SCTE STANDARDS JOURNAL

VOLUME 2 NO. 2 JULY 2022

SCTE TECHNICAL JOURNAL

SCTE
a subsidiary of CableLabs®

SCTE STANDARDS JOURNAL

SCTE
a subsidiary of CableLabs®

SCTE TECHNICAL JOURNAL

VOLUME 2, NUMBER 2

July 2022

Society of Cable Telecommunications Engineers, Inc.
140 Philips Road, Exton, PA 19341-1318

Table of Contents

Editorial Correspondence: If there are errors or omissions to the information provided in this journal, corrections may be sent to our editorial department. Address to: SCTE Journals, SCTE, 140 Philips Road, Exton, PA 19341-1318 or email journals@scte.org.

Submissions: If you have ideas or topics for future journal articles, please let us know. All submissions will be peer reviewed and published at the discretion of SCTE. Electronic submissions are preferred, and should be submitted to SCTE Journals ,journals@scte.org.

Foreword

Every year, the increasing internationalization of cable continues to add new depth and reach to our industry. The opportunity to bring solutions to new markets – or to learn from operators around the world – is part of the spirit of collaboration that makes cable technology so dynamic and exciting.

When SCTE Cable-Tec Expo® returns to Philadelphia Sept. 19-22, one-eighth of the speakers in our Fall Technical Forum sessions will be from overseas. Similarly, three of the seven article contributors to the quarterly SCTE Technical Journal bring a Canadian or a European perspective to our global audience.

Here's what's on tap for the summer issue:

- "Towards Negative Emissions in Cable Networks," by Circonica Circular Energy B.V.'s Diederick Jaspers and Coen Faber and Liberty Global's Sam Khola.
- "The Sustainable Future of Energy Storage for Cooling, Peak Power Balancing, and Renewable Energy," by Energy Cool Apps' Henrik Thorsen and Donné Ros and Tizzon's Ivan Roubos.
- "HFC Network Powering Utilizing Lithium Iron Phosphate Batteries," by Lindsay Broadband's Len Visser.
- "Pathway to a Greener, More Energy Efficient Cable Access Network," by Intel's Rory Sexton, David Coyle and Eric Heaton.
- "A Step-by-Step Methodology for Optimal Access Transformation Planning," by First Principles Innovation's Sudheer Dharanikota and Luc Absillis and Kyrio's Mario Di Dio.
- "MSOX–MSO eXchange: A Converged Data Roaming Platform for 5G, 4G, and Wi-Fi Networks," by CableLabs' John Kim and Omkar Dharmadhikari.
- "Workforce Shortage? Technology to the Rescue," by AFL's Michael Scholten.

"AI Pairing in the Contact Center," a letter to the editor by Afiniti's Fanny Heneine, Theresa Gebert, Madeleine Lamm, and Eleanor Mcdonald, rounds out the contents.

You may have noted that the first four articles on the list continue SCTE's Energy 20/20 thought leadership around powering management and sustainability. "Towards Negative Emissions in Cable Networks" is particularly noteworthy, as it is an update on the solid oxide fuel cell technology that won the Adaptive Power Challenge at Cable-Tec Expo in 2018. We hope you enjoy this and all of the contributions from our cable brethren in this current issue of the SCTE Technical Journal, and we look forward to seeing you when SCTE Cable-Tec Expo returns to a live format – in Philadelphia – with hybrid elements in September.

SCTE STANDARDS JOURNAL

Towards Negative Emissions In Cable Networks

A Technical Paper prepared for SCTE by

Diederik B.C. Jaspers, CTO, Circonica Circular Energy B.V.
4, Robert Boyleweg
Leiden, 2333 CG
The Netherlands
jaspers@circonica.com
+31 628262617

Coen Faber, Project Manager Cable Industry, Circonica Circular Energy B.V.
4, Robert Boyleweg
Leiden, 2333 CG
The Netherlands
faber@circonica.com
+31 653414203

Sam Khola, Director Corporate Responsibility and Sustainability, Liberty Global
53, Boeing Avenue
Schiphol Rijk, 1119 PE
skhola@libertyglobal.com

Cok Arkesteijn, Architect HFC, Liberty Global
53, Boeing Avenue
Schiphol Rijk, 1119 PE
coarkesteijn@libertyglobal.com

Derek DiGiacomo, Senior Director Energy Management Programs, 610-955-7276 SCTE Member
140 Philips Rd
Exton, PA 19341
ddigiacomo@scte.org

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

1. Introduction

As the need for power and energy grows in most industries, the natural world is being robbed not only of its finite resources but also of the ecosystem and climate balance that sustains all known and undiscovered life. This delicate balance is put into danger as we continue extracting resources, release greenhouse gases into the atmosphere, and pollute air, sea, and land through our linear consumption lifestyle. By mining raw material, manufacturing, shipping, consuming, then discarding finished products, we are depleting the only world we have of its resources. In the process we are also tipping such a beautiful wonderworld into an unliveable planet. If we focus the human brain, probably the biggest wonder of the natural evolution, we will be able to solve the growing energy challenge and with it even a lot of the geopolitical conflicts that have held many nations hostage.

Fuel cells could play a big role either in a standalone or hybrid combination with other technologies to solve the energy challenges of today. The first commercial use of fuel cells following the invention of the hydrogen–oxygen fuel cell was in 1932. However, since NASA's use of alkaline fuel cells in the mid-1960s to generate power for satellites and space capsules, fuel cells have been widely used in many other applications. They are used for primary and backup power for commercial, industrial, and residential buildings and in remote or inaccessible areas. They are also used to power transports, including forklifts, automobiles, buses, boats, motorcycles, and submarines. Although fuel cells were invented in 1838 and are the most efficient continuous power source, three factors have impeded widespread use: the abundance of cheap fossil fuels for internal combustion engines (ICEs); the high price tag of fuel cells; and the cost and safety of hydrogen as fuel. However, attention has turned to green hydrogen and fuel cells yet again to help tackle climate change, the most pressing issue of the century humankind faces.

Any innovation that can bridge or revolutionise today's pressing energy questions will be the most important milestone since the industrial revolution. That is arguably also the ultimate achievement of humankind.

The cable sector experiences rising energy costs and energy demand to power the many services seen as essential for many if not all sectors of society and businesses. Under the SCTE Energy 20/20 program, a reduction in energy consumption was targeted in order to achieve its efficiency and sustainability goals.

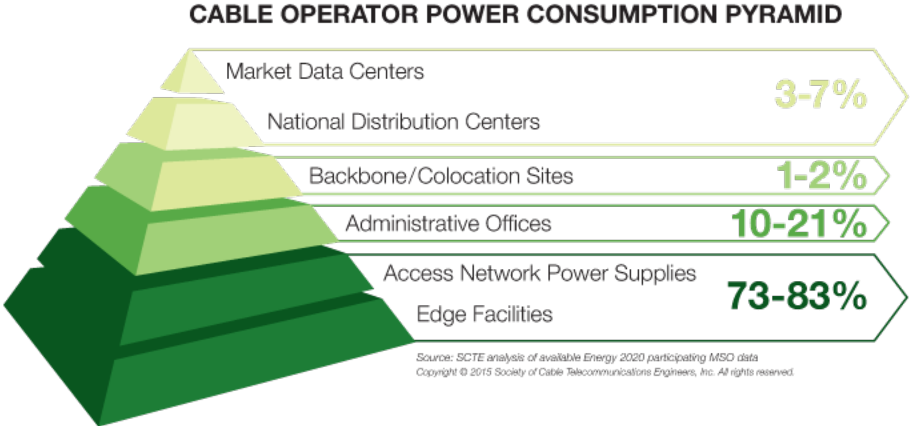


Figure 1 - SCTE Energy20/20 Program – power pyramid of cable operators circa 2015

The goals were set to help focus the minds of the wider industry, including cable operators and the supply chain community, to collectively solve the key material topic of energy. Our industry's main source of emissions is energy and hence the strategic choice to double down efforts.

The cable industry has a fiber rich network reflected in its infrastructure name -- hybrid fiber coax (HFC). It is a distributed network with multiple layers and components; the majority are small sites housed in small facilities and street cabinets located close to the end user. The few larger facilities such as headends and data centers are located in central locations managing the overall network operation. Addressing the bottom row of the pyramid diagram in Figure 1 (2015 version) and in Figure 2 (the latest version) will be important when looking for energy savings and potential emissions reduction strategies. Advances in fuel cell technology could play a key role in specifically addressing this section of the network. Due to small size and ease of scalability the solution could be deployed at the nodes or power injection points where the power supply resides.

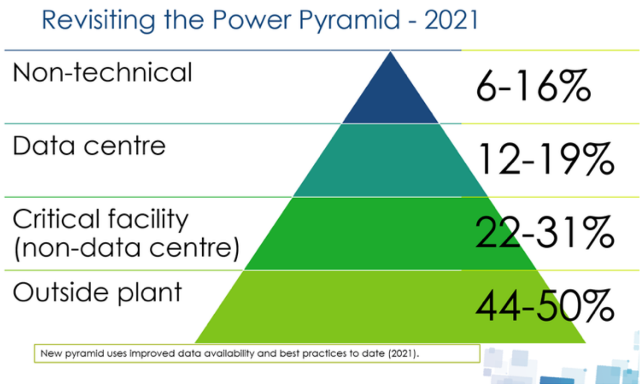


Figure 2 - The revisited power pyramid in 2021

The concept of SOFC HELP fuel cells to achieve negative emissions for cable networks was first presented in 2018, resulting in winning of the Adaptive Power Challenge Prize in the Breakthrough category of the campaign. This was further developed by the fuel cell company Circonica Circular Energy BV.

This white paper describes the possible options towards a negative emission energy system by presenting different practical use cases formulated by the major stakeholders in the cable sector. In this paper we will be doing in-depth analyses for two use cases when deploying solid oxide fuel cell (SOFC) hollow electrodes and loose plates (HELP) fuel cells in the power systems of a cable network.

2. SOFC HELP Fuel Cells

SOFC HELP fuel cells are devices that convert fuels and gases directly into electrical power and heat. They are a bit like batteries that can recharge themselves continuously. The absence of moving parts also makes them silent with less wear and tear than that of combustion-based engines.

The SOFC HELP fuel cell is an innovative circular fuel cell. Its innovative stacking gives it some unique properties, including lifetime of approximately 10 years. It can operate flexibly with multiple fuels and gases (fossil and bio), is modular, and scales to larger power volumes. It has a loose ceramic stacking which enables circularity. It is three times more energy efficient than an internal combustion engine and can reduce approximately 70% of CO_2 when operating on fossil fuels or gasses.

Although fuel cells have existed for a long time and are in use in all US space flights, several factors have blocked widespread application in more commercial applications. The alkaline types designed for space and the polymeric types proposed for the automotive sector are hampered because they use compressed ultrapure hydrogen gas. While the high temperature ceramic type allows feeding with more practical fuels and gases their applications were hampered by high cost and limited lifetime. The SOFC HELP fuel cell has solved these two issues by utilizing existing mass production techniques of injection molding and loose stacking of the fuel cell parts. It is made of one material, ceramic, which can be recycled; it is flexible with fuels and gasses, both fossil and non-fossil; and it can operate with non-purified hydrogen.

An generic feature of SOFC fuel cells is that the heat produced is also of high temperature. This enables useful application of this heat for heating, but also for heat into cooling solutions by using off-the-shelve absorption coolers, hence “usable” heat (see explanation in section 4.4.1). This results in clean power generation at high efficiencies.

The innovative capabilities of stacking of the electrodes makes it possible to design the SOFC HELP fuel cell for different required power volumes. In co-creation with cable operators and their operational managers we identified several use cases and a most preferable power size to start with of 0.4 kW to 1 kW.

Figure 3 - Impressions of SOFC HELP fuel cell with stack of electrodes of 250 W with isolation box.
(Maximum size will be a fuel cell of 1 kW with 4 stacks of 250 W).

3. Towards a Negative Emission Energy System for Cable Networks

3.1. Key Performance Metrics

The main criterion for the degree of sustainability is the net emission reduction of greenhouse gases in the cradle-to-grave approach of the energy carrier.

Other important criteria for cable networks are:

- Uptime and reliability increases of the networks;
- Energy efficiency and energy backup flexibility (operating on different fuels); and
- Circularity of product materials (coax cables, cabinets, fuel cells, etc.) and of CO_2 (capture and re-use).

3.2. Energy-Efficient and Greenhouse Gas (GHG) Emissions-Reducing Energy System

The challenge for cable networks is to come to more energy efficient networks with a lower amount of greenhouse gas emissions. As was mentioned in the introduction and shown with the power pyramids in Figure 1 and Figure 2, there is a large energy consumption in the cable networks at the bottom of the pyramid. The energy consumption consists mainly of fossil energy which creates high greenhouse gas emissions like CO_2, PM and NOx. Backup power is commonly provided by diesel gensets, which also give the negative side effect of noise.
To realize reduction of energy consumption and lower greenhouse gas emission the application of fuel cells seems to be a very promising solution. These could be the hydrogen-fed fuel cells, like protein exchange membrane (PEM) and alkaline fuel cell (AFC), or the SOFC-type fuel cell which generates power at a high efficiency from various fuels (fossil or bio). The high fuel efficiency of an SOFC-type fuel cell inherently can result in a significant reduction of emissions of CO_2 per kWh produced, and they also operate noiselessly.

3.3. Architecture of Cable Network Energy System

Due to the pandemic resulting from the COVID-19 virus, flexible ways of working, such as working-from-home, have become the new normal. As a result, unobstructed access to information technology is critical now more than ever. Power failures, which cause partial outages of the cable TV network, are therefore even more disruptive. To reduce these disruptions batteries can be placed at various locations to take over supply of electricity in case of a power failure in different facilities such as hubs, headends and street cabinets. Temperatures in street cabinets can vary greatly; high temperatures lead to a sharp decrease in life expectancy and low temperatures can reduce battery capacity. However, *fuel cells* do not have these disadvantages. But the question is – will a fuel cell fit in a street cabinet?

The current situation of a cable energy system is depicted in the Figure 4 in a simplified overview.

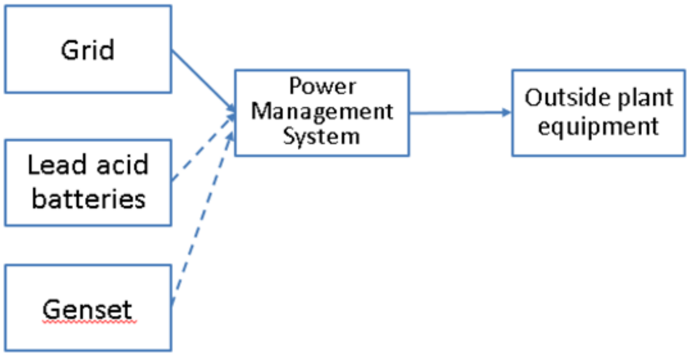


Figure 4 - Base case power system in current situation.

When the grid is active, the power management system regulates power to the outside plant equipment and charges the lead-acid batteries when necessary. In case the grid fails, the power management system switches to the lead-acid batteries to take over the energy supply. But these have limited capacity, so when these batteries get empty, a temporary genset must be connected in order to keep the equipment operational or else the system goes into blackout.

3.3.1. General Description of the Energy Power of a Hybrid Fiber Coax Network in Europe

A hybrid fiber coax network (HFC) is composed of a fiber core network and a coaxial access network. The HFC access network consists of a node with amplifiers downstream from it. These amplifiers get their power through the coax cables from the node location –typically a street cabinet. The power supply for the node and the amplifiers after it delivers on average 400 to 800 W depending on the network structure.

In many cases four batteries are placed in the cabinet to take over the power supply when there is a power failure. In the UK for example, usually four SBSJ70 batteries are used. These 12 V batteries each have a capacity of 64 AH and collectively can take over the power supply for 3.7 hours at a power consumption of 800 W. The total volume of the four batteries is 33*17*18 cm ~41.000 cm^3.

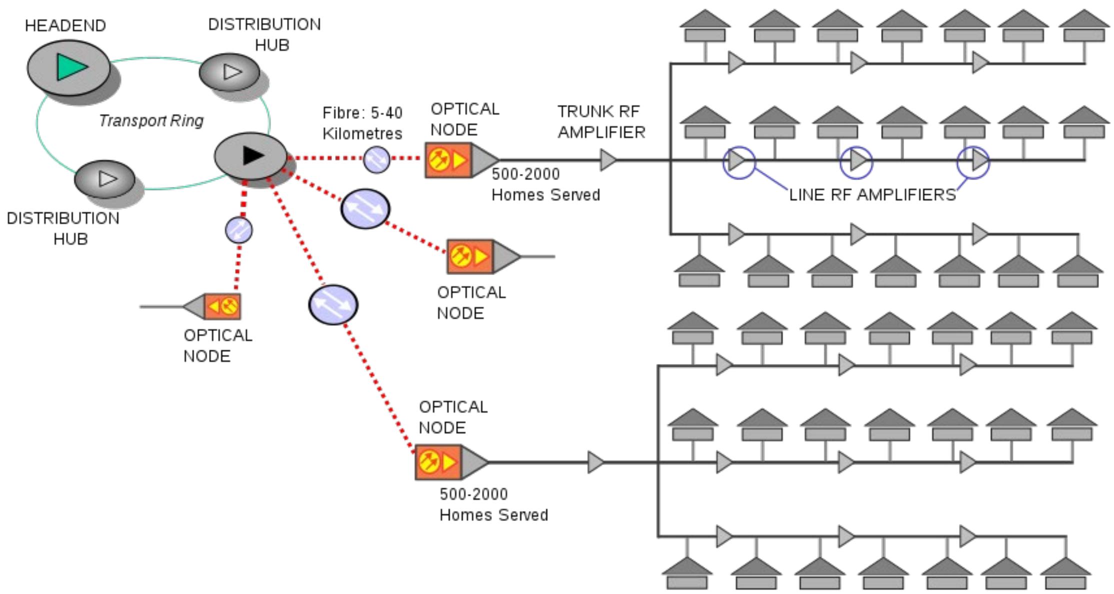


Figure 5 - Example of HFC cable network

There are close to a million of hubs and nodes in European networks.

4. Use Cases of Energy Power System for Cable Networks with SOFC HELP Fuel Cells

In the last years an extensive analysis of feasibility of the usage of SOFC HELP fuel cells for powering cable networks has been done with major cable broadband providers and the SCTE Energy 20/20 team. In Europe, by coordination of Liberty Global, an analysis of use cases for adoption of the SOFC HELP fuel cell has been done in co-creation with responsible European country managers for their respective cable networks in Europe.

The following use cases have been defined and feasibility analyzed for energy power with the application of SOFC HELP fuel cells:

Use case 1 - Inside plant energy optimization (24/7 and the grid as a backup) - critical facilities
Use case 2 - Outside plant energy optimization (replacement of batteries)
Use case 3 - Climate technology (heat for cooling)
Use case 4 - Potential of solutions with CO_2 re-use
Use case 5 - Powering in the home (via home devices)

Use cases 1-3 are possible in coming years with the final development of the SOFC HELP fuel cell and are related to energy power of the existing cable networks. Use cases 4 and 5 are more future oriented and are related with external energy systems of the automotive and home industries. The organizations of these sectors need to be involved in creating solutions for implementation.

The following sections will give insights in the analysis of use cases 2 and 3 for the application of SOFC HELP fuel cells for energy power of cable networks.

4.1. Use Case 2 - Outside Plant Energy Optimization / Backup Power with SOFC Fuel Cell

In this use case the grid feeds the outside plant equipment under normal conditions. But when the grid fails the power system is comprised of a backup power system in the form of an SOFC fuel cell in combination with a small battery for peak shaving and/or bridging of fuel cell start-up time. Fuels for the cell can be comprised of natural gas from the grid supplier or locally stored diesel. The fuel cell acts as a standby instead of the current maintenance-sensitive lead acid batteries, General problems of lead acid batteries are their short lifetime and that they are attractive to steal. Using SOFC fuel cells with a long lifetime can avoid extra rounds of replacement of batteries by service personnel.

In this use case, the amount of kWh of backup is now sized by the fuel tank instead of the battery pack and much longer outage times can be bridged, with four to eight or more hours of backup.

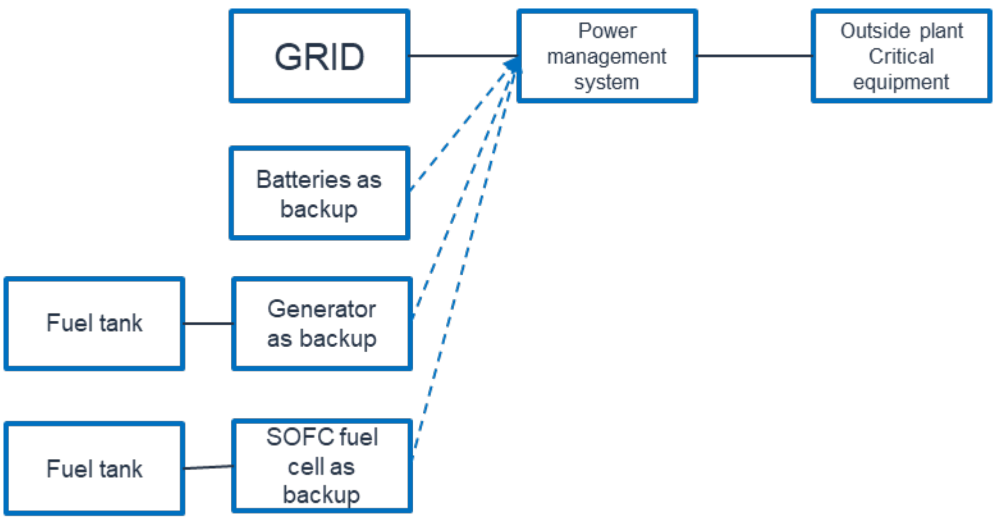


Figure 6 - Outside plant energy optimization (replacement of batteries) with SOFC HELP fuel cell

Table 1 - Hub/node specification in European cable network of energy demand, emissions, and other requirements

Topics	Value
1. Power demand - average (Watt)	400-1000 W
1a Continuous or modulated	Continues
2. Reliable network issue (outage – 24/7)	>99% reliable
3. Heating required (Y/N)	N
4. Cooling required (Y/N)	Sometimes
5. Type of fuels (available and preferred)	All hydrocarbons fuels, like; Diesel, Natural Gas, (bio)Ethanol, Butane and Hydrogen
5a Natural gas grid available (Y/N)	Sometimes
6. Emissions (CO_2, NO_x) critical (Y/N)	Y
7. Noise critical issue (Y/N)	Y

4.1.1. Average Energy Demand of Nodes in European Cable Networks

As customer data consumption increases, the capacity of the network needs to be increased and, in many cases, the network behind the node is split and a node is added to increase the capacity. This results in a lower average power consumption per node. In countries with a lot of deep fiber nodes with a limited number of amplifiers behind each node, the average power consumption per node is relatively low.

Table 2 - Average power of cable networks in Europe (countries A-G)

Network Architecture	Average Power / Node [W]*
A	762
B	97
C	200
D	432
E	412
F	380
G	645

Hubs were created in the past to create more energy supply and reliability in the network. Hubs have more equipment, such as a generator to produce energy, which ultimately make the hubs release a lot of noise as well as emissions. Therefore, permits are needed to put them in certain locations. Generators are traditionally used as back-ups in headends but need to be started and tested a few times a year or even monthly because, they become less dependable if they are not exercised often.

If fuel cells were used in the hubs or nodes they would create no noise and would release less or zero emission (depending on the type of fuel).

The benefits of SOFC HELP fuel cells are significant; specifically with regards to space and business continuity/service uptime when compared to batteries and generators. Table 3 highlights the comparisons.

Table 3 - Comparison of back-up power between batteries, generators and SOFC HELP fuel cells

	Batteries	**Generator**	**SOFC HELP fuel cell**	
	-	Petrol	Diesel	Hydrogen
		Sizing		
Size		Approx. 510 x 290 x 430	350x350x315	350x350x315 660x325x2x3.14
Space for 8 hours back up time	80 dm^3	120 dm^3+ 4.3 dm^3 tank @ ~15% electrical efficiency	1,3 dm^3 tank	7.6 dm^3 @ 700 bar tank
Impact space, extending backup time 2x	High 2x number of batteries	Low more fuel	Low more fuel	Low more fuel
Fits in cabinet	Yes	No	Yes	Yes
		Operational Efficiency		
Reliability	Average	Average	High	High
Lifespan	3-5 years	5-10 years	10+ years	10+ years
Maintenance	High (Regular capacity check)	High (Check oil, cooling and moving parts)	Low	Zero
Sound/noise	Zero	High	Zero	Zero
Efficiency charge/discharge	85%	40%	50%	60%
Start-up time	Immediately available	One minute to start the generator.	A few hours to cold start. Normal	A few hours to cold start. Normal

	Batteries	Generator	SOFC HELP fuel cell	
			operation: Approx. five minutes to ramp up from idle to maximum power. *	operation: Approx. five minutes to ramp up from idle to maximum power. *
Capacity loss during inactivity	Yes	No	No	No
* *Combination (hybrid) with small battery provides direct start-up.*				
Sustainability				
Local Emissions	Very low (H_2)	High CO_2, NO_x, SO_x, PM/soot	Low (only pure CO_2 and water)	Zero (only water)
Able to collect generated CO_2	Not applicable	Difficult and very expensive	Relatively easy	Not applicable
Cooling through absorption	No	No	Yes	Yes

4.1.2. Calculation of Backup Power by SOFC HELP Fuel Cells vs Batteries

A single SOFC HELP fuel cell typically can deliver 15 W at 0.6 V at full power. To deliver 800 W, 54 cells are needed, generating about 32 V. This voltage is readily converted to 48 V with a simple DC-DC converter, causing minimal losses (97% efficient). An advantage is that the DC-DC converter also acts as a voltage stabilizer. If the number of cells is increased, a voltage of 48 V can be reached. This requires 80 cells, which together can deliver up to 1200 W at higher CAPEX costs. The dimensions of a stack (with 80 cells) are 54 cm high with a diameter of 20 cm. The fuel cell takes up a space of 3.14*10*10*54 = 17,000 cm^3. In addition, the local fuel must also be stored. Diesel has an energy density of 10.555 kWh/liter. Assuming the fuel cell has 50% electric efficiency, the fuel cells need 5 kWh or 0.5 liters of Diesel for 3.1 hours of 800 W backup time.

However, if we double the backup time, we see the big advantage of the high energy content of diesel. In the case of the fuel cell, 0.5 liters of extra diesel fuel is enough for 6.2 total hours backup time. To achieve the same backup time with batteries, the number of batteries must be doubled. The volume will then increase by 33,000 cm^3. Therefore, especially with high power and long back up times, the amount of space needed with fuel cells is much lower due to the very high energy density of diesel or of another liquid fuel.

4.1.3. Flexibility of SOFC HELP Fuel Cells and Performance with Hydrogen

As explained in section 2, the SOFC HELP fuel cells are flexible with fuels and gas for operation. This gives several options for optimization and cleaner solutions of the energy system with a fuel cell in a cable network.

For the use case in general. *natural gas with the SOFC HELP fuel cell* is a logical choice when a natural gas grid is available because it has continuous supply and does not need refueling.

Although diesel has the highest energy density, bioethanol is easy to access in lots of markets and, as a biofuel, reduces CO_2 footprint. It is applicable in many use cases for cable networks.

For the SOFC HELP fuel cell, diesel can readily be replaced by biodiesel, providing high impact on emission reduction, improving progress towards carbon neutral cable networks.

The SOFC HELP fuel cell shows superior performance with hydrogen if available.

4.1.4. Example Application of SOFC Fuel Cell with Hydrogen

Hydrogen has an energy content of 120 MJ/kg which equals 33.3 kWh_{chem}/kg.
For 3 kWh_e, 0.15 kg of compressed hydrogen is needed at 60% SOFC HELP fuel cell efficiency. Hydrogen is stored at various pressures of 300 bar up to 700 bar. In transportation, the energy density is more critical than in small stationary applications and 700 bar is used, but a disadvantage of this high pressure is that the system requires a more expensive compressor and higher energy consumption for compression. Storage of 0.15 kg Hydrogen at 700 bar requires approximately 4 liter and 7 liter at 300 bar.

Use of green hydrogen produced from sustainable power sources reduces the CO_2 footprint.

4.1.5. Conclusions Use Case 2 - Outside Plant Energy Optimization with SOFC HELP Fuel Cells as a Backup

SOFC HELP fuel cells as a backup for outside equipment powered by risk prone grids (with more than two to four hours outage a day) is an advantage and looks like a promising use case.

In reliable cable networks with a 24/7 grid, using SOFC HELP fuel cells as a backup instead of gensets is very promising, especially for hubs. An extra advantage can be the silence of fuel cells compared with the noise of combustion based gensets.

After first analysis for reliable cable networks with 24/7 reliable grid the use case for a SOFC HELP fuel cell as a backup in nodes and cabinets is less feasible economically as the fuel cell equipment will be extra costs which will hardly be used with a reliable grid.

Another important advantage of the SOFC HELP fuel cell for powering cable networks is the significant reduction of GHG emissions, especially with the use of biofuels and even more with (green) hydrogen.

4.2. Maintenance Advantages

Fuel cells require very low maintenance. Also, depending on tank sizes, less frequent checking and filling of fuel is needed.

Since SOFC HELP fuel cell stacks contain no moving parts, maintenance is light if any and the cells have a very long lifetime. The blower for the supply of air and the micro-fuel pump have moving parts but are engaged/active only when in backup mode.

4.3. Circularity/ Recycling of SOFC HELP Fuel Cell Stacks

Generic conventional SOFC fuel cells are sinter bonded. One of the unique properties of the SOFC HELP fuel cell is the loose stacking (see section 2) which is possible due to the absence of sinter bonding. In addition to prevention of thermo-mechanical stresses the loose stacking readily enables recycling of the stacks by simply removing the hollow electrodes, electrolytes, and interconnect sheets from the stack at end-of-life. Grinding and advanced milling bring materials to the active state for re-use as basic ceramic material as input for production into fresh hollow electrodes and electrolytes.

