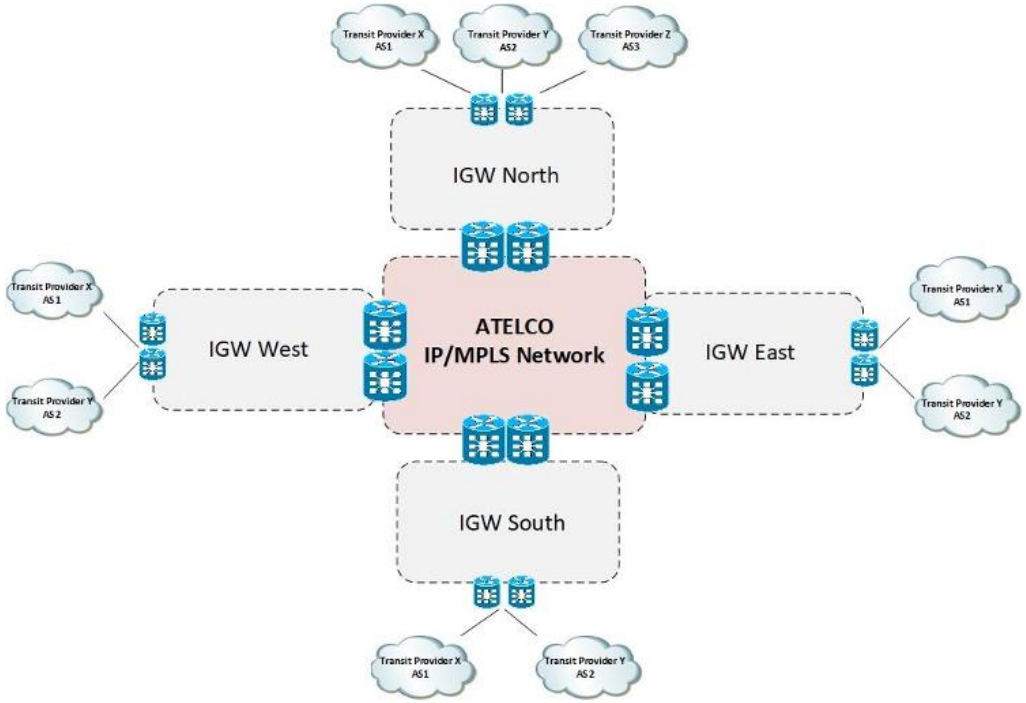


Figure 8-10 *ATELCO's Transit Operator Connections*

ATELCO's Transit Operator connections was shared in the previous chapter. Let's have a closer look. There are three Tier 1 Providers which ATELCO receives IP Transit service within the country. Transit Provider X (AS1), Transit Provider Y (AS2), Transit Provider Z (AS3). AS 1, AS 2, and AS-3 are used here as an example. It doesn't represent any real AS numbers. AS-1 and AS-2 provide an IP Transit service, which means Full Internet routing table in each region. AS-3 also provides IP Transit service with full Internet routing table, but only in the North region.

In the previous chapter, it was explained that the North region is used as a backup to all other regions. In case of failure in any of the regions that brings down both AS-1 and AS-2 connections, the North region will be used as backup.

Global Tier 1 Operators (13 of them as it was explained in earlier chapters) provide IP Transit connectivity in other countries as well. When they provide IP Transit in ATELCO's IGP POP, within the country where ATELCO sells its services, the Global Operator may not have physical infrastructure, such as its own POP location. In this case, the Global Operator commonly pays to the Local Operators infrastructure.

https://t.me/learningnets

For example, they rent a space in ATELCO's datacenters and pay for the colocation to ATELCO. At the other side, ATELCO pays to these Transit Operators for the IP Transit service. So, there are many different business agreements between them, not just Local Providers paying to the Transit Providers. These Tier 1 Operators might have the content within the country, so when ATELCO doesn't have the content that their users request in the network, it can be served from the Tier 1 network within the same country (if that Tier 1 has the Content Provider's servers).

Reaching the content through Tier 1 Operators instead of accessing it via cache engines in its own network would cost more money for ATELCO, because the Tier 1 service provider is the upstream provider for ATELCO and ATELCO has to pay money to reach the content via Tier1. In figure 8-11, ATELCO's users are shown. ATELCO has some Content Providers cache engines inside the network. ATELCO also has Tier 1 Operator connections which have Content Caches in their network too. ATELCO pays Tier 1 Operators to reach the content servers in their network, as well as for the Global Internet.

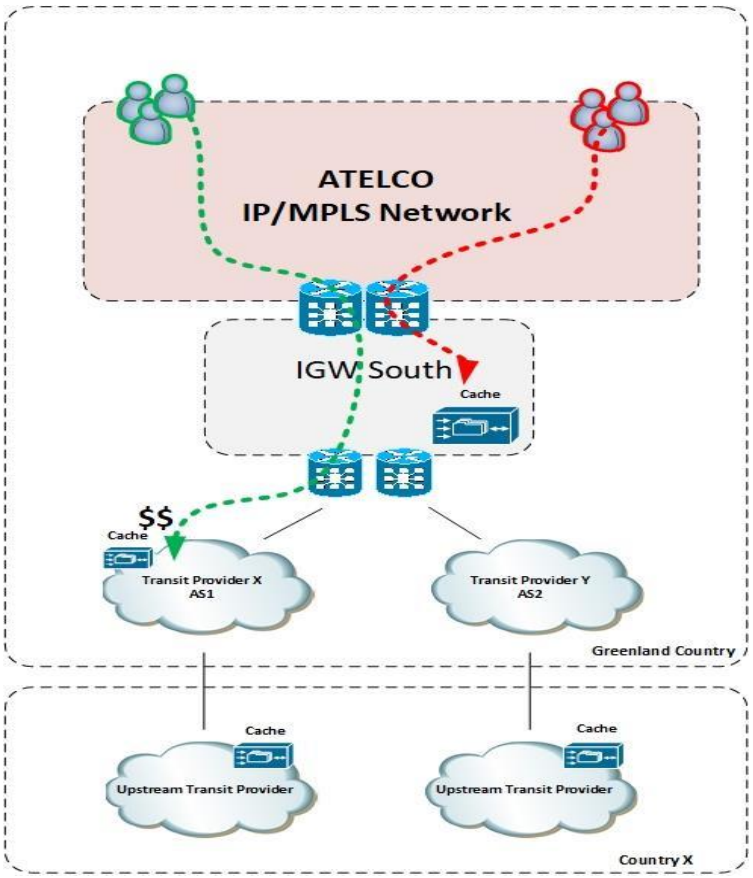


Figure 8-11 *Content Cache Servers and Transit Operator Connections*

https://t.me/learningnets

According to figure 8-11, the Red users (right side of the picture) access the content which is cached inside ATELCO's network. Thus, the Tier 1 Operator network is not used. The Green users (left side of the picture) want to access the content which is not found in the content engines in ATELCO's network, that's why ATELCO sends the user traffic to the upstream network which is its Tier 1 Operator connection. These user's requests are found in the content servers in the Tier 1 Operator which is in the same country. ATELCO still pays for the Tier 1 Operator in this case, but the benefit is lower latency as the content is served from the cache servers within the same country.

If content couldn't be found in the same country but the Tier 1 Operator would need to bring the content from the International locations, then ATELCO still would pay the cost but the latency would increase, which in turn would result in bad user experience.

It was mentioned that IGW and WCL (World Wide Connectivity) Layer was operated by different teams but recently they merged the teams and created a new team called AWN (ATELCO Worldwide Network). There are many Internet Service Providers who have different teams for managing Internet Gateway Layer and World Wide (Global) Connectivity Layer.

Larger and Multi National Internet Service Provider companies tend to have separate teams for these layers. Having different teams to manage different parts or services of the network might add latency when there is a problem with the Interconnections or BGP routing. This is because troubleshooting time increases when multiple people deal with the problem.

But Multinational companies have different teams, because network scalability and technical knowledge requires this. APAC region Interconnection team doesn't involve in the routing problem at the IGW layer in Europe. Also, knowing BGP configuration doesn't mean understanding the Settlement Free Peering, IP Transit and other International terms and the activities as an expert level.

Last but not least, ATELCO's Transit Operators are explained as Global Tier 1 Operators, such as Level 3, TATA etc. These can be different for other ISPs. For example, there might be country wide Tier 1 ISPs which can provide IP Transit service to other Tier 2 ISPs.

In Turkey, there are many ISPs which don't have connectivity to other countries. They receive an IP Transit service from Turkish Telecom which is a national level Tier 1 Operator. But Turkish Telecom is not a Global Tier 1 Operator, as Turkish Telecom receives IP Transit Service from the Global Tier 1 operators outside Turkey. More detail information can be found on www.peeringdb.com

https://t.me/learningnets

Internet Gateway/Shared Services Layer

Internet Gateway/Shared Services Layer of ATELCO connects IP/MPLS to the Global Network. In this layer, they have Internet Gateway Routers and network service nodes such as Firewalls, IDS/IPS. Load Balancer, CGN (Carrier Grade NAT) boxes and the Content Providers cache engines. Also, as it was mentioned, this layer resides in ATELCO's Datacenter. During the scenario, it was told that ATELCO has four datacenters.

Many National Internet Service Providers have much more than four Datacenters. There are Operators which have more than 30 datacenters in the same country. The main idea of having more datacenters is not to have more capacity but being closer to the customers. If the customer is in New York, having a Datacenter in New York for the Service Provider is an advantage. When customers compare the Service Providers to select the one which they will get the service from, proximity is important as they need to pay for the circuit, in order to reach to the datacenter.

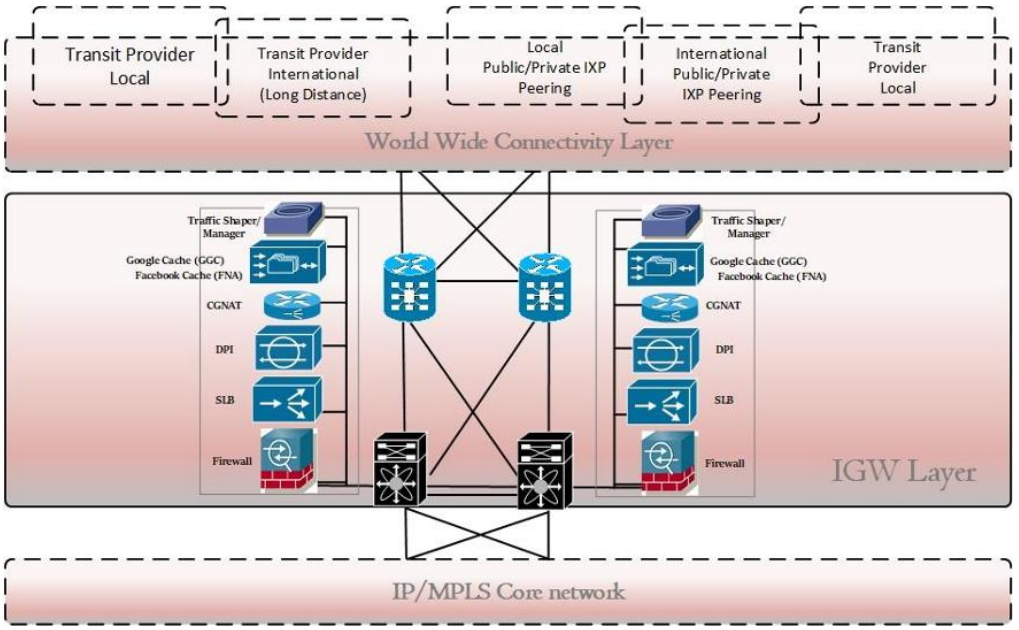


Figure 8-12 *IGW Layer Logical Connectivity in ATELCO network*

Internet Gateway (IGW) and Shared Services Layer in ATELCO's network provide Internet Access and connects the IP/MPLS network to the WCL network. All the service nodes, such as Firewalls, CGNAT, CDN Caches, DNS, Load Balancer, and Parental Control boxes and so on reside in this layer.

https://t.me/learningnets

This layer runs IBGP with ATELCO's internal network (IP/MPLS Service Layer) and runs EBGP with the Worldwide Connectivity Layer (WCL).
ATELCO runs OSPF in the IGW Layer same as in the IP/MPLS Layer. OSPF is used to advertise Route Reflectors and Border Gateway addresses to the Core Network and also advertise the Service Layer Route Reflectors (which reside in the Core) to the IGW Layer.

In order to understand how all these services that reside in this layer are consumed by the users, let's remember how physically IGW POP locations are connected to the network.
IGW POP locations are in the Datacenter. ATELCO has four Datacenters, one in each region in the country.

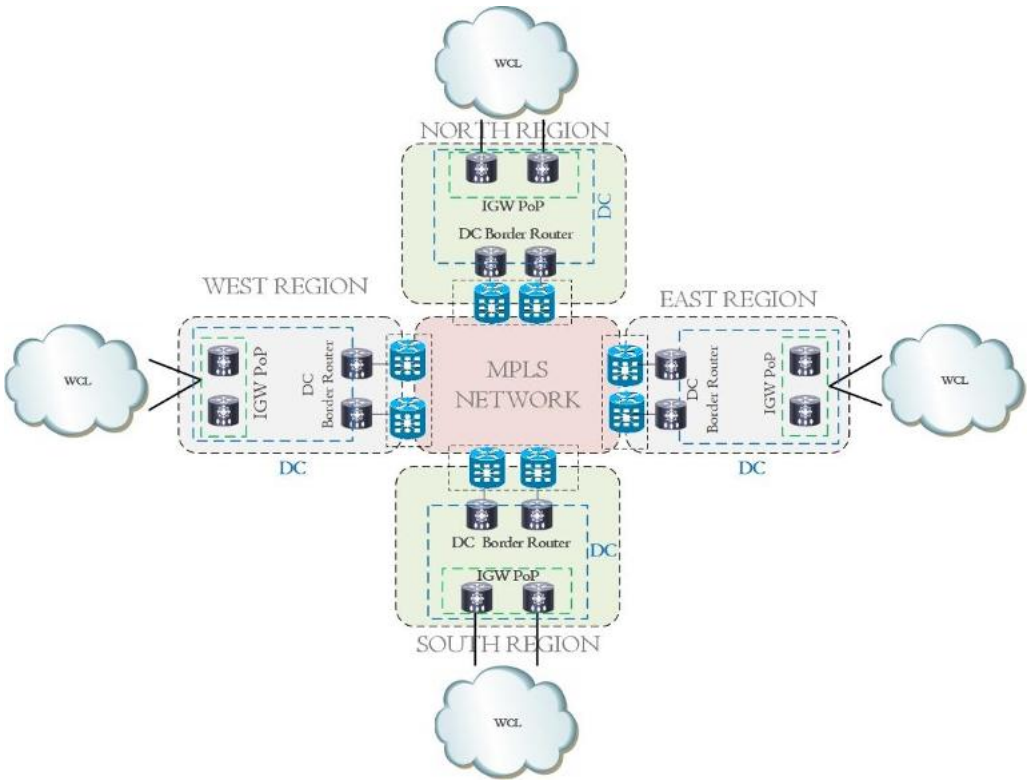


Figure 8-13 *IGW POPs and Datacenter Interconnections over IP/MPLS CORE/Backbone*

In ATELCO's network, Datacenters are connected to each other over IP/MPLS Multi Service network. Datacenter Border Routers are connected to the IP/MPLS Core Routers, thus the users or the devices from each region can access to any other regional Datacenter when it is required.

https://t.me/learningnets

In some ISPs, Datacenters are connected to each other via separate Transport network. This provides better control to the Operator as the Datacenter Interconnect traffic is separated than the Users to Datacenter traffic. But this increases the total cost of ownership of the Datacenter interconnectivity.

Services in ATELCO Network

ATELCO provides Fixed and Mobile Services. As a definition, fixed service typically refers to all of the wired networks that are used for communication services. Communication services usually are HSI (High Speed Internet), VPN, Internet, Voice and Video services. Mobile Service is a transmission service to the users of wireless devices through RF (Radio Frequency) signals, rather than through end to end wire communication.

Fixed Services provided by ATELCO are:

- ADSL
- VDSL v1
- VDSL v2
- FTTH
- Business/Corporate Customers

Mobile Services provided by ATELCO are:

- 3G
- LTE

Earlier in the book, these technologies were explained in detail, thus in this chapter we will explain ATELCO's services deployments and how other ISPs design these services in their network. It was mentioned in the scenario that ATELCO has their BRAS/BNG function in the CORE POP locations, connected to the Aggregation Layer devices.

BNG devices create logically an Edge Layer of the ISP networks, same as on ATELCO's network. It is the default gateway for the DSL and FTTX users, thus it is a very critical node in the network. Both XDSL and FTTX services use the same BNG for the subscriber access control and gateway functionality. BNG/BRAS is used in both XDSL and FTTx services. BNG/BRAS can be deployed as centralized or distributed.

First let's have a look at different BNG design options, after that the detail BNG design of ATELCO will be explained with their BNG topology.

https://t.me/learningnets

Distributed vs. Centralized BNG

In a centralized BNG deployment, BNG is deployed in the Core/Backbone POP location which is the core network of the ISP.

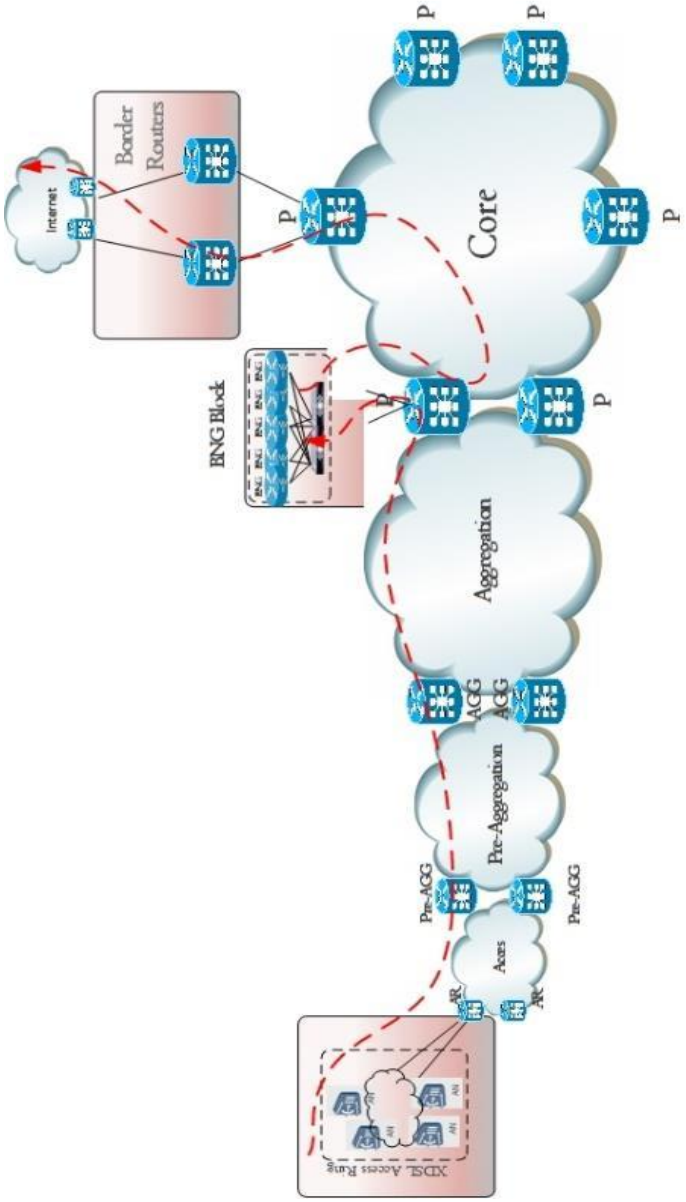


Figure 8-14 *Centralized BNG Deployment*

https://t.me/learningnets

As it is shown in figure 8-14, in a centralized BNG deployment, BNGs are deployed at the CORE layer. For the Fixed service (XDSL, FTTX) residential customer Internet traffic, traffic comes to the Core Layer up to the BNG, and then traffic is sent from BNG to the IGW. BNGs receive default route through the Core routers as the Internet Gateway Border routers are set as next hop.

In a distributed BNG deployment model, BNG is deployed at the Access POP locations or at the Aggregation POP locations.

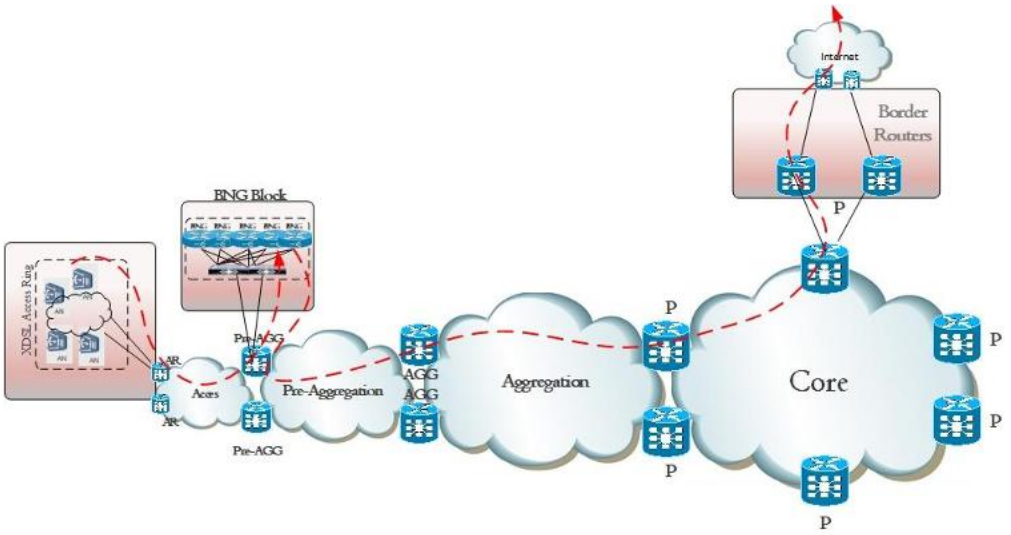


Figure 8-15 *Distributed BNG Deployment*

As the number increases, distributed BNG's can be expensive to deploy but in the case of a failure, since each BNG will have less amount of connection, the number of subscribers who are affected from the failure would be less as well.