4.4. Use Case 3 - Climate Technology – (Cooling or Heating) with SOFC HELP Fuel Cells

For use case 3 it is possible to use the continuous thermal output from the SOFC HELP fuel cell for cooling the equipment of the outside plant. Thermal output from the fuel cell can be converted into cooling duty by a small off-the-shelf absorption pump, which cools electronic equipment in hot climates, increasing the reliability of the equipment. A similar process could be used to cool equipment for inside plants.

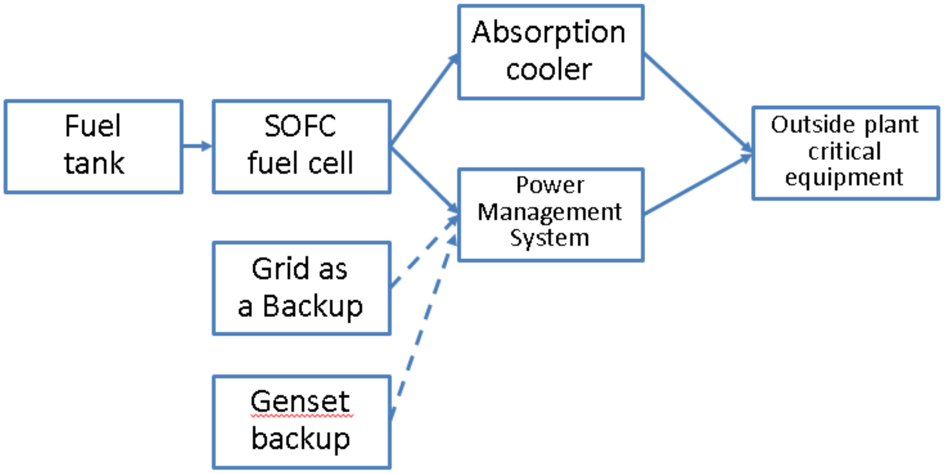


Figure 7 - Energy case power system with SOFC fuel cell with absorption cooler for cooling in warm climate situation

4.4.1. Cooling Scenario

During the hot summer months of the year, the issue of overheating of electronic systems in the outside plant is becoming a common reality, especially in extreme hot tropical regions. In practice, systems are installed to create cooling of the electronic devices using electric fans or air conditioners. These systems require power to operate.

The SOFC HELP fuel cell generates high temperature heat when in operation. This heat of approx. 800-900 °C can easily be captured because it comes out by exhaust tubes. This heat can be transformed into cooling using a longtime existing technology of an absorption cooler. This well-known principle is widely used in outdoor cool boxes powered by an internal propane gas burner. The use of other heat sources like from solar have also been documented.

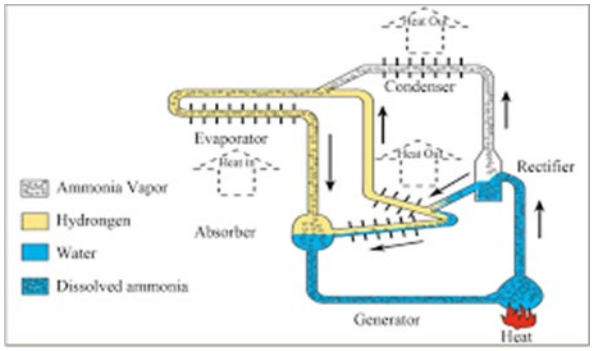


Figure 8 - Refrigeration absorption cycle

Figure 9 - Absorption cool box of Dometic.

For 24/7 cooling 30% of the heat generated by the fuel cell typically will be converted into thermal cooling power, thus 300 W cooling can be generated for every kW of thermal output. In backup mode of the fuel cell there will be little or no cooling output. So, for efficient cooling, 24/7 operation of the SOFC HELP fuel cell is required. The fuel cell could be modulated down to approx. 50% power if less cooling is required.

4.4.2. Heating Scenario

In winter, cold climates can disturb or damage operations of equipment and electronics in outside plant. In cold conditions the heat from the SOFC HELP fuel cell can be used for warming the space of the enclosure of the electronic systems of the hub or node. The temperature of the systems can be managed with the heat of the fuel cell via a regulated ventilator that warms (part of) the air from outside or warms the air from natural air circulation until the required space temperature is reached.

A fuel cell in 24/7 operation gives continuous heat of approx. 800 to 900 ºC and a fuel cell in backup /standby mode gives less heat and a lower temperature of about 700 to 800 ºC.

Arguably the most off-grid, off-road and remote telecommunications infrastructure is the mobile radio network (i.e. cellular). Often enough, the power source running this critical part of the network is not grid electricity but diesel generators. When it comes to cooling such a facility, the energy requirements put yet further strain to the operations and maintenance demand, despite new technology developments using free air cooling during favorable seasons in the year. With the high temperature heat from the SOFC HELP fuel cell utilized to power an absorption cooling device, the fuel cell can provide such a site with both electricity and cooling simultaneously. If the SOFC HELP fuel cell power is used for heating instead of cooling, the heat energy ratio is approximately 1:1.

Further benefits can be realized in terms of space and other operational cost savings due to the modularity of the fuel cell. Cells can be as small as 1 kW power with a small footprint. Additionally, the needed space for the energy storage (fuel tank) can be independently sized and located.

4.4.3. Mobile Network Site Example

For a typical 2, 3 and 4G radio site with a power rating of around 5 to 6 kW, deployment of six SOFC HELP units (current system design is 1kW per unit as per Table 3 above) will provide 24/7 power.

Further, beyond the 6 kW electric power the system also provides between 5-6 kW of thermal power. With off-the-shelf absorption cooling technologies tapping into the thermal power, each site will receive between 1.5 to 1.8 kW cooling power. Similarly, for a comparable 5G site (7.5 kW) eight units will provide the total power and cooling requirements. Such a system will be noiseless without the mechanical wear out of a generator and will eliminate the need for frequent checks.

Figure 10 - Typical power consumption for 2G, 3G and 4G

Figure 11 - Typical power consumption for 5G

4.4.4. Conclusions Use Case 3 – Climate Technology (Cooling and Heating) with SOFC HELP Fuel Cell

The deployment of such a solution with the SOFC HELP fuel cell can provide heat and cooling along with power. A power system with a fuel cell will no longer be the bottleneck for regular maintenance, fuel change, and/or top up. This will result not only in operational cost savings but also can contribute to the reduction or even elimination of the business's scope 1 emissions, the direct greenhouse (GHG) emissions that occur from sources that are controlled or owned by the organization. This environmental benefit should not be underestimated as emissions from generators and air conditioning systems are the most difficult to transition to green technologies. The costs nowadays for available alternative technologies or solutions are very high, and the options to compensate these emissions, for instance carbon credits, are not really preferable options because they do not directly reduce emissions. The SOFC HELP fuel cell is a good possible solution that can help to reduce these emissions in the operations.

The cable industry, known for revolutionary innovations in connectivity and entertainment, is eying another revolution with partners to keep tomorrow's connections services even more reliable and more sustainable.

4.5 Insights of Energy Systems for Cable Networks with CO_2 Circularity

For cable networks which will use the SOFC HELP fuel cell in their energy system, the option of CO_2 recovery will become possible. This is enabled by the fact that after condensation of water vapor the SOFC HELP type fuel cell emits a concentrated CO_2 stream. The concentrated CO_2 gas can readily be stored in a tank using a small compressor or passed over a cartridge with mineral CO_2 binding material. The concentrated CO_2 can be recovered during servicing after refueling of the energy system of the hub or node. This CO_2 can be converted into a fuel named e-methanol in a refinery. E-methanol can, for example, be used in the fuel cell. With this development the power system for the cable network can improve the carbon equation and the cable network can even become carbon negative, because when using bioethanol as a fuel it can achieve net negative emissions.

For use case 4, we envisage an SOFC HELP fuel cell system for cable operators' fleet. A carbon negative emissions concept is described for the use of the SOFC HELP fuel cell in automotive with carbon capture and bioethanol, see [Frontier article].

In future articles, the insights of the other use cases (1, 4 and 5) will be presented. These were also analyzed with cable operators, considering the use of SOFC HELP fuel cells to facilitate carbon negative emissions of cable networks.

- Use case 1 - Inside plant energy optimization (24/7 and the grid as a backup) - critical facilities
- Use case 4 - Potential of solutions with CO_2 re-use
- Use case 5 - Powering in the home (via home devices)

5. Conclusions

The SOFC HELP fuel cell will tackle some of the existing technology challenges that have hampered fuel cell mass deployment. As the cells operate with multiple fuel types they will improve network reliability and tackle the challenge of contaminated fuels in standby fuel tanks. The efficiency of more than 60% results in both energy and cost savings. Operating fuel cells with clean fuel sources does not require any additional product change or maintenance. While the use of clean fuels generates zero emissions the cells are also produced of natural abundant materials and materials that can be fully recycled at end of run life. The SOFC HELP fuel cell had during its early-stage development a collaboration approach of co-engineering and co-creation with the cable sector to receive input on optimal use cases for energy savings and sustainable goals in cable networks.

Since the beginning of the 21st century, digital technology and the power of the Internet has connected the world like never before. It has empowered the masses to access information in an instant wherever one might be. The cable industry, as the pioneers of the media and digital technology, are well aware that their services are increasingly seen as a human right in many parts of the world. While many of the services have helped create a more sustainable world, as demand for the services increase so does the need for energy to power its networks. To tackle the dilemma of the yin and yang the industry has been working together with entrepreneurs, businesses, institutions and others to innovate and to overcome the status quo.

What if we could make the energy challenge a circular process? This may sound like a far-fetched idea but what if we could use energy waste as raw ingredient for energy production? What if we could use existing fuels available today locally and generate electricity free of GHGs? What if we could help make the industry net zero by 2030 while expanding digital services to many more millions of people? What if fuel cells, specifically SOFC HELP, could answer the power needs of the cable network for both standby and 24/7, and the grid as standby power? What if we can tackle the challenge of generating new fuels from energy waste and trapping the GHGs in an energy lifecycle loop? That is when we can truly transform not only the energy needs of our industry but transform our industry as the ultimate contender to solving the climate change challenge as well.

6. Abbreviations and Definitions

6.1. Abbreviations

AFC	alkaline fuel cell
GHG	greenhouse gas
HELP	hollow electrodes and loose plates
HFC	hybrid fiber-coax
ICE	internal combustion engine
kW	kilowatt
kWh	kilowatt hour
NO_x	nitric and nitrogen (di-)oxides
PEM	proton exchange membrane (fuel cell type)
PM	particulate matter
SOFC	solid oxide fuel cell
SO_x	sulfur (di-/tri-)oxides

6.2. Definitions

absorption cooling	Absorption cooling is a cooling method that uses a heat source to provide energy for the cooling system.
alkaline fuel cell	A low temperature fuel cell which runs on pure hydrogen and pure oxygen or purified air only
proton exchange membrane fuel cell	A low temperature fuel cell which runs on pure hydrogen only
solid oxide fuel cell	A high temperature fuel cell which runs on various fuels

7. Bibliography and References

Jaspers, B., Kuo, P., Amladi, A., van Neerbos, W., & Aravind, P. (2021, February 14). *Negative CO_2 emissions for transportation.* https://doi.org/10.31219/osf.io/t8c9p

PCT patent wo 92/06515 (1992, April 16). *Hollow Electrode for an Electrochemical Cell Provided with at Least One Inlet and One Outlet Opening for Gases and also Electrochemical Cell which Contains Such an Electrode.*

SCTE STANDARDS JOURNAL

SCTE a subsidiary of CableLabs®

The Sustainable Future Of Energy Storage For Cooling, Peak Power Balancing And Renewable Energy

The Use Of Phase Change Materials (PCM)

A Technical Paper prepared for SCTE by

Henrik Thorsen, CEO, Energy Cool Aps
Snaremosevej 188-192
Fredericia, Denmark
henrik@energy-cool.com
www.energy-cool.com

Donné Ros, International Sales Representative, Energy Cool Aps
Snaremosevej 188-192
Fredericia, Denmark
dr@energy-cool.com
www.energy-cool.com

Ivan Roubos, CEO, Tizzon
Aartsdijkweg 45
Maasdijk, The Netherlands
Ivan.roubos@tizzon.com

Table of Contents

List of Figures

1. PCM introduction

The use of phase change material (PCM) in telecom, fiber companies, broadcast and data centers is beginning to grow. PCM is a substance that releases/absorbs sufficient energy at phase transition to provide useful heat or cooling.

PCM is the sustainable future of energy storage for cooling, saving up to 95% energy/CO_2 and reducing power consumption at peak hours to increase operational availability and able to be used in conjunction with the use of renewable energy. PCM is used in both active and passive cooling solutions

In Europe over the last 10 years, projects have demonstrated interesting results. The cooling approach during very warm days was shifted to the cooler nights. Very warm days in Europe have also demonstrated the expensive energy costs and instability of the power grid. Renewable energy approaches combined with PCM helped to address both conditions affecting critical facilities

PCM is a well proven, long life and maintenance free solution, used by a variety of large operators all over Europe. Using this type of system also supports the United Nations Sustainable Development goals. Due to its environmental credentials, PCM as part of a triple cool climate management approach (free air, PCM and water cooling), is arguably one of the most sustainable cooling methods in the world today.

Figure 1 – PCM plates used in both active and passive cooling solutions

In a passive solution, with the Cool Peak Bricks, the need for cooling solutions that require power and maintenance is reduced to half the normal requirement of a traditional approach to facility climate management.

1.1. PCM Technology

To expand on the earlier definition, phase change materials are substances with a high heat of fusion that melts and solidifies at a certain temperature and are capable of storing and releasing large amounts of energy. Heat is absorbed or released when the material changes from solid to liquid and vice versa; thus, PCMs are classified as latent heat storage (LHS) units.

The PCM types used in cooling solutions are composed of water and calcium chloride as base material for the PCM-Panels (CaCl2 tech).

Calcium chloride is a natural product and comes from a supplier in Veendam. This supplier extracts magnesium chloride (magnesium salt) which is located in a unique salt layer, about 2000 meters below the ground. This salt layer is the remnant of a primeval sea full of minerals and salts: the Zechstein Sea. This salt layer contains magnesium salt of the purest kind.

The PCM is packaged in high density polyethylene (HDPE) panels. This is a safe material that is highly recyclable and can be melted down. The material consists of carbon and hydrogen.

Complete combustion produces only the non-toxic substances carbon dioxide and water.

To extend the usage of free cooling, during those periods when the outdoor temperature is too high, PCM is used in the air flow of the cooling units.

During periods in which the room temperature can be controlled by direct usage of the outdoor temperature the PCM is bypassed and maintained intact or “saved” for periods where the outdoor air is not usable for cooling.

In case the outdoor temperature has exceeded the level at which it is sufficient for cooling through the free cooling principle, the outside air will be directed through PCM plates.

These PCM plates, having been cooled during the night, during the warmer days will be used for cooling the outdoor free air, which in turn is then cool enough to be used for cooling the technical room/facility.

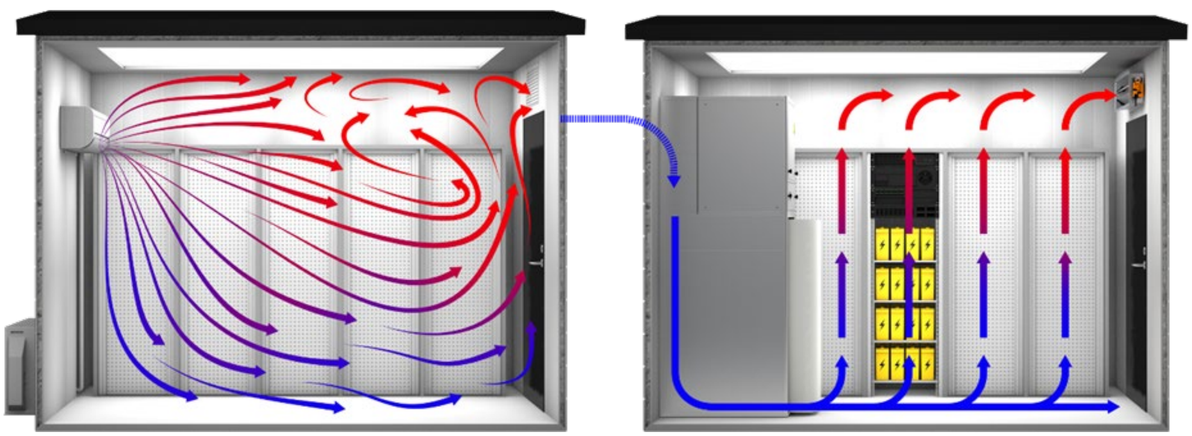


Figure 2 -Traditional A/C system vs displacement cooling

1.2. Triple cooling approach

PCM triple cooling methods saves energy and cost compared to the use of conventional HVAC systems. By using a PCM triple cooling approach, the total operating cost for climate control is reduced by up to 95%. (COP of up to 90, depending on region compared to COP 2-4 for ordinary HVAC systems). COP indicates a system's cooling/heating efficiency. The higher the COP, the more efficiently the system cools or heats.

The PCM triple cooling method is used as when there are 3 different cooling concepts that can be used in 1 solution, depending on the requirements and needs. This approach provides a cooling solution that has effectiveness, storage, operational safety and redundancy.

Triple cooling consists of:

1. Displacement cooling with the use of free air;
2. Energy storage of PCM materials; and
3. Water cooling.

1.3. Passive Cooling For Street Cabinets

Smaller and smaller street cabinets are being used globally for distribution of data to customers. Generally, these cabinets have a small site load, but are often placed on locations where the internal temperature is influenced by direct sun light. Due to the large amount of these locations a cooling solution is required that is plug n' play, has no moving parts and is maintenance free.

Figure 3 -Passive Cabinet Cooling Installtion Example

1.4. Sustainability and cooling systems

The need for reliable air conditioning and refrigeration is increasing rapidly due to temperature rises combined with improved living standards in the developing world. Referencing the International Energy Agency (IEA) projections in air conditioner stock (Figure 5) it is clear the trend as it stands now is bleak.

According to IEA the use of air conditioners and electric fans already accounts for 10% of all global electricity consumption. Therefore, increasing demand for cooling is contributing significantly to climate change. This is the result of the emissions of hydrofluorocarbon (HFCs), carbon dioxide (CO_2), and black carbon from the mostly fossil fuel-based energy that powers air conditioners and other cooling equipment.

With the use of solutions such as phase change materials with free air, the environmental savings include not only the energy reductions enabled by PCM but also the elimination of the need for potent greenhouse gases. Those potent gases, known as hydrofluorocarbons, have in most cases much higher global warming potential than CO_2. According to the United Nations Environment Programme (UNEP) and the International Energy Agency, reducing the production and use of such gases via more climate-friendly cooling systems could save 0.4°C of global warming by 2100.

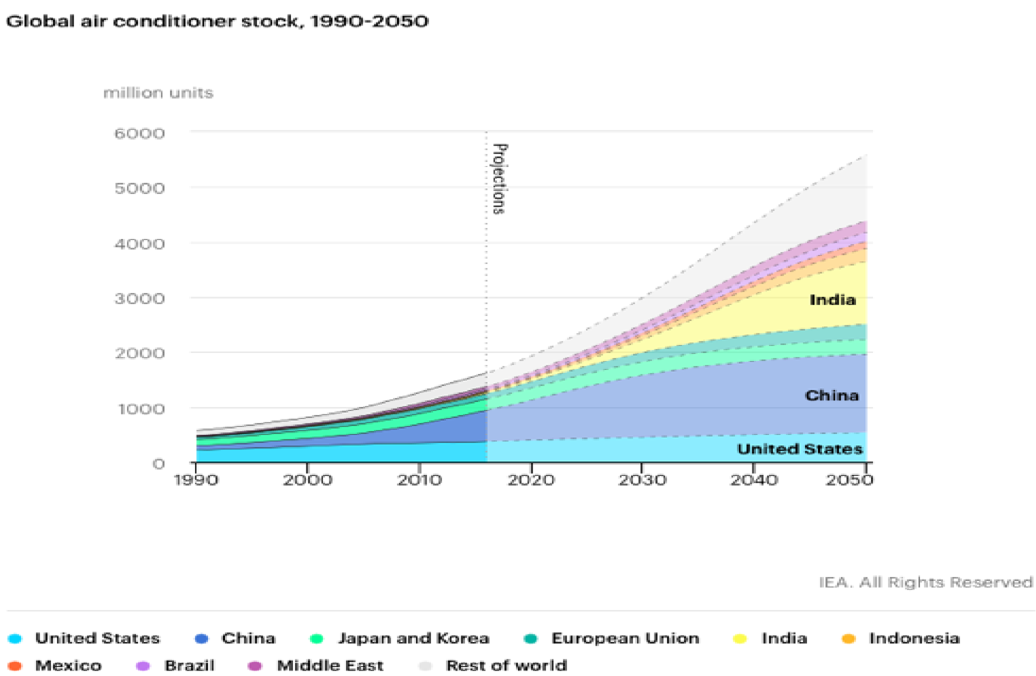


Figure 4 - Air conditioner stock projection through 2050

Here is a sample estimation of the GHG savings from a typical site by deploying PCM and free air compared to traditional split unit cooling systems.

Calculations comparing the month of March in 2016 with direct exchange (DX) with PCM and free air in 2019 show:

		DX	PCM + Free air	
		2016	2019	difference
IT Load	MWh	7,6	7,8	3%
Cooling energy consumpion	MWh	2,66	0,1	-2560%
CO2e (GHG) from grid electricity	kg CO2e	856,254	34,39	-2390%
CO2e (GHG) from refrigirants	kg CO2e	not known	0	not calculated

For the site in question, the typical geographic conditions provided the following:

- Exclusive free cooling in winter season;
- More than 80% free cooling (including PCM use) in spring and autumn; and
- At least 25% free cooling (including PCM) in summer.

2. Why is a PCM triple cooling method the sustainable future?

Rising electricity consumption in the telecommunications industry and increased environmental impact due to the increasing need for communication is a challenge for the industry. The energy systems used for communications must be increasingly supplied by renewable energy if the world is to limit global warming to well below 2 degrees Celsius – preferably to 1.5 degrees – compared to pre-industrial levels, as set out by the Paris Climate Accord.

In countries where the amount of renewable energy supply has grown, energy reduction and consumption still need to occur to help with furthering grid stability. As technology continues to evolve, the demand for energy will continue to be important and will need to be managed to help prevent unnecessary stresses on the electricity grid.

The co-author of this paper, Henrik Thorsen, is the initiator and co-founder of a non-profit European association that is striving to help balance the electricity grid, by means of digitization, storage, and interconnection of energy sectors, so that the electricity grid becomes more flexible and promotes the transition to renewable energy. (Ref. www.center-denmark.com)

As part of this future need to reduce power and balance the electricity grid, the sustainable cooling of the future may be PCM triple cooling. Use of PCM in a triple cool solution generates (storage cooling) savings up to 95% of power consumption for cooling compared to existing HVAC systems with free cooling and A/C. (See Figure 6, example outdoor wall mounted unit). This solution shifts the energy consumption from daytime to nights, thereby increasing the efficiency and helping to balance the power grid.

It is worth noting the worldwide agenda for ESG, carbon footprint and UN Sustainable Developments goals, is something we are all going to relate to more and more.

3. Growing Need to Do Something

The growing need for energy for data centers jeopardizes the EU's climate goal of reducing CO_2 emissions and being CO_2-neutral by 2050. For example, in Denmark, data centers within 10 years will increase the country's total energy consumption by 17%, according to the Danish Council on Climate Change in 2019. Powering data centers is a huge challenge that requires new thinking – globally.

One large data center can consume what corresponds to 4% of Denmark's total electricity consumption.

Typically, the focus is on the very large data centers' massive energy consumption and CO_2 emissions but only a very small part of the research is focused on small and medium-sized data centers. However, research from The United States shows that small and medium-sized data centers account for between a third to a half of the total energy consumption of all data centers.

As a rule, the small and medium-sized data centers are far less efficient than the larger ones, so that a large share can be obtained from the CO_2 accounts if their energy consumption can be reduced.

3.1. PCM Triple Cooling Principle

PCM triple cooling has realized both energy and cost saving on technical installations, where the use of conventional HVAC systems has led to high energy costs. By using PCM triple cool, the total operating cost for climate control can be reduced by up to 95%. (COP, or coefficient of performance, of up to 90, depending on region, compared to COP 2-4 for ordinary HVAC systems.) Again, COP indicates a system's cooling/heating efficiency. The higher the COP, the more efficiently the system cools or heats.

With the use of PCM, a number of benefits can be achieved, besides cooling:

- PCM can be used to shift the highest energy consumption to different periods of the day and away from periods when there is a shortage of (clean) power or when the power cost is higher. The PCM can be utilized during these periods and reduce OPEX and CO_2 emissions.
- PCM can be used as redundancy during power outages, in regions where the temperatures are too high for free cooling alone, instead of having back up power resources (UPS or generators) scaled for traditional cooling systems that have a high-power consumption.
- PCM is used for heating during cold periods, when traditionally free cooling units will be switched off and heaters are used.

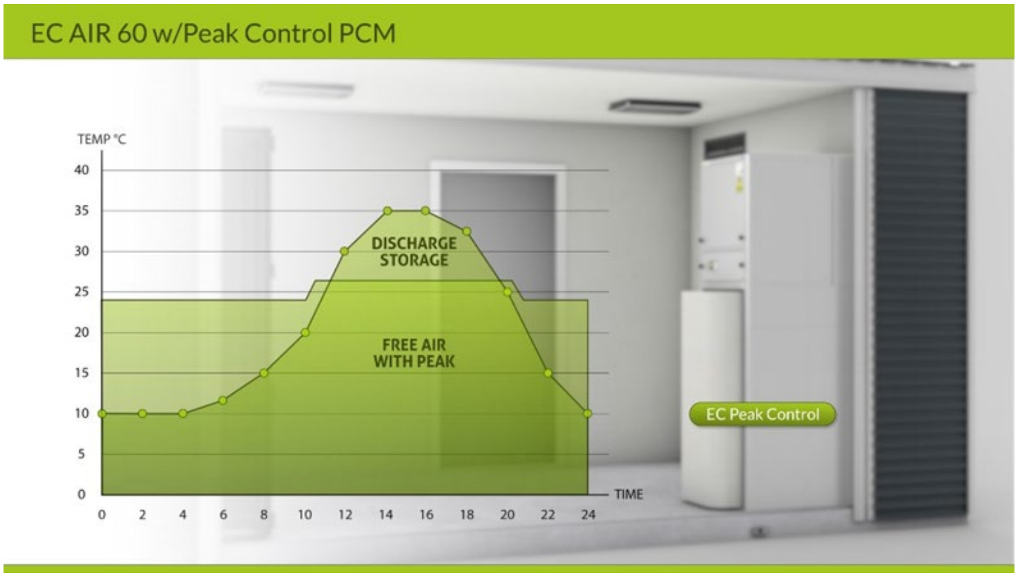


Figure 5 – Storage of heat in PCM

3.2. PCM and Free Displacement Based Cooling Technology

Free cooling solutions cool with the use of outside air. The device creates a fine filtered pressure and leads excess heat away from the electronic equipment. Free cooling units are installed with a motorized fan for optimal air flow and protect the facilities' internal environments from polluting particles. Displacement-based free cooling is a quiet and extremely cost-effective passive cooling method. It is best suited for smaller spaces with stagnant air and allows for complete thermal management. The systems deliver cool air from the outside to the lower part of the room. This forces the warmer air in the room upward, creating a floating cushion of cooler air at the bottom of the room (up to 180 centimeters (71 inches) high). The units work with very limited air flow and low fan speeds. Therefore, this cooling method is a lot more energy efficient than traditional free cooling, uses long-life components, and is almost maintenance free.

3.3. Integrated water cooling

Depending on the regional climate and the requirements of each customer, displacement cooling and PCM cooling units can be equipped with water cooling as well.

For example, in warmer climates where PCM is not sufficient for the total cooling of the technical rooms, water cooling can be implemented.

The water cooling uses the same air flow as the free cooling and PCM cooling.

The primary benefit is that the air flow does not change in the technical room, regardless of whether the system is running on free air, PCM or water cooling. The result is that the air is evenly divided over the server room and there are no hot or cold spots in the server room, such as those that occur with mass ventilation and air-conditioning. This can provide a system that has a COP of 7 or higher, compared to a COP of 3 for an industrial A/C system.

The additional benefit of a water-cooling system is that it is almost maintenance free, and the noise levels are significantly lower than on a traditional A/C system.

The inlet temperature is relatively high compared to traditional cooling systems (12 to 15°C). Due to this a large variety of water inlet solutions can be selected, such as ground sources, water reservoirs, return water from cooling systems, or a traditional chiller, etc., providing an efficient solution.

3.4. Bonus Benefits of PCM Triple Cooling Techiques

3.4.1. Heating During Colder Periods

In regions where the outside temperatures during the winter period get very low, mass ventilations systems are normally shut down and heat pumps are used for cooling.

This prevents too-cold air arriving in the server room that could jeopardize batteries and equipment exposed to too low temperatures. Excessively low temperatures can have a negative effect on the battery capacity on site and can pose a risk for the backup capacity of a facility. Often a heater or A/C unit is used to avoid this situation, but it uses a lot of energy.

PCM is a very good alternative. During cold periods, warm inside air is recirculated over the cold PCM plates and heats them, cooling the air. When the PCM gets too warm, and the room temperature cannot be kept at the desired level, recirculation is closed and the cold outside air is guided through the warm PCM plates. With this the air is heated while the PCM is cooled down. When the PCM is getting too cold the process is reversed again.

3.4.2. Low Maintenance Cost

The additional advantage of displacement is the very low cost of maintenance. The free cooling and PCM systems have extremely long life parts and filter systems because of the low airflow. No parts need specialized technicians for maintenance. Maintenance can be carried out by anyone who has access to the technical room/facility. Depending on the region a filter change is required once or twice a year, taking an average of 5 minutes per unit to change.