As you can see in figure 8-15, the BNG's are placed into the Access network (Access POP) locations. When BNG's are used in the Access network, it is called Distributed BNG model.

In the Distributed BNG model, there might be hundreds of small BNG's in the network, so the operation might be harder to manage compared to the Centralized BNG architecture, but the advantage is small Blast Radius, which means when there is a BNG failure, much less users are affected compared to the Centralize BNG model.

https://t.me/learningnets

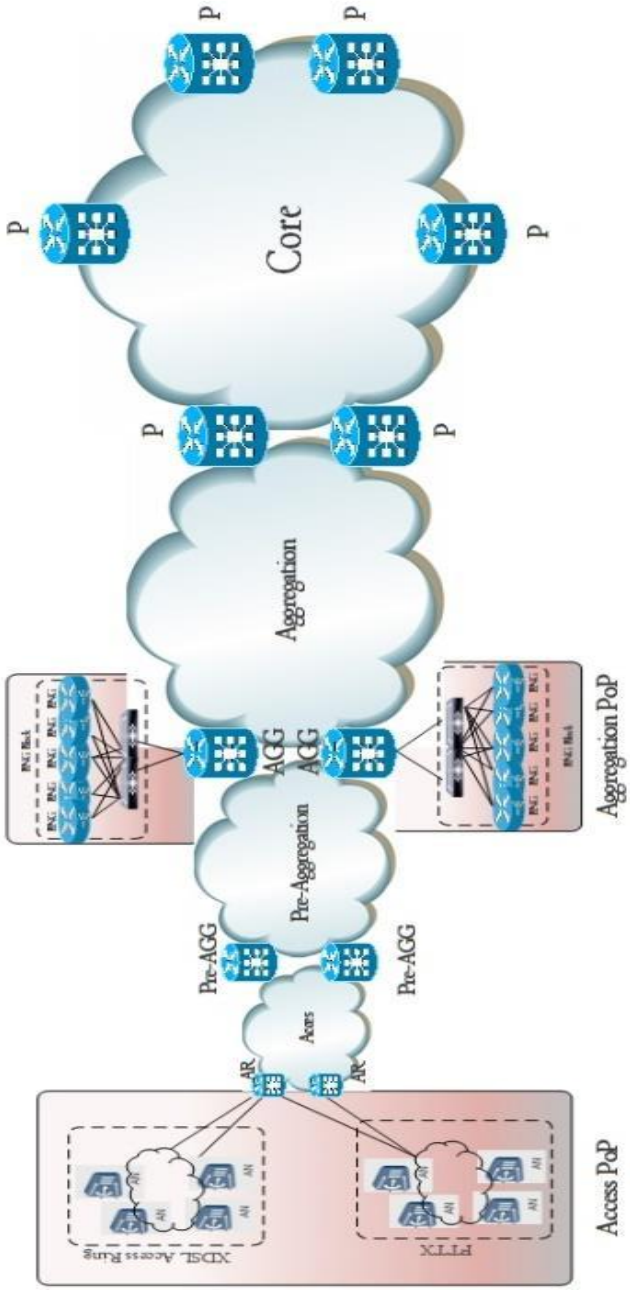


Figure 8-16 *BNG Connectivity in ATELCO Network*

https://t.me/learningnets

In ATELCO's network, there are in total 40 Aggregation POP locations, which are located in four (North, South, East, And West) regions. In each POP, they have 2 BNG's which terminate the user sessions. Their BNG deployment is based on the distributed BNG deployment.

Compared to deploying BNGs in the Access layer, ATELCO has less number of BNG's, but compared to deploying BNGs in the Core layer, they have much more BNGs. Thus we can say probably it is a hybrid model, not exactly centralized or distributed. The Centralized BNG model is more as a scale up model, which large boxes are used that have the highest redundancy features, while in the distributed model, it's more a scale out model which a large number of smaller boxes with less redundancy features (compared to centralized model) are used.

Also, for the BNG feature license, in the distributed model, more unused licenses are seen compared to the centralized model. In the centralized model, based on the traffic flow that all the users PPPOE sessions are terminated on the Central BNG boxes, less number of unused licenses will remain. But in the distributed model, for each box, based on the license number and the number of concurrent users, more unused licenses might be seen.

Table 8-3 *Centralized and Distributed BNG compression table*

Design Type	**Device**	**Location**	**PPPOE session termination**	**Unused License**	**Blast Radius**
Centralized BNG (Type-1)	Single BOX/2 Box for redundancy	SP DC site	On the single box, 2'nd box as redundant/load sharing	Very Less	High
Centralized BNG (Type-2)	Large number of small BNG boxes	SP DC site	users of each city are terminated on separate appliance	Less	medium
Distributed BNG (Type-1)	Regional Based appliance	AGG Locations	Separation based on customers of each region	More	small
Distributed BNG (Type-2)	City Based appliance	Pre-AGG locations	based on each city	Much more	Very small

https://t.me/learningnets

Design Decisions in the Three Parts of ATELCO's Network

In this section, detail explanation of the protocols and the technologies which are used in ATELCO's network will be provided. Also, alternative methods which are used by other Service Providers will be explained. Benefits and the drawbacks of using each protocol and technology will be covered.

1. WCL (Worldwide Connectivity Layer)
2. Internet Gateway/Shared Services Layer
3. IP/MPLS Multi Service Network

In the scenario section, it is shared that ATELCO carries FTTX, XDSL, 3G, LTE, MPLS Layer 2 VPN Point to Point, VPLS and MPLS Layer 3 VPN Traffic of their customers.

Design Decisions in the WCL (World-Wide Connectivity) Layer

In the WCL – World-Wide Connectivity Layer, ATELCO has EBGP connections with other companies. There is no other protocol in this layer. In the Global Internet, routing is done only via BGP (Border Gateway Protocol). Interior Gateway Protocols, such as OSPF, IS-IS or EIGRP are not used. There are many BGP Communities that are supported by the EBGP neighbors.

For example, ATELCO enables BGP Graceful Shutdown community with its BGP neighbors; IP Transit Customers, IP Transit Providers and Settlement Free Peers, to avoid packet loss during the maintenance activities. From the other technologies point of view, Layer 2 technologies at the IXP (Internet Exchange Point) are used such as ARP or Spanning Tree related configurations but we will not cover the details of these technologies.

Design Decisions in the IGW Layer

In the IGW Layer, two protocols are important for ATELCO. These are IGP and BGP. ATELCO runs OSPF as an IGP in the IGW Layer same as in the IP/MPLS Layer. OSPF is used to advertise Route Reflectors and Border Gateway addresses which reside in the IGW to the CORE Network and also advertise Service Layer Route Reflectors which reside in the CORE Network to the IGW Layer.

OSPF design in the IGW layer is very simple. OSPF Non-Backbone area with the same CORE network Process ID, but with different area ID is used in ATELCO's IGW network.
But from the BGP design point of view, there are so many design decisions that we need to understand in ATELCO's network as well as in all ISP networks.

As explained earlier in the previous chapter, ATELCO has 4 Internet exit points, located at the IGW POP sites in the North, South, East and West region in the country, which connects to the transit providers. Transit Provider X with AS1, Transit Provider Y with AS2 and Transit Provider

https://t.me/learningnets

Z with AS3 are the 3 international Transit Providers which provide Internet services to ATELCO's network.

The North Region is connected to all 3 Transit Providers (X,Y,Z) , while the other 3 regions (South, East, West) are connected to only 2 of the Transit Providers (X,Y).

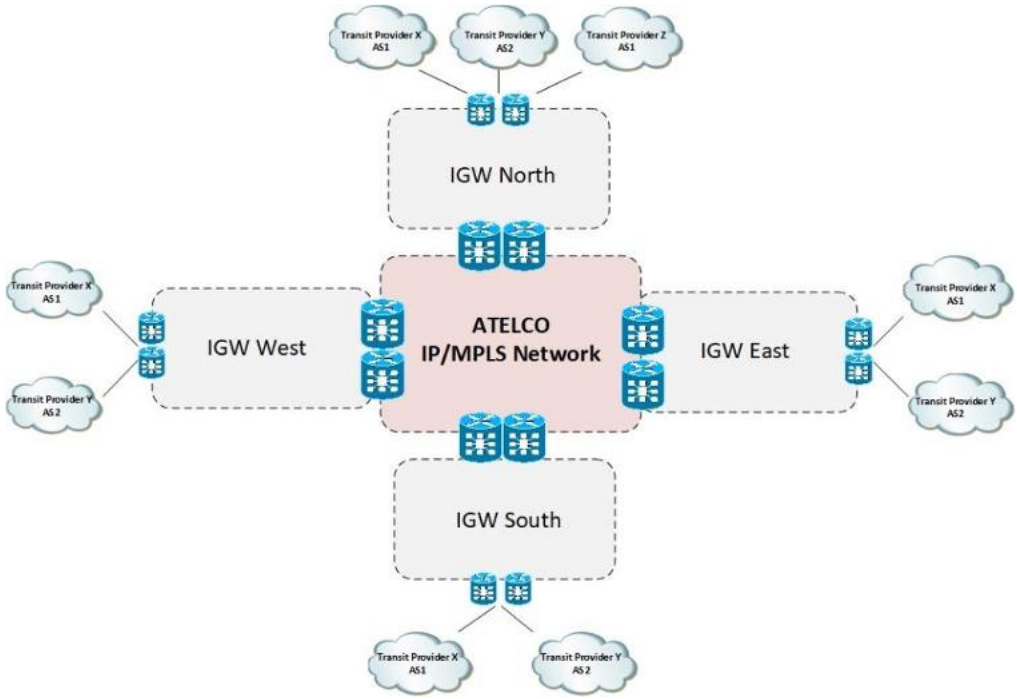


Figure 8-17 *ATELCO IGW Internet Connections – IP Transit Providers*

ATELCO is providing its Internet services to its customers from its 4 IGW POP site regions. All the 3 exit point peers, which are transit providers X, Y, Z are advertising their full Internet routing table to ATELCO.

As mentioned before, ATELCO provides fixed and mobile services such as xDSL, FTTX and LTE. ATELCO is providing these telecom services through various transport layers such as Layer2/Layer3 VPN, VPWS, VPLS, L3VPN, etc. For the customers traffic that is destined outside ATELCOs network (e.g. customers' Internet traffic), ATELCO's strategy is to send its customers packets to the Internet through the shortest path and as soon as possible, with the less hops inside its network.

This policy is done by sending the packet to the nearest exit point which is usually based on IGP metric. The technical reason behind ATELCO's policy is to use the minimum number of internal resources on their network (also known as Hot Potato Routing).

https://t.me/learningnets

As was explained in the previous chapter in the book, ATELCO is using Route Reflectors for its IBGP sessions inside the IGW network. Based on the functionality of Route Reflectors, using RR's in the network usually does not match the Hot Potato Routing strategy of Service Providers networks.

As a reminder, the Route reflectors choose the best path to the exit point based on their perspective, and not the client's perspective. A path to the exit point of the network for a certain prefix can be optimal for the Route Reflector based on its lowest IGP metric to the exit point, but this might not be true from the client's perspective. Route Reflectors only advertise one path as their best path for a prefix and don't advertise any other paths to their clients.

With this Route Reflectors behavior which removes additional BGP advertisements to the control plane of its clients, an issue of suboptimal routing may occur for Route Reflector clients. This is because the Route Reflector client will not have all the available routes and it cannot compare the IGP metric of every path in order to determine the shortest path.

Sub optimality in reflecting the path from the RR to the clients usually happens when the Route Reflector is not topologically near its clients. This sub optimality is more seen when RRs are not in the forwarding path, especially in virtual RR's that are completely out of path.

Route Reflectors work in the following way:

- Route selection is based on their point of view
- RRs propagate only the best path over their sessions by hiding other paths
- This might not be the best path according to the client's point of view
- RRs run best path algorithm and advertise only one update to their clients, which may result in suboptimal routing
- RR's are usually deployed based on exit points in the network

Route Reflectors use the same best-path selection process as normal BGP speakers do. When receiving the same prefix coming from multiple peers, the tiebreaker decision process is done:

1. Highest LOCAL_PREF
2. Locally originated via network/aggregate or redistributed via IGP.
3. Shortest AS-Path
4. Lowest origin type
5. Lowest MED
6. EBGP paths over IBGP
7. Path with the lowest IGP metric to the BGP next hop.

If all the steps before the 7th step are equal, then step 7 will be the deciding factor for the best path for the Route Reflector. So, the preferred path will be the lowest IGP metric to the BGP next hop. By default, Route Reflector's only advertise the best path to their clients, so in case of the

https://t.me/learningnets

tiebreaker explained above, the traffic will be send to the exit point with the lowest cost/shortest path possible.

In-band Route Reflectors usually have a better view of the IGP topology of the network than out-of-band Route Reflectors, so they can better advertise optimal best paths to their clients. Two design decisions were considered by ATELCO for their BGP Route Reflector design.

1. BGP Route Reflectors in North IGW POP
2. BGP Route Reflectors in each individual POP

Let's look at the first solution.

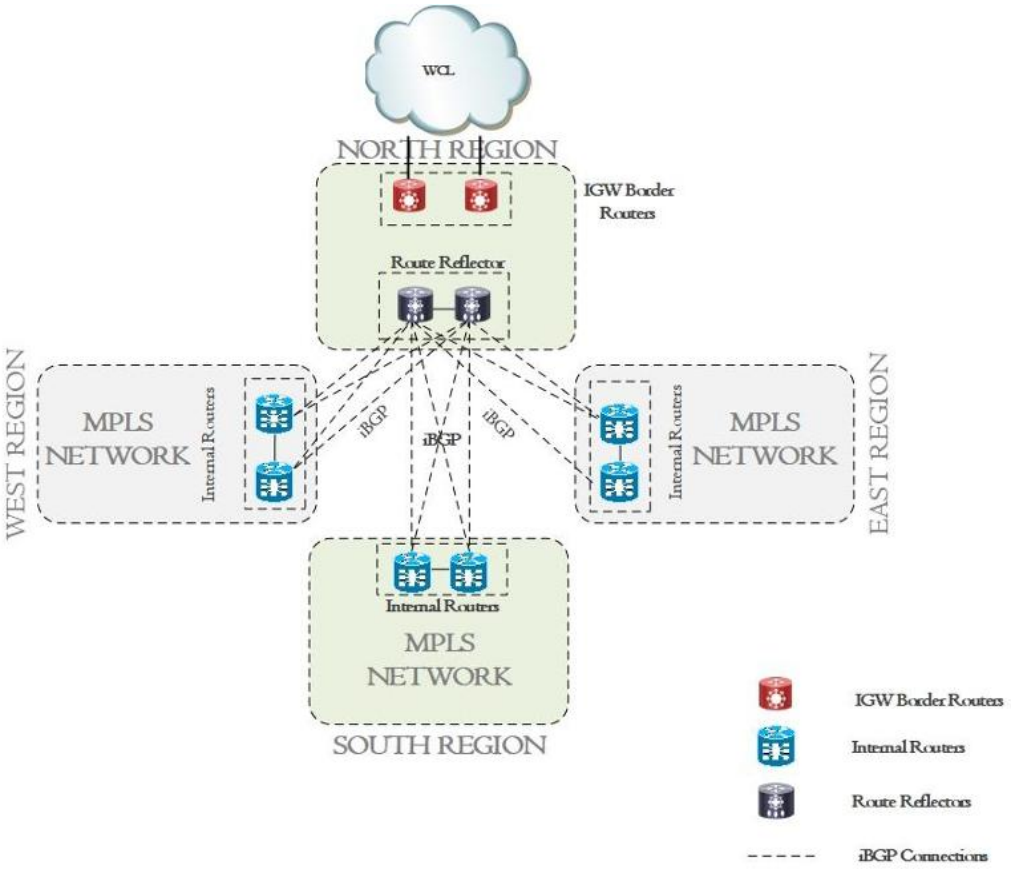


Figure 8-18 *BGP Route Reflectors in North IGW POP Only*

https://t.me/learningnets

The North IGW POP site is the backup of all other IGW POP sites, so more bandwidth capacity is provided through the North IGW POP site.

One of the design architects in ATELCO suggested that placing BGP Route Reflector would reduce the number of IBGP sessions in the network. But this design would create sub optimality for the Internet traffic as for the routers at the other regions would send the traffic to the North region, instead of sending it to their local Internet exit. The reason is that the North region RR selects the North Border Routers as the best exit as and advertises these routers as best path to all other regions.

Centralized BGP Route Reflector placement doesn't always create sub optimal routing, but there are two approaches to overcome sub optimal routing issue when BGP Route Reflector is placed at the centralized location such as placing BGP RR only in the North Region of ATELCO network.

1. BGP Add-Path
2. BGP ORR (Optimal Route Reflection)

Route Reflector Design based on advertising multiple paths (ADD_PATH) to clients for the same prefix

As explained earlier, Route Reflectors create sub optimal routing for their clients by removing some BGP next hop for a given prefix from the control plane. ATELCO's consideration in its design procedure was to use BGP Add-Path to add more information back into the control plane of the Route Reflectors clients by sending more routes for the exit point of the network to them.

BGP Add-Path can be used to send N number of exit point paths to Route Reflector clients. With having multiple paths, clients can select the best path based on their own perspective, so they can get a better control over the choice of exit points of the network. But actually, there is no guarantee that any of those multiple paths would be the most optimal path because not all possible alternative routes might be sent. All the paths need to be sent to provide guaranteed optimality.

To better understand BGP Add-Path, let's have a look at figure 8-19. The customer is sending the prefixes to both PE1 and PE2. Our goal is to send both PE1 and PE2 as the BGP next-hop to PE3.

Without Add-Path, BGP Route Reflector – RR1 would only send one of the PE routers as the best path. With BGP Add-Path capability, RR1 can send both PE1 and PE2 as best paths for the customer prefixes to PE3. PE3 after receiving both paths, can use the multiple paths for BGP PIC, Optimal routing or IBGP Multipath purposes.

https://t.me/learningnets

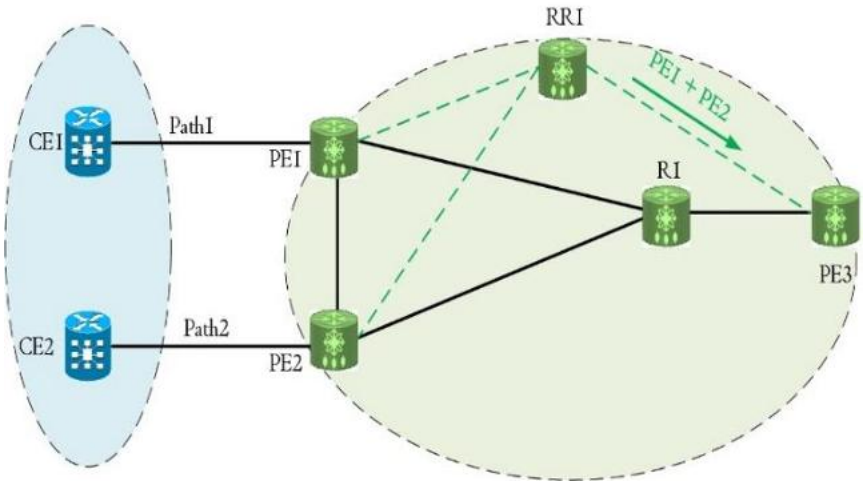


Figure 8-19 *BGP Add-Path*

Reducing routing churns via oscillation, faster convergence, better load sharing and availability are some advantages of using this solution. Improved Path Diversity is another benefit from this solution, which can bring effective BGP level load and fast connectivity restoration (ex. BGP PIC - Prefix Independent Convergence for faster convergence).

By expanding the network to more exit point peering connections, which can result in getting same routes from more peers (especially when receiving full routing tables), more paths and lots of updates are advertised to clients, so the number of BGP announcements will increase for Route Reflector clients, which might lead to significant memory problems and can result in increasing memory requirements on end devices.

Introducing a large number of BGP states to all routers will create a lot of entry on the Route Reflector clients BGP Table. Some clients might not support Add-Path, others that support, might not have enough capacity.

Add-path is a BGP capability, which means it needs to be agreed between RR and RR Client. Add-path capability may require software or hardware upgrade, not only on the Route Reflectors, but also on the Route Reflector Clients. Upgrading both RR and RR Client might take so much time to migrate the BGP Software to one which supports BGP Add-Path feature as there might be so many Edge devices in the network that acts as RR Client.

Based on these issues, this design type was considered as a non-scalable solution for ATELCO's network.

https://t.me/learningnets

BGP ORR (Optimal Route Reflection)

Based on this solution, the RR will do the optimal path selection based on each client's point of view. It runs SPF calculation with their clients as the root of the tree and calculates the cost to the BGP next-hop based on this view. So, the Route Reflectors location would be independent from the selection process of the best-path. So, for the same Prefix, each ingress BGP border router can have a different exit point to the transit providers.

Link-state routing protocol is required in the network for the Route Reflectors to have a complete view of the network topology based on the IGP perspective. No changes are required to be done by the clients.

ORR is applicable only when BGP path selection algorithm is based on IGP metric to BGP next hop, so the path will be the lowest metric for getting the Internet traffic out of the network as soon as possible.

This solution is not an alternative to BGP Add-Path or other methods for Path Diversity, though it is an alternative to provide optimal routing. It can be used together to improve the quality of multiple advertisements, to propagate the route that can be the best path. Also, it can add resiliency and faster re-convergence for the network.

For example, by receiving 4 paths from exit point peers across the network, it will choose the best path plus the 3 other paths based on the IGP cost. So, it's a true way to add resiliency through add-path.

With ORR, at the 1st step, the topology data is acquired via ISIS, OSPF, or BGP-LS. The Route Reflector will then have the entire IGP Topology, so it can run its own computations (SPF) with the client as the root. There could be as many rSPFs (Reverse SPF) running based on the number of RR clients, which can increase the CPU load on the RR.

So, a separate RIB for each of the clients/groups of clients is kept by the RR. BGP NLRI and next-hop changes trigger ORR SPF calculations. Based on each next-hop change, the SPF calculation is triggered on the Route Reflector.

The Route Reflectors should have complete IGP view of the network topology for ORR, so a link-state routing protocol is required to be used in the network. OSPF/IS-IS can be used to build the IGP topology information. In the case of single area OSPF area 0, the RR participating in the IGP process has access to the entire LSDB, so it can determine the total IGP costs between each pair of RR-Clients and the IGW (exit points).

https://t.me/learningnets

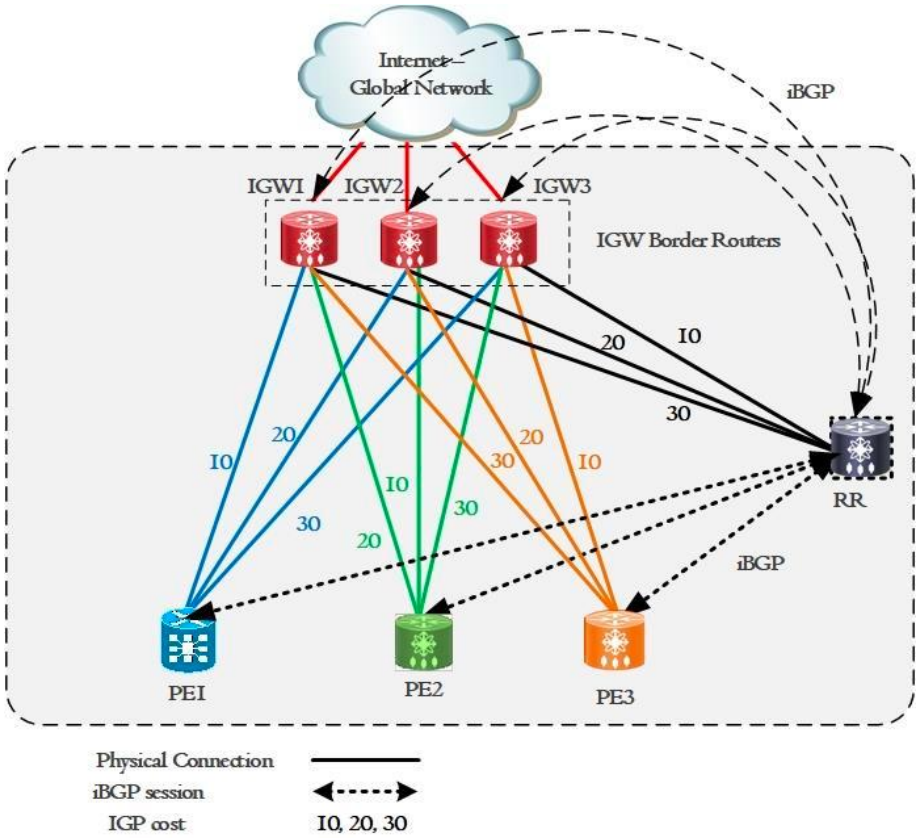


Figure 8-20 *BGP Optimal Route Reflection (ORR)*

In figure 8-20, without ORR, RR would send IGW3 as the best path to all PE devices, because IGP cost of RR to IGW3 is the shortest path. When ORR is deployed, RR sends IGW1 to PE1, IGW2 to PE2, and IGW3 to PE3 as best paths. This provides optimal routing for each and every RR Client. IGP is great for link state distribution within a routing domain or an autonomous system but for link state distribution across routing domains, EGP is required. BGP-LS provides such capability at high scale by carrying the link state information from IGP protocols as part of BGP protocol messages.

Route Reflectors keep track of which route it has sent to each client, so it can resend a new route based on changes in the network topology (BGP/IGP changes reachability). The Route Reflector function is 1 process per route but the ORR function is 1 process per route per client router.

ORR brings the flexibility to place the Route Reflector anywhere in the topology, and brings much better output when is used with ADD-PATH.

https://t.me/learningnets

Different types of ORR (Optimal Route Reflection) Deployments

1. Optimal BGP path selection based on client IGP perspective

   Optimal BGP path selection is done Based on the Client's IGP Perspective, and not the RR's IGP perspective. To reduce the SPF calculation overhead on the RR, Optimization such as partial and incremental SPF can be used.

2. Optimal BGP Path Selection Based on Policy

   This solution is based on User Defined Policy. The clients will always send traffic to a specific exit point of the network regardless of how the topology looks like. For any BGP UPDATE, it first calculates shortest path from the perspective of a virtual IGP location, then policy based path selection, and finally IGP based path selection. For example, one of the Policy methods can be used for the customers who pay more and gets SLA (Can be classified and marked with BGP Communities), so the traffic can be sent to particular Internet region and particular Transit Operator, instead of doing Hot Potato routing.

ATELCO at first considered placing IGW BGP Route Reflectors in a Centralized location, at the North region IGW POP with or without BGP Add-Path and/or BGP ORR as well, but as the last decision, they decided to deploy two IGW Internet Based BGP Route Reflectors in each region. This gives geographical redundancy, as well as provides optimal routing.

Placing BGP Route Reflector in the IGW POP locations in each region would increase the number of BGP Route Reflectors in the network, but provides optimal routing.

https://t.me/learningnets

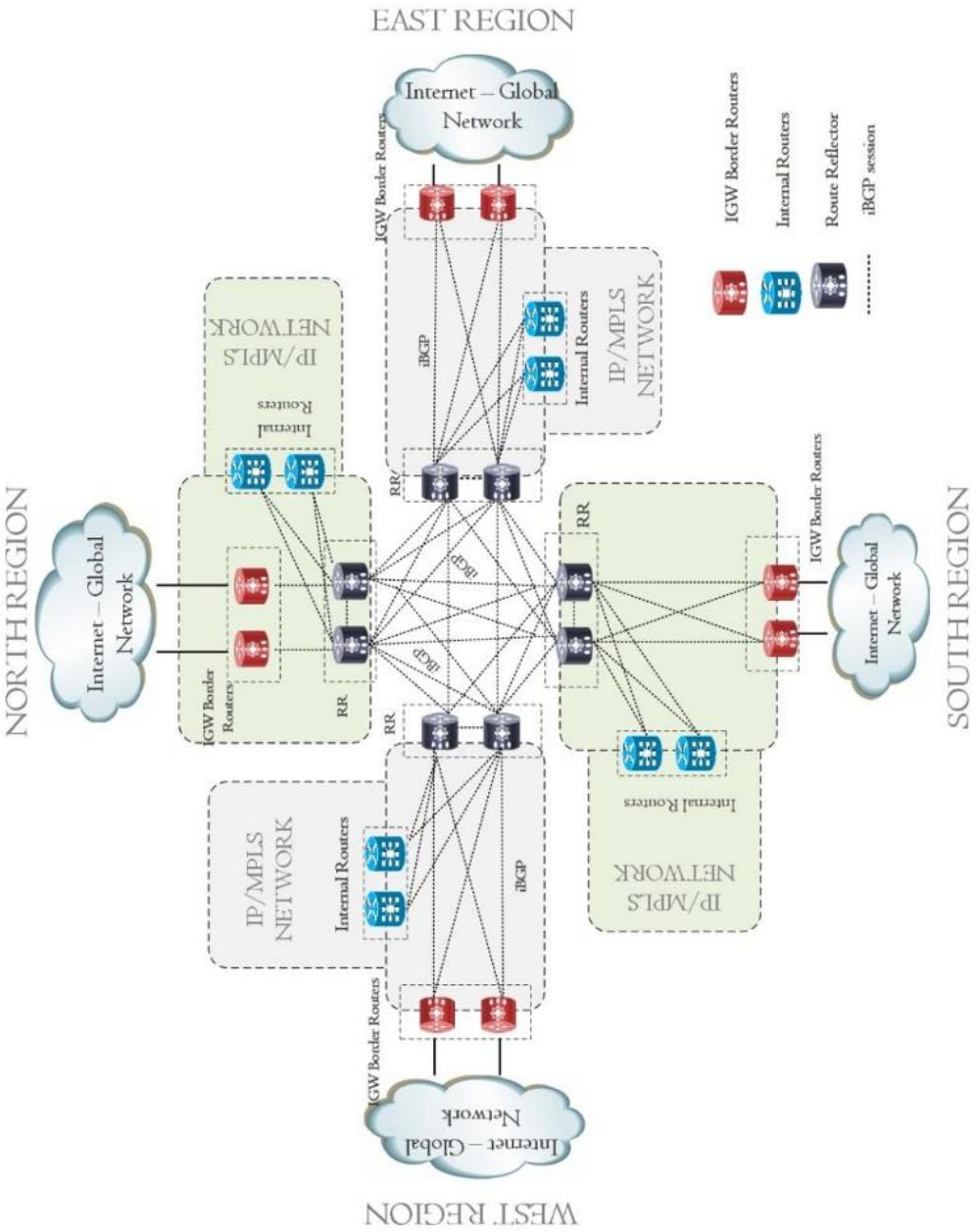


Figure 8-21 *BGP Route Reflectors in each individual POP*
High quality picture can be seen at www.orhanergun.net/SPBook/RR.jpg

https://t.me/learningnets

By placing BGP RR in each IGW POP location, each regional RR selects that particular regional IGW Border Router as the best path for the Internet traffic and sends this exit path to its internal regional routers, so optimal routing is achieved. In figure 8-21, Internal Routers are in the IP/MPLS network.

As shown in figure 8-22, the Pre-Aggregation Layer Routers are the Clients of the Service Route Reflectors, which are located in the Core POP sites. The Service Route Reflectors are the clients of the IGW Route Reflectors. Internet and VPN prefixes are carried through the Service Route Reflectors. Also, every IGW Border Router is the Route Reflector Client of the IGW Route Reflectors.

When IGW RR receives the Internet prefixes from the IGW Border Router, it reflects it to the Internal Routers (Service Route Reflectors) which require Internet Prefixes. ATELCO's IP/MPLS network is connected to the IGW network via CORE (P) routers.

Route Reflectors in the IGW Layer are used to receive Internet prefixes from IGW Border Routers and advertise it to each other and to the Service Route Reflectors which are located in the Core layer of the IP/MPLS network. After Service RR receives the Internet routes, it can advertise the prefixes to wherever it is necessary, such as to the Business Customer PE routers which are terminated at the Pre-Aggregation Layer.

Between the IGW BR (Border Router), IGW RR, Service RR and the Pre-Aggregation Layer Routers (User PE – UPE devices), there is an IBGP session as all of them are using the same BGP AS number which is 65000.

End to end Internet Route Advertisement can be seen from figure 8-22.

https://t.me/learningnets

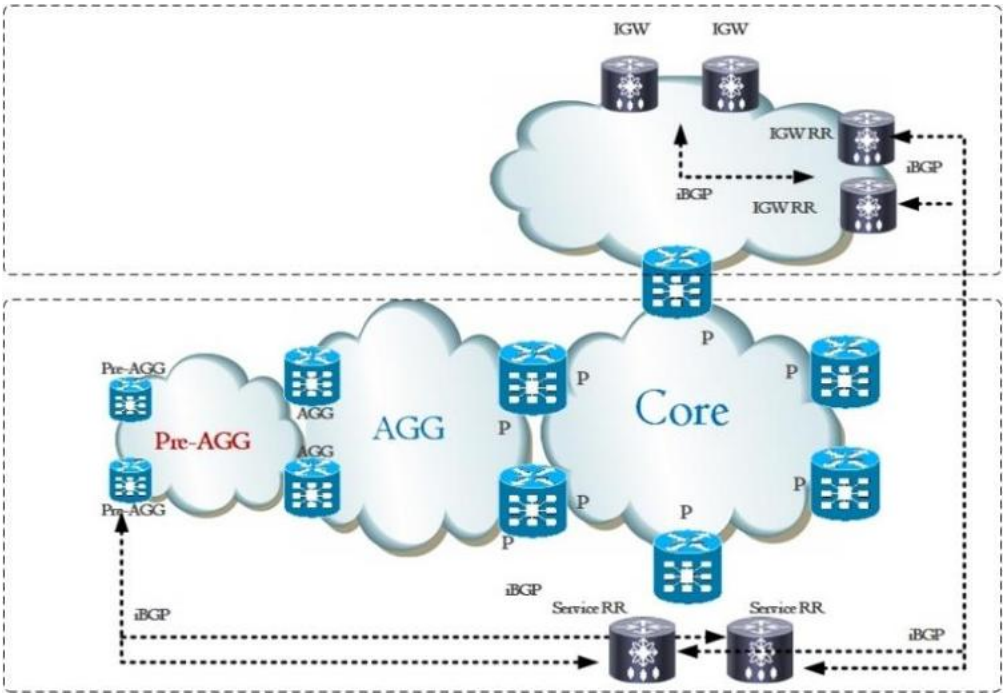


Figure 8-22 *End to End Route Distribution for Internet Prefixes to the Corporate Customers*

In the IGW Layer, Layer 2 Spanning Tree Protocol is used on the Service Switches which terminate devices such as Firewall, DPI, Parental Control, Load Balancers and Cache Engines.

ATELCO and other ISPs deploy MPLS L2 VPN and MPLS L3 VPN for their internal purposes as well. For example, they use MPLS Layer 2 VPN for Datacenter Interconnection purpose and they use MPLS Layer 3 VPN for LTE X2 service.

https://t.me/learningnets

Design Decisions in the IP/MPLS Network

Let's have a look at ATELCO's IP/MPLS Multi Service Architecture.

ATELCO is using Seamless/Unified MPLS Architecture. This architecture is widely known by network engineers who are working in Mobile Operators. ATELCO as a converged Internet Service Provider which provides both Fixed and Mobile Services followed the same approach and deployed Seamless/Unified MPLS architecture. Before the specific deployment of ATELCO's Seamless MPLS network, let's have a look at how this architecture works. Why it is used, what are the advantages and the disadvantages and so on.

SEAMLESS/UNIFIED MPLS

Seamless/Unified MPLS provides the architectural baseline for creating a scalable, resilient, and manageable network infrastructure. Seamless MPLS architecture is used to create very large-scale MPLS networks. It reduces the operational touch points for service creation.

Seamless MPLS architecture is best suited to the very large scale Service Provider or Mobile Operator networks that have 1000s or 10s of thousands of access nodes and very large aggregation networks. IP traffic increases rapidly due to video, cloud, mobile Internet, multimedia services and so on. To cope with the growth rate of IP Traffic, capacity should be increased but at the same time operational simplicity should be maintained.

Since there might be 1000s to 10s of thousands of devices in the Access, Aggregation and Core network of large Service Providers or Mobile Operators; extending MPLS into the Access networks comes with the main problems:

- Large flat routing designs adversely affect the stability and convergence time of the IGP.
- Resource problems on the low-end devices or some access nodes in the large-scale networks.

In Seamless MPLS, the Access, Aggregation and Core networks are portioned in different IP/MPLS domains. Segmentation between the Aggregation and the Core networks can be based on Single AS (Autonomous System) design or Multi AS. Partitioning Access, Aggregation and Core network layers into isolated IGP domains helps reduce the size of routing and forwarding tables on individual routers in these domains. This provides better stability and faster convergence.

In Seamless MPLS, LDP is used for label distribution to build MPLS LSPs within each independent IGP domain. This enables a device inside an Access, Aggregation, or Core domain to have reachability via intra-domain LDP LSPs to any other device in the same domain.

Reachability across domains is achieved using RFC 3107 (BGP+Label). RFC 3107 is obsoleted by RFC 8277 in 2017. BGP is used as an Inter domain label distribution protocol. Hierarchical LSP is created with BGP. This allows the IGP link state database in each isolated domain to remain as small as possible, while all external reachability information are carried via BGP, which is designed to carry millions of routes.

https://t.me/learningnets

In Single AS Multi-Area Seamless MPLS design, IBGP labeled unicast is used to build inter-domain LSPs. In Multi-AS Seamless MPLS design, EBGP labeled unicast is used to build inter domain (Aggregation, Core domains) LSPs between the AS.

EBGP labeled unicast is used to extend the end-to-end LSP across the AS boundary. There are at least five different Seamless MPLS models based on the access type and network size. Network designers can use any of the following models based on the requirements.

These models are:

1. Flat LDP Aggregation and Core
2. Labeled BGP access with flat LDP Aggregation and Core
3. Labeled BGP Aggregation and Core
4. Labeled BGP Access, Aggregation and Core
5. Labeled BGP Aggregation and Core, IGP Redistribution into Access

https://t.me/learningnets

SEAMLESS MPLS DEPLOYMENT MODELS

1. Flat LDP Aggregation and Core Model

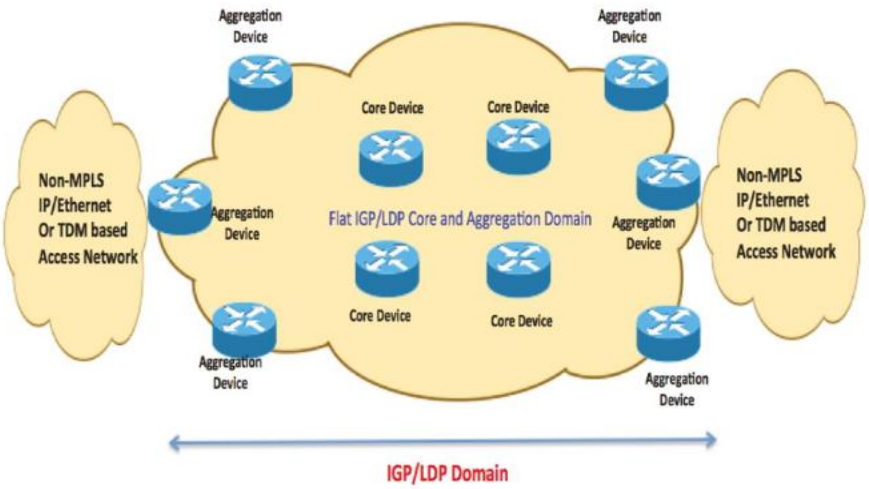


Figure 8-23 *Flat LDP in Aggregation & Core Network–Non-MPLS Access Model*

This Seamless MPLS model is applicable to small networks. Flat Aggregation and Core Network means, both Aggregation and Core networks are in a single IGP domain, for example in a single OSPF Area. In this model, the Access network can be; a non-MPLS IP/Ethernet or TDM based network. The network is usually small in size. There is no MPLS in the Access network. There is no Hierarchical BGP LSP as well.

Also, small-scale Aggregation network is assumed to be composed of Core and Aggregation nodes that are integrated in a single IGP/LDP domain consisting of less than 500 to 1000 nodes. Since there is no segmentation between the network layers, a flat LDP LSP provides end-to-end reachability across the network.

https://t.me/learningnets

2. Labeled BGP Access with flat LDP Aggregation and CORE Network Model

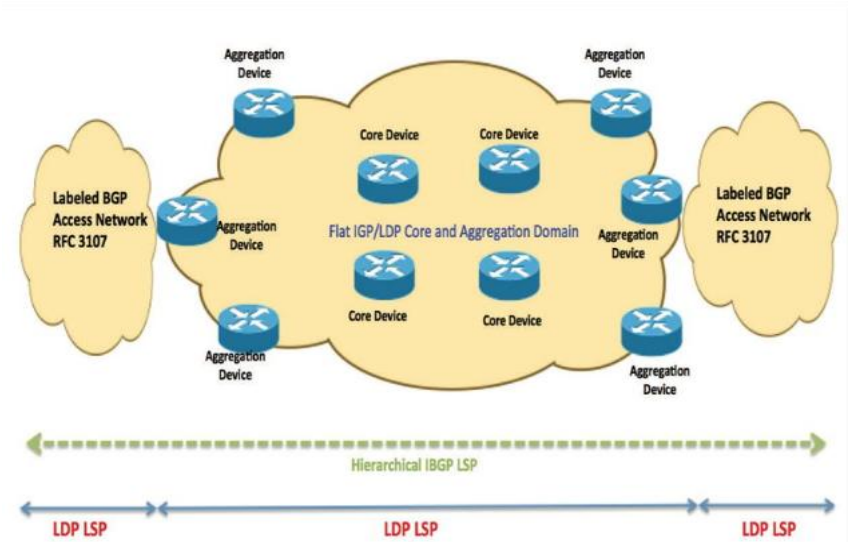


Figure 8-24 *Flat Aggregation and Core with Labeled BGP Access Model*

This model applies to small networks as well. Flat Aggregation and Core Network means, both Aggregation and Core networks are in a single IGP domain, for example a single OSPF Area. This type of architecture assumes an MPLS enabled Access network and small scale Aggregation and Core networks.

Small-scale Aggregation network is assumed to be composed of Core and Aggregation nodes that are integrated in a single IGP/LDP domain consisting of less than 500 to 1000 nodes. This model is very similar to Flat Aggregation and Core with Non-MPLS Access model, but in this model; Access network nodes also run IP/MPLS.

The Access network runs a separate IGP domain from the Aggregation and Core. The separation can be enabled either by making the Access network part of a different IGP area from the Aggregation and Core nodes, or by running a different IGP process on the Aggregation nodes.

The Access network is integrated with the Aggregation and Core network with labeled BGP LSPs, so the Aggregation nodes act as ABRs and perform BGP next-hop-self (NHS) function to extend the IBGP hierarchical LSP across the network.

https://t.me/learningnets

3. Labeled BGP in the Aggregation And Core Network Model

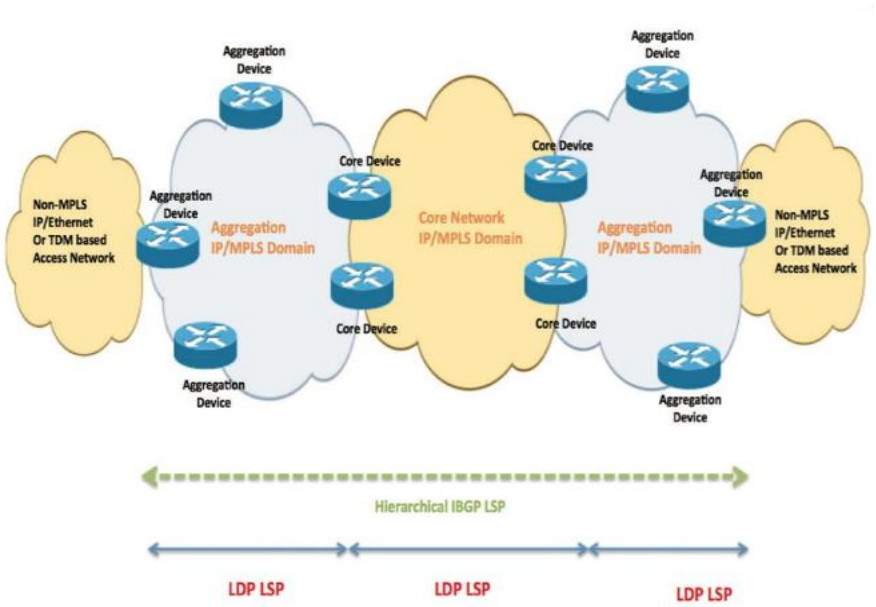


Figure 8-25 *Labeled BGP Aggregation and Core, Non-MPLS Access Model*

This model applies to medium to large scale networks. This model assumes a Non MPLS Access network and the fairly large Aggregation and Core networks. In this model, the network is organized by segmenting the Aggregation and Core networks into independent IGP domains.