3.4.3. Reduced noise level

The additional advantage of displacement and the use of these systems is noise reduction. The free cooling and PCM systems have extremely low noise levels and in normal operation can hardly be heard.

3.4.4. Cooling During Power Failure

As the system is operated by use of low air volumes and speed regulated fans, power consumption is extremely low. A typical system, with cooling capacity up to around 5KW, has an average power consumption of less than 60W.

This means the systems can be connected without any need for upgrade to the existing UPS power supply backup. Even during periods when the outside temperatures are too high for free cooling, when an A/C system typically would be required, the PCM can be used with the same power consumption as free cooling. The room will be cooled even during power failure, assuming enough PCM capacity is deployed.

Phase change material can help prevent longer down-time during power outages or air conditioner failure in warm regions. In turn, PCM will allow more time for rectification of the problem and will prevent services from being interrupted, which would lead to loss of revenue/customer satisfaction.

3.4.5. Lifetime – Cradle to Cradle

The triple cool approach has been developed with a focus on optimal performance and reliability, reducing the maintenance to a minimum and the lowest possible energy consumption.

The entire cooling system components are completely recyclable, PCM is cradle to cradle, and the products have a very long lifetime, something that has been proven over the last 10 years.

PCM has at lifetime of least 25+ years, whereas the cooling units (air handling units AHU) have a lifetime of 10+ years.

The end of life and replacement of this solution cannot be compared to traditional systems. Traditional systems need to be dismantled at the end of life and replaced with new ones, whereas a triple cool approach with PCM is re-used and only a few moving parts must be replaced. Therefore, CAPEX cost and CO_2 (carbon) footprint are significantly reduced at end of life as well.

With the triple cool strategy, the active cooling system/air conditioners will be running less and the amount of AC start/stops will be limited greatly. The lifetime of the air conditioners will be extended greatly and lower CAPEX investments will be required over the course of the systems useful life.

Figure 6 - The United Nations Sustainable Developments goals

3.4.6. Battery Temperature Control

One of the most critical items, specifically on cell sites, is battery temperature.

In general, the equipment placed here can handle higher temperatures, but batteries have the best performance, and lowest degradation, when kept below 24°C during warm periods.

Batteries are normally placed low, up to around 120 centimeters (47 Inches) high, while active equipment is placed up to 180 centimeters (71 inches) high.

Displacement has the benefit, as the cold air is distributed on the floor and is warmed up by the internal heat load, that the floor temperature is always lower than the room temperature at 180 centimeters (71 inches) high.

With displacement one avoids having hot and cold spots, something very common when using mass ventilation of traditional A/C systems.

3.4.7. Humidity

Humidity associated with free cooling is often considered as a risk factor. When free cooling is applied cold air is distributed in the technical facility. As the cold, moist air meets the warm, dry air in the cabin the cold air is warmed up and the humidity drops. Therefore, there is no risk of condensation on the electronic equipment.

3.5. Case study – PCM Triple Cooling Principle in Technical Facilities

During outside air temperatures between +18°C and 25°C, cooling is normally provided by outside air only.

With temperatures above 25°C the phase change material will melt as soon as the outside air temperature rises above the melting temperature. This absorbs the energy and cools down the system air within the ASHRAE range.

Once the outside temperature is lower than that of the phase change material, the system control reverses the process by using the outdoor air source to cool the PCM. The phase change material solidifies as soon as the outside temperature is lower than the phase change material. The absorbed energy is once again ready to release into the system air.

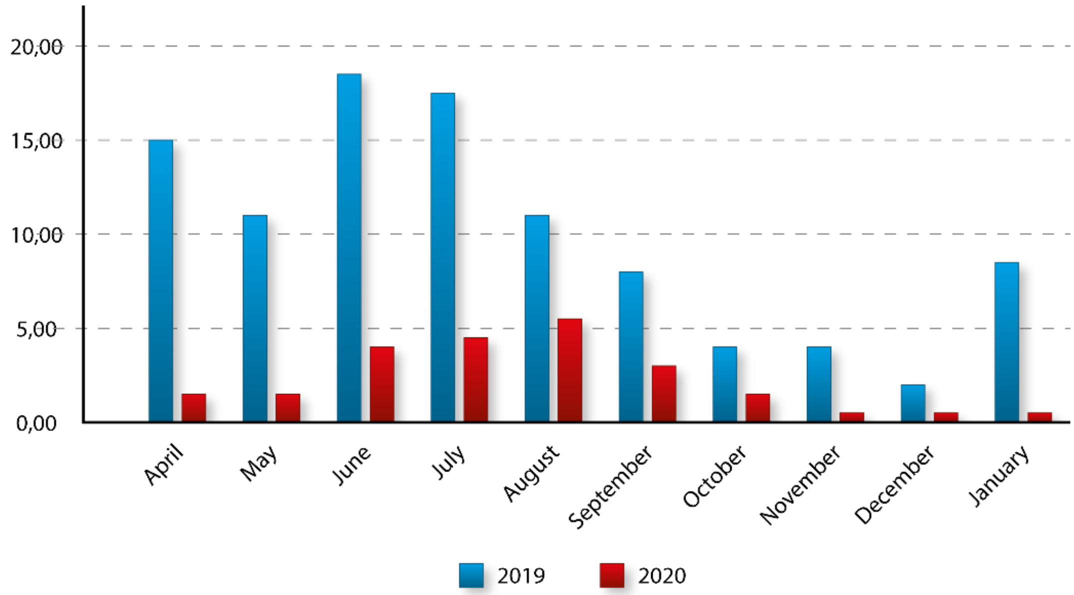


Figure 7 - Power Consumption Cooling 2019 And 2020 Standard Cooling Vs Triple Cooling Application

The use of phase change material and outside air- and water-cooling results in a constant adjustment and lowering of the system air temperature and humidity, according to ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). This process also reduces the carbon footprint and increases the energy efficiency and uptime.

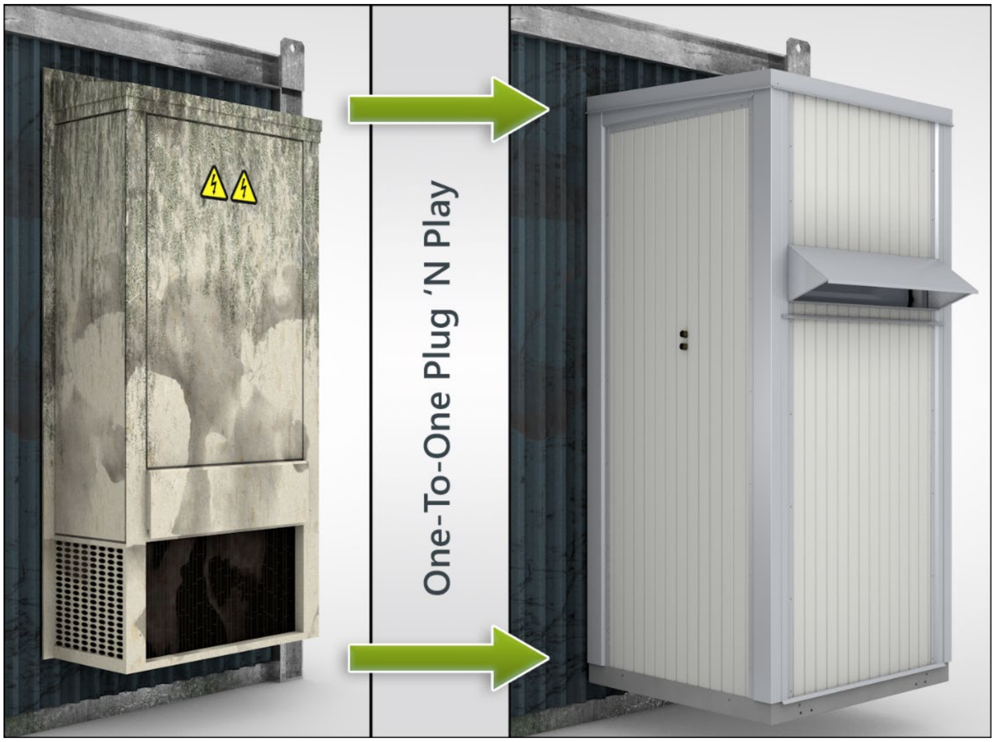


Figure 8 - Replacement of traditional cooling HVAC with triple cool method

Following is a use case that involved a hub facility that was converted to temperature regulation that included phase change material and outside air. Note that yearly average temperature in the year prior to conversion, 2019, was 1.7 degrees lower than it was in 2020. Nevertheless, the converted facility yielded:

- Higher availability;
- Redundancy;
- Significant noise reduction;
- Reduction of energy consumption for cooling systems from 72 MW/year to less than 18MW/year with a higher average temperature; and
- Improved PUE from 1.24 to 1.053.

Figure 9 – Turnkey data center solution with triple cooling components

A similar solution was implemented as a turnkey solution at another region. The solution has been operational now for five years without any interruption and with a large cost reduction for both power and maintenance.

Reports show the following improvements:

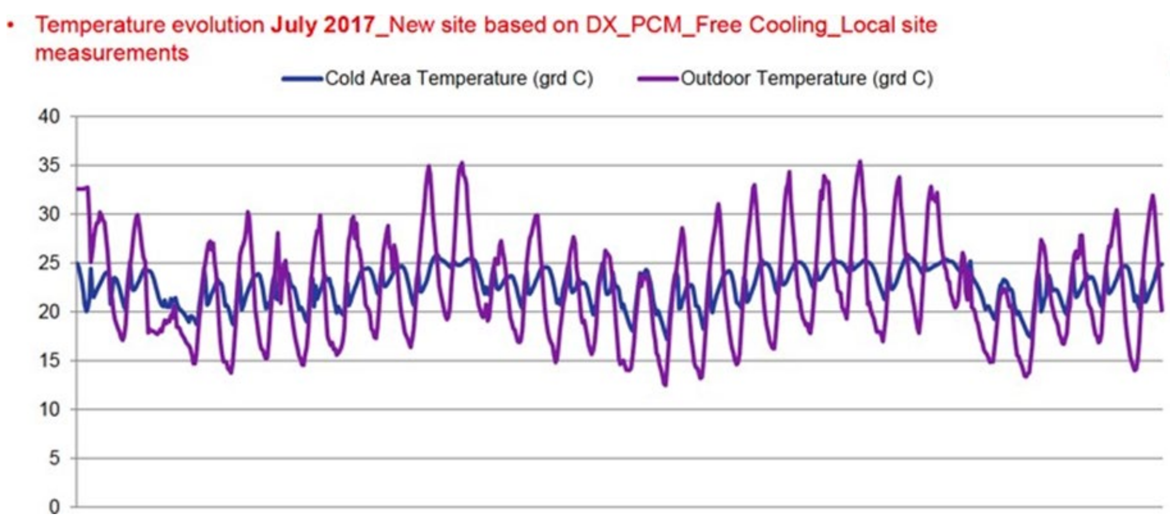
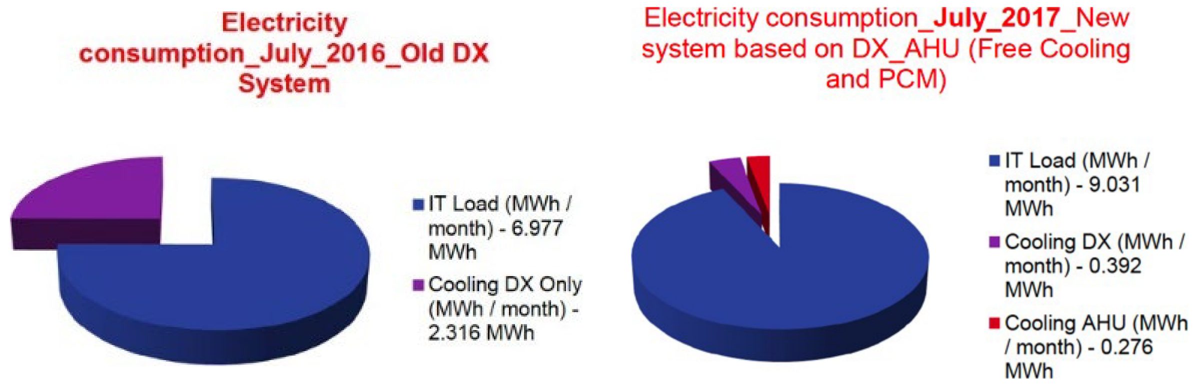


Figure 10 - Outside temperatures vs. room temperatures using PCM and free cooling



PUEc=1.33 (July 2016) PUEc=1.074 (July 2017)

Figure 11 - Results from 2016 (old DX) and 2017 (triple cool application) reported by owner

Converting the EDGE facilities to phase change material and outside air realized the following benefits, despite a higher IT-load in the year following conversion:

- Much higher availability;
- Scalable shelter solution;
- Modular cooling system;
- Reduction of energy consumption for cooling systems from 2,3kW/h average to 0,7kW/h average, even with an increase of the IT-load; and
- Increase of PUE from 1.33 to 1.074;
- Reduction of maintenance.

3.6. PCM and critical facilities

Using PCM for Critical Facilities offers a range of benefits compared to existing cooling solutions.

Due to the low energy consumption the active PCM solution can be connected to existing backup systems, without any need to re-dimension them.

The system with PCM gives the same cooling benefits as an A/C system, but with a fraction of the energy consumption. This means that during outages the cooling can be guaranteed, which is critical to keep the facilities up and running. Without cooling, in most cases, the equipment will soon get overheated and power down, resulting in a loss of services.

At the same time, PCM can be used in the same way when fire is detected or weather conditions require a close down of fresh air into the facility and a possible outage on external power. These could be situations like tornadoes/hurricanes, fire and smoke in the surroundings of the facilities, or any other occurrence that results in a loss of power or requires the facility to be closed from external influences.

3.7. Recording Results

Recording the results is done in two ways:

Equipment owners record all the power consumptions on locations. This is divided into total power consumption, power consumption for cooling, and UPS power consumption.

Cloud recording. All information about the performance of the cooling system is stored in the cloud which monitors the system performance, the amount of free cooling, and PCM operation, as well as troubleshooting the cooling system remotely.

The cloud system, in cooperation with the client, is also used to optimize the performance of the cooling and secure more PCM operations, when the technical site is operational.

4. Conclusion

This paper introduced a proven solution which is among the most cost-efficient systems available today.

Large operators in Europe have been implementing these solutions with energy and emissions reductions up to 95% compared to traditional cooling systems. Systems operate successfully for many years in climate conditions from -25^0 C to +45^0 C. PCM systems can be installed in new building and as a rebuild in existing technical sites, even when the site is operational. This can be done without interrupting the main operation of the site.

The solution features low energy consumption and cradle to cradle sustainability credentials that minimize the impact on environment, reduce OPEX and overall CAPEX for cooling system over the site lifetime for operators, and help to achieve emissions reduction goals.

5. Abbreviations

A/C	air conditioning
CAPEX	capital expenses
COP	coefficient of performance
DX Cooling	direct expansion cooling
EDGE	Enhanced Data rates for GSM Evolution
ESG	environmental, social and governance
HVAC	heating, ventilation and air conditioning
KW	kilowatt
OPEX	operating expenses
PCM	phase change materials
PUE	power usage effectiveness
UPS	uninterruptable power supply

6. Bibliography and references

Liberty Global UPC Alba Iulia

Sustainable Development Goals (SDGs)

https://energyinnovation.org/2020/03/17/how-much-energy-do-data-centers-really-use/

https://www.dtu.dk/english/news/all-news/nyhed?id=%7BC53891B3-DF71-4FD9-AB74-3D9C3BC2423B%7D

HFC Network Powering Using Lithium Iron Phosphate Batteries

History, Science, Comparable Attributes and Use Cases

A Technical Paper prepared for SCTE by

Len Visser, Product Line Manager Power Solutions, Lindsay Broadband Inc., SCTE Member
2-2035 Fisher Drive
Peterborough, ON K9J6X6 Canada
lvisser@lindsaybb.com
(705) 742-1350 x241

Table of Contents

List of Figures

1. Introduction

Network powering is a complex imperative. In today's evolving environment, powering solutions need to be clean, energy-efficient, safe, reliable, and sustainable. Deployment scenarios extend across hybrid fiber/coax (HFC) networks, small cells, and Wi-Fi access points, as well as long-duration utility power outages and disaster recovery using fixed and portable solutions.

Powering solutions also span a range of considerations and goals, including:

- Device and system efficiencies – to reduce energy consumption;
- Clean, environmentally-friendly devices – to reduce the carbon footprint;
- Reliable devices – to provide continuous service, minimize maintenance and extend lifespans;
- Renewable and sustainable energy sources – to reduce fossil fuel use;
- Safe devices – to remove hazards to personnel and surroundings; and
- Emergency and disaster-recovery suitability – to ensure portability, quick set-up, smooth operation, and long run-time

Broadband operators have typically used valve-regulated lead acid (VRLA) batteries to meet their power requirements. To help achieve today's expanded objectives, they now have other options, including lithium iron phosphate (LiFePO4, or LFP). Based on a highly stable chemistry, LFP batteries offer premium performance while being environmentally friendly. They are safe, with no corrosive substances or fire and explosion hazards. They use fewer materials and contain no rare earth metals or groundwater-contaminating compounds, avoiding the drawbacks of VRLA and other leading lithium battery technologies, such as lithium nickel manganese cobalt oxide (NMC). They allow network operators to maximize run-time with minimal battery counts and minimal maintenance. In this paper, we will provide additional background on LFP batteries and chemistry, review their attributes, and discuss three use cases.

2. Battery Background

As in most areas of science and industry, the technology of batteries has developed more rapidly in the past 40 years than in the prior 120. Invented in the mid 19th century, the lead-acid battery remained largely unchanged for about 100 years. The first type of VRLA battery (gel cell) emerged in 1957 and the second (absorbent glass mat or AGM) was patented 15 years later. Then in the late 1970s, scientists began developing what would become the lithium-ion battery.

Several other technologies also appeared during this timeline. Wet-cell nickel-cadmium (NiCd) batteries, invented at the end of the 19th century, later became widely used in portable power tools and other electronic devices, until being supplanted by nickel-metal hydride (NiMH) batteries, which emerged commercially in the 1990s. The VRLA and the lithium-ion families, however, remain the dominant segments. As the global scientists driving lithium-ion technology continued to experiment (eventually winning the Nobel Prize for their discoveries) [1], they focused on the chemical makeup of the battery's positive terminal, or cathode. Two major technologies that emerged were NMC and LFP.

Looking forward, some analysts expect LFP batteries to overtake NMC as early as 2028.[2] Two leading markets are transportation and utility-scale stationary energy storage. Demand comes from manufacturers of low-speed vehicles (LSVs), automated guided vehicles (AGVs) and now, with Tesla's decision in 2021, high-speed electric vehicles (EVs).[3] The renewable energy sector is also leaning on LFPs to capture energy during periods of excess production. Communications and broadband service providers represent additional demand for LFPs, but before looking at these use cases, let's examine the underlying technology and assess its comparative advantages.

3. LFP Technical Overview and Attributes

Like other lithium-ion batteries, LFPs discharge and recharge energy through lithium ions that move from a negative electrode (anode) through an electrolyte to a positive electrode (cathode) and back again. Lithium-ion cells typically use a graphitic carbon compound at the anode and an intercalated lithium compound on the cathode.

Why does LFP work so well? Experts point to its thermal and structural stability, its aversion to water molecules, and its exceptional performance. Largely responsible for that performance is LFP's crystalline lattice, or olivine-like structure, which features an "extremely flat charge/discharge profile." [4] The application of LFP olivine material was the fruit of a decade of intensive research and development, leading to commercialization and widespread use of these batteries beginning around 2012.[5] The upshot is they are today among the longest-lived batteries ever developed. Manufacturers typically rate LFP batteries for 5,000-plus charge/discharge cycles because iron phosphate's robust crystalline structure does not break down under repeated packing and unpacking of the lithium ions.

Supplementing the chemistry is a battery management system (BMS), which constantly monitors key operational parameters during charging and discharging using outputs from sensors and gives the actual status of voltages, currents, and temperatures within the battery as well as the state of charge (SoC). A microcontroller manages information from the sensing circuitry and makes decisions with the received information using application-specific algorithms that are digitally encoded into the microcontroller. For more details on LFP and how it compares to VRLA and NMC chemistries, see Section 3 and its subsection discussions of its reliable, green, safe, efficient, and socially responsible attributes.

3.1. Reliability

3.1.1. Lifespan

LFP ($LiFePO_4$) batteries live ten times longer than VRLA and three times longer than NMC. (See Figure 1.) This means a higher reliability over a longer period. A longer lifespan also increases reliability, reducing labor and replacement costs. (For the relevance of longer lifespans to mandated network reliability, see section below on use cases.)

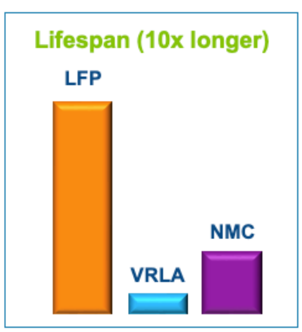


Figure 1 – Battery Lifespan: LFP, VRLA, NMC

3.1.2. Energy Density

LFP batteries have a higher energy density (≈140 Wh/kg) compared to ≈ 45 Wh/kg for VRLA. (See Figure 2.) The lead-acid battery is an aqueous system. The single cell voltage is nominally 2V during discharge. Lead is a heavy metal; its specific capacity is only 44Ah/kg. In comparison, the LFP cell is a non-aqueous system, having 3.2V as its nominal voltage during discharge. Its specific capacity is more than 145Ah/kg. Therefore, the gravimetric energy density of an LFP battery is three times higher than that of VRLA. An additional bonus is that because the LFP is a non-aqueous battery, there are no spill hazards, and the battery can be installed on its side, upside down or standing up.

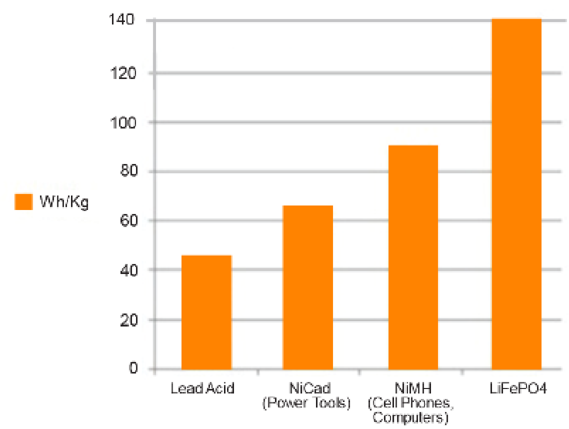


Figure 2 – Energy Density: Lead Acid (VRLA), NiCad, NiMH, LiFePO4 (LFP) (Wh/Kg)

LFP batteries also provide more power for the same battery weight, resulting in a longer run-time. More energy storage in the same physical space equates to three-times longer run-time in terms of energy density alone. Also, LFP batteries are one-third the weight and one-third the size of VRLA in terms of energy density (See Figure 3.). This means, there is a much higher capacity-battery in the same physical form-factor, allowing greater configuration flexibility in the same enclosure.

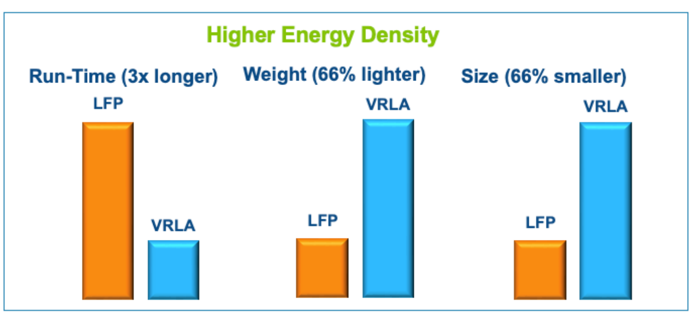


Figure 3 – Energy Density: LFP vs. VRLA (Run-Time, Weight, Size)

3.1.3. Charge Time

Due to its very low internal resistance, LFP has a charge rate four times faster than VRLA. (See Figure 4.) This means that during multiple successive outages, LFP batteries used in network standby power supplies come up to full charge faster, resulting in longer run-times and greater plant reliability when applied, for instance, in an HFC network. During recurring utility power outages, LFP batteries can often fully recharge before the next outage.

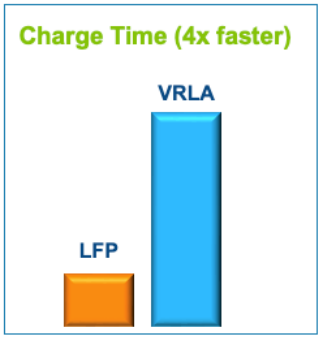


Figure 4 – Charge Time: LiFePO4 vs. VRLA

3.1.4. Depth of Discharge

LFP provides longer run-time for emergency use. Depth of discharge (DoD) refers to the percentage of a battery that has been discharged relative to its capacity. For LFP, 80 percent of its capacity is usable energy vs. only 40–50 percent for the typical VRLA (See Figure 5.).

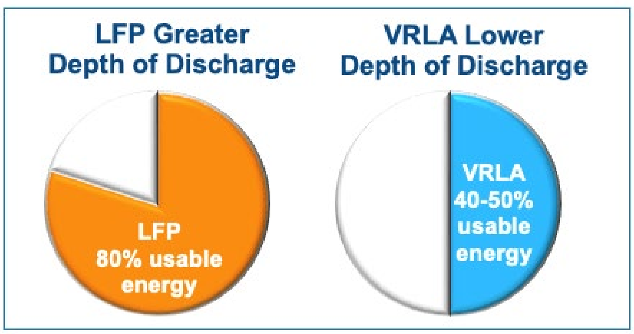


Figure 5 – Depth of Discharge: LFP vs. VRLA

3.1.5. Discharge Rates

High discharge rates have minimal effect on LFP battery lifespans, whereas it greatly reduces VRLA lifespans. LFP battery capacity is independent of the discharge rate, whereas a VRLA battery is reduced to only 60 percent of its rated capacity (equaling less run-time) and its lifespan is significantly shortened.

LFP battery discharge voltage is almost constant, thereby keeping the discharge current constant, e.g., for the same amount of battery power delivery to a standby power supply inverter. In contrast, VRLA battery discharge voltage decays as run-time progresses, resulting in increasing discharge current to maintain the same battery power delivery to the power supply inverter (See Figure 6.). Consider a flashlight that dims over time because battery voltage is decreasing as the battery discharges. If the battery was LFP, the bulb would be bright from beginning to end. Once discharged, the battery would simply not turn the light on.

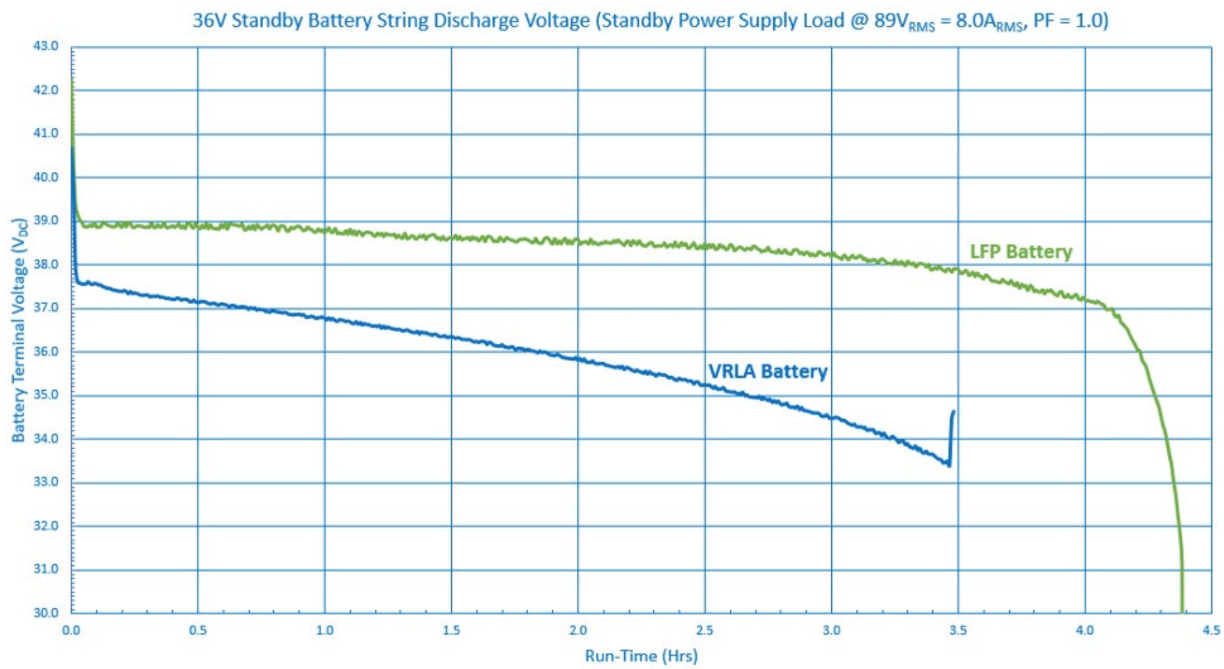


Figure 6 – Discharge Test: LFP vs. VRLA

3.1.6. Maintenance

Due to lack of corrosion concerns, periodic battery interconnect cable changes are not required in LFP batteries. This results in a lower OpEx for maintenance and higher reliability in normal and emergency use. When used in AC generators, LFPs eliminate start-up mechanical problems, such as carburetor and sparkplug fouling, stale fuel, etc. (For more on generators, see discussion of use cases below.)

3.2. Green/Environmentally Friendly

3.2.1. Contaminant-free

LFPs have no lead, mercury, cadmium, cobalt, or rare-earth metals to pollute and enter ground water.

3.2.2. Recyclable

LFPs are 100 percent recyclable, compared to only 60 percent recyclable materials in VRLA.