The segmentation between the Core and Aggregation domains could be based on a Single-AS multi-area design or a Multi-AS. In the Single-AS multi-area option, the separation can be done by making the Aggregation network part of a different IGP area from the Core network, or by running different IGP processes on the Core ABR nodes. Access network can be based on IP, Ethernet, TDM or Microwave links. LDP is used to build intra-area LSP within each domain.

The Aggregation and Core domains are integrated with labeled BGP LSPs. In the Single-AS multi-area option, the Core devices as an ABR perform BGP NHS function to extend the IBGP hierarchical LSP across the Aggregation and Core domains.

https://t.me/learningnets

4. Labeled BGP Access, Aggregation And Core Model

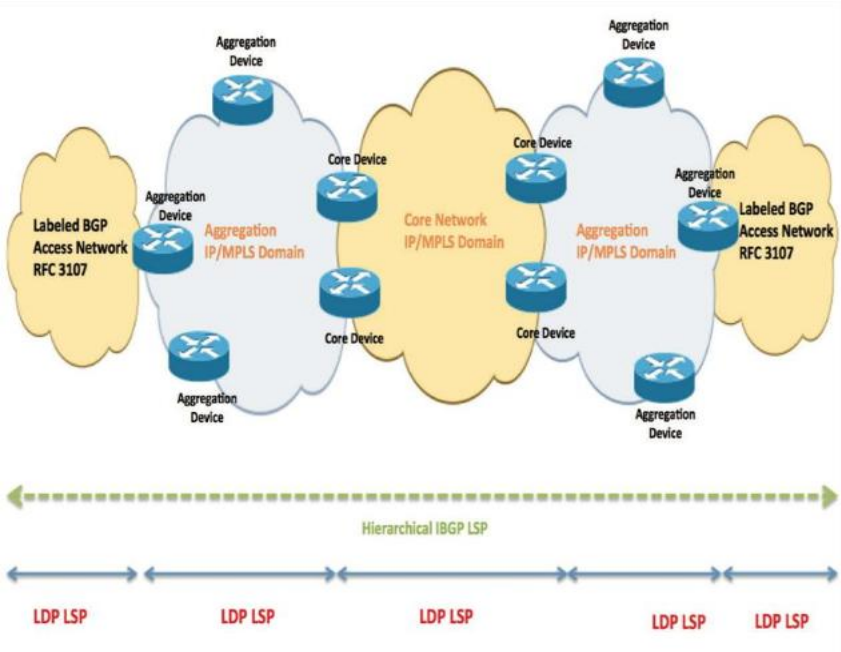


Figure 8-26 *Labeled BGP Access, Aggregation and Core Seamless MPLS model*

This model can be used in very large-scale networks. This model assumes an MPLS enabled Access network with large Aggregation and Core networks. The network infrastructure is organized by segmenting the Access, Aggregation and Core into independent IGP domains. The segmentation between the Access, Aggregation and Core network domains could be based on a Single-AS multi area design or a Multi-AS based design.

The difference of this model from the 'Labeled BGP Aggregation and Core' model is, Access network nodes run MPLS as well and RFC 3107, so BGP LU (Labeled BGP) is extended up to the Access network nodes.

https://t.me/learningnets

5. Labeled BGP Aggregation and Core with IGP Redistribution into Access Network Model

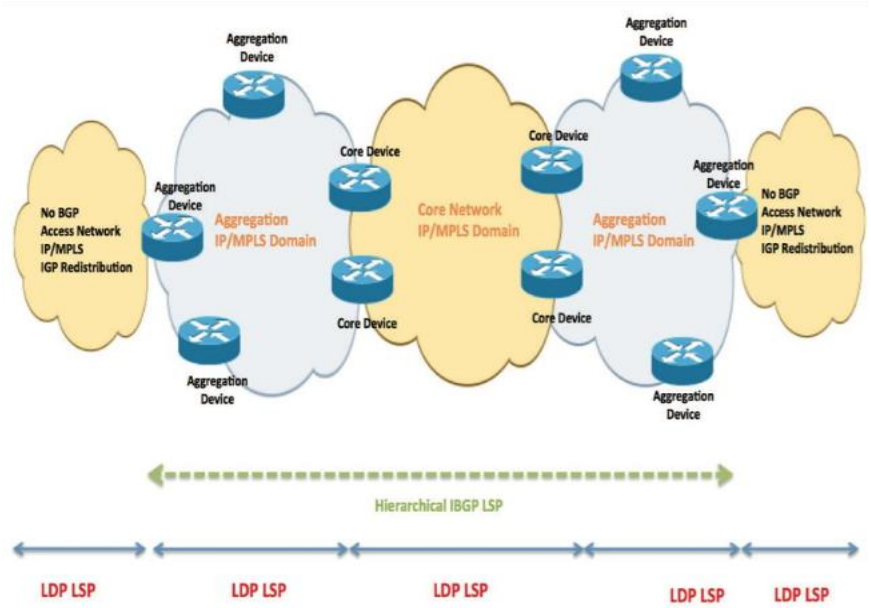


Figure 8-27 *Labeled BGP Aggregation and Core with IGP Redistribution into Access Network*

In this model, the Access network runs MPLS but not with labeled BGP. LDP LSPs are created in the Access network.
This model applies to very large networks. The network infrastructure organization in this architecture is similar to 'Labeled BGP Access, Aggregation and Core '. Redistribution is performed on the Aggregation nodes. Access network loopback addresses are redistributed into the network. There is no hierarchical end-to-end BGP LSP in this model.

ATELCO has Seamless MPLS architecture. Seamless MPLS is used currently only for the Mobile Service in ATELCO's network. As it was given in the scenario part, ATELCO is using BGP AS 65000 in IP/MPLS Multi Service Network, IGW and WCL layers.

Using different AS in the Backbone network and the Internet Gateway layer is commonly seen in today's ISP networks. But for simplicity, ATELCO has chosen to use the same AS number in its different parts of the network.

It's important to know that there are many Mobile Operators that use different BGP AS numbers in their Packet Core Network and IP Core Network. Also, they use different BGP AS's for their outside connectivity in IGW layer, that's why it is a common design having 2 or 3 different BGP AS numbers in the Mobile Network Operators.

https://t.me/learningnets

As it was explained before in the scenario section, ATELCO is not using BGP in the Access Layer devices for fixed services but they have BGP and MPLS all the way to the Access devices (Cell Site Routers) for the Mobile service.

They have MPLS in the Access Layer rings for the Mobile service. They only have Layer2 802.1q without MPLS or BGP for the DSLAM and GPON rings which means that for the fixed service, they have only Layer 2 in the Access network as it was mentioned many times so far.

Let's have a look at the detail of ATELCO's Seamless MPLS design and for each service, how the end to end packet flow happens in the network.

ATELCO's Seamless MPLS design

Mobile Service IGP, BGP and MPLS Design

In the Access, Pre-Aggregation, Aggregation and Core Domain, ATELCO is using OSPF as the IGP routing protocol. In the Core and Aggregation network, ATELCO has OSPF Process ID 10 with the same OSPF Area. This design places all Aggregation and Core Routers in the same IGP domain. Since there are much less Aggregation and Core Routers in the network, compared to Pre-AGG and Access, scalability is not a concern by placing Aggregation and CORE routers in the same OSPF Area.

As it can be seen from figure 8-28, Pre-Aggregation domain has different OSPF process than any other part of the network.

In order to exchange prefixes between different OSPF Process ID's, manual redistribution is required. Our goal is to carry the devices loopback which will be used for MPLS LSP, via BGP, and not with OSPF, thus IGP redistribution will not be enabled in ATELCO's network.

MPLS is enabled in the Access, Pre-Aggregation, Aggregation and CORE network. Within each layer (Access, Pre-AGG, AGG and CORE), LDP is used for intra domain MPLS LSP. BGP is used to connect the domains end to end.

Let's explain the above sentence more as it is very important to understand the scalability aspect of this design. Access routers BGP next-hop is the Pre-Aggregation routers, Pre-Aggregation routers BGP next-hop is the Aggregation routers. For the Aggregation routers, although they are within the same IGP domain with the core routers, their BGP next-hop is the CORE routers.

LDP based MPLS LSP is terminated at the BGP Next-hop. Thus, in figure 8-26, LDP LSP starting and termination point for each domain is shown separately.

BGP LSP starts from the CSR (Cell Site Router) and it is end to end between the CSR routers. But in many deployments, CSR has an interaction with another CSR in the LTE architecture for X2 traffic and CSR interacts with the Mobile Packet Core for the S1 traffic.

https://t.me/learningnets

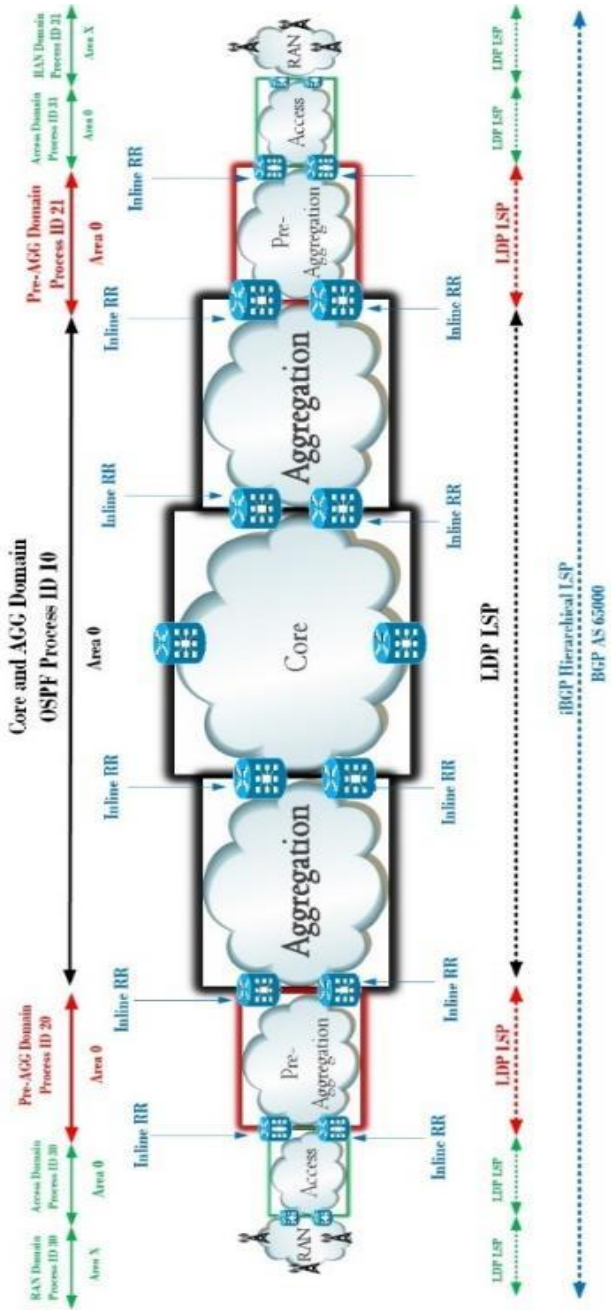


Figure 8-28 *Mobile Service IGP, BGP and MPLS Design*

https://t.me/learningnets

Fixed Service IGP, BGP and MPLS Design

The Difference between Mobile and Fixed Service from the IGP, BGP and MPLS point of view is only the Access part of their IP/MPLS network. Access is purely based on Layer 2 in ATELCO network. DSLAM or OLT rings are terminated at the AR (Access Router). As mentioned previously in this chapter, Mobile Cell Site Router devices are terminated at the Mobile Access Routers, while DSLAMs and OLTs are terminated at the Fixed Access routers.

This is done to separate the functionality of the devices. Different groups in the company manage different Access devices. Also, this separation provides fate sharing. When Mobile AR fails for any reason, it doesn't affect fixed services. For the Fixed service, Core and Aggregation networks are in the same Area and are using the same OSPF Process ID.

Pre-Aggregation layer has a different OSPF Process ID than Aggregation and CORE network of ATELCO. This is similar to Mobile Service in ATELCO's design. Different Pre-Aggregation domains have different OSPF Process ID's.

For example, in U.S, Pre-Aggregation layer routers in New York use a different OSPF Process ID than Pre-Aggregation layer routers in Chicago. Having MPLS as a unified control plane for both Fixed and the Mobile services provides other advantages to ATELCO network.

1. Troubleshooting is one of the advantages of Unified MPLS. There is no complex interaction such as PWHE (Pseudo wire Head End Termination), LSP Stitching and so on.

2. FRR is another advantage. IP FRR mechanisms such as LFA and Remote LFA can protect LDP traffic in case of failure and provide 50ms convergence time.

3. Less number of devices needs to be configured for the service. For example, if CSR needs to communicate with the Core device, only two points, Core device and CSR, need to be configured.

Carrying the loopback interface via BGP, instead of IGP is commonly known as RFC 3107 or BGP-LU (Labeled Unicast) approach. There is a newer architecture which is based on Segment Routing which aims removing BGP LU and simplifying service creation, providing more coverage for LDP based MPLS network through TI-LFA (Topology Independent Loop Free Alternate), removing LDP based LSPs from the network completely and so on. I encourage readers to look at Segment Routing based MPLS architecture in the Evolving Technologies in the Service Provider Networks chapter as Segment Routing brings many benefits to the Service Providers, Enterprises and the Mobile Operator networks.

https://t.me/learningnets

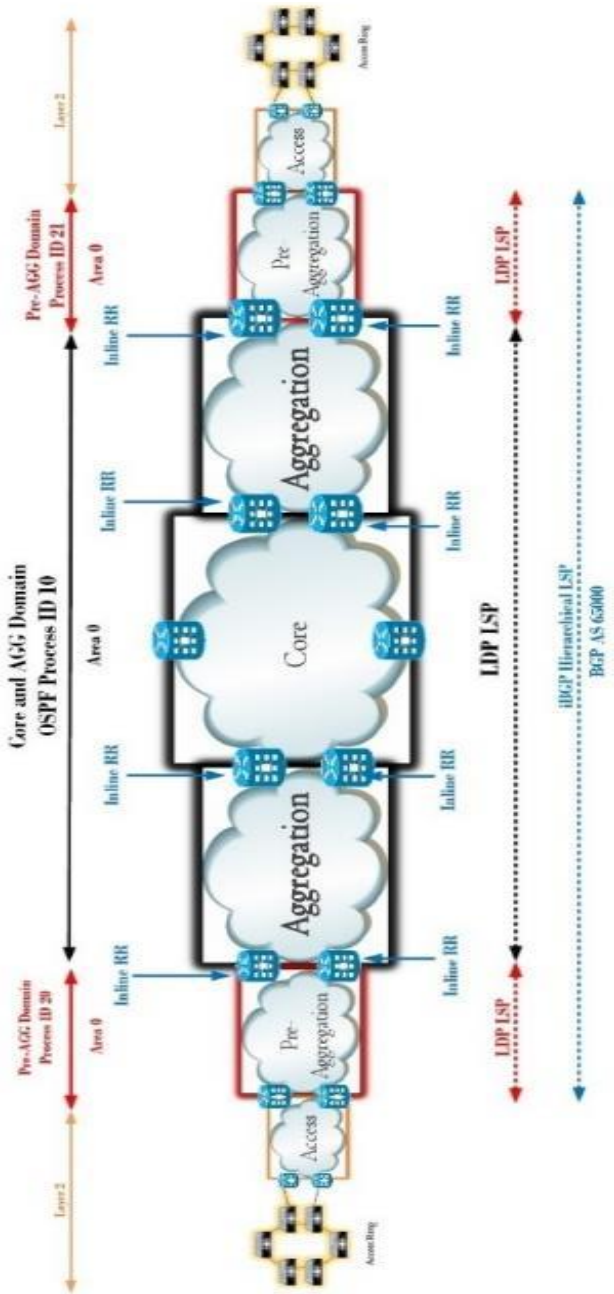


Figure 8-29 *Fixed Service IGP, BGP and LDP Design*

https://t.me/learningnets

ATELCO Security Policy

During the previous chapter (ATELCO network overview), it was given that due to the corporate security policy in every layer, except the Access layer, OSPF MD5 and LDP MD5 authentication will be enabled on ATELCO network.

OSPF and LDP MD5 authentication provides transport security for these protocols. MD5 authentication uses an encoded MD5 checksum that is included in the transmitted packet. The receiving routing device uses an authentication key (password) to verify the packet. You define an MD5 key for each interface. If MD5 is enabled on an interface, that interface accepts routing updates only if MD5 authentication succeeds. Otherwise, updates are rejected. The routing device only accepts OSPFv2 packets sent using the same key identifier (ID) that is defined for that interface.

ATELCO enables authentication between two LDP peers, which verifies each segment sent on the TCP connection between the peers. They configure authentication on both LDP peers using the same password; otherwise, the peer session is not established. Although ATELCO is using MD5, IETF published an RFC showing the problem by using MD5.

MD5 hashes are no longer considered cryptographically secure, and they should not be used for cryptographic authentication. In 2011, the IETF published RFC 6151, "Updated Security Considerations for the MD5 Message-Digest and the HMAC-MD5 Algorithms," which cited a number of recent attacks against MD5 hashes, especially one that generated hash collisions in a minute or less on a standard notebook and another that could generate a collision in as little as 10 seconds on a 2.66 GHz Pentium 4 system. As a result, the IETF suggested that new protocol designs should not use MD5 at all.

ATELCO has Flowspec in their networks to prevent DDOS attacks which are targeted to their customers. For the BGP protocol, Flowspec rules are sent by the Border Routers to each Edge Router through Router Reflectors.

Flowspec is defined in RFC 5575. New AFI and SAFI is defined for BGP Flowspec. AFI/SAFI: 1/133: “Unicast Traffic Filtering Applications” Instead of blackholing entire IP address (RTBH works in this way).

https://t.me/learningnets

There are many fields in the flow which can be used for filtering:

- Type 1: Destination Prefix
- Type 2: Source Prefix
- Type 3: IP Protocol
- Type 4: Source or Destination Port
- Type 5: Destination Port
- Type 6: Source Port
- Type 7: ICMP Type
- Type 8: ICMP Code
- Type 9: TCP Flags
- Type 10: Packet Length
- Type 11: DSCP
- Type 12: Fragment Encoding

Flow routes are automatically validated against unicast routing information or via routing policy. If customer is advertising flowspec routes which ATELCO has a route in the routing table, firewall filter is created based on match and action criteria.

After identifying the flow, any of the following actions can be taken:

- Rate limit to 0 (Drop the traffic)
- Remarking
- Redirect to VRF (Route Target)
- Sampling the Traffic (Monitoring)

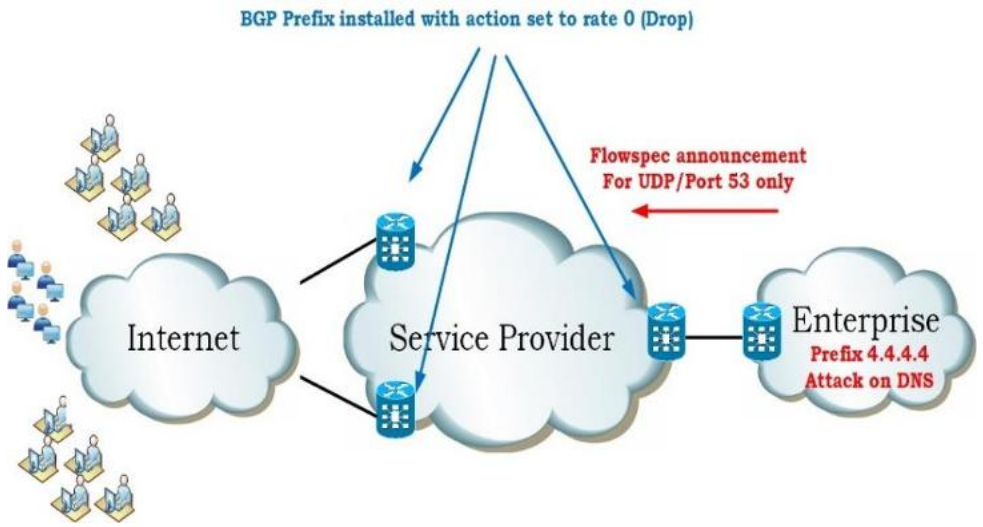


Figure 8-30 *BGP Flowspec*

https://t.me/learningnets

BGP Flowspec is supported by Commercial vendors and the Open Source implementation. Commercial vendors, Arbor Peakflow, Juniper DDOS Secure, Alcatel Lucent, Juniper JUNOS, Cisco on ASR and CSR. BGP Flowspec, compared to other methods such as manual filtering, Destination or Source based RTBH, is more granular as it is not dropping all the traffic towards an IP address, but maybe just a single port on the IP address.

BGP Flowspec can filter the traffic based on many other fields in the IP Packet, but still provides same level of automation as RTBH. Automation means, distributing the filtering policy rules to the edge nodes in the network, based on the matching criteria. Service Providers allow Customers to advertise Flowspec routes. When attack starts towards the customer (victim) IP address, customer initiates the filter, not just entire IP address, but based on the attack type (DNS attack, NTP attack etc.) only UDP 53 port can be filtered for the victim's IP address.