3.2.3. Less Waste

Fewer raw materials (from fewer batteries required) for the same energy density reduces waste. A longer lifespan also equates to fewer batteries used over time, translating to more reductions in waste.

3.2.4. No Fossil Fuel

When used in AC generator applications, LFPs replace fossil fuel-powered generators:.this enables conformity to anti-idling regulations.

3.3. Safety

3.3.1. No Outgassing

Based on a safe chemistry, LFPs are the most stable lithium-type battery on the market. LFPs do not suffer from the outgasing of harmful acidic gases and explosive hydrogen that is present with lead-acid during charging.

3.3.2. No Sulphuric Acid-Based Electrolyte

The absence of sulphuric acid-based electrolytes eliminates corrosion on connectors and cables. LFPs eliminate the need for rubber gloves, aprons, and personal protective equipment (PPE) to prevent chemical burns. They eliminate the need for corrosion-inhibiting grease on battery terminals, as well as corrosion itself, along with technician exposure to corrosive substances when replacing cables and connectors.

3.3.3. Non-Combustible

As indicated in lab tests, news stories and user-generated videos, some lithium battery chemistries, including NMC, can be very volatile and flammable. Newer LFP chemistry is inherently non-combustible and non-explosive. It is chemically and structurally stable with no chance of thermal runaway. It is safe to transport. Charged and discharged states of LFP are physically similar and robust, allowing ions to remain stable during oxygen-fluctuating charge cycles, which again prevents thermal runaway. If mishandled during charging or discharging, or if subjected to collisions or short circuiting, the battery won't explode or catch fire. Internal decomposition occurs when the battery temperature reaches the 900°F to 1100°F range (500°C to 600°C). This absence of flammability hazards reduces the need for special personnel training.

3.4. Efficiency

3.4.1. No Float Charging

LFP chemistry does not require a floating charge or de-sulphation pulses applied to the battery to maintain its capacity at or near the full charge. Its absence during charging and maintenance cycles reduces operational energy use.

3.4.2. Temperature Effects

All batteries suffer from the effects of extreme temperatures; however, LFP suffers less from extreme temperatures. Its operating temperatures for charge are 32°F to 113°F (0°C to +45°C), and discharge from -4°F to 140°F (-20°C to +60°C). At 0°F (18°C), it can be discharged to 70 percent of rated capacity, whereas VRLA can only be discharged to 45 percent of rated capacity. The BMS monitors performance of the LFP cells and limits degradation from its optimal operating temperature of 68°F (20°C).

For extremely cold environments, optional internal BMS-controlled cell strip heaters maintain the cells between 60°F to 75°F (16°C to +24°C) and allow operation at lower ambient temperatures. Charging is then possible down to -13°F (-25°C). Lead acid batteries can be discharged and charged at a lower temperature limit, but in a less than fully charged battery, the electrolyte becomes more like water, which is more likely to freeze than when it is fully charged, resulting in permanent battery damage. LFPs operate better than VRLAs at hot temperatures due to the higher lithium ionic conductivity and can be discharged up to 140°F (+60°C); VRLA can only be discharged efficiently up to 113°F (+45°C).

3.4.3. Low Self-Discharge Rate

Very low internal resistance prevents wasted energy required to “top up” batteries.

3.5. Social Responsibility

3.5.1. No Cobalt

Over 70 percent of the world’s supply of cobalt comes from the Democratic Republic of Congo, where cobalt mining often involves human rights violations. Unlike NMC batteries, LFP uses no cobalt.

4. Application Use Cases

The long life, low maintenance, high efficiency, and unparalleled safety of LFP batteries make them applicable in a wide range of use cases. In addition to their roles in transportation and utility-scale stationary energy storage, they can lend vital support to communications networks. Power loss or continual sags in voltage (rolling brownouts), whether caused by fires, floods, hurricanes, terrorism, or other disruptive forces, require immediate remediation. LFP battery-based solutions are well-suited to fill that gap through standby power solutions, emergency power trailers, and portable AC generators.

4.1. Network Standby Power Supplies

Operators are looking for ways to increase standby run-times. One driver is regulation. Along with adopting several resiliency strategies for wireless and wireline service providers, the California Public Utilities Commission (CPUC), which has tremendous influence on utility regulation in the U.S., has prioritized backup power duration of at least 72 hours.[6] To achieve this level throughout the life of a battery, operators need to assess risks given realistic run-times for various battery chemistries.

One consideration is the number of battery strings vs. an acceptable depth of discharge (DoD). VLRA has a very mediocre DoD but can be pushed higher, which would allow for fewer battery strings, but the negative impact of a higher DoD on reliability over time can be significant. Temperature is another factor. Cold temperatures reduce any battery’s ability to function properly, although LFP can handle cold better, if fitted with internal cell heaters, than VLRA batteries. As noted above, LFP operates better than VRLA at hot temperatures and can be discharged up to 140°F (+60°C), considerably higher than the efficient discharge limit for VRLA.

Compared to traditional VLRA batteries, LFP batteries offer many advantages in standby scenarios. They are safer, higher performance, compact, and energy-dense, with 80-90 percent DoD and a very long lifespan to maximize run-time with minimal battery counts and minimal maintenance. By contrast, network operators must perform quarterly maintenance on VRLA batteries or suffer critical reductions in network readiness. As a result, LFP can provide reliable and long-lasting network standby run-times during power outages. The reduction in the battery footprint required also directly relates to a reduction in pad and cabinet size, which minimizes visual “pollution.” The contrast with standard VRLA is striking: An LFP standby system can achieve 72-hour run-times with $12V_{DC}$ high-capacity LFP batteries using four times fewer batteries. In a modest plant leg scenario (with an 8 Amp PF=0.85 current draw), an operator would need 28 $36V_{DC}$ 100Ah VRLA battery strings (84 batteries) at a DoD of 60 percent, vs. 6 $36V_{DC}$ 344Ah LFP battery strings (18 batteries) at 80 percent DoD.

In addition to the dramatic reduction of batteries required, an LFP standby solution has the added benefits of a 10x longer lifespan; significantly faster recovery/recharge time, which is critical to remaining operational during multiple outages; and an environmentally friendly battery chemistry that eliminates required maintenance. The elimination of any float charging requirement for LFP battery strings also reduces the operational use of energy.

4.2. Disaster Recovery Emergency Power Trailers

During natural disasters and rolling brownouts, rapidly deployable, portable, high energy-density backup power systems are also needed to power communications equipment in the field during recovery periods. Combined with high-efficiency inverters, LFP batteries can be combined to create sustainable, all-inclusive backup power trailers. Portable small-footprint trailers offer long run-time AC and DC power outputs and are capable of powering critical equipment, such as:

- 4G and 5G small cells
- Wi-Fi access points
- MDU communications equipment
- DAS networks
- HFC networks
- Fiber aggregation sites with active powering

Both 120 and $240V_{AC}$ 60Hz AC sine wave power as well as 12, 24 and $48V_{DC}$ power can be provided using 4-8 high-capacity 14kWh LFP batteries. This yields 56 to 112kWh of energy storage with 90 percent of that stored energy available as usable energy. Multiple 15A, 20A and 30A, $120/240V_{AC}$ receptacles can provide vast amounts of power from the highly efficient inverters. They can be recharged via solar panels or flexible solar ground mats, or individually in warehouse settings via external charging stations. Run-time examples using an 84kWh trailer include a typical HFC leg (1080W), 2.2 days; 5G small cell (420W), 4.9 days; and Wi-Fi AP (64W), 32 days.

4.3. Portable AC Generators

Another application that pairs LFP batteries with high-efficiency inverters are smaller AC generators, which can be used at critical locations during power outages or at job sites in any conditions when field technicians need portable power. This LFP-inverter combination provides a green alternative to portable fossil fuel-powered AC generators, without compromising power or performance.

Anti-idling compliant and safe for indoor use, LFP-based AC generators feature large amounts of energy storage. They deliver high-quality, clean, regulated $120V_{AC}$, 60Hz sine wave power, free from the typical stepped sine wave and electrical noise (hash) that fossil fuel generators produce. This provides "normal" stabilized AC power that load devices are designed to use, increasing their efficiency, and reducing possible device malfunction or damage. This results in an electrically clean, operationally silent, low-carbon footprint AC supply without undesired CO or CO_2 greenhouse gas emissions.

Small hand-carried units (1kWh) are an ideal source of power for test instruments, fiber optic splicers, and other tools. Larger units (6.2kWh) are easily transported in a technician's vehicle and can be used to power network power supplies, small-cell radios, Wi-Fi APs, etc. These AC generators are a sustainable form of energy with multiple charging methods, including AC mains rapid charger; efficient maximum power point tracking (MPPT) solar charge controller with conventional or flexible solar panel arrays; and simultaneous AC plus solar charging for very fast re-charging.

5. Final Considerations

Energy storage continues to evolve. While lithium iron phosphate (LiFePO4, or LFP) is one of latest commercially available technologies, this is an active area of research and development. New approaches will emerge, some of which may not even fall within the traditional battery category.[7] The push away from carbon-based energy and toward renewable sources is one driver. How well these products perform specific functions within an HFC network has been a key focus of this discussion.

The evidence is strong. In addition to the green and socially responsible attributes of LFP batteries, they outperform the alternatives in terms of reliability, safety, and efficiency. Of course, no technology does it all. While LFP has three times as much energy density as VRLA, as noted above, other lithium batteries, including NMC, surpass LFP in this metric. That in part explains why NMC batteries do well in the global motive power market, as forklifts and other heavy machinery look for all the power they can get. Some LFP strengths could also appear as drawbacks to some. Just as a boat owner may hesitate to purchase an LFP battery with a ten-year lifespan, the owner of an HFC network looking to sell within a few years may be less motivated to invest in LFP.

Any mention of investment raises financial questions. While economics is beyond the scope of this paper, a few points are in order. On the one hand, LFP requires no precious metals, such as nickel or cobalt, which is a cost advantage in comparison to other lithium batteries. Yet LFP's long lifespan, among other benefits, creates a price premium, especially over VRLA. Operators should conduct their own analysis, including total cost of ownership (TCO). The key takeaway here is that LFP is a powerful technology, well-suited for operators looking to migrate away from a legacy base while addressing their persistent — and in many cases growing — need for standby power, disaster recovery and portable generators.

6. Abbreviations

AC	alternating current
AGM	absorbent glass mat
AGV	automated guided vehicle

Ah	amp hour
AP	access point
BMS	battery management system
CO	carbon monoxide
CO_2	carbon dioxide
CPUC	California Public Utilities Commission
DC	direct current
DoD	depth of discharge
EV	electric vehicle
HFC	hybrid fiber/coax
kg	kilogram
LiFePO4	lithium iron phosphate
LFP	lithium iron phosphate
LSV	low-speed vehicles
MPPT	maximum power point transfer
NiCd	nickel cadmium
NiMH	nickel-metal hydride
NMC	lithium nickel manganese cobalt oxide
OpEx	operating expenditure
PF	power factor
PPE	personal protection equipment
RMS	root mean square
SCTE	Society of Cable Telecommunications Engineers
SoC	state of charge
TCO	total cost of ownership
V	voltage
VRLA	valve-regulated lead acid
Wh	watt hour

7. Bibliography and References

[1] "The Nobel Prize in Chemistry 2019 was awarded jointly to John B. Goodenough, M. Stanley Whittingham and Akira Yoshino "for the development of lithium-ion batteries." The Nobel Prize in Chemistry 2019. NobelPrize.org.

[2] "Global lithium-ion battery capacity to rise five-fold by 2030," 22 March, 2022. Wood Mackenzie. Woodmac.com/press-releases

[3] "Tesla will change the type of battery cells it uses in all its standard-range cars," 20 Oct. 2021. Cnbc.com.

[4] "Lithium Iron Phosphate: Olivine Material for High Power Li-Ion Batteries," Christian Julien, et al., *Research & Development in Material Science*, December 2017.

[5] "Overview of olivines in lithium batteries for green transportation and energy storage," K. Zaghib, et al., *Journal of Solid State Electrochemistry*, vol. 16, 2012.

[6] Communications Network Resiliency, California Public Utilities Commission. Cpuc.ca.gov.

[7] "Flywheel Energy Storage Replacement For Lead-Acid Batteries in CATV Network Stand-By Power Supplies," William D. Bauer, *Journal of Energy Management*, vol. 5, no. 1, March 2020

SCTE STANDARDS JOURNAL

Pathway to a Greener, More Energy Efficient Cable Access Network

A Technical Paper prepared for SCTE by

Rory Sexton, Software Enabling and Optimization Engineer, Intel Corporation – DPG/NESG, SCTE Member 69748
2200 Mission College Blvd,
Santa Clara, CA 95054
rory.sexton@intel.com
+353 61 477 400

David Coyle, Senior Software Enabling and Optimization Engineer, Intel Corporation – DPG/NESG, SCTE Member 65339
2200 Mission College Blvd,
Santa Clara, CA 95054,
david.coyle@intel.com
+353 61 477 400

Eric Heaton, Platform Solutions Architect, Intel Corporation – DPG/NESG, SCTE Member 39527
2200 Mission College Blvd,
Santa Clara, CA 95054,
eric.d.heaton@intel.com
+1 (408) 765 3447

Table of Contents

List of Figures

1. Introduction

The standardization of Distributed Access Architecture (DAA) for Data over Cable Service Interface Specifications (DOCSIS) and further advancements in the flexible MAC (media access control) architecture (FMA) standard have enabled the transition to a software-centric cable network infrastructure [1]. One option that the DAA and FMA architectures allow is for the DOCSIS MAC software to be deployed as a virtual network function (VNF) on general purpose x86 servers in a multiple-system operator (MSO) headend as a virtualized cable modem termination system (vCMTS), while the DOCSIS physical (PHY) layer is housed in a street-unit near subscribers.

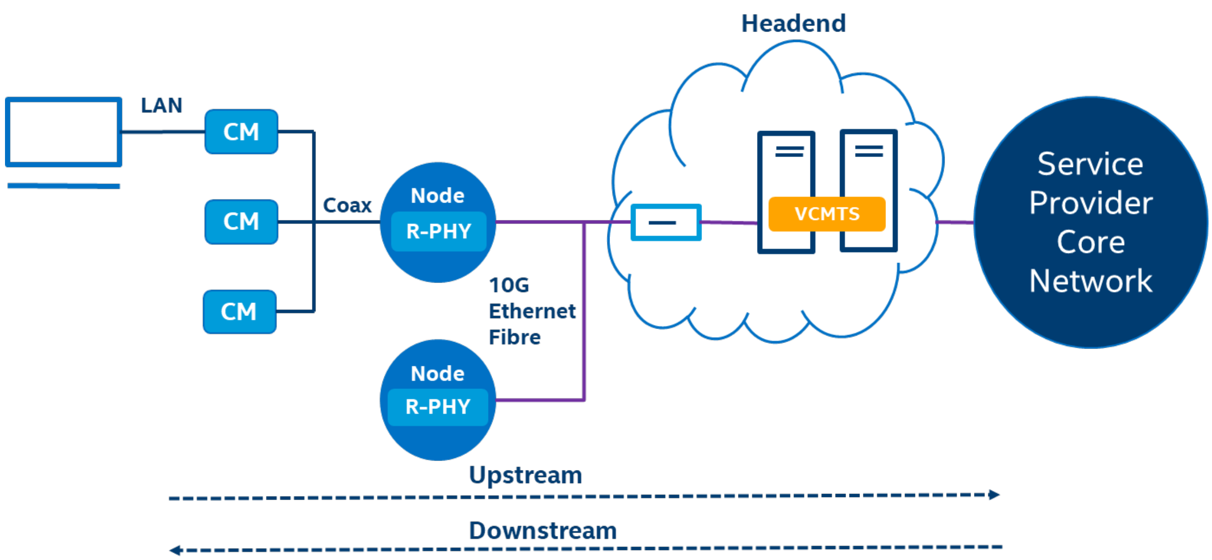


Figure 1 - Cable access network

The DOCSIS MAC component of a vCMTS is typically built on top of technologies such as Data Plane Development Kit (DPDK) [2] or Vector Packet Processing (VPP) [3], both of which are open-source projects that accelerate packet processing workloads running on a variety of central processing unit (CPU) architectures. One of the main principles of DPDK and VPP that help with this packet processing acceleration are poll-mode drivers (PMDs). These are user-space drivers that rely on continuously polling an Ethernet interface for incoming packets, thereby removing any unnecessary interrupt processing overhead, and ultimately boosting performance.

Coinciding with this move to Network function virtualization (NFV), there is also a major push across all aspects of society and industry to reduce our carbon footprints. The cable industry is, and should be, no exception. As NFV and vCMTS deployments become more and more prolific across MSO networks, there have never been more opportunities to find energy efficiencies on a per-server, per-rack, or per-site (headend) basis. This is largely achievable through the use of newly available telemetry and controls to manage hardware and software elements in real time. In fact, it is quite common for the operating system (OS) itself to automatically reduce operating power when it detects that the demand on CPU cores allows entry to a power-optimized state in a given piece of equipment. However, it has traditionally been difficult to identify these periods of lower demand for applications that optimize for performance and relentlessly poll the hardware for packets or events using DPDK PMDs - like the DOCSIS MAC component of a vCMTS. These types of applications appear to fully utilize the CPU cores regardless of the network load they are processing due to their polling nature. This behavior prevents built-in hardware and OS capabilities from effectively controlling power elements for DOCSIS MAC implementations. Alternative techniques are therefore required to achieve energy efficiencies on such deployments.

This paper will discuss the hardware capabilities present on the latest generations of x86 servers (namely C-states and P-states) and new techniques available to flawlessly match the network load at the lowest possible power for these types of data plane applications. It will also address some common misconceptions regarding the usage of some of these capabilities and techniques. In doing so, the paper aims to set out a pathway towards a greener and more energy efficient vCMTS. Through further exploration and detailed lab benchmarking, quantifiable benefits for different system/test configurations will provide actionable recommendations for operators and their vendors creating and deploying NFV solutions in the edge or access network.

2. Measurements

For the purpose of this paper, the CPU power draw was measured on a dual processor server running a reference implementation of a vCMTS DOCSIS MAC data plane. The server contained 2 Intel® Xeon® Scalable Gold 6338N CPUs, each of which has 32 cores. The Intel® vCMTS reference data-plane [4] was used as the workload on the platform. Measurements included in the subsequent sections are based on this hardware and software in conjunction with the power management techniques described in detail in the paper. Similar power savings are achievable on various other hardware CPUs and software stacks.

3. C-states

3.1. Overview of C-states

C-states are power savings states that reduce power consumption on a per-core basis, by turning off specific portions of a core. Entering a power optimized C-state leads to power savings. However, C-states also require that execution of instructions on the core be stopped while the core resides in the power saving state [5].

Several C-States are supported on x86 server processors. Deeper C-states provide greater power savings as more functions of the core are disabled. However, they also have a higher exit latency. While the increased power savings of deeper C-states are desirable, the increased exit latency results in the core taking longer to start executing instructions upon exit from the power saving state. As more functions of the core are powered off in deeper C-states, there can also be an impact even when the core has restarted instruction execution. For example, data may get flushed from caches in the deepest C-state (C6); this data may have to be reloaded into cache as the core resumes execution. While the power savings provided by C-states may be significant, these associated costs also need to be considered when deciding to enable or disable them.

The following are the traditional C-states on an x86 server:

- C0 is the active core state where the core is executing instructions. In C0 the core is considered fully turned on.
- C1 and C1E (C1 Enhanced) are light C-states. In these states the main CPU clock is stopped via software and the CPU voltage is reduced.

- C6 is the deepest C-state offering the greatest power savings. In C6 the cores L1/L2 cache and last level cache (LLC) are also flushed. Disabling these additional functions increases energy efficiency.

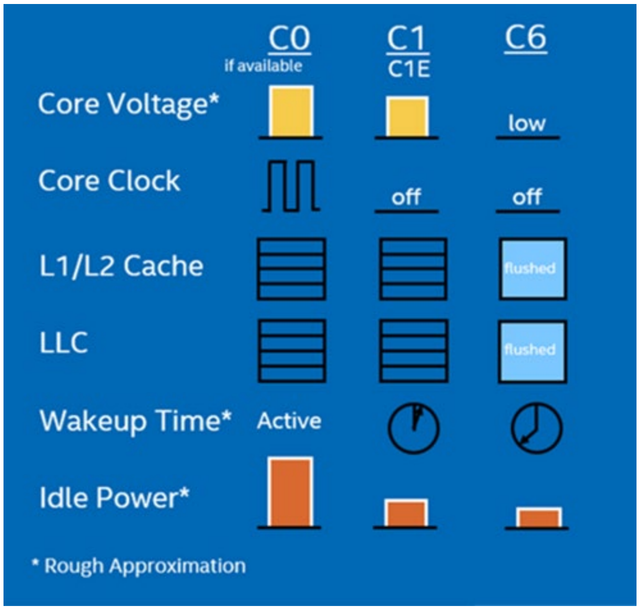


Figure 2 - C-state features

C-states can be controlled either autonomously by hardware, or the OS. If controlled by hardware, they are controlled solely by the CPU power control unit. The OS can also control C-states through the Advanced Configuration and Power Interface (ACPI) governor by the Linux scheduler [6]. In both cases, C-states are triggered by reduced load on CPUs, detected at either a hardware or OS level using the Linux scheduler.

The latest x86 CPU architectures also provide additional, and extremely light, C-states - C0.1 and C0.2 [5] [7]. These new power saving states allow for user-space application software to directly request the core to enter power optimized states, with negligible exit latencies on supported CPUs. C0.1 and C0.2 provide optimum flexibility, carefully balancing power saving C-states without the cost of wakeup times.

3.2. Enabling Traditional C-states (C1, C1E, C6)

Due to the exit latencies of C1, C1E and C6 states and the expectation that they will negatively affect performance, it has generally been considered best practice in the cable industry to disable these C-states entirely for workloads requiring high-performance and determinism. Contrary to this, if used correctly, these C-states can provide significant power savings for such workloads without any negative impact to performance or operation.

Take, for example, a vCMTS deployment on a common off-the-shelf (COTS) server. If C-states are enabled on such a deployment, cores that are not fully utilized will be placed in power saving C-states. While this is true for cores running threads managed by the Linux scheduler, it is not true for cores running DOCSIS MAC threads that continuously poll the hardware for packets or events via a DPDK PMD. These cores will not enter any idle C-state as they will be detected by both the CPU power control unit and the OS to be fully utilized, regardless of the level of network load that they are processing. As a result, these performance critical polling threads will not be affected, while cores running non-performance critical threads will transition between C-states. For example, 20 cores of an x86 CPU may be used for performance optimized polling threads and a further 12 cores for control-plane, infrastructure and failover. With such a deployment and at peak network load, we saw power savings of up to 10% by simply enabling C-states. When the system was made entirely redundant during lower periods of activity, the C-states gave savings of up to 70% of overall CPU power. Figure 3 shows how each CPU core in a dual CPU x86 server would be used in such a deployment. Enabling C-states on the 2 blue (infrastructure) and 10 yellow (control-plane) cores on each CPU allows the platform to save power.

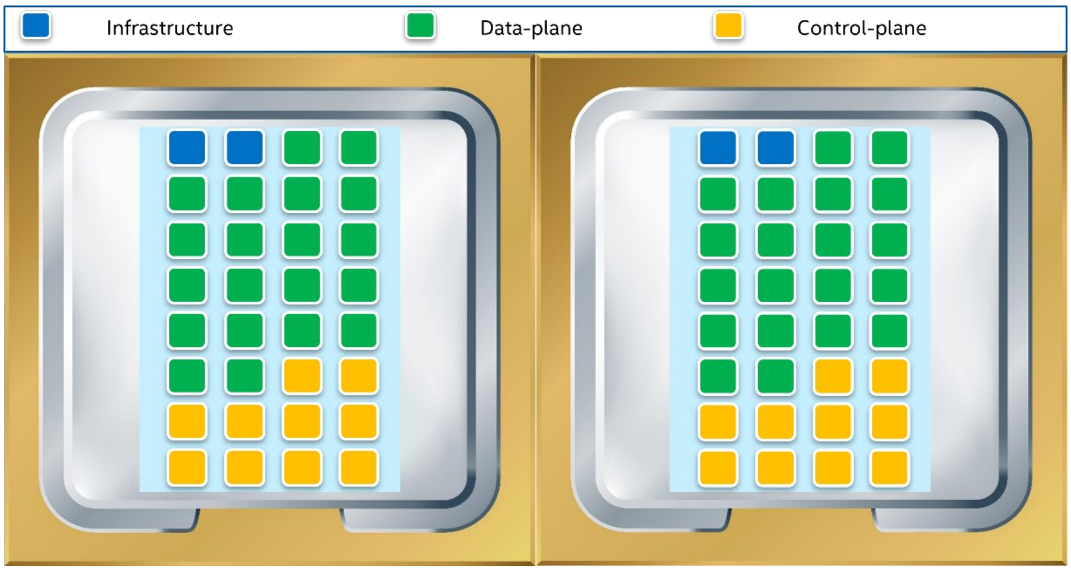


Figure 3 - Core utilization in example deployment

For greater peace of mind that applications will be unaffected in terms of operation and performance, C-states can be enabled or disabled at a per-core level [8]. Each C-state (C1, C1E and C6) can also be enabled or disabled exclusively at a per-core level allowing for fine-grained control over which C-states will be used. This becomes significant when we consider that C1 gives good power savings with an exit latency of just a few microseconds, while the C6 exit latency is tens of microseconds (as reported by the ACPI kernel module on some CPUs). Enabling C1 and disabling C6 can be a good tradeoff between achieving power savings and not impacting the operation, performance, and determinism of certain control-plane threads if required. As power becomes increasingly important, identifying previous misconceptions that are costing operators watts is a valuable exercise. Before disabling C-states, their potential power savings should first be explored by operators and vendors.

3.3. Enabling C-states Efficiencies on DOCSIS MAC Dataplane Cores

While the previous section addressed the use of C-states for threads that do not relentlessly poll hardware for packets, performance optimized DOCSIS MAC threads experiencing periods of lower network load can also benefit from the use of C-states. The main obstacle is that both CPU hardware and the OS are unable to distinguish between high and low load for applications that use optimized polling models. Regardless of the network load they are processing, these types of applications appear to fully utilize the CPU by continuously polling even when they are not receiving packets. This challenge can be overcome by new techniques involving the detection of low loads by the application's DPDK PMD, the component best placed to determine the true real-time load of the application.

Latest CPUs support a WAITPKG instruction set which allows user-space applications to put the core into one of the two previously mentioned, power-optimized C0.1 or C0.2 states. C0.1 has a faster exit latency than C0.2, and both exit much faster than C1 or C6 [7]. Two instructions, in particular, are of interest:

- UMONITOR: Sets up an address range to be monitored by hardware for writes, and activates the monitor
- UMWAIT: Instructs the core to stop instruction execution until the monitored address range is written to. The core can enter the C0.1 or C0.2 states or switch to a hyper-thread sibling.

Support for these instructions has been added recently to several DPDK Ethernet PMDs via its power management application programming interface (API) [9]. When enabled, these PMDs monitor the number of packets received each time the network interface card (NIC) is polled. After a configurable number of empty reads, the PMD issues the UMONITOR instruction to activate a monitor on the next receive descriptor address of the NIC. It then issues the UMWAIT instruction, allowing the core to enter one of the aforementioned power-optimized states. As soon as the address is written to (signaling there is a new packet available), the core is woken up instantly and it continues processing traffic. Unlike C1 and C6, these lighter C-states provide power savings but are more suited to performance critical cores, thanks to their negligible exit latencies.

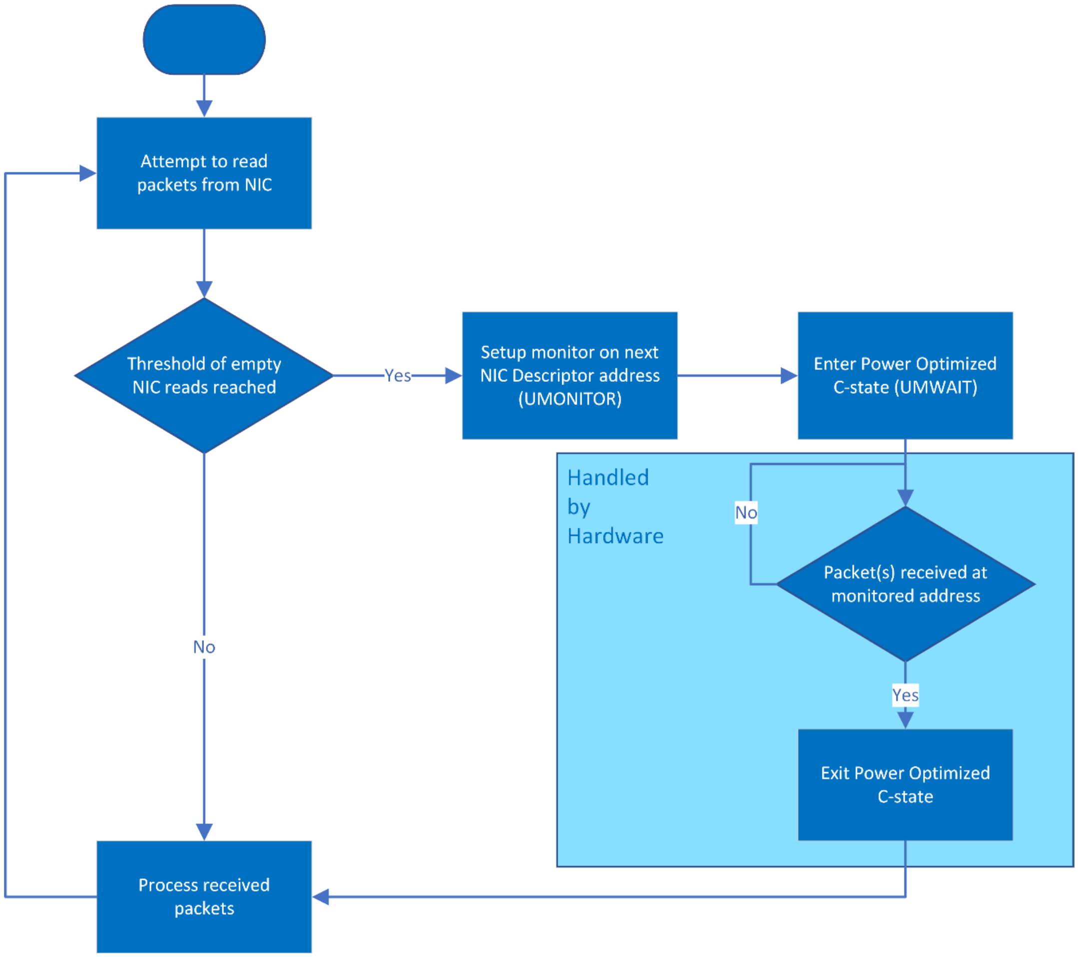


Figure 4 - UMONITOR, UMWAIT flow chart

This technique can be easily integrated into an application as it is implemented within the PMDs and requires only minor modifications to the application itself to enable the feature. The example code within DPDK can be adapted for any PMD, making this new power saving technique suitable for any user-space application driven by a PMD. Other virtualized workloads have seen a CPU power saving in the region of 2-8%, varying with the network load, and it is expected that a DPDK-based vCMTS deployment would achieve similar energy efficiencies.

4. P-states

4.1. Overview of P-states

Performance states, or P-states, is the term used to describe a specific frequency and voltage that a core operates at while executing instructions. Power can be saved by the reduction in operating frequency and voltage achieved by using lower P-states. P-states allow for the frequency and voltage to be controlled on a per-core basis. Unlike C-states, the execution of instructions continues while P-states are adjusted on a core. Thus, P-states provide a mechanism by which to save power while the core continues to operate, albeit at a lower frequency and voltage reducing the instruction execution rate [5].

The associated frequency of P-states varies greatly across different CPUs. Due to this, they are commonly referred to by names. P-states are named P1 to Pn, where P1 is the guaranteed base frequency that all cores can run at and Pn is the lowest frequency P-state providing greatest power savings. The operating frequency of P-states is measured in MHz. An example of P-state distribution is shown in Figure 5, ranging from P1 to P15 in 100MHz decrements.

Figure 5 - Example P-state distribution

P-states can be controlled using hardware-controlled power states (HWP), meaning they are controlled entirely by hardware based on the individual load of each core and OS hints. Alternatively, they can be controlled by direct requests from the OS using the intel_pstate or acpi_cpufreq driver. While HWP gives more autonomy in how P-states are controlled, it can, again, be difficult for the hardware to differentiate between peak loads and periods of low demand due to the polling nature of workloads such as a vCMTS. Direct OS control is more useful in such deployments as operators and vendors can use custom software to accurately detect the network load and match the P-states in an energy efficient manner.

P-states can have an associated transition latency during which core execution is temporarily paused as the core changes from one state to another. This transition latency was once a limiting factor in terms of their usage. However, due to major improvements in the architecture and reductions in this latency, it is now at a point where the potential impact on workload latency and jitter are much less apparent, though it remains a consideration in solutions.

4.2. Potential Power Savings using P-state Tuning Techniques

As the P-state is lowered across cores, the CPU power draw reduces accordingly. The potential savings will vary depending on the number of cores being scaled down and the extent to which the P-states are reduced on those cores. Figure 6 below shows the results of an experiment we ran, giving an estimation of the potential power savings for a typical vCMTS deployment on an x86 server where the P-state is lowered to different levels on 40 cores, each running a polling-based implementation of a DOCSIS MAC.

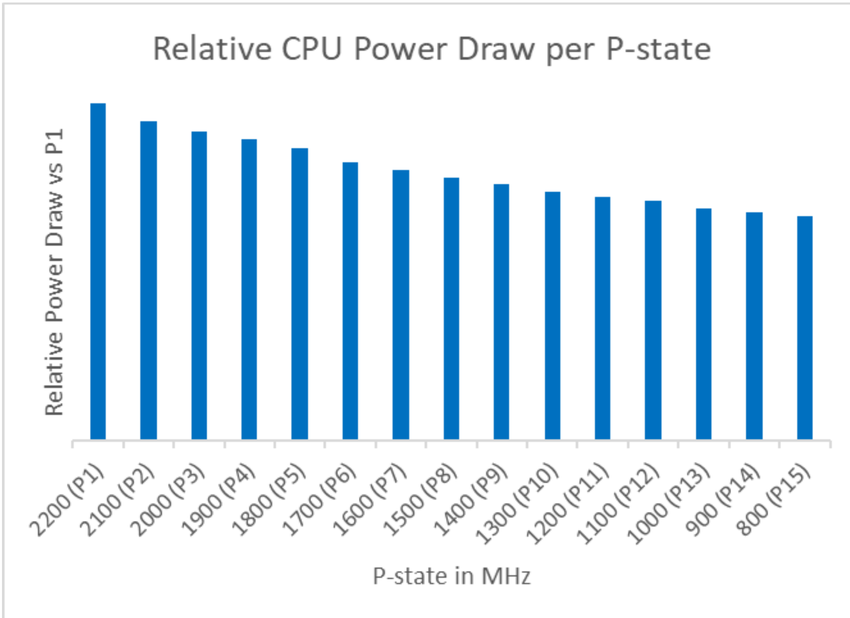


Figure 6 - Relative CPU power draw per P-state

From this initial exploration, we can conservatively predict that the savings achievable by tuning P-states sits in the envelope somewhere between 5% to 30% of CPU power draw. Energy efficiencies within this range are expected for all the P-state tuning techniques discussed in the subsequent sections.

4.3. P-state Tuning Based on Predictable Traffic Patterns

The most basic method to achieve power savings with P-states is to configure a pre-determined core frequency capable of handling the expected network load. For the most part, network load follows a reasonably predictable pattern over a 24-hour period. By studying varying levels of network load and the P-state requirements of a specific DOCSIS MAC implementation, operators and vendors can make accurate estimations of what P-state to set cores to for selected periods of a 24-hour timeframe. Pre-adjusting the frequency of cores in such a manner is sure to provide power savings, particularly at nighttime when networks are usually under-utilized.

This rudimentary approach, however, is not without considerable pitfalls. Pre-configuring the P-state leaves the operator susceptible to unexpected increases in network load atypical of a normal 24-hour period. Such increases in load could be caused by unforeseen events or certain social holidays and would result in degradation of service for the end subscriber. The lack of real-time metrics used in this P-state tuning technique give it clear limitations.

4.4. In-band P-state Tuning

As was the case with C-states, the application’s DPDK PMD is also very well placed to make decisions on the P-state of the cores. The PMD can monitor the load it is feeding to the application and scale the frequency of the core accordingly in a “just-in-time” fashion. As the first reception point of packets, the PMD is the first place capable of detecting an increase in load and quickly scaling up the frequency of the core. Such a technique is an example of in-band P-state tuning, where the decision-making and action to scale the core up or down is contained entirely within the user-space vCMTS application. The exact algorithms used to calculate the P-state setting can vary. The advised approach is to conservatively scale down frequency and save power but aggressively scale up frequency to reduce the likelihood of a degradation in service when an unexpected burst of packets is detected. Once the informed P-state decision is made, the PMD adjusts the P-state via the Linux kernel system file system "sysfs". Again, DPDK contains examples of how this in-band frequency scaling technique can be used on many different CPUs [9].

Although P-state transition latency has been greatly reduced, the response time between the increase in network load and the P-state being increased remains the most important design consideration when implementing in-band P-state tuning techniques. NIC descriptors generally only contain enough space in their buffers to hold at most a few milliseconds worth of packets, thus a response time in this region is required to prevent packet loss. In-band techniques are capable of reactions within this time frame as the PMD or user-space application itself is responsible for detecting the increase in network load, making a P-state adjustment decision, and applying the determined P-state.

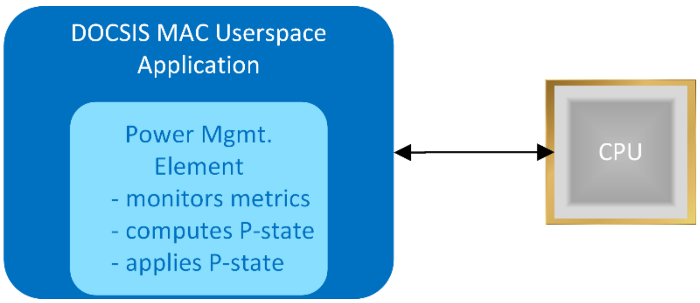


Figure 7 - In-band P-state tuning

4.5. Out-of-band P-state Tuning

Out-of-band P-state tuning refers to P-state tuning techniques that are managed by an entity outside of the user-space application itself. In the case of a vCMTS DOCSIS MAC implementation, real-time telemetry can be delivered to an external agent. Such an approach not only separates the power management logic from the application itself but also allows for a single agent to control power elements of the entire platform. This is a distinct advantage when placing containerized DOCSIS MAC implementations in orchestrated environments. In a container-based infrastructure, privileged permissions are required to apply power management controls. Restricting such permissions to a single power agent on a node simplifies the placement of network functions and ensures the entity applying power controls is aware of the entire platform.

Similar to in-band P-state tuning, P-state reaction time is a key factor when implementing out-of-band tuning. A fast response time in the region of a few milliseconds becomes more difficult to achieve when P-states are being managed by an external entity. This is in large part due to the requirement for telemetry input for the P-state decision to propagate to the agent. A low-latency mechanism is, as a result, necessary for an effective solution.

Solutions can be based on two categories of telemetry. The first is application specific metrics including, but not limited to, active subscriber statistics, throughput per active cable service group, and average size of PMD dequeues. These types of metrics provide an exact insight of the network load of a vCMTS

DOCSIS MAC implementation. With awareness of these metrics, a power management agent can apply informed P-state actions. The telemetry stack used is of prime importance to ensure metrics can be delivered to the external agent within the aforementioned timeframe.

Alternatively, the external agent can achieve energy efficiencies by monitoring platform specific metrics on a per-core basis and applying suitable P-state changes. While HWP and the OS are unable to distinguish between varying network load on polling workloads, custom software can be used to do exactly that by monitoring detailed core metrics. Instruction rates, cache utilization and branch distribution are examples of per-core statistics that can provide an accurate representation of the true network load of a core. By first measuring these statistics at known levels of network load during a training period, an accurate determination of network load can then be computed by the agent as it closely monitors real-time metrics.

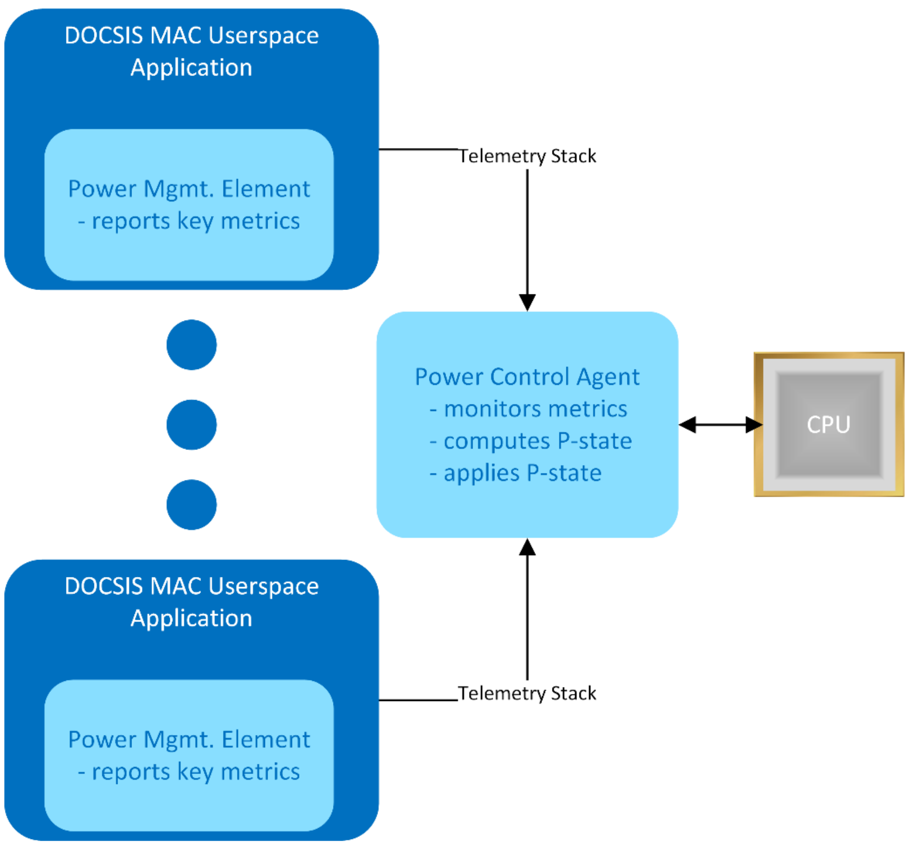


Figure 8 - Out-of-band P-state tuning

5. Conclusion

As operators accelerate the deployment of NFV workloads, including vCMTS, opportunities to do so in a more energy efficient manner must be capitalized upon. This is best enabled via the newly available techniques and controls discussed in this paper. Both power saving C-states and P-states play a major role in the pathway to a greener, more environmentally friendly vCMTS deployment on COTS x86 servers. In the experiments we ran, legacy and newly available C-states provided CPU power savings of up to 10% under high network load and up to 70% reduction in CPU power draw on an idle system. P-state tuning techniques also showed significant energy savings of up to 30% of CPU power draw on a vCMTS deployment. We strongly recommend operators and vendors perform detailed analysis of the power management controls available to them and carefully consider their integration either within the vCMTS application itself or as a separate software agent. By developing tailored algorithms and undergoing lab benchmarking operators and vendors will begin to develop a greater understanding of the power elements under their control with the view to maximizing energy efficiencies on their vCMTS deployments.

6. Abbreviations and Definitions

ACPI	advanced configuration and power interface
API	application programming interface
COTS	common off-the-shelf
CPU	central processing unit
DAA	Distributed Access Architecture
DOCSIS	Data Over Cable Service Interface Specification
DPDK	data plane development kit
FMA	flexible MAC architecture
HWP	hardware-controlled power states
LLC	last level cache
MAC	media access control
MSO	multiple-system operator
NIC	network interface card
NFV	network functions virtualization
OS	operating system
PHY	physical
PMD	poll mode driver
sysfs	virtual file system provided by the Linux kernel
vCMTS	virtualized cable modem termination system
VNF	virtual network function
VPP	vector packet processing

7. Bibliography and References

[1] B. Ryan, M. O'Hanlon, D. Coyle, R. Sexton and S. Ravisundar, “Maximizing vCMTS Data Plane Performance with 3rd Gen Intel® Xeon® Scalable Processor Architecture,” [Online]. Available: https://networkbuilders.intel.com/solutionslibrary/maximizing-vcmts-data-plane-performance-with-3rd-gen-intel-xeon-scalable-processor-architecture.

[2] “DPDK (Data Plane Development Kit),” Linux Foundation Projects, [Online]. Available: https://www.dpdk.org/.

[3] “FD.io - The World’s Secure Networking Data Plane,” Linux Foundation Projects, [Online]. Available: https://fd.io/.

[4] Intel Corporation, “Intel vCMTS Reference Dataplane,” [Online]. Available: https://www.intel.com/content/www/us/en/developer/topic-technology/open/vcmts-reference-dataplane/overview.html.

[5] K. Devey, D. Hunt and C. MacNamara, “Power Management - Technology Overview,” [Online]. Available: https://builders.intel.com/docs/networkbuilders/power-management-technology-overview-technology-guide.pdf.

[6] The Kernel Development Community, “CPU Idle Time Management,” [Online]. Available: https://www.kernel.org/doc/html/v5.0/admin-guide/pm/cpuidle.html#.
[7] Intel Corporation, “Intel 64 and IA-32 Architectures Software Developer's Manual,” [Online]. Available: https://software.intel.com/content/www/us/en/develop/download/intel-64-and-ia-32-architectures-software-developers-manual-volume-2b-instruction-set-reference-m-u.html.
[8] “Comms Power Management Github,” Intel Corporation, [Online]. Available: https://github.com/intel/CommsPowerManagement.
[9] “DPDK Power Management,” Linux Foundation Projects, [Online]. Available: https://doc.dpdk.org/guides/prog_guide/power_man.html.

SCTE STANDARDS JOURNAL

A Step By Step Methodology For Optimal Access Transformation Planning

A Technical Paper prepared for SCTE by

Sudheer Dharanikota, Chief Executive Officer,
First Principles Innovations, LLC, SCTE Member
111 Fieldbrook Ct.
Cary, NC 27519
sudheer@fpinno.com
+1-919-961-6175

Luc Absillis, Chief Operating Officer,
First Principles Innovations, LLC
111 Fieldbrook Ct.
Cary, NC 27519
luc@fpinno.com
+1-919-345-2655

Mario Di Dio, Vice President,
Kyrio,
858 Coal Creek Circle
Louisville, CO 80027
m.didio@kyrio.com
+1-303-661-3749

Table of Contents

List of Figures

1. Executive Summary

Broadband access networks are constantly evolving to keep pace with ever-growing subscriber demand, competitive pressure, and the requirement of launching new revenue-generating services. The access transformation plan for implementing this evolution drives most of the future investments and future operational complexity, and can be a gating factor for revenue opportunities. Building an optimal access transformation plan is a key factor in an operator's business success.

Building such a plan is a complex collaborative process. This paper introduces a multi-step methodology to break down the complexity and accommodate the uncertainty of future assumptions. It showcases simple examples and highlights the decision-making process from a financial point of view. We show how different business requirements such as product roadmaps and budgetary constraints will influence the optimal solution to turn into a realistic executable solution. However, even though the financial implications of the access plan have been considered, a similar comprehensive analysis needs to be conducted from the operational, service, and technological risks points of view.

2. Introduction

Cable operators are going through major access network transformations. These are driven by one of the following: competitive pressures [1]; targeted response, fixed wireless confusions [2]; 10G evolution goals [3]; or recent wins of Rural Digital Opportunity Fund (RDOF) funding [4] from the government. To address these needs operators are considering many access technology transformation options [5], [6], [7]. All these transformation options come at an expense from the financial, operational, and capabilities points of view. We have presented many papers on how to analyze different access transformation options; see, for instance, [8], [9], and [10]. The burning question to all operators is which is the right transformation strategy? This simple question that industry leaders are trying to answer is very complex and involved. One key takeaway from all the access transformation papers is that building an access network transformation strategy is a data-driven interactive process that cannot be mimicked with a single algorithm. The process starts with creating in-depth insights on all transformation options through evaluating multiple transformation strategies, optimization algorithms, and what-if scenarios. Using these insights, an informed decision can be made on the transformation strategy that best fits the corporate goals. Our goal in this paper is to provide a comprehensive mechanism on how to arrive at such a plan from a *financial requirements* point of view.

The fundamental questions one needs to ask include:

- What are the clearly defined transformation goals?
- How do we evaluate and compare the transformational plans uniformly?
- How can we be sure that this is the best transformation strategy amongst available options?
- How do we align the company strategy into an optimal solution?

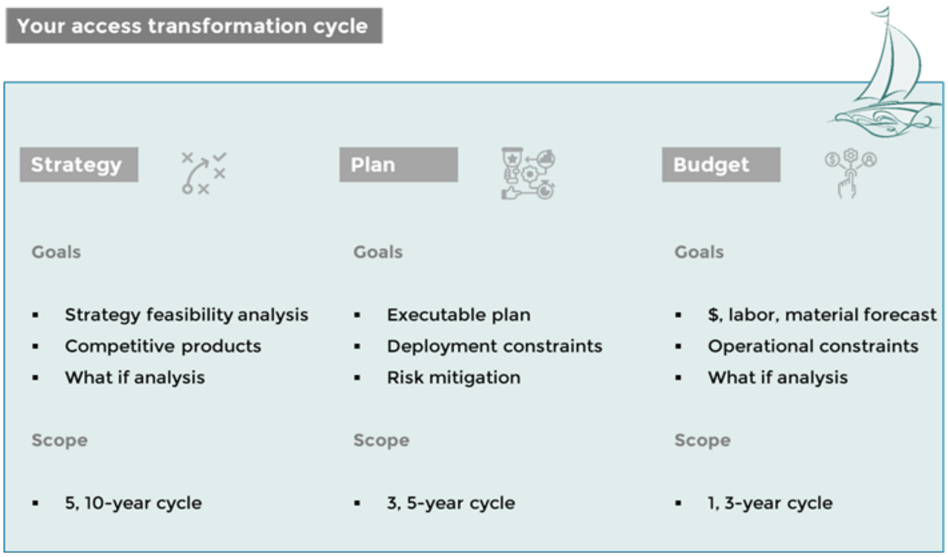


Figure 1 – A typical access transformation goals and scope

To understand the financial implications of the transformation choices, one needs to understand the access transformation cycle. This is elaborated in [8], [11]. As shown in Figure 1, when performing access transformation, one needs to consider the strategic, planning, and budgeting needs. Typically, **strategic planning** is done with a long-term vision in mind over a five- to 10-year horizon. While doing such long-range planning there will be many risks/unknowns (financial, operational, and technical) that the team needs to consider. Hence, they need to evaluate these through different *what-if* scenarios. The strategic plan should result in an **executable plan** that mitigates risks such as financial budget limitations, resource availability, etc., by establishing limits. Typically, these are done over three- to five-year cycles. Such an executable plan drives the yearly **budgetary planning** cycles more focused on the operational challenges such as market-level spending, labor, and material challenges.

As we have emphasized, access transformation planning is a multi-dimensional analysis that should consider realistic scenarios. In this paper, we focus on the financial aspects of such strategic planning and show how one can derive an optimal transformation strategy.

3. High-Level Process

3.1. SOFT Framework

Before diving into the financial transformation optimization, we want to present the other aspects of the access transformation plan that have implications (as explained in Figure 2) for the organization. We call this the services, operations, finance, and technology (SOFT) framework.

Creating an access transformation plan that is **executable** and **meets the goals** is essential for success in this hyper-competitive environment. To define what meets the goals, the planning team should consider:

- The financial implications that include the revenue through product offerings and deployment costs;
- The technology choice implications that include the risks and capabilities of newer technologies, and the deployment challenges;

- The operational implications that include operating the networks and managing customers; and
- Finally, the service implications that include keeping the customers happy and retaining them while offering next-generation services.

Figure 2 – Access transformation analysis dimensions

For a proper 360° planning, analysis in all four dimensions of SOFT is essential. But in this paper, we focus on the financial optimality aspects of the transformation planning.

3.2. Understanding Total Investment

For any access transformation planning, knowing where your access network is now and what is your target state at the end of the planning period, as shown in Figure 3, is essential. To reach the target access network state, cable operators have many transformation upgrade options as explained in [12] and summarized in Figure 3. All these upgrade options cost different amounts and will have different investment timings (that is, different net present value or NPV of the investment). In some cases, the target state can be achieved with a fundamentally different end-state technology option (such as FDX or FTTH). The financial decision between these upgrade options can be made based on the total NPV or total cost required for these upgrade options. There may be many viable upgrade paths to reach a target

SCTE STANDARDS JOURNAL

state. From the financial point of view, the operator needs to consider minimizing **the total investment** in making the right choice.

Figure 3 – Access transformation upgrad options to reach a target state

An **investment difference** between two upgrade options – $option_1$, and $option_2$ – is the NPV cost difference of the upgrade cost overlay – that is, NPV ($option_1$) or NPV ($option_2$) – for the operator reaching from the current state to the target future state.

For example, as shown in Figure 3, an operator who has deployed D3.1 with Node + 5 status on an 800 MHz plant and a sub-split configuration wants to reach a symmetrical 10 Gbps capable solution in 10 years. The question here is what is the least expensive investment option the operator needs to take? Let's say the operator is considering the following two upgrade options:

- $Option_1$: *Reach N+0 FDX*: Upgrade to 1.2 GHz plant, N+0, and eventually to FDX
- $Option_2$: *Reach XGS-PON*: Upgrade to XGS-PON based FTTH

The investment difference in selecting $option_1$ versus $option_2$ will be NPV (*Reach N+0 FDX*) – NPV (*Reach XGS-PON*) for the whole operator network over the 10 year period.

Our goal in this paper is to determine an optimal plan that an operator should use to reach their target state. This goes beyond comparing a few scenarios. It involves finding the right order of upgrades to reach the target state with the optimal NPV cost. Note: Keep in mind that a financially optimal plan may not be an optimally executable or operational plan. In this paper, we focus only on financial optimality.

3.3. Transformation Methodologies

The key part of building any access transformation plan is prioritizing the upgrade options one wants to consider. To create an optimal solution, we needed to include all possible upgrade options for each state of a node. In this paper, we included the possibility of different types of upgrades a cable operator might consider. These options include:

Spectrum upgrades:

- Increase the upstream capacity through mid-split, high-split, and full-duplex options
- Increase the downstream capacity by adding additional OFDM blocks or adding full-duplex blocks (can include the increase of the overall plant capacity by upgrading to 1.2 GHz or 1.8 GHz)
- Increase the capability of fiber-to-the-home from XGSPON to NGPON2 or 100G PON

Sub reduction upgrades:

- Reduce the number of subscribers per node through node splits
- Reduce the number of subscribers per node through fiber deep options with N+0 FDX nodes
- Convert the HFC nodes to fiber-to-the-home nodes

One of the low-cost options available for the cable operators is managing the spectrum allocation on the plant (between the node and the home). As shown in Figure 4, the operator can use the shared spectrum amongst homes on a node that can be effectively used as an upgrade lever[1]. The first spectrum lever is the total available spectrum per node. Typically, cable networks have a downstream upper frequency limit of 750 MHz, ~ 800 MHz, or 1 GHz now. There is a lot of work in progress to support 1.2 GHz and 1.8 GHz plants. The other option is to carve out the spectrum for upstream and downstream usage judiciously. Most of the operators are using sub-split (5 MHz to 42 MHz) for upstream usage. The future upstream spectrum options available for the operators now are mid-split (5 MHz to 85 MHz) or high-split (5 MHz to 204 MHz). These upstream and downstream bandwidths are used for carrying different DOCSIS technologies, as shown in the figure. Note that it is not the intention of this paper to go into the details of the spectrum allocation logistics, but to demonstrate the available options to operators. Note also that concurrent usage of the same spectrum in the upstream and downstream direction using FDX is not shown in the figure.

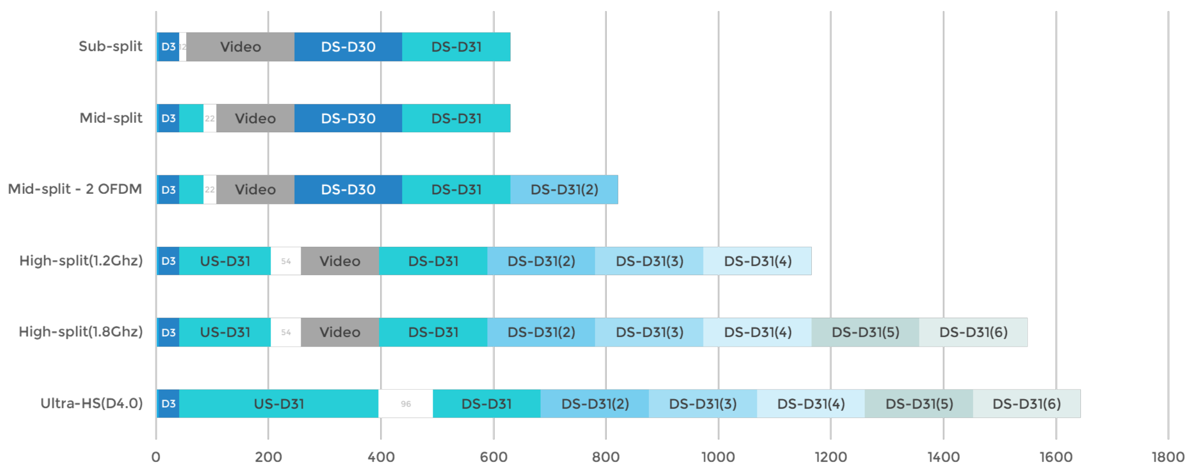


Figure 4 - Different spectrum allocation options that are available to operators

Figure 5 shows a graphical representation of all the upgrade options considered in the transformation plan used for this paper, including the above-mentioned spectrum options. The analyses used in this paper

[1] The costs of moving from current to future state spectrum options are considered in our analysis.

consider a six-year quarterly plan[2]. The ellipses in the graph represent the technology state of a node at a given time. The arcs in the graph represent the valid options between the technology states. Note that in

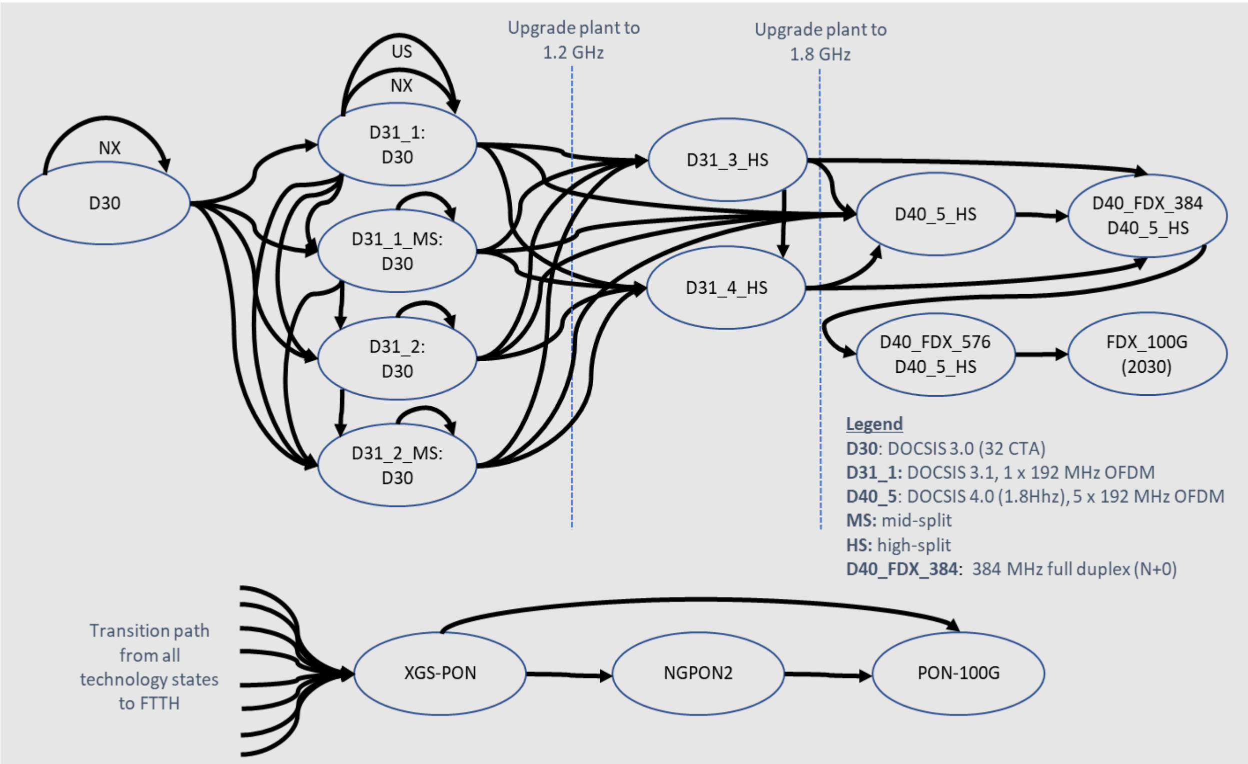


Figure 5 - Different HFC and FTTH access upgrade options considered in this paper

the figure we did not provide the priority of choosing an option, which is also essential to pick the right transitional option when more than one future technology state can solve the current upgrade needs.