Also, ATELCO has been configuring AS-Path filters for their Business/Corporate customers to prevent some type of BGP based misconfigurations and attacks but they are researching other solutions to prevent possible future BGP based incidents.

Configuring only AS-Path filter doesn't prevent many known attacks. It doesn't prevent BGP route leaks which can be the result of customer's misconfiguration. Two commonly known BGP information security problems are Route Leaks and BGP Hijacks.

BGP Route Leaks can create blackhole, extra latency, packet loss, thus will have bad effect for the customer experience. The result of a route leak can be redirection of traffic through an unintended path that may enable eavesdropping or traffic analysis, or simply blackhole network traffic as there is no enough capacity on the network which leaked the prefixes.

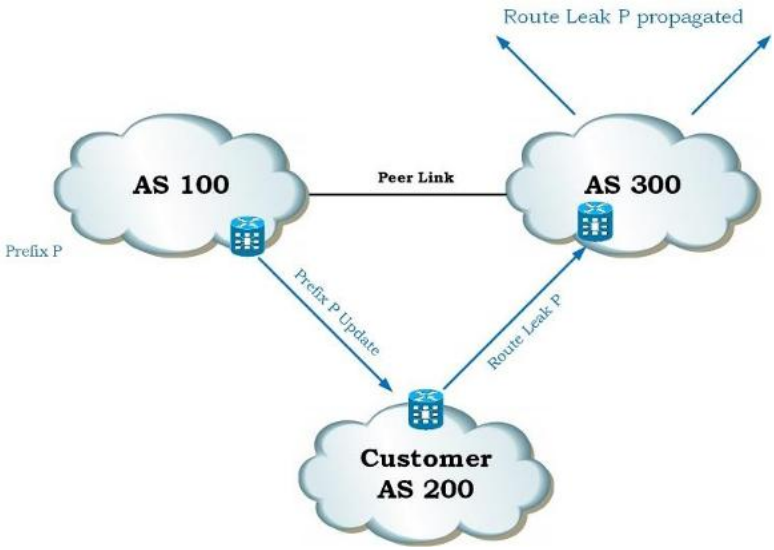


Figure 8-31 *BGP Route Leak*

https://t.me/learningnets

RFC 7908 highlights the Problem Definition and Classification of BGP Route Leaks. In figure 8-31, a multihomed customer AS learns a route from one upstream ISP and simply propagates it to another upstream ISP.

It should be noted that leaks of this type are often accidental (not malicious). The leak often succeeds (the leaked update is accepted and propagated), because the second ISP prefers customer announcement over peer announcement of the same prefix.

In figure 8-31, there is a peer link between AS 100 and AS300. AS 300 learns Prefix P which is behind AS100 from direct link as well as from its customer AS 200. AS 300 prefers the route through the customer for Prefix P, because Service Providers increase local preference value for the same prefix if they learn from their customer. Although on the AS 300 BGP routing table, AS-Path length for Prefix P is shorter towards AS100, while the local preference value is compared before AS-path attribute when the same prefix is learned from multiple BGP neighbors.

If AS 300 would have a filter to accept only allowed prefixes from AS 200, AS 200 wouldn't be able to leak the prefixes from AS 100. But for the large scale networks this type of filtering is almost impossible, because customer might be multihomed and they may receive their prefixes from another upstream AS. So in this case, each time the customer has a new prefix, they need to communicate with their Service Providers.

BGP Hijack is shown in figure 8-32. BGP Hijacks can be prevented by Route Origin Validation, but BGP Route Leaks, cannot be prevented by Origin Validation. BGPSEC is the current IETF effort for Path protection. BGPSEC builds on RPKI by adding cryptographic signatures to BGP messages. It requires each AS to digitally sign each of its BGP messages.

The signature on a BGPSEC message covers (1) the prefix and AS-level path; (2) the AS number of the AS receiving the BGPSEC message; and includes (3) all the signed messages received from the previous AS's on the path. Almost there is no deployment yet with BGPSEC. It Requires BGP messages to be changed, router hardware to be changed as well.

Figure 8-32 depicts BGP Hijacks, specifically Sub-prefix hijack. AS4 advertises the more specific route '10.10.0.0/24' which belongs to AS1. But AS1 doesn't advertise specific prefixes to its BGP neighbors. Instead AS1 advertises 10.10.0.0/16 which is less specific routing advertisement. Thus, all AS's; AS2, AS 3 and AS5 prefer AS4 which is the attacker for the 10.10.0.0/24 subnet. Because based on routing, more specific routes are always preferred to less specific routes, even though less specific routes have shorter BGP AS-Path length.

https://t.me/learningnets

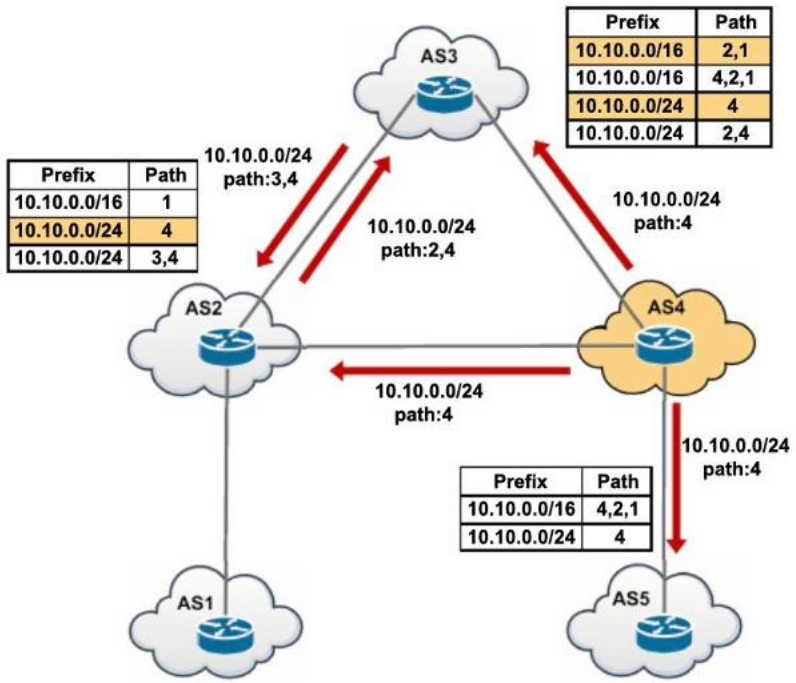


Figure 8-32 *BGP Sub-Prefix Hijack*

Route Origin Validation which can be provided by RPKI can prevent issues. As an example, with RPKI, AS 2, AS 3 and AS5 can validate that actual owner of 10.10.0.0/24 subnet is AS1, thus they can reject the advertisement from AS 4. ATELCO currently has only AS-Path filtering for BGP information security in their network but they should start deploying RPKI for Route Validation.

They should start considering to place max-prefix limits in order to prevent accidental full Internet routing table leak from their customers or settlement free peers.

RPKI will provide Origin Validation to them, Max-Prefix limits will protect their networks from large number of prefixes to be leaked from customer or peers. In addition to these methods, they should start following the recent enhancements with path validation through BGPSEC.

https://t.me/learningnets

Summary

In this chapter, detail information about ATELCO's network was shared. In the previous chapter, the overall infrastructure, services which they provide, their worldwide locations and the connectivity to other networks were provided. In this chapter, technologies and designs were covered in more detail, and alternative solutions were provided.

Also, there was some suggestions to them. Networks are evolving, transit service cost is declining, Internet becoming a Flat architecture, data which is carried over networks is growing; thus, it is important to follow the recent changes in the industry and when it is necessary, planning the network changes.

Because of this, scalability of the network is important, otherwise long term planning, testing and validation will be necessary for the Service Providers. Service Providers don't want to be late to announce their products to market. Also un-scalable, un-flexible networks will always be a problem for their services to be deployed and announced to their customers.

https://t.me/learningnets

Chapter 9

Chapter-9 Evolving Technologies in the Service Provider Networks

Introduction

Service Providers encounter various challenges to provide next generation services to accommodate fast-paced demands of the market. Furthermore, by the introduction of 5G, video traffic growth, IoT and cloud services, combined with services requiring ubiquitous connectivity from Access to Core, Service Providers require unprecedented level of flexibility, elasticity and scalability in the network infrastructure.

In this chapter, we are going to introduce new approaches to design highly scalable Service Provider networks by means of new technologies such as Segment-Routing (SR), Fast Reroute with TI-LFA, PCE, Egress Peer Engineering, EVPN, PBB-EVPN, BGP in Massively Scalable Datacenter, NFV and Multicast BIER.

https://t.me/learningnets

Service Provider Design Using Segment Routing

Service Providers must choose a very flexible design that meets any to any connectivity requirements, without compromising in stability and availability. In the divide-and-conquer strategy in which the Core, Aggregation, and Access domains are partitioned in different IGP domains, formerly used by Unified/Seamless MPLS, and as presented previously in the ATELCO Scenario in the book, it reduces the size of routing and forwarding tables within each domain, so it provides better stability and faster convergence.

Traditionally, Unified MPLS used LDP or RSVP-TE to build LSP within the IGP domain and used BGP-LU (RFC 3107) for inter-Domain LSPs. Segment Routing reduces the number of required protocols in a Service Provider network by adding simple extensions to IGP protocols such as ISIS or OSPF that can assign and distribute labels to build LSP within each IGP domain. This enables a device inside an Access, Aggregation, or Core domain to have reachability through intra-domain SR LSPs to any other device in the same region. In the next pages, we will see in some scenarios that it is better to eliminate BGP-LU for faster convergence and simplicity of the network. Programmability based network architecture based on Segment-Routing will add SLA awareness into the network and provides better network scaling.

Segment Routing Introduction

Based on the definition of IETF (RFC 8402), Segment Routing (briefly SR) leverages the source routing paradigm. This definition is the most notable feature of SR that everything happens in Head-end node based on an ordered list of instructions, called “segment”. A segment can be local or global within the SR domain and often referred as Segment Identifier (SID). SR can work on MPLS dataplane. A segment is encoded as an MPLS label and ordered segments are encoded as a stack of labels.

Processing of segments starts from top and after completion of the top-most segment, it is popped from the stack. Also, SR can work on IPv6 data plane with the new type of routing header. A segment is encoded as an IPv6 address and ordered segments are encoded as a stack of IPv6 addresses in the routing header. Segment Routing reduces the number of protocols needed in a Service Provider Network. Simple extensions to traditional IGP protocols like ISIS or OSPF, provide full Intra-Domain Routing and Forwarding Information over a label switched infrastructure, along with Fast Re-Route (TI-LFA) capabilities.

This is an enhancement in comparing SR with Label Distribution Protocol (LDP), that SR capable IGP node advertises segments for its attached prefixes and adjacencies through IGP header instead of using another protocol.

https://t.me/learningnets

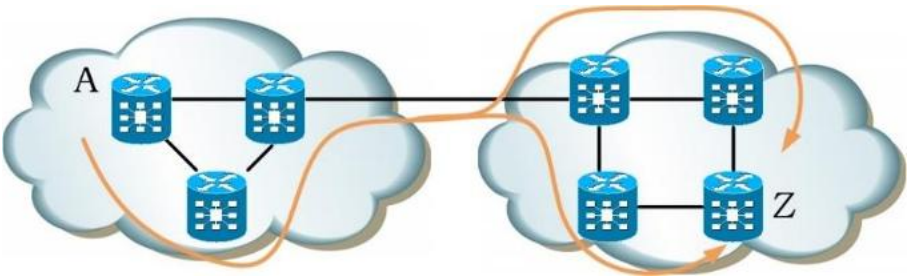


Figure 9-1 *Segment Routing can steer the traffic in any direction*

With Segment Routing, Interior gateway protocol (IGP) distributes two types of segments: prefix segments and adjacency segments. Each router (node) and each link (adjacency) has an associated segment identifier (SID). A prefix SID is associated with an IP prefix.

The prefix SID is manually configured from the segment routing global block (SRGB) range of labels, and is distributed by IS-IS or OSPF. The prefix segment steers the traffic along the shortest path to its destination. A node SID is a special type of prefix SID that identifies a specific node. It is configured under the loopback interface with the loopback address of the node as the prefix and it must be globally unique.

An adjacency segment is identified by a dynamic label called an adjacency SID, which represents a specific adjacency, such as egress interface, to a neighboring router. The adjacency SID is distributed by OSPF or IS-IS. The adjacency segment steers the traffic to a specific adjacency and it must be locally unique.

Traffic Engineering using SR

Segment Routing steers a network traffic into SR Policy that contains an ordered list of segments. A SR Policy is a framework that enables instantiation of an ordered SID list on a node for implementing a source routing and it is uniquely identified through a tuple (headend, color, and endpoint). SR policy also can be used for Fast Reroute (FRR) or Operations, Administration, and Maintenance (OAM) purposes.

In comparison of SRTE with RSVP-TE, advantages of SRTE are Multi-domain support by using PCEP for compute, Equal Cost Multi Path and automated traffic steering. Also, there is a component called Binding-SID (B-SID) that is fundamental to SR and it may involve a list of SIDs and it is bound to SR Policy for greater scalability.

https://t.me/learningnets

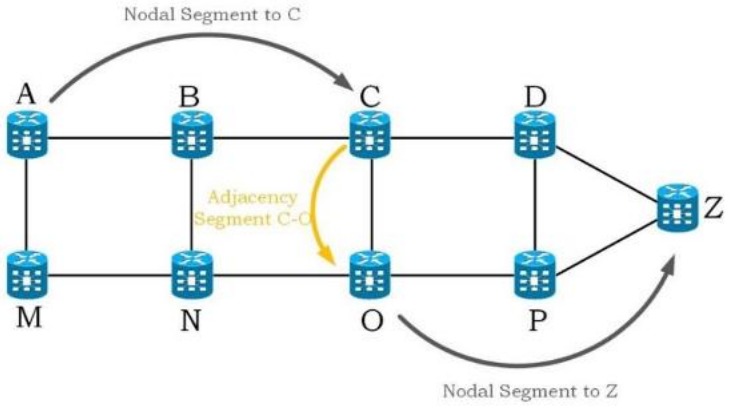


Figure 9-2 *Segment Routing provides Source Routing using Segments*

Moreover, a feature of SRTE On-Demand Next-hop (ODN) that automatically instantiates an SR Policy to a BGP next-hop when required and steers traffic automatically inside the SR Policy. In this case, BGP learns a route via a tail-end (Egress PE) node with color, and the headend node (Ingress PE) authorizes the on-demand SR Policy path and maps the defined color and end-point to a local dynamic SR Policy path.

PCEP and Segment Routing

When using a controller to compute a path dynamically through multi-domain network (Multi-Area or Multi-AS), a control-plane protocol called PCEP (Path Computation Element Protocol) is used between the routers and the controller.
Path Computation Element has two main components:

1. PCE
2. PCC

PCE - Path Computation Element is defined in RFC 4655. Path Computation Element (PCE) is an entity that is capable of computing a network path or route based on a network graph, and takes the administrators constraint into an account during the computation. The PCE can be located within a network node or component, on an out-of-network server, etc.

For example, a PCE would be able to compute the Traffic Engineering LSP path by operating on the TED (Traffic Engineering Database) and considering bandwidth and other constraints applicable to the TE LSP. Traffic Engineering Database (TED) is filled up by the link state protocols such as OSPF and ISIS.

https://t.me/learningnets

Path Computation Element (PCE) is a compute server for calculating the path through nodes in multiple domains to find out the best path for SRTE and send an ordered SID list to a headend node to reach its destination. Path Computation Client (PCC) is a node that requests a path with specified detail from PCE to reach its destination. These two components are using PCEP (Path Computation Element Protocol) to make a stateful connection with each other.

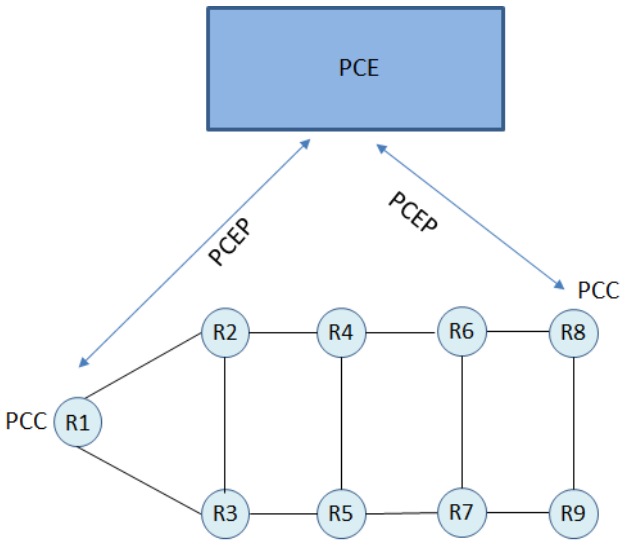


Figure 9-3 *PCE Architecture*

END to END Segment routing (Single BGP-AS)

This section focuses on the sample operator network design based on Segment routing multi IGP domain network and application of new features such as PCE controller. The access node is a service edge node or PE node which the service layer starts or terminates here.

The end-to-end path is established using SDN traffic controller using PCEP Protocol or BGP-LU as shown in figure 9-4. The end-to-end inter-domain network path is programmed through controller and is selected based on the customer SLA, such as “low latency” path.

https://t.me/learningnets

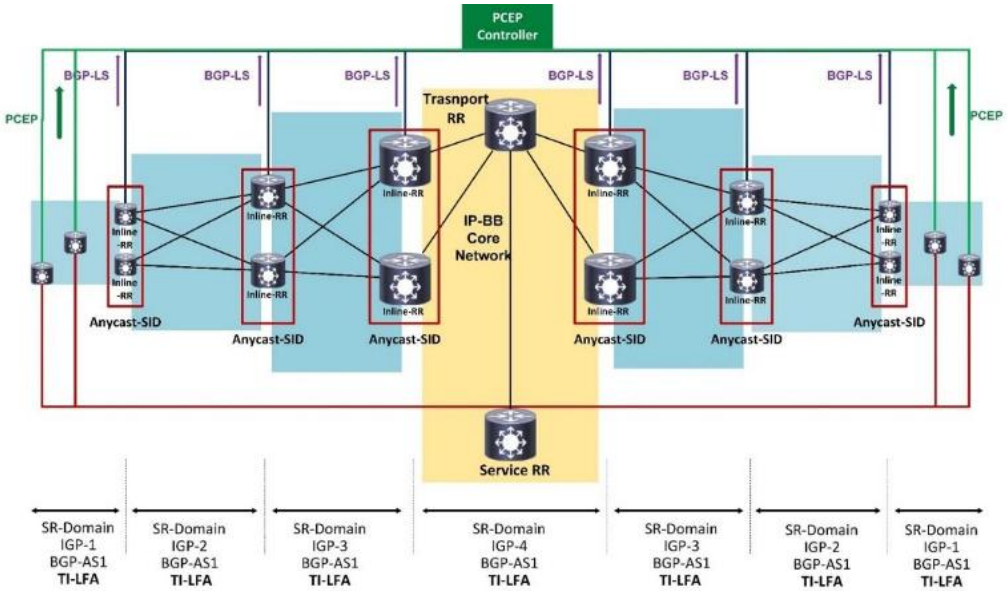


Figure 9-4 *END to END Segment routing (Single BGP-A*S)

The characteristics of this network are as follow:

- There are different network domains, with separate IGP running in each domain
- TI-LFA for fast convergence in each IGP domain
- Each domain is connected to another domain using two border routers which are inline RR's, get the same Anycast-SID and the same IP address for high availability and can be used for load balancing.
- There are two types of RR's, transport RR and Service RR
- BGP-PIC is used for BGP Fast Reroute (Dataplane convergence)

Note: End-to-end transport path can be achieved by BGP-LU without SDN controller or SRTE-Policy (SR-ODN) by the received Segment-List from PCEP Controller that collects network topology information from different domains using BGP-LS.

Note: End-to-end transport can be achieved by BGP-LU or SDN-driven path by the received Segment-List from PCE Controller

https://t.me/learningnets

> **Note:** For Anycast-SID, additional signaling protocols are not required, as the network operator is able to simply allocate the same Prefix SID (thus an Anycast-SID) to a pair of nodes typically acting as ABRs (Border Routers located between domains).

Major components of the given topology are:

Domain Border Routers

Routers which are located between the boundaries of domains and play as:

- RR-Client for upstream router and RR-Server for downstream routers.
- Distribute IGP Topology information to the controller using BGP-LS address-family
- Handle traffic flow between domains

Transport RR

A BGP Route Reflector for underlay traffic which is used for the IPv4/v6 address family, so the BGP Route Reflector is called as an IP Route Reflector or for short, Transport RR.

Service RR

For overlay services, each node participating in BGP-based service termination has two BGP sessions with Domain Specific S-RRs which can be located in each domain, or Central S-RR which can be located in the core and reflects VPNv4, VPNv6, L2VPN, EVPN. For Redundancy reasons, there are at least 2 S-RRs.

PCE Controller

This transport option is based on SID-List that PCE controller provides. Each domain has its own IGP/SR, and two IGP border routers in each domain using BGP LS to distribute topology, bandwidth, reliability, latency, SRLG and other transport states of the IGP domain to the SDN controller.

The SDN controller by gathering topology data and current state of the network from different domains, builds the end-to-end best path and alternate disjoint path that satisfies a given service requirement and sends the corresponding segment list to the service edge router. SR-PCE can be Domain Specific which will be located in each domain or Central SR-PCE which can be located in the Core.

https://t.me/learningnets

Service Node or PE Routers

VPN Service is defined here and VPN traffic starts and terminates at this point. Service Node connects to the SDN-Controller using PCEP Protocol and plays the role of PCC (Path computation client) and the controller takes the role of PCE (Path computation engine).

PCC, for the given traffic requests to get the best path to the Egress router, sends a request-Message to the controller using PCEP protocol and gets a Segment-List which steers the traffic to the TE tunnel using transport label to the Egress router.

TI-LFA with Segment Routing

To speed up the convergence process, TI-LFA provides precalculated-preinstalled backup path based on post-convergence state of the network against the link or node failure in any IGP network. TI-LFA is completely Stateless and does not require any additional signaling mechanism such as RSVP-TE (Path/Reserve Message), therefore each node in the IGP calculates a primary and a backup path automatically and independently based on the IGP topology.