3.4. Different Upgrade Strategies

The upgrades shown in Figure 5 represent the cardinality of the possible. An operator can manually specify the priorities of the upgrades, for example in a tool such as AP-Jibe [13]. This provides the best control of the upgrade strategy but can be very cumbersome and requires an in-depth understanding of the option priority impact.

Exhaustive optimization

The most apparent optimization criteria for a network transformation are the total investment cost or the NPV of the total investment cost. With all the possible upgrade paths available in the model, it is possible for each node to **exhaustively** calculate all the viable upgrade paths and pick the path that offers the least investment (cost or NPV). For this paper, we defined two different exhaustive criteria to determine what the *best solution* means:

---

[2] The costs used for all upgrade actions in this example are based on the default cost included in the AP-Jibe tool. These costs are averages based on extensive industry research by our team.

- **The lowest NPV**: In this exhaustive optimization criteria, for every node in the network, pick the upgrade path that keeps the node compliant with the needs and has the lowest total NPV for all the upgrade costs incurred during the analysis period.
- **The lowest cost**: In this exhaustive optimization criteria, for every node in the network, pick the upgrade path that keeps the node compliant with the needs and has the lowest total cost for all the upgrades incurred during the analysis period.

The advantage of such brute force optimization or exhaustive optimization is that we will find the least cost upgrade path at the node level. On the other hand, the optimal is for the exact set of inputs and evaluation period. As explained in [14], calculation of the optimal upgrade path for 10 years versus looking at the first 10 years of an optimal path calculated for 15 years may present drastically different results. In addition, the solution may provide a different upgrade path at a node level and may not be an optimal solution from the technology, operations, and service point of view.

Business criteria based optimization

As an alternative to the manual upgrade strategy or an expensive exhaustive optimization, the operator can use different corporate strategies as guiding principles to pick the best upgrade paths. The following are the three common classes of corporate strategies (Refer to ***Greedy versus Exhaustive algorithms*** sidebar):

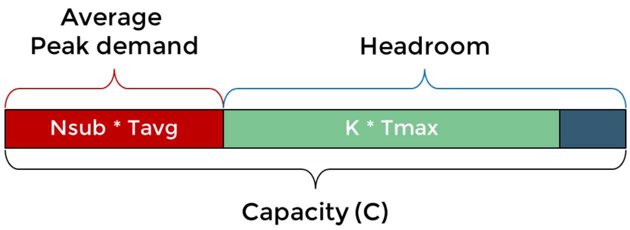
Kick the can down the road: In this strategy, when a node needs to be upgraded, the preference is to pick the lowest cost option that satisfies the upgrade requirements. This is the *lowest cost* greedy optimization algorithm.

> **Greedy versus exhaustive algorithms**
>
> Greedy algorithms: These are the algorithms used to select the optimal upgrade option based on the selected criteria. The option is selected without any knowledge of the future needs.
>
> Exhaustive algorithms: These algorithms evaluate all viable upgrade paths for the full duration of the analysis and pick the optimal path based on optimization criteria such as minimize total cost.

Minimize network upgrade actions (capacity-based): In this strategy, when a node needs to be upgraded, the preference is to pick the viable option that provides the most added capacity to the node. That is because, intuitively, upgrade options that add the highest capacity will survive upgrades the longest. Using this option mitigates the risk of unforeseen demand growth increases (e.g., the COVID-19 pandemic impact). This is the *highest capacity* greedy optimization algorithm.

Least cost per capacity: A middle ground strategy tries to lower the network upgrade frequency and provide some growth risk mitigation without always using the biggest upgrade step. This upgrade strategy picks the viable upgrade option with the least cost per added bit of capacity. This is the *least cost per bit* greedy optimization algorithm.

Recently operators have been focusing more on network evolution strategies that combine demand growth with Quality of Experience (QoE) triggers [15], also referred to as K-factor triggers, based on the formula illustrated in Figure 6: C >= Nsub * Tavg + K * Tmax



C	The service group capacity
N_{sub}	Number of subscribers in a service group
T_{avg}	Average peak time consumption per sub
K	Access network QoE factor
T_{max}	Maximum speed tier offering

Figure 6 - Quality of Experience based capacity allocation

K-factor triggers are also commonly used to incorporate future speed tier roadmap requirements into the access transformation plan.

Upgrades based on K-factor triggers are driven by the available headroom (see Figure 6) on an interface or service group rather than the capacity available per subscriber on the network. It, therefore, makes sense to include alternative upgrade strategies that focus on headroom rather than capacity:

Minimize network upgrade actions (headroom-based): In this strategy, the preference is to pick the option that adds the most headroom to a node's interface. The maximum speed tier an operator can deploy on an access link is bounded by the available headroom. A higher headroom strategy is the best hedge against competitive threats. This is the *highest headroom* greedy optimization algorithm.

Least cost per headroom: In this strategy, the preference is to pick the option with the least cost per added bit of headroom. This is the *least cost per headroom* greedy optimization algorithm.

If the operator can compare results for these strategies side by side, it creates valuable insights and allows them to quickly refine your transformation plan and pick a strategy that is closest to your vision.

A greedy algorithm by nature considers local optimization without considering the future implications of the decisions. For this reason, the operator may not get the lowest cost solution, but it provides alignment with the company strategy and more importantly, the upgrade decisions will be uniform across the organization.

3.5. Six Steps to Reach an Optimal Plan

Finally, let's create a process for reaching an optimal access network transformation plan

(Step 1) *Define the current and target status of the transformation plan*: As explained before in the paper, a transformation plan must be defined with a clear target state. Also, it is equally important to get the current state of the network as accurate as possible, preferably at the node level. The status of the node should include the deployed technology, homes passed composition information, distribution mileage information, and possibly the location of the node. It is also essential to forecast the bandwidth demand growth as accurately as possible.

(Step 2) *Create all the transformation requirements*: A transformation is driven by the transformation requirements or the drivers such as the product roadmaps, budgetary constraints, technology availability assumptions, etc.

(Step 3) *Find the transformation options, their costs, and the resource requirements*: Before running the planning exercise, gather details on the upgrade options that are available to you in terms of their capabilities, costs that will be incurred to make the transitions, and the resource requirements to make the transitions.
(Step 4) *Apply transformation business rules*: Apply different transformation business rules to your upgrade strategies such as market-level strategies to address competition, and preferred K-factor values while committing to the product offering, etc.
(Step 5) *Run different scenarios with different upgrade strategies and optimization methodologies*: Use the above rules across multiple optimization criteria and get node-level details. Aggregate them and use them to compare with the other scenarios.
(Step 6) *Compare the results against the evaluation criteria*: Compare different optimization strategies, discuss amongst the leaders, and decide what makes sense for your company goals.

We elaborate on this process with the help of the examples in the following sections.

4. Developing Optimal Solution

Before performing the optimal transformation planning, the operator needs to collect the following high-level strategic information:

- The current network status at least at a node level;
- The competition such as their network status, upgrade options, etc.; and
- The target state goals.

In the following subsection, we provide a couple of examples that we use in this paper to demonstrate optimal transformation planning. Even though we could consider enterprise-level planning, we considered a single market in each case to go deeper into the analysis. In a real operator's analysis, multiple such markets with potentially differing strategies may be applied.

4.1. Example 1: Meet the Rural Competition

A cable operator that is serving a rural market with some fiber overbuilder competition is assessing its transformation strategy. Here is some information that we would consider in this example. Note that this section covers Steps 1 – 4 of the optimal transformation planning process. Also note that we do not go into the details of the cost and resources needs as highlighted in Step 3, as it is out of the scope of the paper.

- The market statistics:
    - Homes: ~150K (SFU 65%, MDU 22%, businesses 13%)
    - Subscribers: ~120K (~50K DOCSIS 3.1, ~70K DOCSIS 3.0)
    - Network status: 800 MHz plant
    - Node spectrum configuration:
        - Downstream DOCSIS 3.0 – 32 QAM, DOCSIS 3.1 - 1x192 MHz OFDM, video - 40 QAM channels
        - Upstream – Sub-split (up to 42 MHz)
- The competition for this operator is mostly the local fiber overbuilders who are deploying FTTH
- The operator's goals include:
    - Compete with FTTH by offering 1/1 Gbps by 2026 and meet the new broadband definition

- Accomplish these offerings with the least cost to reach the target state
- The customer demand growth assumed for this market is
  - SFU and MDU heavy nodes upstream 30%, downstream 35%
  - Business heavy nodes upstream 32%, downstream 42%
- The planned top tier product roadmap is shown in the table

Year	Downstream BW	Upstream BW
2022	1 Gbps	35 Mbps
2023	1 Gbps	100 Mbps
2024	1 Gbps	500 Mbps
2026	1 Gbps	1 Gbps

- The operator budgetary constraints are not considered in this example
- Upgrade options
  - In an 800 MHz plant, any upgrade to more than one block of 192 MHz OFDM is not possible
  - When an upgrade of the plant is needed the operator chooses to jump to 1.2 GHz with high-split

4.1.1. Scenario Optimization Analysis

After completing Steps 1– 4, the operator needs to use a planning tool, such as AP-Jibe [13] as a next step (Step 5) to get a sense of the organic network evolution needs. Typically, the organic network evolution is driven by the customer demand growth profiles. For this example, the operator may identify the market technology transitions as shown in Figure 5.

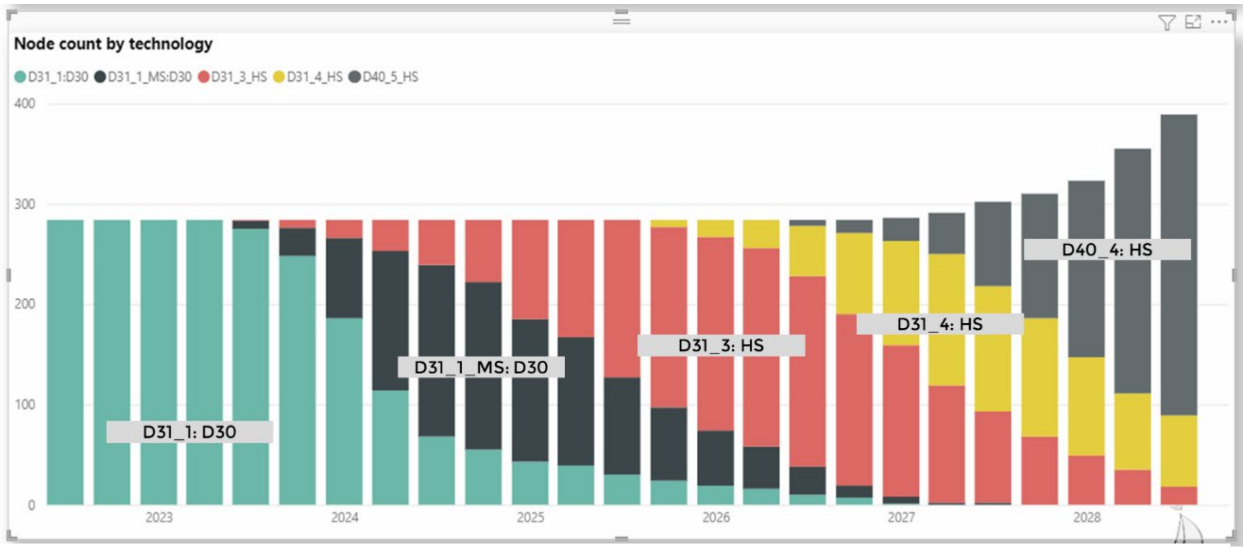


Figure 7 - Example 1 organic demand growth based access upgrade analysis

As part of Step 5, we also recommend for the operator to include business requirements such as the product needs, budgetary constraints (not included in this example), etc. This gives the operator a real picture of the node upgrades, as shown in Figure 7. Note that the product needs in the upstream direction are triggering mid-split and high-split faster (as shown in Figure 7) than the organic growth as shown in Figure 8. This is the time an operator should reconsider any changes to the business requirements.

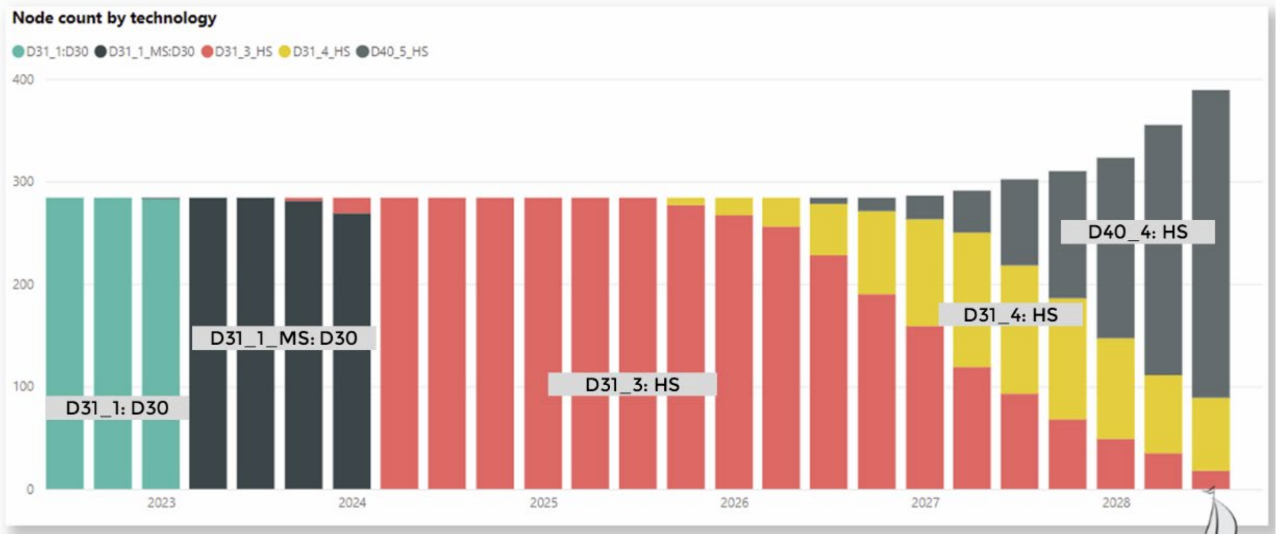


Figure 9 - Example 1 business constraints with the organic growth based access upgrade analysis

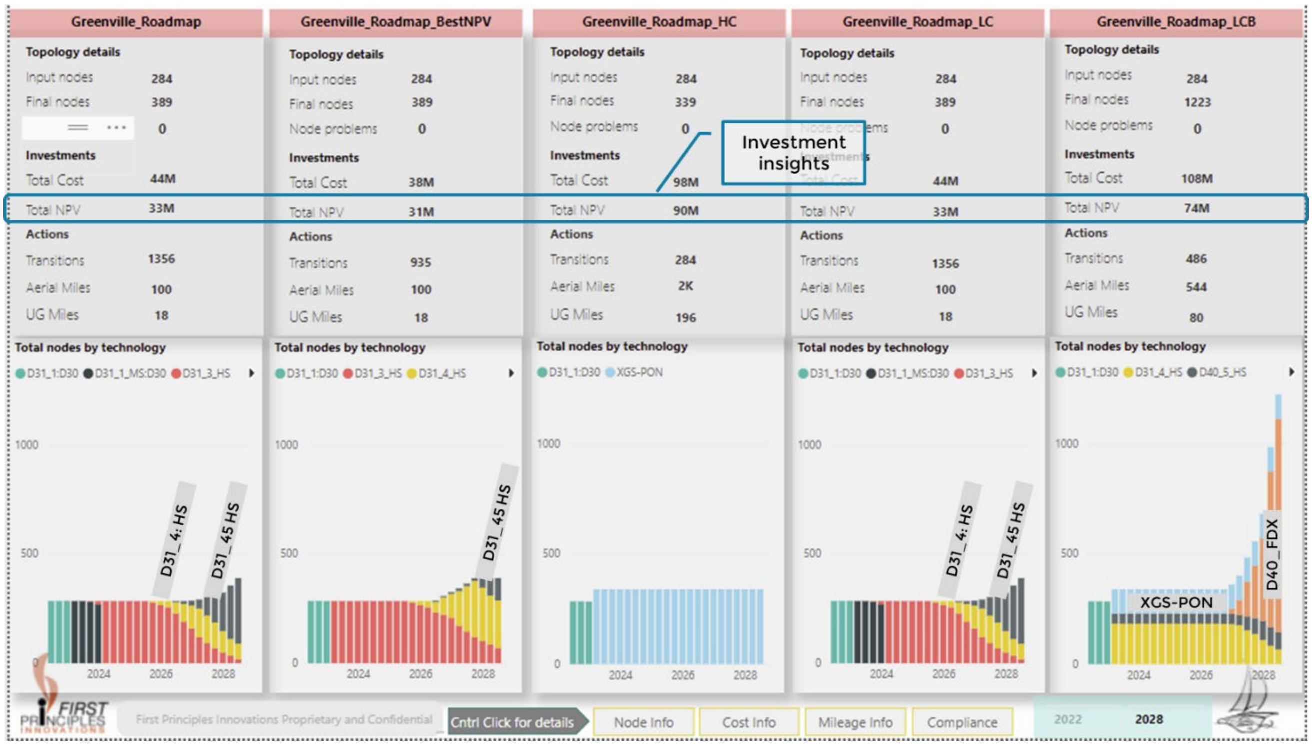


Figure 8 - Comparison of exhaustive and business criteria based optimal solutions for example 1

Once the baseline upgrade plan is created, it is time to find an optimal solution *that meets the business goals*. Refer to [14] for a detailed discussion on the optimization options. As shown in Figure 9 an operator can compare these options side by side to understand the implications. The basic roadmap gives the non-optimal solution. The best NPV provides the exhaustive optimization of the upgrade strategy without looking into the business strategies, whereas the next three provide different business strategies based on optimization. In Figure 9 the business strategies included in the comparison are: *Minimize network upgrade actions* (Greenville_Roadmap_HC); *Kick the can down the road* (Greenville_Roadmap_HC); and *Least cost per capacity* (Greenville_Roadmap_LCB). A quick comparison shows that not doing the optimality analysis will have an investment difference of $2 M ($33 M - $31 M). The highest capacity option (HC) will have a $59 M ($90 M - $31 M) investment difference. The lowest cost option (LC) will on the other hand still have an investment difference of $2 M ($33 M - $31 M) but will have a significant overall cost of $44 M. Making such side by side comparison gives the leaders the financial impact of their choices or not making such an analysis.

Figure 10 - Demonstration of a bias created by the exhaustive analysis with bounded endpoint

Note that the network upgrades do not stop after the six years of the analysis that we are performing in this example. Hence, conducting a six-year optimal analysis is going to create a bias towards low-cost options at the end of the analysis period. To avoid such bias, we recommend analyzing longer periods, such as 10 years for six years, and picking the first six years, as shown in Figure 10. Also, note that a low-cost example does not always mean an executable plan. In this paper, we focused only on the financial dimension, but not the other dimensions such as operations to analyze the 360° optimal solutions.

All these insights can be used to recommend an upgrade strategy to the leadership team, as shown in Figure 11. The recommendation includes the insight from the best NPV solution to forgo the upgrade to mid-split in the early years to realize the cost and NPV benefit. In the later years, the recommendation is to diverge from the best NPV solution to favor future safeness by selecting upgrades to D4.0 over node split actions to fulfill the upgrade requirements.

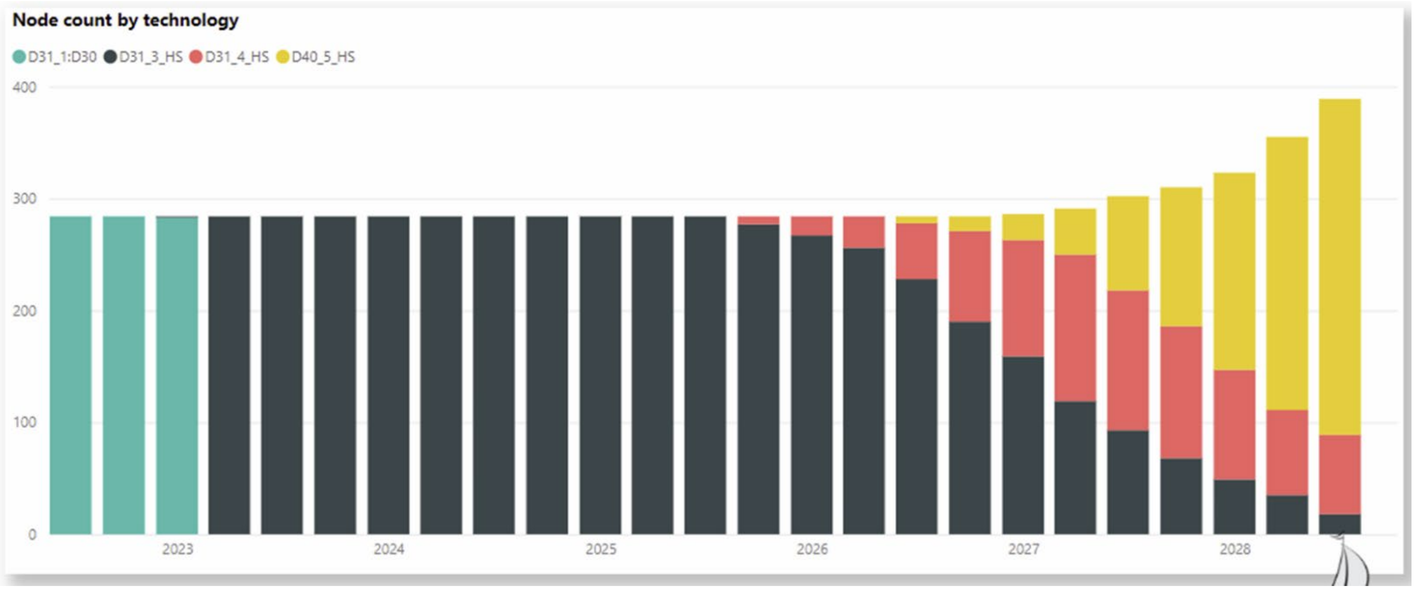


Figure 11 - Example 1 optimized access transformation plan that meets the roadmap constraints

4.2. Example 2: Targeted Deployment to Meet Aggressive Competition

A cable operator that is serving an extremely competitive urban market is assessing its transformation strategy. Here is some information that we would consider in this example. Note that this section covers Steps 1 – 4 of the optimal transformation planning process. Also note that we do not go into the details of the cost and resources needs as highlighted in Step 3, as it is out of the scope of the paper.

- The market statistics:
  - Homes: ~502K (SFU 65%, MDU 18%, businesses 17%)
  - Subscribers: ~251K (~201K DOCSIS 3.1, ~50K DOCSIS 3.0)
  - Network status: Recently upgraded to 1 GHz plant
  - Node spectrum configuration:
    - Downstream DOCSIS 3.0 – 32 QAM, DOCSIS 3.1 1 x 192Mhz OFDM with 40 QAM channels for video
    - Upstream – Sub-split (up to 42 MHz)
- The competitors for this operator are the telcos and the fiber overbuilders who are aggressively deploying FTTH and pushing symmetrical speeds. Their opportunities include gaining new customers with targeted better products and retaining their existing customer base.
- The operator's goals include:
  - Compete with FTTH by offering 5/2 Gbps by 2026 and meet the new broadband definition

- Accomplish this with the least cost to reach the target state but also be able to offer the best future-safe products
- The customer demand growth assumed for this market is
  - SFU and MDU heavy nodes upstream 30%, downstream 40%
  - Business heavy nodes upstream 32%, downstream 42%
- The planned top tier product roadmap is shown in the table

Year	Downstream BW	Upstream BW
2022	1 Gbps	100 Mbps
2023	1 Gbps	500 Mbps
2024	1 Gbps	1 Gbps
2026	5 Gbps	2 Gbps

- The operator's budgetary constraints include: 75M in 2022 with a 5% incremental budget per year
- Upgrade options – all options in Figure 5 are being considered

4.2.1. Scenario Optimization Analysis

Note: We have repeated some of the explanations that we used in Example 1 in this example also for the sake of clarity. We suggest the readers pay attention to the subtle differences.

After completing Steps 1– 4, the operator needs to use a planning tool, such as AP-Jibe [13], as a next step (Step 5) to get a sense of the organic network evolution needs. Typically, the organic network evolution is driven by the customer demand growth profiles. For this example, the operator may identify the market technology transitions as shown in Figure 12.

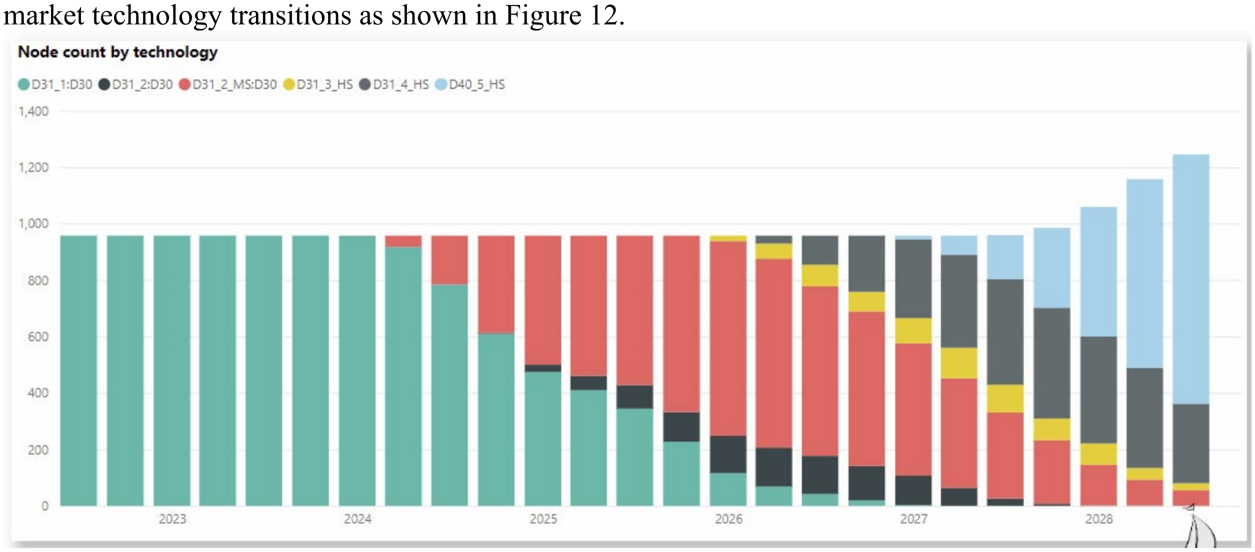


Figure 12 - Example 2 organic demand growth based access upgrade analysis

As part of Step 5, we also recommend for the operator to include business requirements such as the product needs, budgetary constraints, etc. This gives the operator a real picture of the node upgrades, as shown in Figure 12. Note that the product needs in the upstream direction are triggering the mid-split, high-spit, and major transition to FDX when 5G symmetrical (as shown in Figure 13) is introduced rather than the organic growth as shown in Figure 12. This is the time an operator should reconsider any changes to the business requirements.

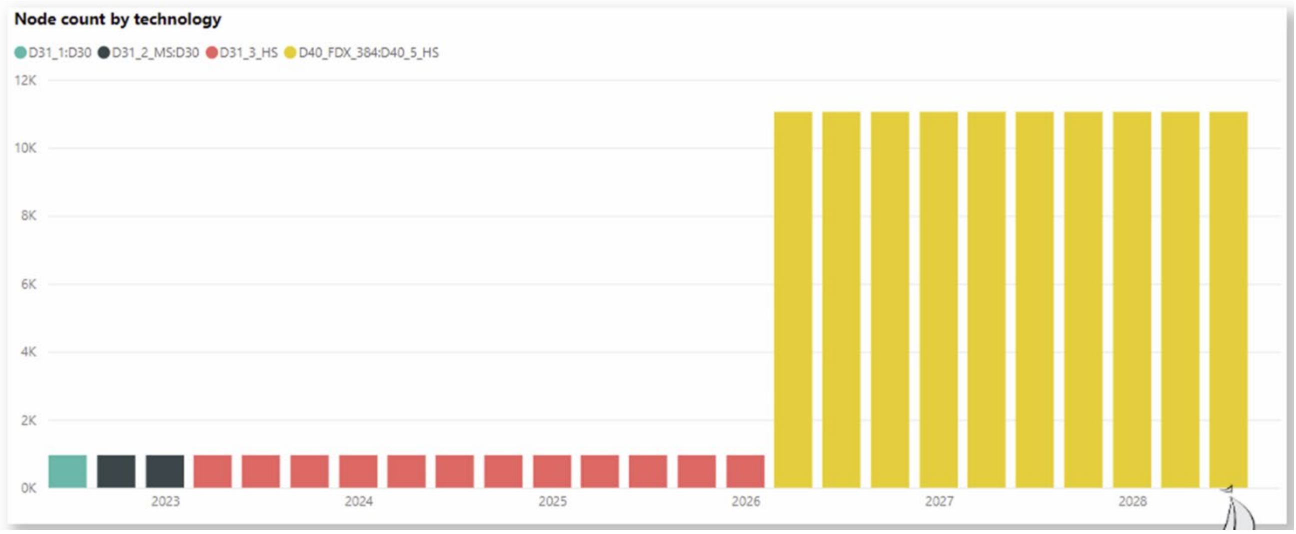


Figure 14 - Example 2 roadmap constraints with the organic growth based access upgrade analysis

Once the baseline upgrade plan is created, it is time to find an optimal solution *that meets the business goals*. Refer to [14] for a detailed discussion on the optimization options. As shown in Figure 14 the operator can compare these options side by side to understand the implications. The basic roadmap gives the non-optimal solution. The best NPV provides the exhaustive optimization of the upgrade strategy

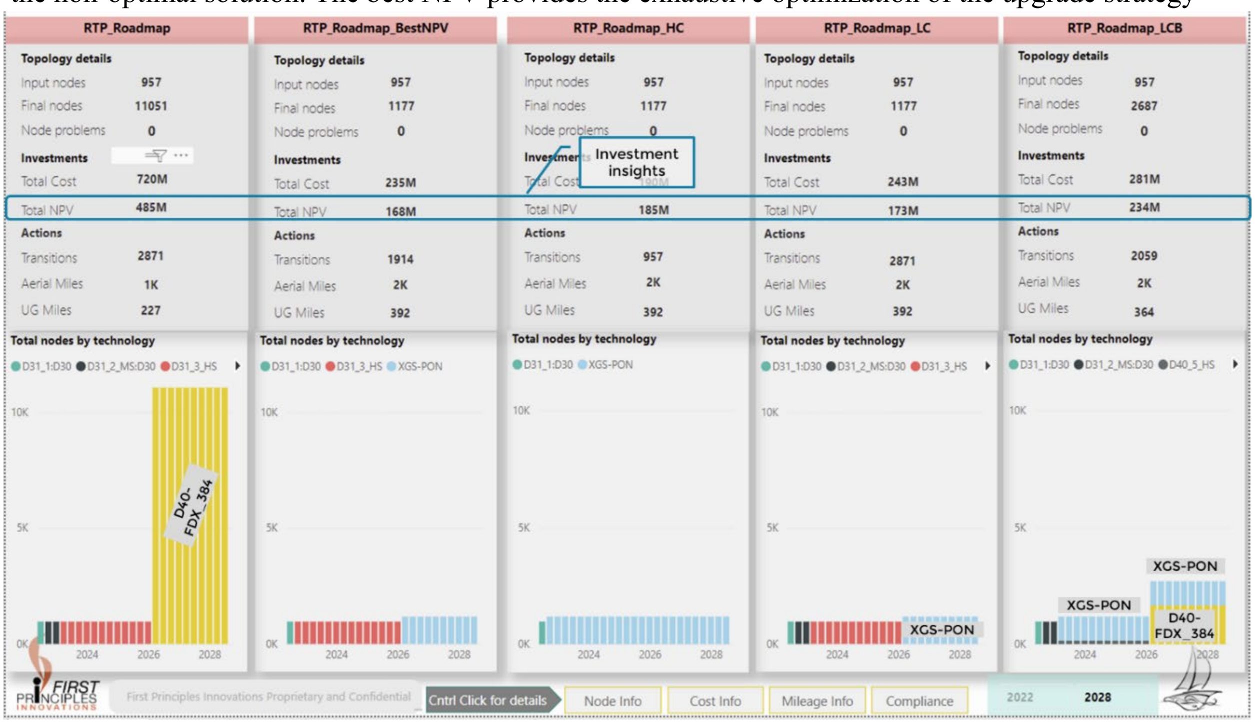


Figure 13 - Comparison of exhaustive and business criteria based optimal solutions for example 2

without looking into the business strategies, whereas the next three provide different business strategies based on optimization. In Figure 14 the business strategies included in the comparison are: *Minimize network upgrade actions* (RTP_Roadmap_HC); *Kick the can down the road* (RTP_Roadmap_HC); and *Least cost per capacity* (RTP_Roadmap_LCB). A quick comparison shows that not doing the optimality analysis will have a significant investment difference of $317 M ($485 M - $168 M). The highest capacity option (HC) will have a $17 M ($185 M - $168 M) investment difference but will have a lower overall cost of $190 M. The lowest cost option (LC) will on the other hand still have an investment difference of $5 M ($173 M - $ 168 M) but will have a significant overall cost of $243 M. Making such side by side comparison gives the leaders the financial impact of their choices or of not making such an analysis.

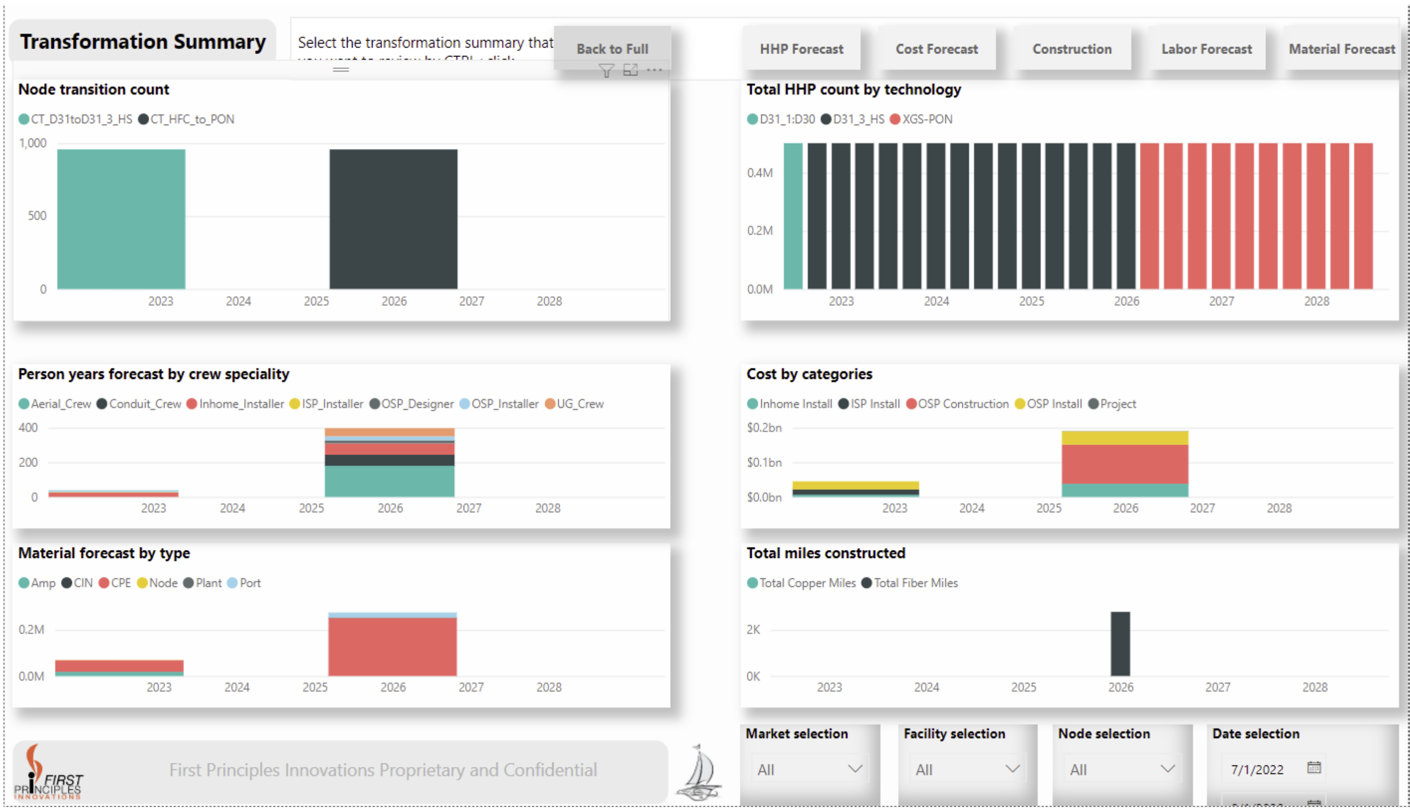


Figure 15 - Comparing SOFT parameters before committing to the final plan is essential

Before committing to a plan, we recommend the operator performs a 360° view on the drivers of all the SOFT parameters such as costs, resources, etc. For example, as shown in Figure 15, it is clear that upgrading the full market to FTTH in one period will run into budget and resource issues. To mitigate the financial risk, realistic yearly budget caps for the market can be overlayed on the solution. Such budgetary restrictions to spread the activities, as shown in Figure 16, even though more realistic to implement, will delay node upgrades that can cause issues with compliance (due to roadmap requirements) or even with the health of the node (due to congestion). This forces the operator to debate the priorities of the nodes to determine which nodes can be delayed in the upgrades, as shown in Figure 17. At the end of this optimal prioritization exercise, the operator needs to determine a more realistic executable transformation option.

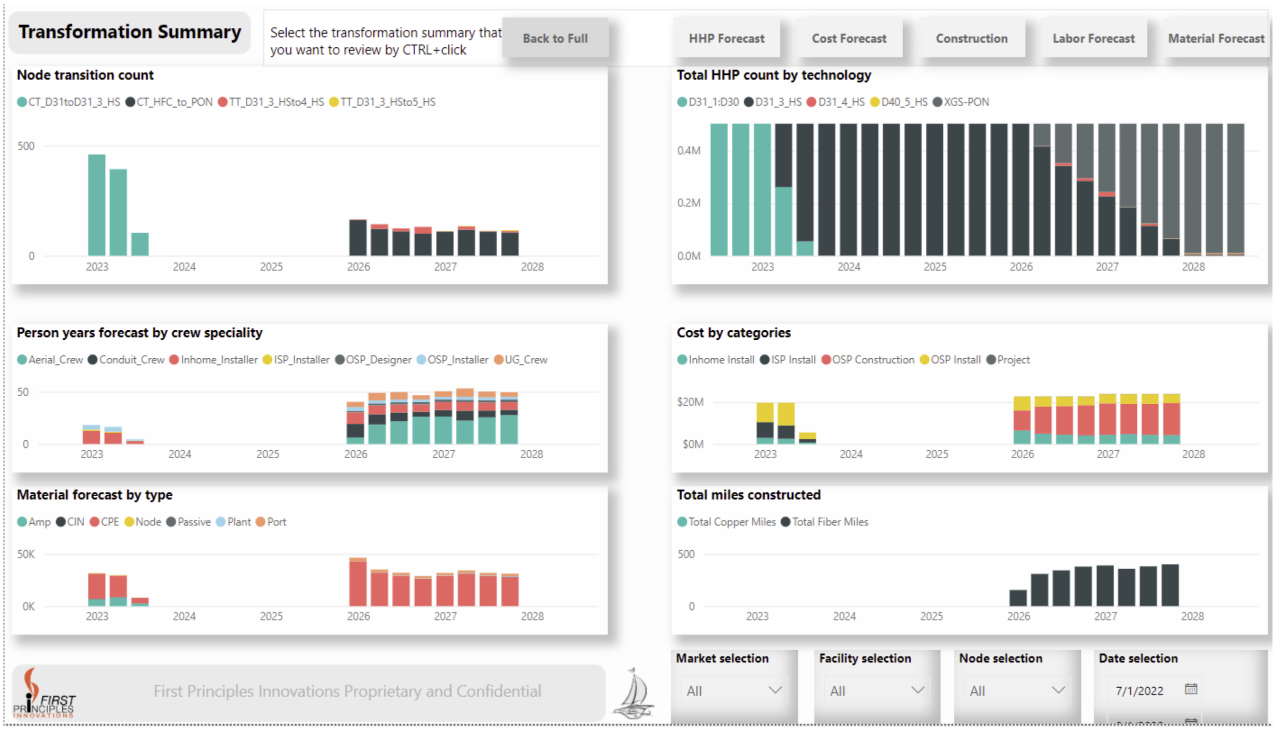


Figure 16 - Impact of budgetary constraints on activities, resources, construction and cost

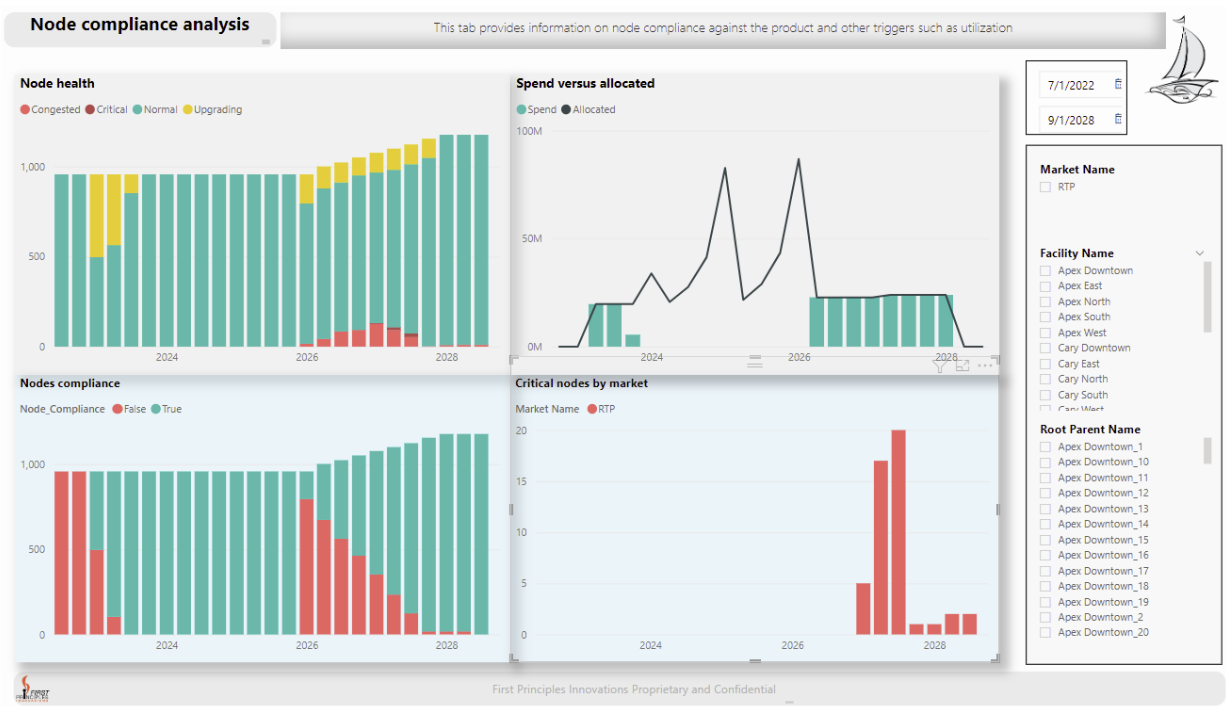


Figure 17 - Node compliance and node health when budgetary constraints are applied

5. Conclusions and recommendations

With a plethora of technology options available to cable operators, it is no surprise that many upgrade paths can be found that fulfill their future service (target state) requirements. However, as seen from the simple examples in this paper, the cost of the solutions can be vastly different. Therefore putting in the effort to find an optimal solution is an absolute must.

Building the optimal access transformation solution is a six-step process but is not a simple calculation, as demonstrated in this paper. The key to the process of creating an optimal solution is generating detailed insights to get the 360° views of SOFT implications (only financial implications are considered in this paper) due to the upgrades. Finding the details of the upgrade options is only part of the problem. The next step is to determine which of these upgrades is optimal.

From a financial perspective, the most important insight comes from calculation of the mathematically optimal path. However, it does not provide the complete picture and should always be complemented by an in-depth analysis of business criteria upgrade strategy, as shown in this paper.

The optimal solution created from a financial perspective is not the end of the road. As shown in the second example, overlaying the plan with realistic financial constraints is a necessary first step toward an implementable solution.

Lastly, it is important to consider risks and constraints from all domains in the SOFT framework to refine the plan to a point where all stakeholders in the organization can be onboard.

6. Bibliography and References

[1] AT&T Press Release, *AT&T Becomes Fastest Major Internet Provider, Delivering Fiber with up to 5-Gigs of Speed*, Jan 2022, available here
[2] Frank Rayal, Sudheer Dharanikota, Fixed Wireless Access in Cable Operator Context – A Performance and Spend Analysis, SCTE CaleTec-Expo, Oct 2021, available here
[3] Luc Absillis, *How to reach 10G systematically?* FPI application note, Aug 2019, available here
[4] *Implementing the Rural Digital Opportunity Fund (RDOF) Auction*, FCC, available here
[5] Belal Hamzeh, *CableLabs Completes Full Duplex DOCSIS Specification*, available here
[6] *DOCSIS 4.0 technology*, CableLabs, available here
[7] *FSAN Roadmap,* ITU standards group, available here
[8] Sudheer Dharanikota, Luc Absillis, Mark Welsko, *Capital and Operational spend aspects in Cable Operator 10G evolution*, SCTE Journal, Sept 2020, available here
[9] Luc Absillis, *A multi-trigger what-if analysis approach to network transformation*, Apr 2020, FPI white paper, available here
[10] Sudheer Dharanikota, Luc Absillis, Rajesh Abbi, *Cable Operator Operational Considerations To Reach 10G Access Networks*, SCTE Journal, July 2021, available here
[11] Sudheer Dharanikota, *Understanding the basics of transformation in telecom, Duke Tech Solutions white paper*, June 2017, available here
[12] Luc Absillis, Sudheer Dharanikota, *Brownfield access network planning*, FPI white paper, available here

[13] *Access Planning Jibe (AP-Jibe), A Cable operator access transformation planning toolset*, From FPI, refer here

[14] Luc Absillis, Sudheer Dharanikota, *Creating an optimal access transformation plan*, FPI white paper, available here

[15] Tom Cloonan et. al, *Simulating the impact of QoE on per-service group HSD bandwidth capacity requirements*, CommScope white paper, available here

SCTE STANDARDS JOURNAL

MSOX - MSO eXchange

A Converged Data Roaming Platform for 5G, 4G, and Wi-Fi Networks

A Technical Paper prepared for SCTE by

John Kim, Director of Connectivity, CableLabs, SCTE Member
858 Coal Creek Circle
Louisville, CO 80027
John.kim@cablelabs.com
303-661-3473

Omkar Dharmadhikari, Lead Mobile Architect, CableLabs, SCTE Member
858 Coal Creek Circle
Louisville, CO 80027
o.dharmadhikari@cablelabs.com
303-661-3875

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

1. Introduction

CableLabs believes convergence will be the driving force in making a truly connected world a reality. CableLabs' view on convergence is expansive and covers multiple areas of convergence. One of these areas is service provider (inter-operator) convergence, which is a framework for proving a consistent and seamless user experience across multiple operator networks and is featured by security, policy, and service portability among operators. The framework can range from opening mutually accessible network interfaces to deeper integration, such as network sharing. Ultimately, service provider convergence allows operators to scale their networks and create opportunities beyond their respective coverage footprint and/or core competencies by seamlessly and efficiently leveraging one another's network resources.

MSO eXchange (MSOX) is a building block for enabling service provider convergence. MSOX is a unified roaming framework that interconnects Multiple System Operators' (MSOs) mobile and Wi-Fi wireless deployments and allows MSOs to converge each other's wireless networks.

The main motivation for the creation of the MSOX framework is fueled by a scenario in which MSOs offer mobile services as Mobile Virtual Network Operators (MVNOs), relying on the network coverage from Mobile Network Operator (MNO) hosts. With the advent of smartphones and widespread deployment of mobile technologies like 4G LTE (and now 5G NR), wireless connectivity is becoming an integral part of connectivity service offerings for MSOs. Several MSOs that lack mobile network infrastructure rely on the MVNO model to supplement their connectivity offerings with mobile. The recent CBRS auction outcome indicates that, at minimum, some U.S. MSOs are exploring an option of building out their own, albeit targeted, mobile networks. When deployed, these networks can be used in conjunction with Wi-Fi to offload MVNO subscriber traffic from the MNO network. This offload approach lessens the MSO's reliance on the MNO network, with an ultimate goal of improving the MVNO's economics. CableLabs' recent study mapped out novel MVNO architectures for converging MSO and MNO wireless networks to provide increased control and visibility and enhanced user experience for MSOs. However, these architectures did not address how MSOs can leverage each other's wireless network coverage when their mobile subscribers travel outside their market boundaries.

The proposed MSOX concept provides a mechanism for MSOs to roam into one another's wireless networks outside their respective markets, when available, rather than relying on the MNO network. As MVNOs start deploying their own mobile radio networks, the facilitation of roaming among these networks will improve the economics of their wireless offerings and enable differentiated services. With adjacent coverage footprints from several MSOs, MSOX can enable seamless session transfers across inter-MSO coverage boundaries for improved user experiences. MSOX can also be used to revamp and reinvigorate inter-MSO Wi-Fi roaming by addressing converged mobile and Wi-Fi roaming scenarios. In short, MSOs can leverage MSOX to use one another's wireless access networks in a seamless manner to provide a competitive alternative to MNOs.

Beyond the roaming aspects, MSOX can serve as a launching pad for additional MSO-powered services. For example, MSOX can host shared mobile network elements for MSOs who want to minimize the cost of deploying and managing such network functions. MSOX can be leveraged to offer enterprise-grade services, such as private mobile networks, by offering the ability to interconnect multiple geographically distributed sites.

The focus of this paper is to provide a high-level technical concept, architecture, and functional description of MSOX; business and regulatory aspects of realizing MSOX are beyond the scope of this paper. The rest of the paper is organized as follows. Section 2 provides a high-level overview of mobile data (4G and 5G) and Wi-Fi roaming architectures defined by the 3rd Generation Partnership Project (3GPP) and the Wireless Broadband Alliance (WBA), as well as operational guidelines and requirements defined by The GSM Association (GSMA). Section 3 introduces the concept of MSOX in detail, including the main motivations for its creation and expected benefits. Section 4 describes the MSOX-supported functions and roaming use cases.

2. Overview of 4G/5G Mobile Data and Wi-Fi Roaming

This section provides a high-level overview of roaming architectures and associated interfaces that have been specified for mobile and Wi-Fi data roaming.

2.1. 4G/5G Mobile Data Roaming

3GPP has specified roaming architectures for both 4G and 5G standards. It specifies the roaming interfaces and network elements that are involved in supporting the roaming functionality. GSMA further codifies the requirements for the implementation and deployment of these 3GPP-specified architectures. GSMA has specified a series of Permanent Reference Documents (PRDs) to facilitate roaming across operators. These documents include a minimum set of technical requirements to facilitate roaming interoperability and interworking, IP traffic routing, formats for billing records, and a template for roaming business agreements between the visited and home operators.

For data roaming, depending on the desired Internet accessibility, 3GPP has specified two architecture options.

- Home Routed (HR): Internet and application access is via the home network.
- Local Breakout (LBO): Internet and application access is via the visited network.

It should be noted that the implementation of bilateral or multilateral roaming agreements does not impact the below-defined architectures as defined by 3GPP and GSMA. The functions and interfaces remain unchanged in both cases.

2.1.1. 4G Roaming

Figure 1 depicts the roaming architectures specified for 4G in 3GPP.

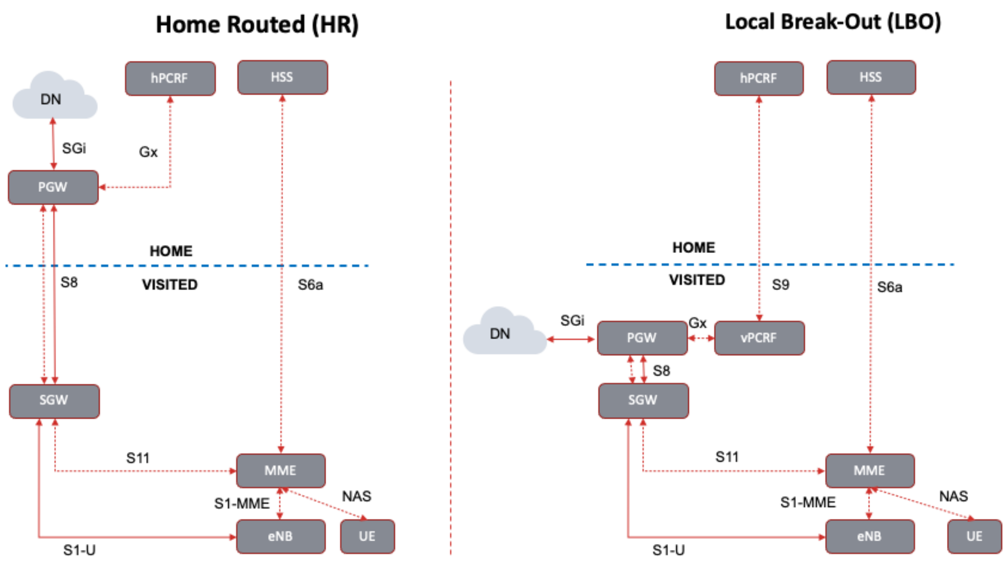


Figure 1 - 4G roaming architecture options

There are three main roaming interfaces specified as part of 4G roaming: S6a, S8, and S9. The S6a interface is used to retrieve the authentication credentials of the subscriber roaming to the visited network and is required in both HR and LBO roaming architectures. S8 is required only in the HR architecture to set up and anchor 3GPP Protocol Data Unit (PDU) sessions in the home network. S9 is required only in the LBO architecture to retrieve user-specific policies for the 3GPP PDU sessions anchored in the visited P-GW.

Implementation of HR and LBO architectures for a roaming subscriber can be on a per Access Point Name (APN) basis. For a given subscriber, it is possible that HR can be set up for one APN and LBO for another APN. An example of such implementation would be where voice APN utilizes LBO while the default data APN utilizes HR.

2.1.2. 5G Roaming

Figure 2 depicts the roaming architectures specified for 5G within 3GPP.

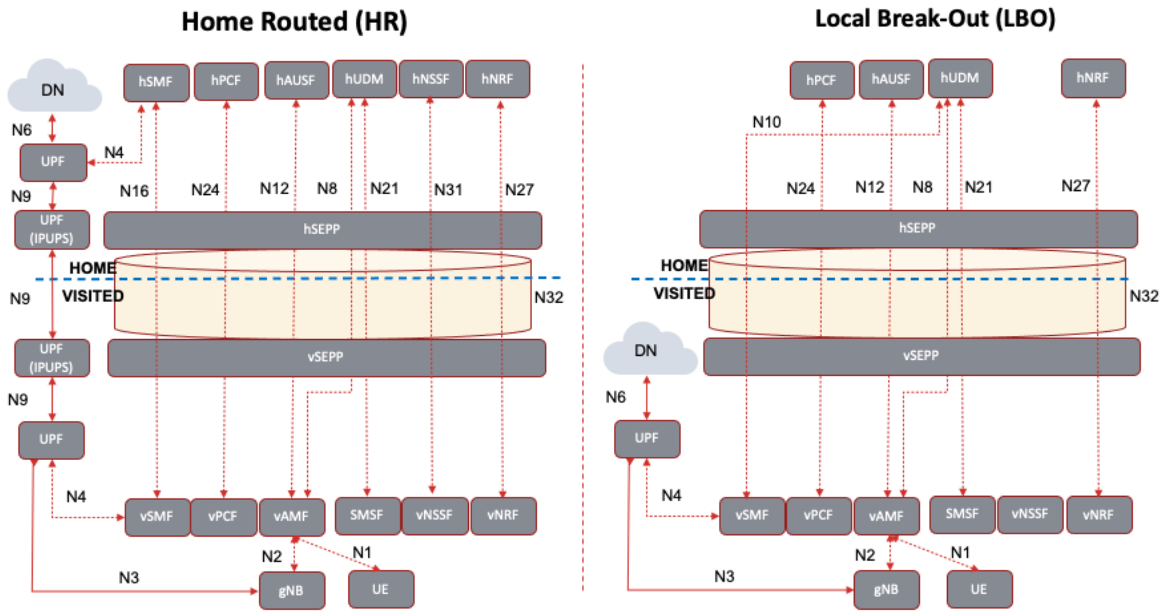


Figure 2 - 5G roaming architecture options

In 5G, like 4G, both HR and LBO roaming architectures are specified by 3GPP. However, compared to 4G, there are three major architecture changes when it comes to roaming.

- Unlike 4G, the user is authenticated by the home network. In 4G, the home subscriber service (HSS) delivers the authentication vector to the visited mobile management entity (MME), which, in turn, authenticates the user. In 5G, the authentication server function (AUSF) in the home network is responsible for authenticating the user, not the access & mobility management function (AMF) in the visited network.
- A new security function, security edge protection proxy (SEPP), is specified. SEPP enables secured relaying of service-based control plane interfaces between the 5G functions of the visited and home networks.
- Signaling interfaces between network functions of the visited and home networks are based on HTTP2 instead of Diameter.

SEPP was introduced to enable end-to-end security between visited and home operators. It is very similar to Diameter end-to-end signaling security (DESS). The N32 interface between the visited and home SEPPs consists of two sub-interfaces: control plane (N32-c) and forwarding plane (N32-f). The control plane interface, N32-c, is short lived and is used to exchange security credentials on a bilateral basis. The forwarding plane interface, N32-f, is then used to exchange the interface signaling between the two ends using the credentials negotiated via the control plane interface.