TI-LFA can find the optimal path all the time in any kind of topologies and covers all kind of topologies like large ring topologies. Using segment routing, there is no need to use Targeted LDP for rLFA. In comparison with RSVP-TE FRR, in SR there is no need to maintain the control plane state in the network, so state information is carried in the packet header itself. OSPF opaque LSA and ISIS TLV's are used for this purpose.

Common benefits of TI-LFA are:

- There is no need to keep the state of the network like RSVP-TE FRR
- TI-LFA makes the smaller label stack for repair path
- There is no need for complex configuration, many things can happen automatically
- TI-LFA uses Post convergence path (RSVP-TE FRR cannot guarantee post convergence optimal path)
- There is no need to establish targeted LDP with remote routers. This is required for Remote LFA.
- By using the segment list in the head-end router, the backup path is created.

https://t.me/learningnets

Egress Peer Engineering

Monetary cost, latency and packet loss are important parameters for the Quality of User Experience for the customers; In the Service Provider networks, traffic engineering can be done by optimizing any of the above parameters.

BGP NLRIs don't provide the information about the cost, latency or loss of the path or exit point for the destinations. In figure 9-5, traditional egress peer engineering and modern way of egress peer engineering are shown. But before that, it is important to explain a specific terminology which is used among the Service Providers. The data-plane interconnection link between different networks is called Network to Network Interconnection (NNI).

Control-plane (EBGP) direct connection between two AS's allows Internet traffic to travel between the two, usually as part of a formal agreement called peering.

This peering can be settlement free based or settlement based (Ex: IP Transit). Settlement Free Interconnection and IP Transit Connections were explained in the earlier chapters. The selection of the best exit link for a given destination prefix selection and the enforcement of this selection on a network is not a simple task. This is because the decision for one prefix might impact other traffic by changing the utilization of the NNI link and potentially leading to overload.

Traditionally, SPs use a policy to manipulate the BGP attributes contained in NLRIs received from a peer. This policy-based manipulation is usually performed on the Egress ASBR, but sometimes also on a route reflector (RR) and the Ingress ASBR. This traditional technique provides some level of flexibility and control on how traffic leaves the SP and AS.

However, it is also limited by the BGP path selection algorithm and the fact that the results apply to all traffic for given prefix, regardless of the traffic's origin (Doesn't matter which Ingress ASBR sends the traffic).

https://t.me/learningnets

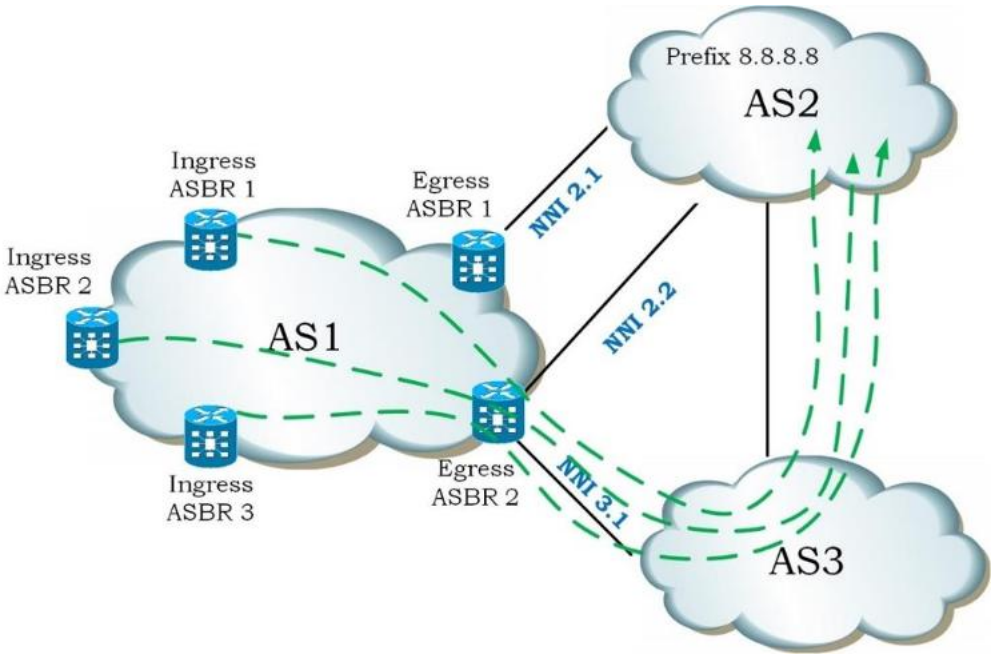


Figure 9-5 *Traditional Egress Peer Engineering with BGP*

In figure 9-5, assume prefix 8.8.8.8 is originated in AS 2, while AS2 and AS3 are the peers. AS1 learns the prefix from both AS2 and AS3 through NNI 2.1, NNI 2.2 and NNI 3.1. If on Egress ASBR2, the Local preference is higher for 8.8.8.8 towards AS3, NNI 3.1 link is used by every Ingress ASBR in the network. You don't have the ability to send different Egress Peer Link per Ingress ASBR.

In fact, even if BGP ORR (Optimal Route Reflection) is used, when the policy is created by using Local Preference or any BGP path attribute, Ingress ASBRs will always use the same BGP egress link for the given destination.

https://t.me/learningnets

Modern – Better way of EPE

If EPE had the ability to distribute traffic among several egress links, based on not only the destination address, but also by considering the ingress ASBR (or ingress port etc.), this would provide much finer granularity and also bandwidth management. This would be especially true if EPE were combined with traffic statistics and centralized optimization (Controller).

The EPE solution should direct traffic for a given prefix that enters the network on a particular Ingress ASBR to a particular egress NNI (Egress Link) on a particular Egress ASBR. Better Policy for utilization of both internal and inter-domain resources (NNI Bandwidth etc.) can be as follow for AS 1

- [Ingress ASBR1, 8.8.8.8] to NNI 2.1
- [Ingress ASBR2, 8.8.8.8] to NNI 2.2
- [Ingress ASBR3, 8.8.8.8] to NNI 3.1

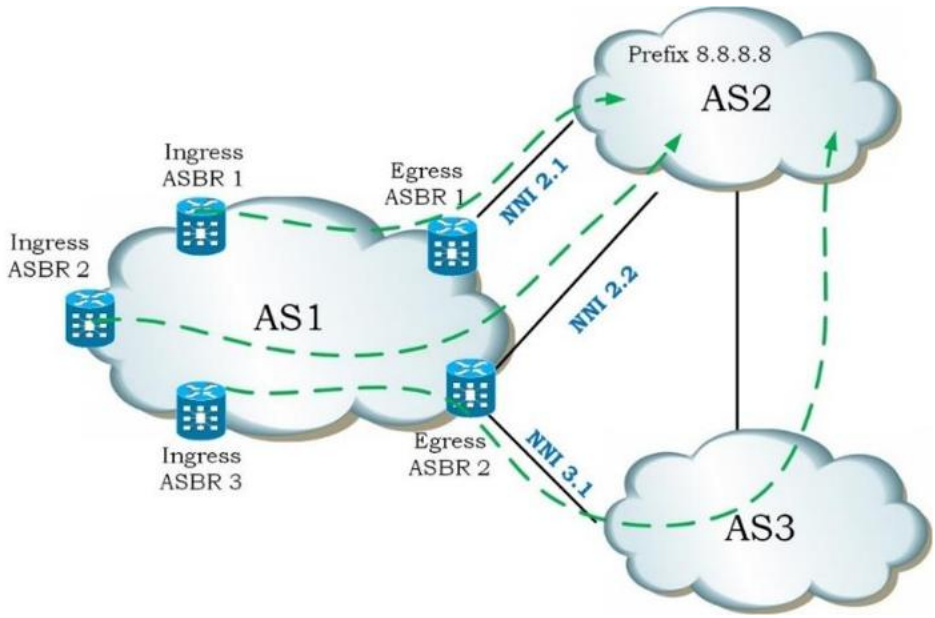


Figure 9-6 *Better BGP Traffic Engineering Policy for utilization of both internal and inter-domain resources*

https://t.me/learningnets

Modern EPE Requirements

The utilization data for each NNI is necessary to send the traffic to the correct Egress node and link. This data is provided by traditional or modern telemetry infrastructure (for example, SNMP interface statistics). The reachability information for destination IP prefixes is required and this information is provided by EBGP advertisement from peer ASs.

The fine-grain partitioning of egress traffic into "flows" with information about the traffic volume carried by each would be great to redirect different size of the flows to different egress link. Flow based traffic engineering can provide maximum level of flexibility.

An Egress Peering Engineering (EPE) controller executes some logic to map these "flows" to NNI in a globally optimal way (Distributed TE cannot provide global optimality, Bin Packing, Dead-Lock are known problems in distributed traffic engineering without controller).

Another requirement for the modern Egress Peer Engineering is a network infrastructure that allows forwarding traffic from an ingress AS Border Router (ASBR) to the designated egress NNI, as determined by the EPE controller.

A BGP EPE enabled egress PE node may advertise Segment Routing SIDs corresponding to its attached peers. They enable source routing for inter-domain paths. The controller learns the BGP peer SIDs and the external topology of the egress border router through BGP-LS routes. The controller can program an ingress node to steer traffic to a destination through the egress node and peer node using BGP labeled unicast (BGP-LU).

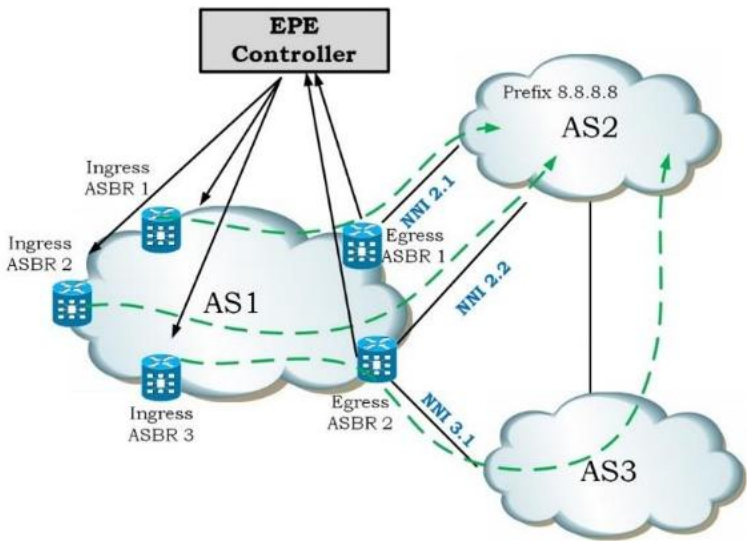


Figure 9-7 *EPE Controller can inject the end to end path per Ingress ASBR to different links on egress ASBRs*

https://t.me/learningnets

Segment routing and LDP Internetworking

In this part, we focus on the mechanisms through which SR interworks with LDP in cases where a mix of SR-capable and non-SR-capable routers co- exist within the same network and more precisely in the same routing domain.

Segment Routing, can be used on top of the MPLS data plane without any modification. Segment Routing control plane can co-exist with current label distribution protocols such as LDP.

This section focuses on the sample network design with some domains running traditional LDP and while other domains migrated to SR as both types of domains have to be able to work together.

Mapping Server

To provide Internetworking in the direction of LDP to SR, no additional signaling or state is required. However, in the direction of SR to LDP Internetworking, one node must advertise prefix-SIDs on behalf of non-SR nodes, thereby allowing non-LDP routers to send and receive labeled traffic from LDP-only routers. To achieve this operation, there is a role named Mapping Server.

Segment Routing Mapping Server assigns prefix-SIDs to prefixes owned by non-SR-capable routers as well as to prefixes owned by SR capable nodes.

Let’s assume we have a scenario that is separated by two domains, as one of them uses LDP and another uses SR. Based on figure 9-8, traffic wants to go from P7 to P1, but P7 doesn’t run LDP and wants to assign label for P1. It received a SID 16001 from mapping server to reach P1. Then it adds label 16001 and sends it through the nodes.

When traffic is received by P3, it looks up for destination and finds LDP label assigned for P1, then it swaps Prefix-SID with LDP label and sends traffic to P2 and it sends traffic to P1 with appropriate label.

https://t.me/learningnets

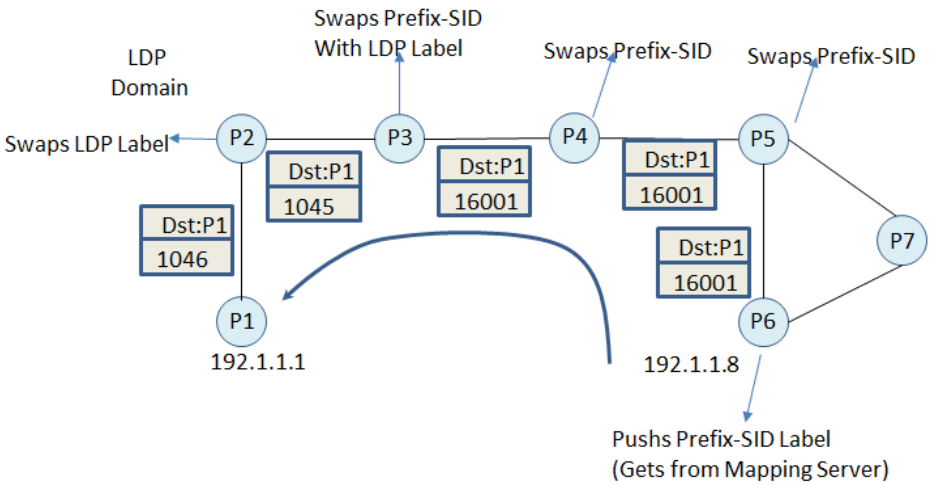


Figure 9-8 *Segment Routing Mapping Server*

Ethernet VPN (EVPN)

Ethernet VPN (EVPN) is a standard protocol and is response to limitations of VPLS when it comes to redundancy, multicast optimization, multipathing and provisioning simplicity. Also, legacy L2 technologies (VPLS, PBB) still rely on flooding and learning to build Layer 2 forwarding database.

EVPN works in Multi-homing scenarios in Active-Active mode and with BGP split-horizon label, egress PEs use this label to perform split-horizon filtering over attachment. In BGP-EVPN, MAC distribution between PEs does not happen in data plane (Like VPLS using Flood and learn), but it's done via control plane advertisement using BGP.

Regarding to VPLS, All-Active redundancy models are not deployable as VPLS technology lacks the capability of preventing L2 loops. For example, BUM traffic sourced from the CE is flooded throughout the VPLS Core and is received by all PEs, which in turn floods it to all attached CEs. Whereas in contrast, EVPN can solve this problem by means of BGP-Based Control Planes.

MAC distribution with BGP offers greater control over Flood and learn process, such as restricting who learns what, and the ability to apply policies. It uses MP-BGP for this purpose.

In EVPN, PEs advertise the MAC addresses learned from their connected CEs, along with an MPLS label, to other PEs in the control plane using Multiprotocol BGP (MP-BGP). Control-plane learning enables load balancing of traffic to and from CEs that are multi-homed to multiple PEs and it also improves convergence times in the event of certain network failures. This is known as Flow-based load balancing in the Industry.

https://t.me/learningnets

Please note that EVPN is a layer 2 and layer 3 VPN technology, which uses MP-BGP as the Control Plane Protocol while MPLS and VXLAN can be the Data plane protocol for EVPN. In fact EVPN control plane has various data planes like MPLS, NVO (VXLAN, NVGRE), PBB and VPWS. In order to carry traffic across the network, it is required that tunneling protocol is being specified. VXLAN is one of the options, but it is not the only one.

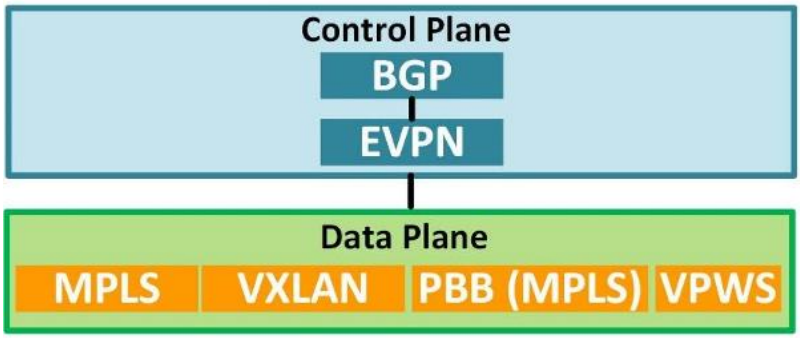


Figure 9-9 *EVPN and its Data Plane technologies*

Note:
As LDP and Segment-Routing act as transport layer or underlay, VPLS, MPLS L3 VPN and EVPN are overlay services based on underlay technologies, therefore any overlay technology can be carried by any technology on underlay technology. For example it is possible to provide L3VPN or EVPN services based on LDP or SR or RSVP-TE in transport layer.

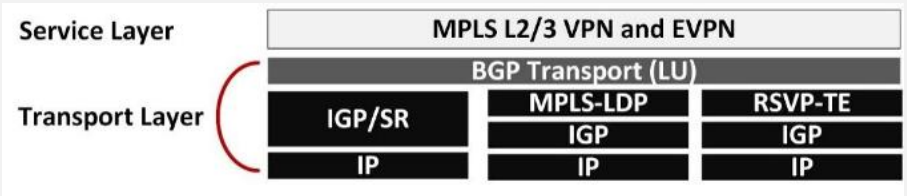


Figure 9-10 *Underlay and Overlay protocols in layer 2 and layer 3 VPN's*

https://t.me/learningnets

ARP Suppression

One of the scalability challenges in large scale Layer 2 networks is BUM traffic between the data centers. Most of the solutions use Flood and Learn dataplane learning, but when it comes to scaling, number of broadcasts for ARP resolution is a challenge.

To address this challenge, EVPN enabled PEs to cache ARP messages as they act as ARP proxy for locally attached hosts, thereby preventing repeated ARP broadcasts over the network.

When an end-host sends an ARP request for another end-host IP address, PE locally intercepts the ARP request and checks for the ARP-resolved IP address in its ARP suppression cache table. If it finds a match, the PE sends an ARP response on behalf of the remote end host.

If the PE doesn't have the ARP-resolved IP address in its ARP suppression table, it floods the ARP request to the PEs. After the PE learns about the MAC and IP addresses of the ARP requested host, the information is distributed through the MP-BGP EVPN control plane to all other PEs. Any subsequent ARP requests do not need to be flooded.

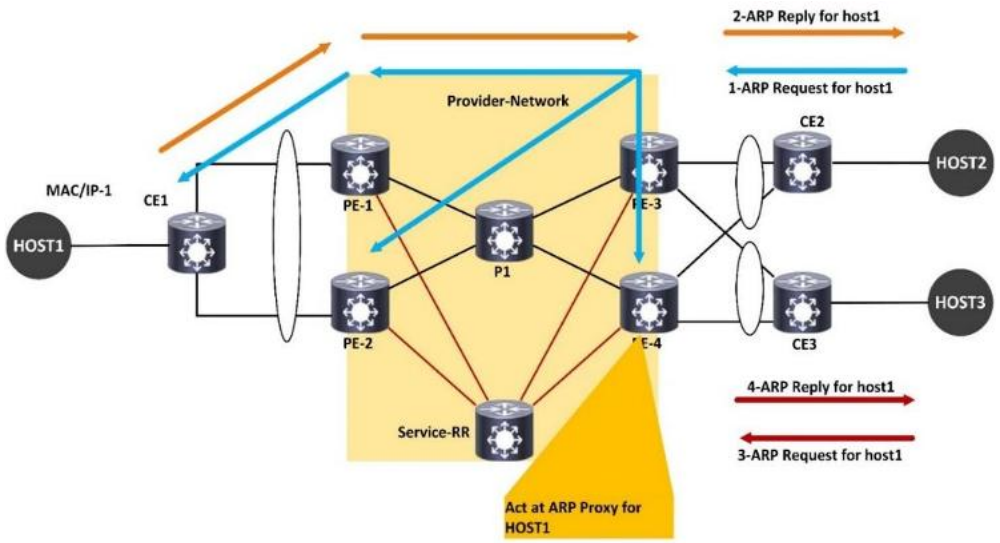


Figure 9-11 *ARP Suppression with EVPN*

https://t.me/learningnets

Provider Backbone Bridging Ethernet VPN (PBB-EVPN)

PBB (IEEE 802.1ah), known as "Mac-in-Mac" is used for routing over a provider's network allowing interconnection of multiple Provider Bridge Networks without losing each customer's VLANs.

EVPN based PEs run Multi-Protocol BGP to advertise & learn customer MAC address (c-MACs) over cores to overcome flood & learning limitation of VPLS.

A common problem in EVPN technology Control plane MAC distribution poses scalability concerns in certain environments, and this brings PBB EVPN to the table. In PBB EVPN, a typically small number of B-MACs (one B-MAC per instance and PE) are discovered through EVPN control plane, and the customer MACs (C-MACs) are learned in the forwarding plane.

In fact, in PBB instead of BGP, routing happens for each Customer mac address to reach the destination, all C-MAC behind a PE will bind to one B-MAC of each PE. Thus to reach the destination, BGP cares about B-MAC of the PEs not C-MACs behind the PEs.

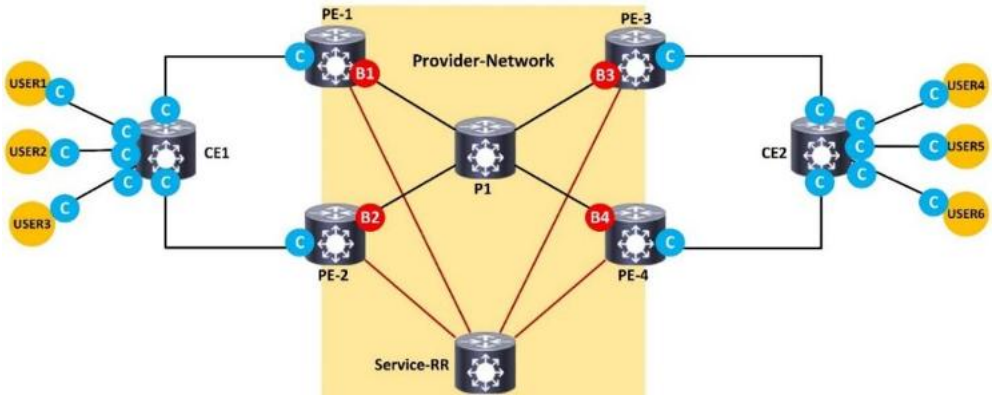


Figure 9-12 *Provider Backbone Bridging Ethernet VPN (PBB-EVPN)*

EVPN PE learns customer MAC-Address using data plane learning, then advertises it using BGP to RR or another PE.

https://t.me/learningnets

NFV – Network Function Virtualization

Network Functions Virtualization (NFV) founded by European Telecommunication Standard Institute (ETSI) with Industry Specification Group (ISG) that contains seven of the world's leading telecom network operators. A challenge of large scale telecom networks is increasing variety of proprietary hardware and launching new services that may demand installation of new hardware. This challenge requires additional floor space, power, cooling, and more maintenance. With evolving virtualization technologies in this decade, NFV focuses on addressing the telecom problems by implementing network functions into software that can run on server hardware or hypervisors.

Furthermore, by using NFV, installing new equipment is eliminated and it will be related to health of underlay servers and result is lower CAPEX and OPEX.
There are many benefits when operators use NFV in today's networks. One of them is Reducing time-to-market to deploy new services to support changing business requirements and market opportunities on new services.

Decoupling physical network equipment from the functions that runs on them will help telecom companies to consolidate of network equipment types onto servers, storages, and switches which are in data centers. In NFV architecture, responsibility for handling specific network functions (e.g. IPSEC/SSL VPN) that run in one or more VM, is Virtual Network Function (VNF).

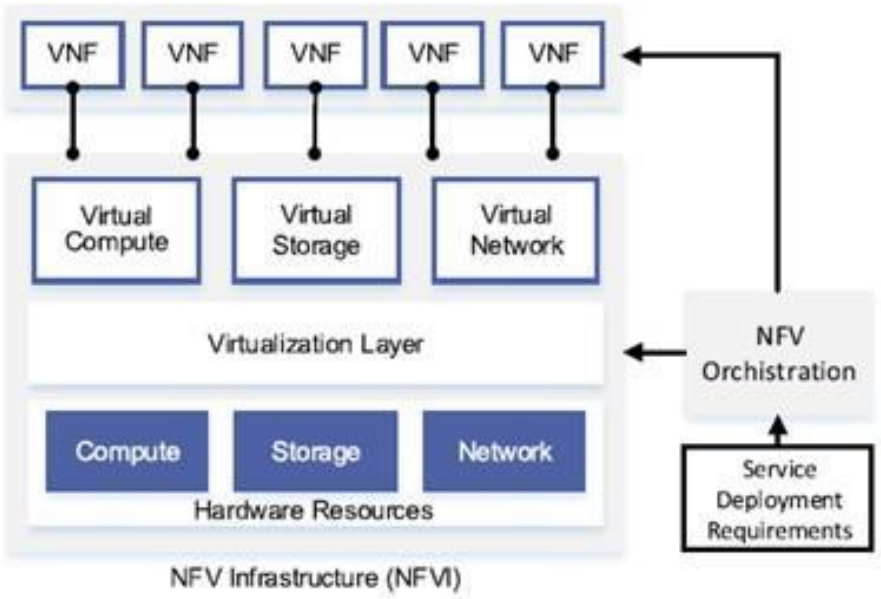


Figure 9-13 *NFV Infrastructure*

https://t.me/learningnets

As figure 9-13 depicts, the whole system of NFV that contains physical and virtual components, is called NFV Infrastructure (NFVI). NFVI can be different based on deployments and vision of a Service Provider. For example, NFVI can build upon Docker or maybe any kind of hypervisor or mixing both of them.

Service Providers may use their own OSS/BSS to provision their infrastructures and boost service hosting to their customers and users. Based on this approach, there should be other protocols and components that help Service Providers to build their end-to-end full automated services with using NFV. To meet this demand, ETSI released a framework that shows functional blocks and reference points in the NFV framework.

The main reference points and execution reference points are shown by solid lines and are in the scope of NFV. These are potential targets for standardization. The dotted reference points are available in present deployments but might need extensions for handling network function virtualization. However, the dotted reference points are not the main focus of NFV at present. Figure 9-14 illustrates ETSI NFV framework architecture that is taken from ETSI document.

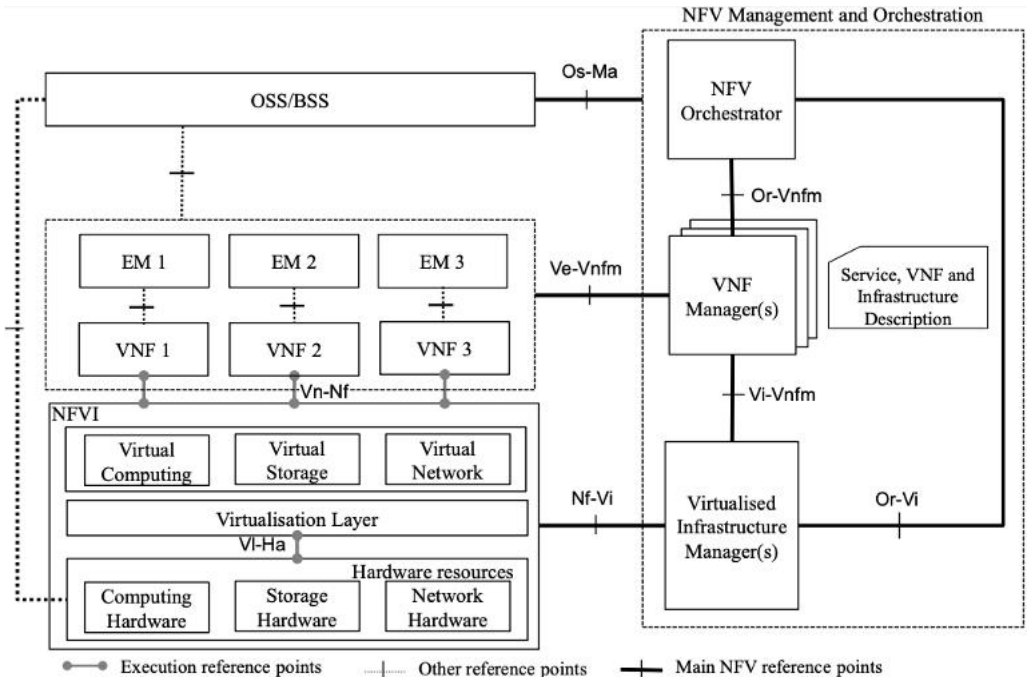


Figure 9-14 *ETSI NFV framework architecture (source: www.etsi.org)*

https://t.me/learningnets

A key component in the NFV architectural framework is the virtualization layer. This layer abstracts and logically partitions physical hardware resources and anchors between the VNF and the underlying virtualized infrastructure.

The primary tool to realize the virtualization layer would be the hypervisors. The NFV architectural framework should accommodate a diverse range of hypervisors. On top of such a virtualization layer, the primary means of VNF deployment would be instantiating it in one or more VMs.

Therefore, the virtualization layer shall provide open and standard interfaces towards the hardware resources as well as the VNF deployment container, e.g. VMs, in order to ensure independence among the hardware resources, the virtualization layer and the VNF instances. VNF portability shall be supported over such a heterogeneous virtualization layer.

The decoupling of a VNF from the underlying hardware resources presents new management challenges. Such challenges include end-to-end service to end-to-end NFV network mapping, instantiating VNFs at appropriate locations to realize the intended service, allocating and scaling hardware resources to the VNFs, keeping track of VNF instances location, etc. Such decoupling also presents challenges in determining faults and correlating them for a successful recovery over the network. While designing the NFV Management and Orchestration, such challenges need to be addressed.

In order to perform its task, the NFV Management and Orchestration should work with existing management systems such as OSS/BSS, hardware resource management system, CMS used as a Virtualized Infrastructure Manager, etc. and augment their ability in managing virtualization-specific issues. Also, SDN (Software Defined Network) can bring agile and lower provisioning time to the network alongside of NFV.

NFV Platforms

As described in the previous topic, virtualization plays an eye-catching role in data center and by adding NFV over virtualization, network components may change. Also, NFV solves problems of Datacenters, Enterprises, and Service Providers' networks. Generally, there are two types of NFV implementations: Open Source and Commercial Products. In the following section with some examples, we'll describe benefits of NFV and some of the use cases.

Open platform for NFV (OPNFV) facilitates the development and evolution of NFV components across various open source ecosystems. Virtual network functions range from mobile deployments, where mobile gateways like SGW and PGW and related functions (MME, HLR, PCRF, etc.) are deployed as VNFs, to deployments with virtual customer premise equipment (CPE), tunneling gateways, firewalls or application level gateways and filters to test and diagnostic equipment.

https://t.me/learningnets

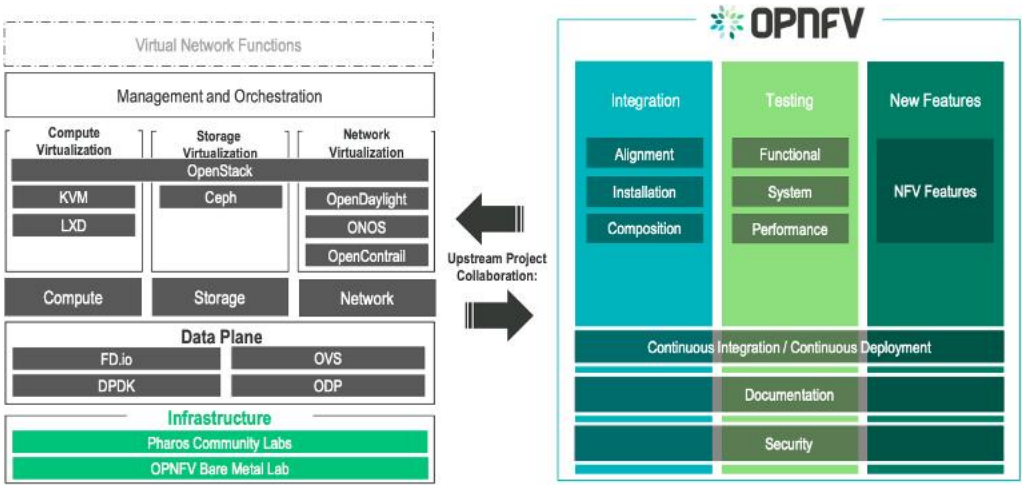


Figure 9-15 *Open source components placed in original ETSI NFV framework standard (source: https://www.opnfv.org)*

Figure 9-15 shows that how open source components are placed in original ETSI NFV framework standard. With OpenStack as IaaS platform, all of main components (compute, storage, and network) can manage and provision in single panel. On the other hand, with open source data planes, customization to data plane is possible, based on requirements and evolving.

There are commercial versions of NFV that helps Service Providers to build their own NFV network with the support of vendors behind it. This type of NFVI is simpler to manage but your network will always affiliate with their road maps and visions. In addition, they may enforce you to choose their recommended platforms and hardware. They may allow open source components to add in the NFVI either.

Regardless of these problems, let's take a look on commercial NFVIs. In figure 9-16 which is a Cisco solution for NFVI illustrated, there are plenty of components used in every layer of infrastructure and some of them are Cisco proprietary and some of them are open sources. In physical infrastructure, there is Cisco's compute, network, and storage devices. However, in virtual infrastructure, it uses either open source components or vendor supported components.

For example, cisco uses Red Hat OpenStack Platform (OSP) for virtual infrastructure layer and Red Hat Ceph (a unified distributed storage system) for using storage and SD-Storage purpose.

https://t.me/learningnets

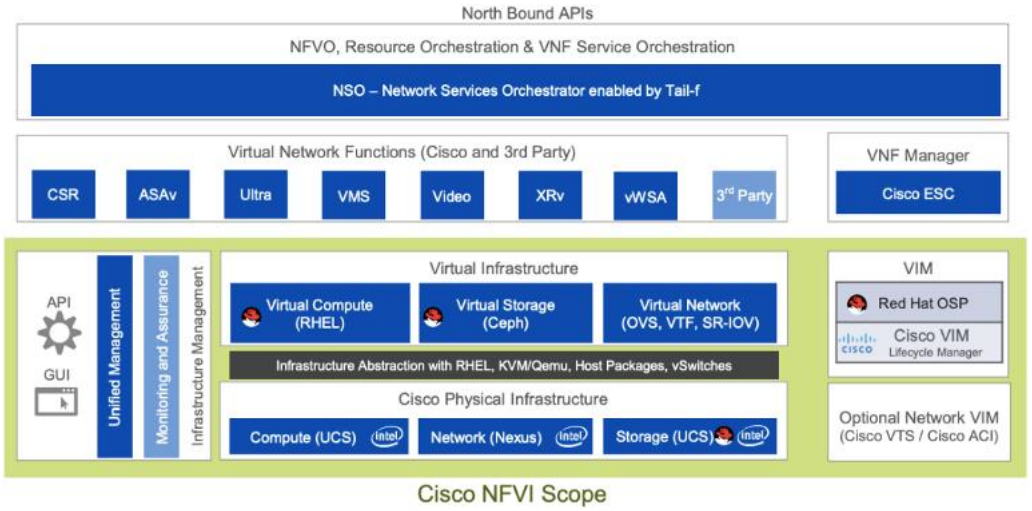


Figure 9-16 *Cisco NFVI Scope (source: www.cisco.com)*

However, in virtual network layer, it uses OVS (open virtual switch) which is an open source software. In VNF layer, Cisco uses its own products like XRv, CSR, ASAv, etc. beside third party products that they are also vendor related. For orchestration layer, Cisco uses Network Service Orchestrator (NSO) that is enabled by Tail-f for addressing MANO and SDN.

Besides, Cisco makes some limitations on technologies like VXLAN. If you need VXLAN, you should follow their instruction and components, then based on the situation you can have VLAN or VXLAN in your network.

NFV Use Cases

NFV in Mobile operators

With NFV, mobile operators can implement their components in virtual manner. Telecom Exchanges/Central Offices that are connected to Aggregation/Pre-Aggregation, contain many VNFs that are needed for existing or future network. Components such as vPE, CRAN, vBNG, etc. are deployed on regional data centers or mini data centers to stretch and provide functionality near to customers. Also, customer equipment has benefits of easy and zero-touch provisioning that decreases complexity and knowledge of implementation.

Cloud-RAN is another topic that forces large mobile operators to convert their physical components such as BBU (Base-Band Unit) to virtual components and place them in central or regional data centers.

https://t.me/learningnets

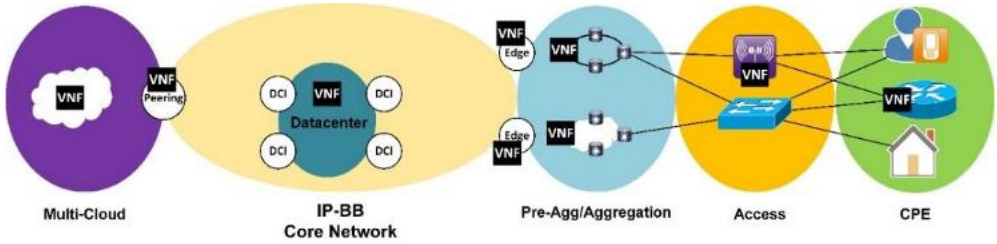


Figure 9-17 *OPNFV*

BGP Router Reflector with NFV

Besides, imagine that whenever load of one component reaches alarm threshold, one copy of same instance can be created by NFVI and load of the component can change to normal.

Another useful case is using Virtual Route Reflector (vRR) instead of dedicated physical RR. In traditional networks, RRs are dedicated physical devices and all services (Address Family) are implemented on them, so if 4 address families are required by Service Provider, with considering HA, they implement 8 dedicated physical hardware. They can implement different services on the same RR but it would create fate-sharing. Thus, when there is a problem with one service, problem effects the other services which are deployed on the same RR.

Also, if the number of routes increases, RR's hardware can be upgraded with RAM and CPU to address the scale. However, with vRR implementation, Service Providers will create per BGP service vRR on only two physical devices. On the other hand, easy provisioning of them is the key benefit in case of failure.

Alongside, because vRR is a VM, virtualization technologies provide dynamic mobility for the VM so it can move from one geographical location to another, instead of an operational service being interrupted. It is quite effective, because not only it increases the availability but also flexibility of your network in the event of failure. In the meantime, while upgrading your service or related packages, you don't need to worry about interruption of your service as you can make a snap shot of your original version or move it to another data center while you face temporary issue in one datacenter.

https://t.me/learningnets

Using BGP for Routing in Large-Scale Data Centers

A Data center (DC) architect can use BGP to build large-scale data centers (DC) that hosts a few applications that are distributed across thousands of servers and scale from department to Internet scale audiences. RFC 7938 has been published to cover operational and some design practices for using BGP in the Datacenter.

Unlike a traditional data center, which hosts applications deployed in silos, the RFC 7938 data center is characterized by a few very large applications that are distributed over geographically-distributed homogenous pools of compute and storage. In essence, these data centers behave less like hosting environments and more like highly optimized computers.

The industry refers to these data centers as warehouse-scale computers (WSC) or MSDC (Massively Scalable Datacenter). Such data centers, also known as "hyper-scale" or "warehouse-scale" data centers, have a unique attribute of supporting over a hundred thousand servers. In order to accommodate networks of this scale, operators are revisiting networking designs and platforms to address this need.

The RFC 7938 is based on operational experience with data centers built to support large-scale distributed software infrastructure, such as a web search engine. The primary requirements in such an environment are operational simplicity and network stability so that a small group of people can effectively support a significantly sized network.

Experimentation and extensive testing have shown that External BGP (EBGP) is well suited as a stand-alone routing protocol for these types of data center applications.

BGP in the Datacenter as IGP – Why not other IGPs?

OSPF and IS-IS were considered and used by some of the Web-scale companies initially before their migration to BGP. But, due to lack of Multiprotocol support by OSPF, lack of good Open source implementation for link state protocols (OSPF and IS-IS) in general and most importantly flooding scope of the link state protocols and blast radius impact of link state protocols, these issues were the biggest factors of using “BGP as IGP” in the Datacenters.

Routing Protocol and Topology Requirements in the Large-Scale Datacenters based on RFC 7938

Massively Scalable Datacenters or commonly known as Massively Scale, Hyper Scale or Warehouse-Scale Datacenters can have 100s of thousands of servers.

https://t.me/learningnets

These datacenters require

Select a topology that can be scaled "horizontally" by adding more links and network devices of the same type without requiring upgrades to the network elements themselves. This requires CLOS topology.

CLOS Topology

In the CLOS topology, there are Leaf and Spine Switches. There are no shortcut links between Leaf or Spine switches.



Figure 9-18 *CLOS Topology*

After understanding what CLOS topology is and why it is necessary for the MSDC (Massively Scalable Datacenters), let's continue with routing protocol requirements in the MSDC Datacenters. Narrow set of software features/protocols supported by a multitude of networking equipment vendors (Open Source Implementation is preferred, ExaBGP, FRRouting etc.) is desired.

One of the requirements of the routing protocol in Massively Scalable Data Center is that it should have a simple implementation in terms of programming code complexity and ease of operational support (State Machine of BGP vs. IGP protocols). Another requirement from the routing protocol is to minimize the failure domain of equipment or protocol issues as much as possible. (IGP Flooding scope, Periodic database refresh vs. BGP Incremental Update)

Traffic Engineering capability is an important requirement for the protocol in this type of Data Center. Small and Big flows can be placed onto different links between the switches (Using BGP community and the centralized controller).

Also in the hyper scale DC requirements, having protocol synchronization (Inter-dependency between protocols) is not wanted, thus having a BGP as single protocol is an important parameter to have simple and less OPEX design.

https://t.me/learningnets

Counter Arguments for using BGP in Datacenter as Routing Protocol

There are some counter arguments for using BGP in the Datacenter as a Routing Protocol. We will examine each one of them separately.

- BGP is perceived as a "WAN-only protocol “and not often considered for enterprise or data center applications
- BGP is believed to have a "much slower" routing convergence compared to IGPs
- BGP is perceived to require significant configuration overhead and does not support neighbor auto-discovery

Counter Arguments for BGP in DC – It is for WAN

BGP is perceived as a "WAN-only protocol “and not often considered for enterprise or data center applications. This is partly true, thus there are some tweaks for BGP to use in the Datacenter as Routing Protocol.

In the WAN networks, expectation from BGP is stability, in the Datacenter, stability is important but rapid notification and convergence is more important. We will look at how convergence speed can be increased and which tweaks are required for BGP to have it in the Datacenter as Routing Protocol.

Counter Arguments for BGP in DC – It is Slow

BGP is believed to have a "much slower" routing convergence compared to IGPs, timers can be tuned, BGP can converge much faster than conventional thoughts. For Fast BGP convergence, there are in general three timers**;** Minimum Route Advertisement Interval, Keepalive and Hold Timers

MRAI Timer

BGP has MRAI (Minimum Route Advertisement Interval) per neighbor. Events within this minimum interval window are collected together and sent at one shot when the minimum interval expires. This is essential for the most stable code, but it also helps prevent unnecessary processing in the event of multiple updates within a short duration such as link flaps.

The default value for this interval is 30 seconds for EBGP peers, and 0 seconds for IBGP peers. However, waiting 30 seconds between updates is not necessary for a densely-connected network such as CLOS topology. 0 is the more appropriate choice for MRAI timers in EBGP in the DC, because we’re not dealing with routers across administrative domains when we deploy EBGP in the Datacenter

https://t.me/learningnets

Keepalive and Hold Timer

By default, the keepalive timer is 60 seconds and the hold timer is 180 seconds. This means that a node sends a keepalive message for a session every minute. If the peer does not see a single keepalive message for three minutes, it declares the session dead. By default, for EBGP sessions for which the peer is a single routing hop away, if the link fails, this is detected and the session is reset immediately.

What the keepalive and hold timers do is to catch any software errors while the link is up but has become one-way due to an error, such as in cabling issue.

Some operators enable Bidirectional Forwarding Detection (BFD) for sub-second detection of errors due to cable issues. However, to catch errors in the BGP process itself, you need to adjust these timers. Inside the data center, three minutes is too much. Recommended values configured inside the data center are 3 seconds for keepalive and 9 seconds for the hold timer.

Lack of BGP Neighbor Auto-Discovery

BGP has been requiring manual neighborship process. This would be operationally very hard when there are 10s of thousands of devices in the data center which we need to enable BGP on top. Neighbor Auto-Discovery can make the BGP neighborship process dynamic, thus provides much lower OPEX for the data centers when they want to deploy BGP as their Routing Protocol. There are two IETF drafts now for BGP neighbor auto discovery:

1. BGP LLDP Peer Discovery
2. BGP Neighbor Discovery Draft

Counter Arguments for BGP in DC – BGP LLDP Peer Discovery

In BGP, neighbor adjacency is configured manually, by putting a neighbor IP address and ASN into the BGP. Neighbor Auto Discovery is a desired future to reduce OPEX when BGP is used inside the Datacenter. This IETF draft can change the behavior of manual adjacency setup, allowing the BGP adjacency on the point-to-point links to be established automatically, using the LLDP protocol. LLDP is an Industry Standard.

It is intended to replace proprietary Cisco Discovery Protocol (CDP). With LLDP, network devices such as switches operating on Layer 2 (Link layer) of OSI model can collect upper layer information of neighboring device, such as IP address, OS version etc. BGP peering discovery using LLDP could be used in large Layer 3 data centers where EBGP is being used as a single routing protocol. Deployment of BGP with enabled BGP peering discovery using LLDP in large data centers in the future would significantly lower the BGP configuration overhead

https://t.me/learningnets

Counter Arguments for BGP in DC – BGP Neighbor Discovery Draft

This approach suggests to change the way of BGP hello message. Instead of using other protocols such as LLDP, the draft introduces a new BGP Hello message. The message is sent periodically on the interfaces where BGP neighbor auto-discovery is enabled to the multicast IP address using UDP port 179. The hello message contains ASN of the sender along with IP address, router id etc.

BGP Path Hunting

When BGP is used in the DC, based on the AS allocation, it might suffer from the BGP Path Hunting behavior. BGP Path Hunting will slow down the convergence when there is a failure in the network, link or node. Let's have a look at the details of BGP Path Hunting before starting the BGP ASN allocation.

Without topology information, a router does not know the physical link state of every other node in the network, it doesn't know whether the route is truly gone (because the node at the end went down itself) or is reachable via some other path. That's why a router proceeds to hunt down reachability to the destination via all its other available paths. This is called path hunting

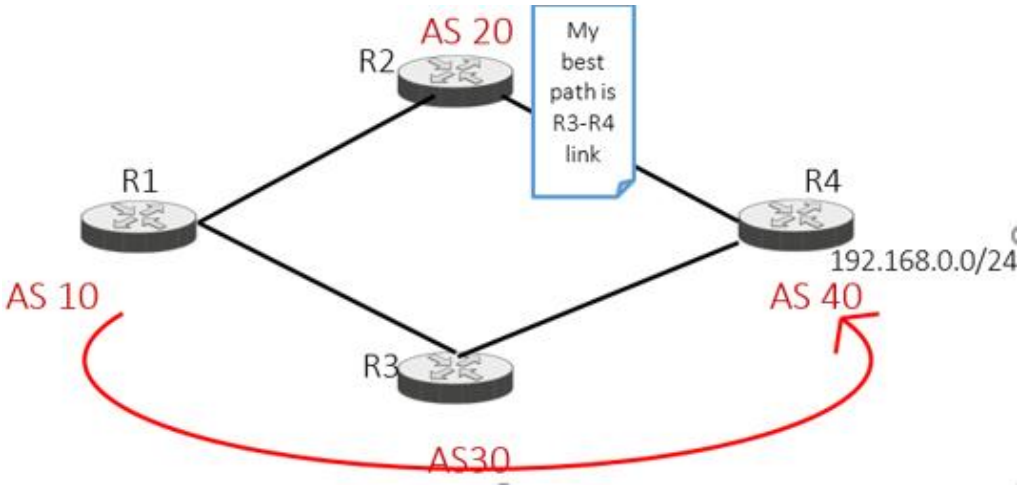


Figure 9-19 *Path Hunting Topology*

In figure 9-19, let's assume R1 selected best path to 192.168.0.0/24 as R3. R1 advertises [R1 R3 R4] As-Path to R2. R2 accepts the advertisement but doesn't use it, as R2 has shorter path to 192.168.0.0/24. Now, when Router R4 fails as it is shown in figure 9-20, R2 loses its best path to 192.168.0.0/24, and so it re-computes its best path via R1, AS_PATH [R1, R3, R4] and sends this message to R1. R2 also sends a route withdrawal message for 192.168.0.0/24 to R1. When R3's withdrawal to route 192.168.0.0/24 reaches R1, R1 also withdraws its route to 192.168.0.0/24 and sends its withdrawal to R2.

https://t.me/learningnets

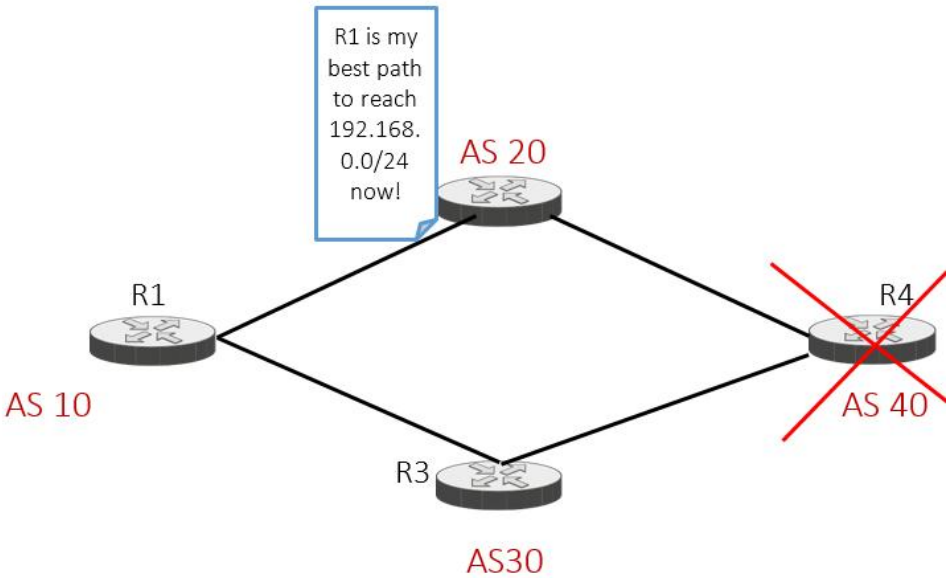


Figure 9-20 *Path Hunting Topology*

It is important to know that, EBGP AS number allocation will trigger path hunting when there is a failure to the destination. Path Hunting will slow down the convergence which is not good for the Datacenter BGP. We will next look at how ASN allocation should happen to avoid BGP Path Hunting behavior when EBGP is used inside the Datacenter

https://t.me/learningnets

ASN Numbering Schema when EBGP is used inside Datacenter

In general, RFC 7938 and some other real world BGP deployments inside the Datacenter follow the below ASN allocation schema when they use EBGP as their IGP protocol inside the DC.

- All ToR (Top of Rack) switches (Sometimes referred as Tier 3 switches) are assigned their own ASN, unique ASN
- Leaf Switches (Tier 2 Switches) inside a POD have a same ASN, but leaves in different PODs have a unique ASN. POD is sometimes referred as Cluster
- Spines share a common ASN, Spine switches sometimes referred as Tier 1 devices

Figure 9-21 represent the recommended AS number allocation schema.

Recommended BGP ASN Allocation Schema for 3 Tier CLOS Networks

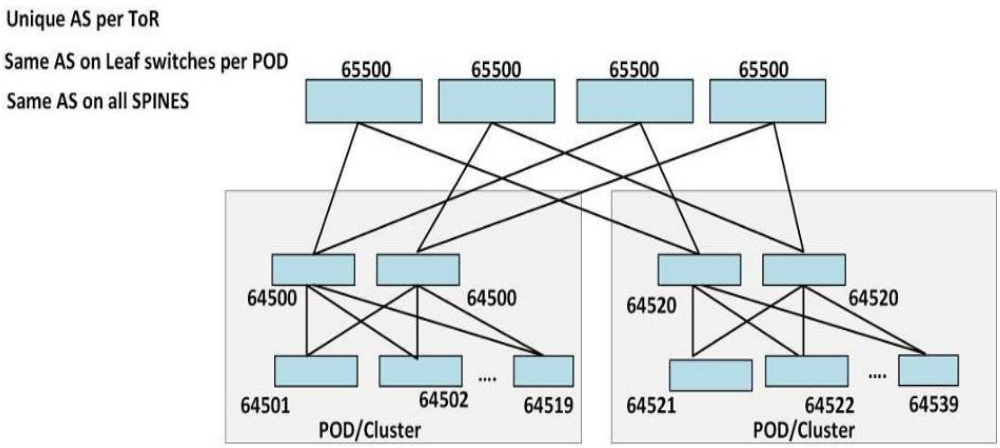


Figure 9-21 *Recommended BGP ASN Allocation Schema for 3 Tier CLOS Networks*

https://t.me/learningnets

Possible Problems with BGP ASN Allocation Schema

We might have some problems with the above ASN allocation schema.
If 2 byte ASN space is used, Private ASN range is recommended as there might be a mistake to leak Public ASN to the Peers or Transit. Since Operator knows most of the public AS numbers of the large companies, separating AS number for DC usage is better for troubleshooting. The problem with using 2 byte ASN Private ASN space is that 2 byte ASN space private range is limited to 1023 AS. There might be much more TOR switches in the DC than 1023 AS. In this case, we have two Options:

Either use 4 byte ASN or assign same AS numbers on different POD/Cluster's TOR switches. 4 byte ASN is not still supported by some BGP implementations thus it might limit the vendor selection of DC equipment. Also, using 4 byte ASNs puts additional protocol complexity. If same AS numbers are used in the different Cluster TOR switches, when there is a traffic between separate Cluster TOR, BGP "allow-as in" feature is required as it is shown in figure 9-22.

Recommended BGP ASN Allocation – Using 2 Byte Private ASN

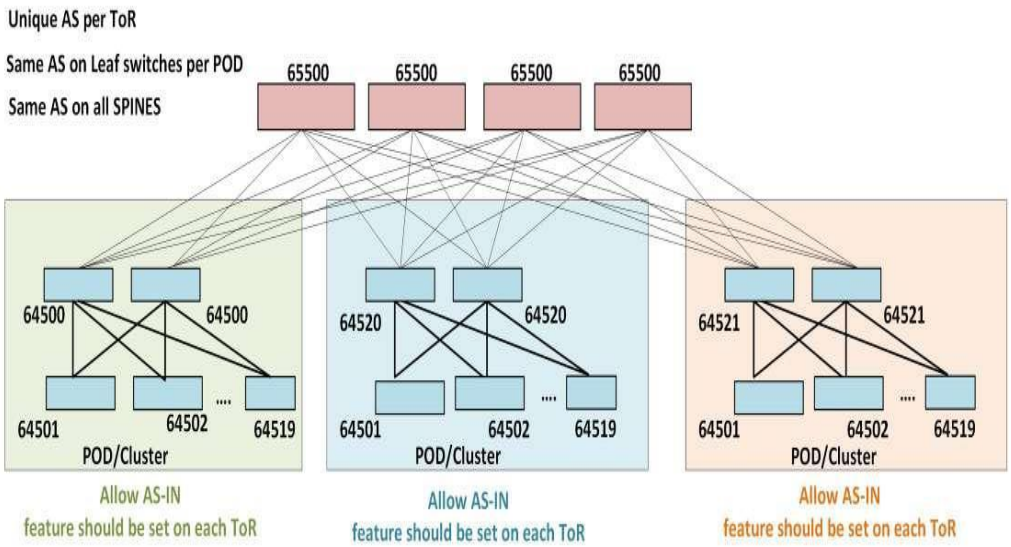


Figure 9-22 *BGP ASN Allocation with 2 Byte Private ASN on 3 stage CLOS topologies*

https://t.me/learningnets

Multicast BIER (RFC8279)

Bit Index Explicit Replication (BIER) is an architecture that provides optimal multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain any multicast related per-flow state. BIER also does not require any explicit tree-building protocol for its operation. So, it removes the need of PIM, MLDP, P2MP LSPs RSVP etc.

A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. The BIER header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by setting the bits that correspond to those routers in the BIER header. The obvious advantage of BIER is that there is no per flow multicast state in the core of the network and there is no tree building protocol that sets up tree on demand based on users joining a multicast flow.

In that sense, BIER is potentially applicable to many services where Multicast is used. Many Service Providers currently investigating how BIER would be applicable for their network, what would be their migration process and which advantages they can get from BIER deployment.

By using BIER header, multicast is not sent to the nodes that do not need to receive the multicast traffic. That’s why, multicast follows an optimal path within the BIER domain. Transit nodes don’t maintain the per-flow state and as it is mentioned above, no other multicast protocol is needed. BIER simplifies multicast operation as no dedicated multicast control protocol for BIER is needed while the existing protocols such as IGP (IS-IS, OSPF) or BGP can be leveraged. BIER uses new type of the forwarding lookup (Bit Index Forwarding Table). It can be implemented by software or hardware changes. Hardware upgrade requirement can be a challenge for BIER but when it is solved, BIER can be the single de-facto protocol for Multicast.

https://t.me/learningnets

ACRONYMS AND ABBREVIATION

AMS-IX: Amsterdam Internet Exchange Point
AON: Active Optical Network
AR: Access Router
AS: Autonomous System
ASBR: Autonomous System Boundary Router
ASN: Autonomous System Number
ATM: Asynchronous Transfer Mode
ATS: Automatic Transfer Switches
AWN: ATELCO Worldwide Network
BFD: Bidirectional Forwarding Detection
BGP: Border Gateway Protocol
BIER: Bit Indexed Explicit Replication
BNG: Broadband Network Gateway
BPL: Broadband over Power Line
BR: Border Router
BRAS: Broadband Remote Access Server
BSS: Business Support System
BUM: Broadcast Unknown Unicast Multicast
CAPEX: Capital Expenditures
CPE: Customer Premises Equipment
CDN: Content Delivery Network
CDP: Cisco Discovery Protocol
CGNAT: Carrier Grade Network Address Translation
CLS: Cable Landing Station
CMTSL: Cable Modem Termination System
CO: Central Office
CSP: Cloud Service Provider
CSR: Cell Site Router
CWDM: Coarse Wavelength Division Multiplexing
DDOS: Distributed Denial of Service
DNS: Domain Name System
DOCSIS: Data Over Cable Service Interface Specification
DSL: Digital Subscriber Line
DSLAM: Digital Subscriber Line Access Multiplexer
DWDM: Dense Wavelength Division Multiplexing
EBGP: External Border Gateway Protocol
EPE: Egress Peer Engineering

https://t.me/learningnets

EPON: Ethernet Passive Optical Network
ETSI: European Telecommunication Standard Institute
EVPN: Ethernet Virtual Private Network
FCC: Federal Communications Commission
FRR: Fast Reroute
FTTX: Fiber to the X (Home, Premises, Curb...)
GEO: Geostationary Earth Orbit
GGC: Google Global Cache
GGSN: Gateway GPRS Support Node
GPON: Gigabit Passive Optical Network
GPRS: General Packet Radio Service
GRE: Generic Routing Encapsulation
GRX: GPRS Roaming Exchange
GSM: Global System for Mobile
HFC: Hybrid Fiber-Coaxial
HK-IX: Hong Kong Internet Exchange Point
HVAC: Heating, Venting, Air Conditioning
IAAS: Infrastructure as a Service
IBGP: Internal Border Gateway Protocol
IEEE: Institute of Electrical and Electronics Engineers
IETF: Internet Engineering Task Force
IGW: Internet Gateway
IPLC: International Private Leased Circuit
IPTV: IP Television
IRU: Indefeasible Right of Use
IS-IS: Intermediate System to Intermediate System
ISP: Internet Service Provider
ITU: International Telecommunication Union
IXP: Internet Exchange Point
LAC: L2TP Access Concentrator
LAN: Local Area Network
LDP: Label Distribution Protocol
LEO: Low Earth Orbit
LER: Label Edge Router
LFA: Loop Free Alternate
LINX: London Internet Exchange Point
LLDP: Link Layer Discovery Protocol
LNS: L2TP Network Server
LSA: Link State Advertisement

https://t.me/learningnets

LSN: Large Scale Network Address Translation
LSP: Label Switch Path
LSR: Label Switch Router
LTE: Long Term Evolution
L2TP: Layer 2 Tunneling Protocol
LU: Labeled Unicast
MANO: Management and Orchestration
MEO: Medium Earth Orbit
MIM: Man in the Middle
MLPE: Multi-Lateral Peering Exchange
MME: Mobility Management Entity
MMF: Multi-Mode Fiber
MNO: Mobile Network Operator
Mobile AR: Mobile Access Router
MP-BGP: Multi-Protocol Border Gateway Protocol
MPLS: Multi-Protocol Label Switching
MRAI: Minimum Route Advertisement Interval
MSAN: Multi-Service Access Node
MSC: Mobile Switching Center
MSDC: Massively Scale Datacenter
MSO: Multiple Systems Operator
MTU: Maximum Transmission Unit
MVNO: Mobile Virtual Network Operator
NAT: Network Address Translation
NFV: Network Functions Virtualization
NFVI: Network Functions Virtualization Infrastructure
NMS: Network Management System
NNI: Network to Network Interconnection
NOC: Network Operation Center
NTP: Network Time Protocol
OCP: Oracle Cloud Platform
ODN: Optical Distribution Network
OLT: Optical Line Terminal
ONU: Optical Network Unit
OPEX: Operational Expenditures
ORR: Optimal Route Reflection
OSP: Openstack Platform
OSPF: Open Shortest Path First
OSS: Operations Support System

https://t.me/learningnets

OTT: Over the Top
PAAS: Platform as a Service
PBB: Provider Backbone Bridging
PCC: Path Computation Client
PCE: Path Computation Element
PCEP: Path Computation Element Protocol
PCH: Packet Clearing House
PE: Provider Edge
PFE: Power Feed Equipment
PGW: Packet Gateway
PIC: Prefix Independent Convergence
PNI: Private Network Interconnection
PON: Passive Optical Network
POP: Point of Presence
P2P: Point to Point
P2MP: Point to Multi Point
PSTN: Public Switched Telephone Network
PTP: Precision Time Protocol
PW: Pseudo Wire
QoS: Quality of Service
RPKI: Resource Public Key Infrastructure
RR: Route Reflector
RSVP: Resource Reservation Protocol
RTBH: Remotely Triggered Blackhole
RTT: Round-trip Time
SAAS: Software as a Service
SDH: Synchronous Digital Hierarchy
SDN: Software Defined Network
SFI: Settlement Free Interconnection
SGSN: Serving GPRS Support Node
SGW: Serving Gateway
SID: Segment Identifier
SLA: Service Level Agreement
SLTE: Submarine Line Terminal Equipment
SMF: Single Mode Fiber
SONET: Synchronous Optical Network
SP: Service Provider
SR: Segment Routing
SRTE: Segment Routing Traffic Engineering

https://t.me/learningnets

SSM: Source Specific Multicast
SYNCH-E: Synchronous Ethernet
TCP: Transmission Control Protocol
TDM: Time Division Multiplexing
TE: Traffic Engineering
TI-LFA: Topology Independent Loop Free Alternate
TLD: Top Level Domain
TOD: Time of Day
TOR: Top of Rack
UDP: User Datagram Protocol
UMTS: Universal Mobile Telecommunication System
UPS: Uninterruptable Power Supply
VDSL: Very High Speed Digital Subscriber Line
VLAN: Virtual Local Area Network
VNF: Virtual Network Function
VPLS: Virtual Private Lan Service
VPN: Virtual Private Network
VPWS: Virtual Private Wire Service
VSAT: Very Small Aperture Terminal
VTEP: VXLAN Tunnel End Point
VXLAN: Virtual Extensible Local Area Network
WAF: Web Application Firewall
WAN: Wide Area Network
WCL: Worldwide Connectivity Layer
WDM: Wavelength Division Multiplexing
WISP: Wireless Internet Service Provider
WWAN: Wireless Wide Area Network °°°

https://t.me/learningnets

Service Provider Networks - Design Architecture Perspective

About this book

- This book will give you a High Level of overview of the Service Provider Network Design and Architecture.
- It talks about the unique aspects of Service Provider networks, different types of Service Providers and the business relationships between them.
- It covers the Service Providers services, different access last mile offerings and transport networks, and their subscribers and services.
- Technical explanation about different types of Fixed and Mobile network services and the Service Provider physical locations are also explained.
- You will see the Big Picture of Service Provider Networks.

After understanding the Service Provider Concepts and Technologies, a fictitious National Service Provider network, named ATELCO will be introduced, to give you a more view of the technologies, protocols, services and end to end traffic flow in detail. And at last the Evolving Technologies used in Service Providers and Massively Scale Datacenters will be seen.

orhanergun.net

About the Author



Orhan Ergun, CCIE/CCDE Trainer, Author, Network Design Advisor and Cisco Champion 2019. Orhan Ergun is award winning Computer Network Architect, CCDE Trainer and Author. has well known industry certificates CCIE #26567 and CCDE #20140017. has more than 17 years of networking experience and has been working on many medium and large-scale network design and deployment projects for Enterprise and Service Provider networks.He has been providing consultancy services to African, Middle East and some Turkish Service Providers and Mobile Operators for many years. Orhan has been providing Cisco network design training such as CCDE, Pre-CCDE, Service Provider Design and many Advanced technologies for many years, and created best CCDE Training Program to share his network design experience and knowledge with the networking community.Orhan is sharing his articles and thoughts on his blog www.orhanergun.net. All the training and consultancy services related information can be found from his website. Orhan has a Training and Consultancy company located in Istanbul, Turkey.

OrhanErgun
orhanergun.net

Scan me

https://t.me/learningnets