Additionally, 3GPP has specified protocol for N32 interconnect security (PRINS), which allows an intermediary (e.g., IPX provider) to modify and signal to the receiver that some elements of the unencrypted signaling message content have been modified, and to notify the receiver to verify the integrity of the modifications.

The inter-PLMN user plane security (IPUPS) function is used to associate the session-related signaling connection between the V SMF and the H SMF over the N16/N32 to the GTP U over the N9 to ensure only authorized user traffic is allowed to ingress/egress the two networks. The session management functions (SMFs) in the visited and home networks are responsible for setting up the IPUPS in their respective networks using the N4 interface. Depending on individual operator preferences, the same or a different user plane function (UPF) from the serving UPF can be used for IPUPS in the visited and/or home networks.

2.2. Wi-Fi Roaming

Depending on whether the 3GPP network is leveraged for Wi-Fi roaming or not, there are two prevalent deployed architectures of Wi-Fi roaming: 3GPP integrated Wi-Fi roaming and WRIX/WBA OpenRoaming™.

2.2.1. 3GPP Integrated Wi-Fi Roaming

Connectivity between a user's mobile device and the 3GPP core network via Wi-Fi access is specified as an overlay solution to minimize impact to the Wi-Fi access specifications. To facilitate access to mobile services via Wi-Fi while roaming, 3GPP has identified several different architecture options depending on:

1. Whether Wi-Fi and 4G/5G networks belong to the same visited operator or different operators;
2. Whether Wi-Fi access can be trusted or untrusted by the mobile operator; and
3. The location of the data session anchor.

To facilitate flexibility in the location of the data session anchor, like 4G/5G roaming architectures, 3GPP has specified HR and LBO architectures for roaming across Wi-Fi networks. Figure 3 depicts the HR and LBO architecture options for Wi-Fi roaming between 4G and Wi-Fi. Options 1 and 2 depict the HR and LBO roaming architectures for untrusted Wi-Fi access, and Options 3 and 4 depict the roaming architectures for trusted Wi-Fi access. The main difference between the trusted and untrusted Wi-Fi access is whether the transport of user equipment (UE)-related signaling is secured using the evolved Packet Data Gateway (ePDG) function or not. ePDG also serves as a security gateway protecting the 4G network from outside attacks. The UE establishes a secured tunnel with the ePDG prior to exchanging the credentials to authenticate and set up PDU sessions.

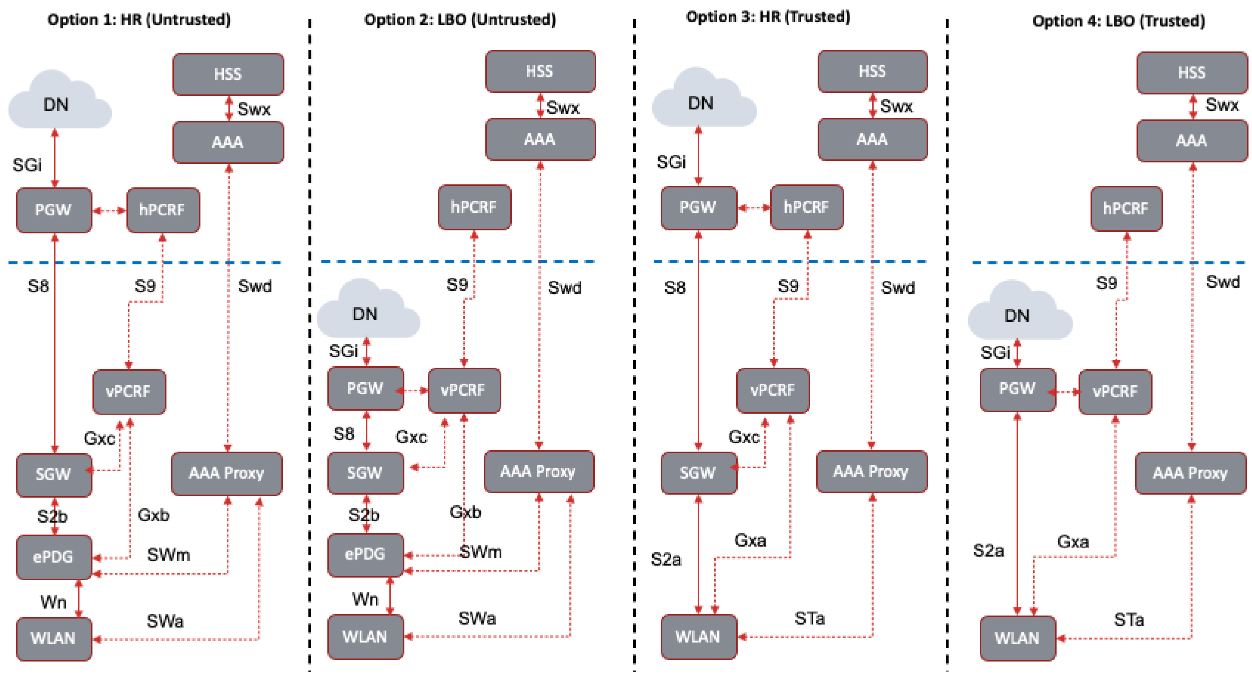


Figure 3 - Wi-Fi/4G roaming architectures specified in 3GPP

Generally, home-routed roaming architectures for trusted and untrusted Wi-Fi are used for packet data network (PDN) connections associated with voice over Wi-Fi and those requiring seamless mobility across Wi-Fi and 4G/5G. LBO architecture is preferred for PDU sessions that are best effort in nature and do not require a common anchor point.

For trusted Wi-Fi networks, EAP-AKA' has been specified for mutual authentication, and EAP-AKA is used for untrusted Wi-Fi access. In the case of untrusted Wi-Fi access, it is assumed that the user has successfully established a connection with the Wi-Fi using local or third-party credentials.

Figure 4 depicts a subset of Wi-Fi roaming architectures for 5G. A key difference between the 4G and the 5G roaming options is that 4G roaming requires ePDG to be located in the visited country, and 5G allows the N3IWF to also be located in the home country (Figure 4, Option 2). If WLAN is considered to be trusted for its subscribers, the requirements for the visited network provider to deploy an N3IWF are minimized. It also gives the home network provider greater control in enabling multi-access PDU session establishment for HR sessions and minimizes the need for the visited network provider to support related capabilities within its networks.

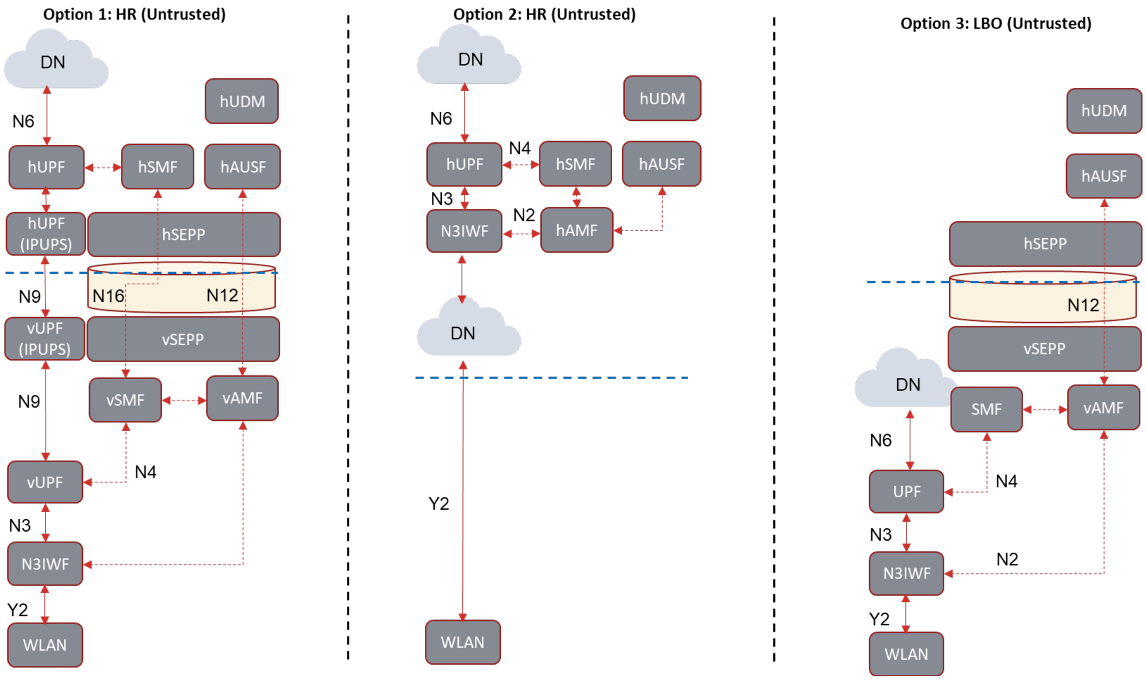


Figure 4 - Untrusted Wi-Fi/5G roaming architecture options

3GPP has specified integration of a trusted Wi-Fi access network into the 5G Core (5GC) via a 5G network function, trusted network gateway function (TNGF), as shown in Figure 5. TNGF is like N3IWF in that it terminates N2 and N3 interfaces. In addition, it also terminates EAP-5G signaling and behaves as an authenticator when the UE attempts to register to the 5GC via the trusted Wi-Fi network.

SCTE STANDARDS JOURNAL

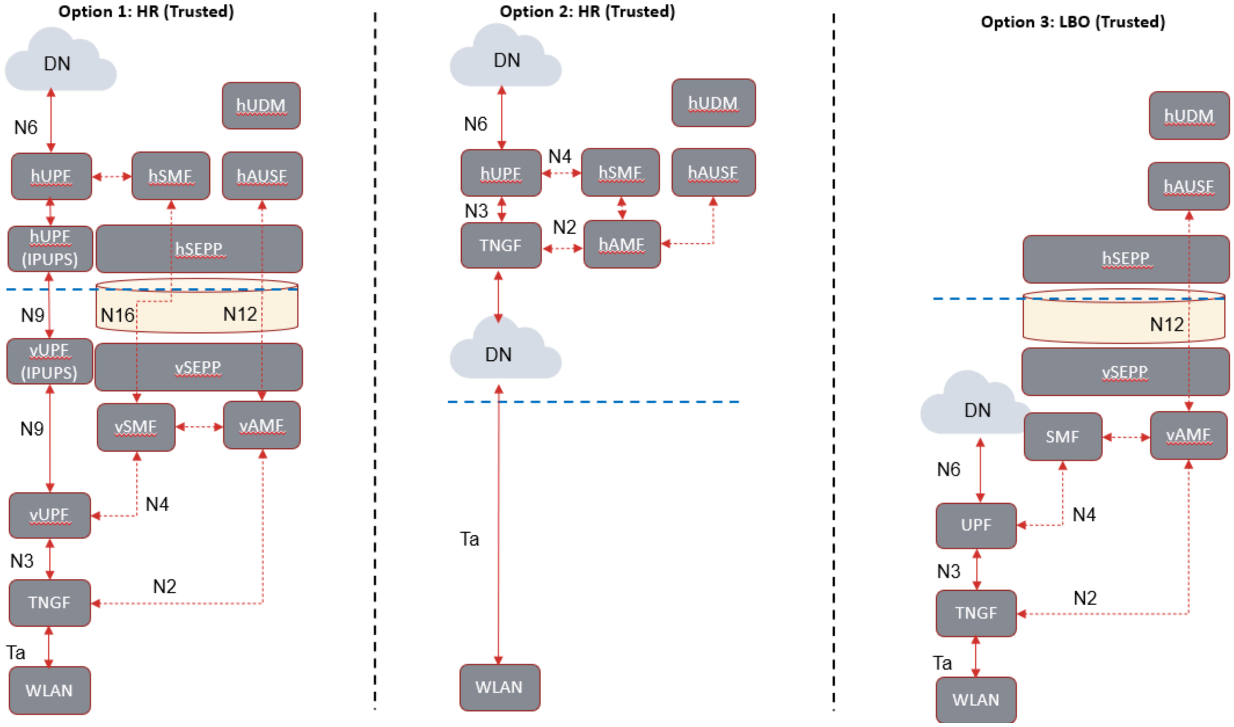


Figure 5 - Trusted Wi-Fi/5G roaming architecture options

2.2.2. *WRIX/OpenRoaming*

3GPP Wi-Fi roaming assumes that the Wi-Fi is tightly coupled with the 3GPP core network to leverage the mobile services to which the user has subscribed through their home network provider. It relies on 3GPP-specified credentials to authorize the service (e.g., Wi-Fi calling). However, 3GPP does not address how the device gets connected to an untrusted Wi-Fi network in order to reach the secured gateway (ePDG or N3IWF) to get access to the subscribed mobile services. In addition, there are several devices that have only Wi-Fi connectivity options. To fulfill wide-area Wi-Fi accessibility for the users, WBA has specified architecture, specifications, and guidelines to facilitate roaming across Wi-Fi networks with minimal user intervention. Figure 6 depicts the roaming architecture framework specified by WBA.

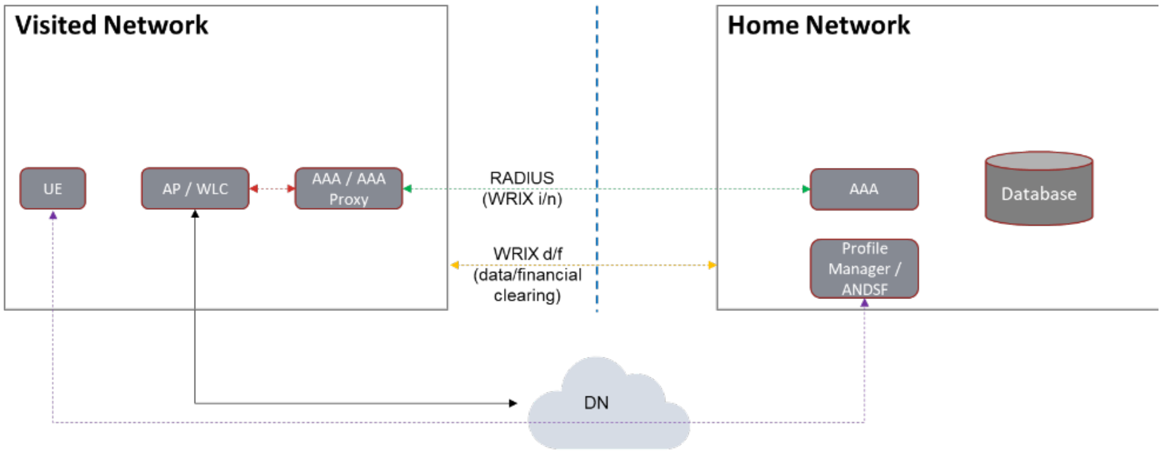


Figure 6 - WBA's Wi-Fi roaming intermediary eXchange

Passpoint is used to enable discovery of Wi-Fi networks. It also relies on the Remote Authentication Dial-In User Service (RADIUS) Protocol for the interface between the visited and home networks to authenticate and authorize Wi-Fi devices and exchange accounting records. To facilitate these exchanges and financial reconciliation, WBA has specified several specifications and guidelines, referred to as Wi-Fi roaming intermediary eXchange (WRIX) specifications. The specifications cover network interconnection and Usage Detail Record (UDR) formats and interfaces to facilitate data and financial clearinghouse implementation. WBA has also published a technical exchange document (TED) to communicate key information required to implement a roaming relationship and has created the commercial business exchange document (CBED), a template for exchanging billing information and simplifying the settlement process.

Since its development, one of the challenges to the adoption of WRIX has been attracting and onboarding smaller Wi-Fi network operators to negotiate commercial agreements and participate in it. Many Wi-Fi network operators have relied on third-party identifiers (e.g., Facebook/Google login) or require a complex sign-in procedure (e.g., loyalty member identifier plus time-limited passwords) for the users to access the Wi-Fi networks, which has resulted in a fragmented user experience. Complicated onboarding (sign-in to access available Wi-Fi), the lack of ability to negotiate/guarantee minimum Quality of Service (QoS) for the connection, the need for manual network configurations, security/privacy concerns, and the proliferation of 4G coverage all have kept the promise of wide-area Wi-Fi roaming accessibility unfulfilled.

Through the OpenRoaming initiative, WBA is trying to overcome some of the challenges outlined above by establishing a simplified policy framework for participation, settlement, and QoS over Wi-Fi that can scale across deployments of millions of Wi-Fi networks without requiring extensive network expertise and configuration. OpenRoaming also establishes a dynamic discovery (using DNS) of the home RADIUS server and secured connectivity setup between the visited and home network providers' RADIUS servers (RADSEC) to ensure user privacy. OpenRoaming builds upon the Passpoint and WRIX specifications to simplify the ability of small Wi-Fi network operators to participate without the overhead of negotiating commercial agreements or implementing elaborate network configurations. Figure 7 depicts the OpenRoaming architecture framework.

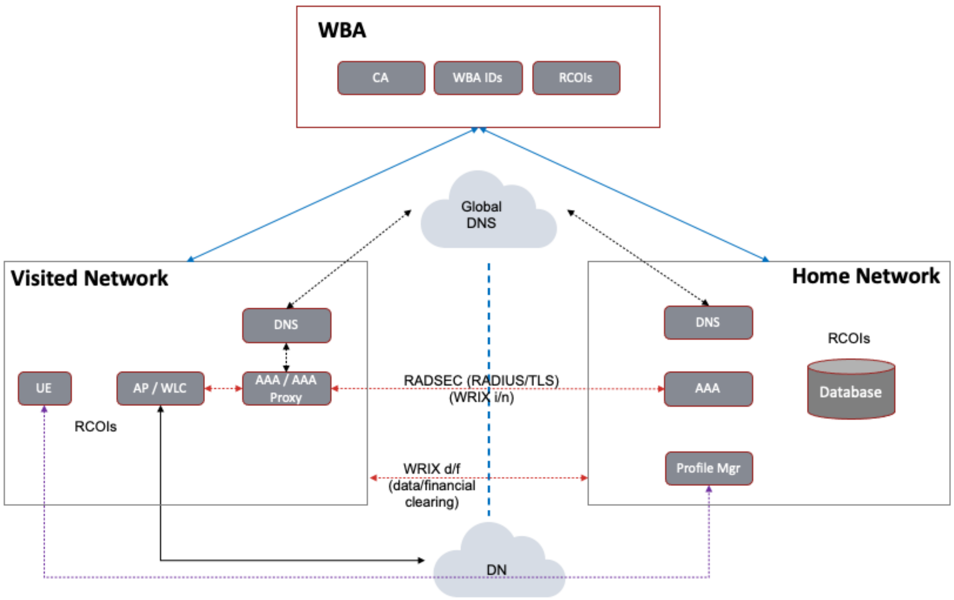


Figure 7 - OpenRoaming architecture framework

More specifically, it streamlines inter-network connectivity through a centralized administration and allocation of public key infrastructure (PKI) certificates and identifiers (e.g., WBAID, RCOIs) by WBA to mutually authenticate visited and home network providers. OpenRoaming mandates the use of secured RADIUS (RADSEC) standardized by the Internet Engineering Task Force (IETF) to securely exchange RADIUS Protocol messages between the visited and home networks. It also specifies dynamic discovery of the RADIUS server through DNS to minimize static configurations within the visited network.

Finally, to minimize configurations within the visited network, OpenRoaming specifies a standard set of roaming consortium organization identifier RCOIs (both settlement free and settlement required) that the visited network can configure and make available to the device via IEEE 802.11 ANQP. RCOIs are provisioned to allow the identity provider (IdP) to set the acceptable level of QoS. The home operator provisions the user device with a profile with one or more of these RCOIs. Based on the RCOIs supported by the Wi-Fi network in its proximity, the user device will automatically connect and authenticate itself with the home network via the visited network. OpenRoaming has the capability for the visited network to advertise (through the configured RCOIs) whether it supports anonymized user identities or not.

The OpenRoaming framework continues to leverage previously established WRIX specifications to facilitate intermediaries (e.g., roaming hubs) to use interconnection, data, and financial clearinghouse capabilities to scale Wi-Fi roaming and further broaden participation by operators of smaller/bespoke Wi-Fi deployments.

2.3. Operationalizing Roaming

3GPP is focused on the technical aspects (interfaces between the visited and home network functions), and GSMA is focused on creating the following elements:

1. Commercial agreement templates to facilitate streamlined exchange of roaming terms between operators, either directly or via intermediaries;
2. Technical information templates to facilitate a standardized way of exchanging critical network information required to enable roaming;
3. Specifications to facilitate data clearing and settlements;
4. Minimum technical requirements to facilitate interoperability and consistent service experience for roaming users; and
5. Test specifications to streamline end-to-end testing of roaming interfaces and capabilities.

To meet the needs of operators both small and large, GSMA has standardized agreements, specifications, and technical requirements to facilitate both bilateral and multilateral roaming.

In bilateral roaming, two network operators agree to allow their users to roam over each other's networks. Bilateral roaming arrangements are typically found between large mobile operators that have the necessary operational expertise, resources, and capabilities to implement, operate, and enforce a multitude of such individual agreements.

Multilateral roaming (roaming hub) is a managed service provided by an intermediary that enables roaming relationships with many operators through a single roaming agreement and a single connection with the intermediary. Each network provider can still choose to allow or disallow users of a specific operator having a roaming agreement with the intermediary. Generally, multilateral roaming agreements reduce an operator's operational overhead related to managing, operating, and enforcing many separate bilateral agreements. Multilateral roaming agreements also allow network operators to reduce the time needed to enable roaming services for their customers. Multilateral roaming arrangements are typically used by small- and medium-sized mobile operators that generally have neither the resources nor the expertise to implement a large number of bilateral roaming agreements.

To facilitate bilateral and multilateral roaming, GSMA has created roaming handbooks BA.40 and BA.60, which provide introduction and entry points to roaming processes and explain where to find more details on rules and procedures. GSMA has specified binding and non-binding requirements to facilitate bilateral and multilateral roaming agreements. PRDs, such as AA.12, AA.13, and AA.14, have been specified for bilateral roaming agreements, whereas AA.73 and AA.74 have been specified for roaming agreements where hubs (i.e., intermediaries) are involved. Figure 8 provides a visual list of key GSMA documents.

SCTE STANDARDS JOURNAL

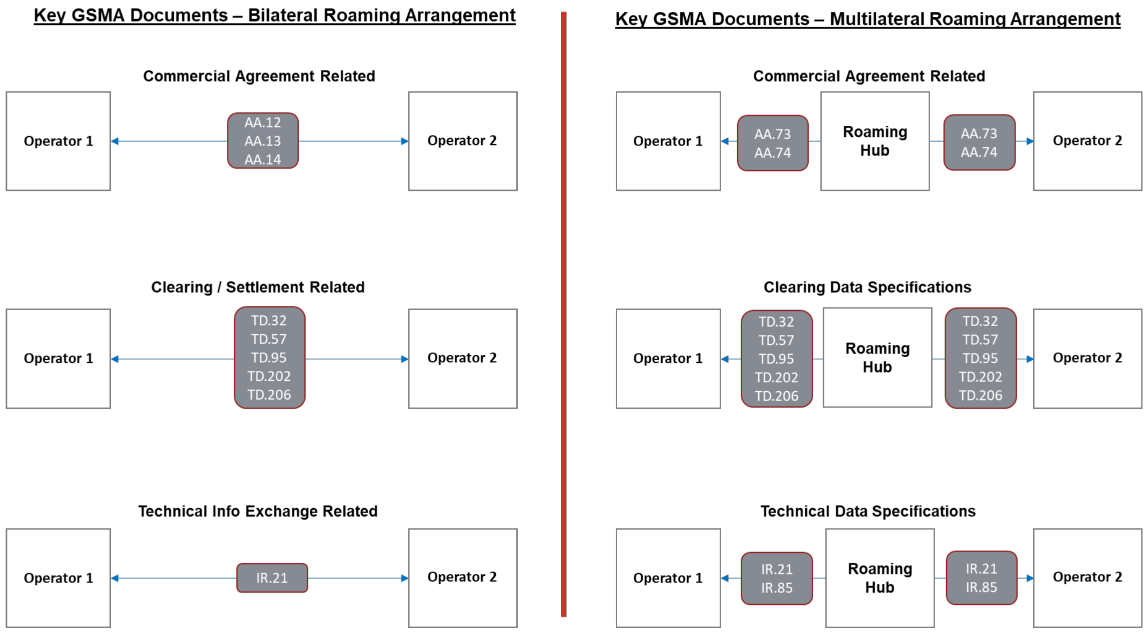


Figure 8 - Key GSMA documents

GSMA has also defined various relevant PRDs (e.g., TD.36, TD.95, and TD.202) and processes to enable clearing and settlement between roaming partners. Transferred Account Procedures (TAP) standards (TD.57 and TD.32) are used to facilitate the exchange of roaming call detail records for retail billing for voice and SMS-centric services. With the evolution of mobile network traffic to be more data centric and the prevalence of home-routed architecture to enable visibility of user traffic for the home operator, it is less necessary to gather device-specific usage from the visited network to facilitate invoicing of roaming charges to the individual subscribers. Also, the TAP standard has not been able to fully support the latest network evolutions and innovations in terms of roaming services (e.g., IoT roaming) for sustainable and efficient monetization of these next-generation services. As a result, a new billing approach for roaming based on wholesale has been developed within GSMA under the umbrella of Billing and Charging Evolutions (BCE) and has been documented in TD.202 and TD.206. It enables wholesale-based reporting and/or invoicing for inter-operator data roaming and reduces the operational overhead of performing service-based billing using TAP records. BCE has also specified processes for the two parties to streamline disputes and reconcile wholesale charges in TD.204, as well as the financial settlement process in TD.95. Additional documents related to commercial agreement and settlement methods (BA.04), SLA guidelines (BA.51), and fraud management (FF.21) are also available.

Finally, GSMA PRDs IR.21 and IR.85 are also exchanged between the operators and the roaming hub. These documents contain technical data that enables interconnect and traffic routing between the networks. Examples of the information contained in these documents include routing information, network elements information, APNs and WLAN Information, a contact list, and a list of service hubs. Additionally, GSMA has developed minimum feature requirements and test specifications (not depicted in the Figure 8) related to TADIG (Transferred Account Data Interchange Group) and IREG (International Roaming Expert Group) testing, which are needed before enabling inbound and outbound roaming.

2.3.1. Internetwork packet eXchange (IPX) and roaming hubs

In addition to specifying processes, procedures, agreements, and file formats, GSMA has also developed architectural guidelines and requirements for an intermediary to facilitate interconnection between service providers. This intermediary is generally referred to as an Internetwork Packet eXchange. An IPX provides end-to-end quality of service assurance and supports interconnect payments. To provide enhanced services and features, unlike the Internet, the IPX can be service aware. It is a successor to the General Packet Radio Service (GPRS) Roaming eXchange (GRX), which was created to fulfil the need of data connectivity between GSM/GPRS operators via a closed and secured environment. As 3GPP evolved to support additional services such as Multimedia Messaging Service (MMS), Wi-Fi roaming authentication, and Short Message Service (SMS) over IP, GPRS Tunnelling Protocol (GTP) was enhanced to support these services when roaming, thereby enabling IPX providers to provide roaming for these services.

An IPX is composed of IP infrastructure (e.g., routers, switches, and security gateways) and application/protocol-aware proxies to facilitate QoS-enabled service-aware connectivity between the networks of visited and home network operators. An IPX can connect two network operators directly or indirectly via another IPX. The IPX, between two service providers, can be established via one of three connectivity options: transport only, bilateral service transit, or multilateral service hub.

The transport-only connectivity option (Figure 9) enables bilateral interconnections between service providers without service awareness but with guaranteed QoS or traffic prioritization. This option allows two service providers to use the IP network to interconnect directly with a peer without any involvement at the service level from the IPX provider. The charging is based on data volumes on a wholesale retail basis.

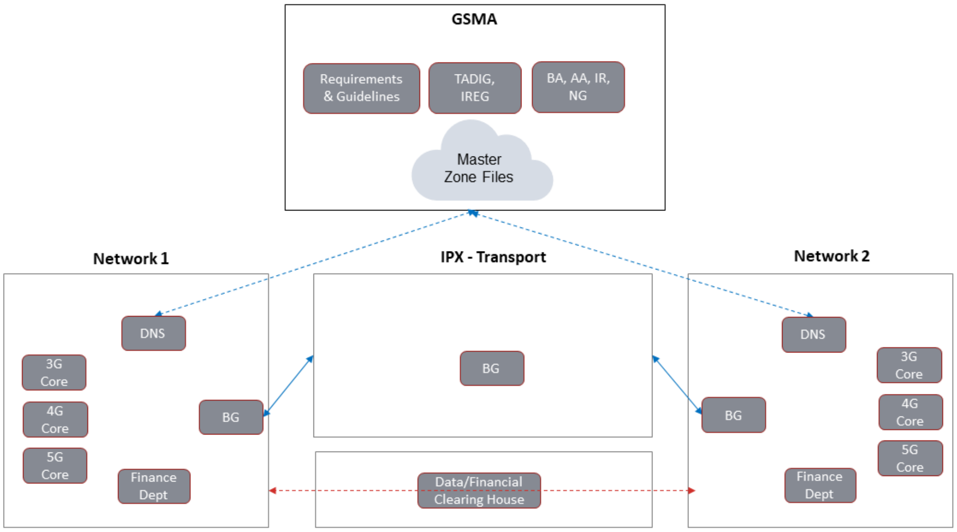


Figure 9 - IPX transport model

The bilateral service transit connectivity option (Figure 10) enables bilateral relations between service providers using both the IPX transport layer and the service layer. The service provider uses the proxy functionality provided by the IPX for voice/signaling while maintaining a direct commercial relationship with the other service providers. In this case, the IPX provider provides a service level agreement (SLA) on a service basis. The transit option charging model is usually based on the service usage. In most cases, the underlying IP network does not provide an end-to-end IP connectivity between service providers, but it is used to connect the service providers with the IPX proxies.

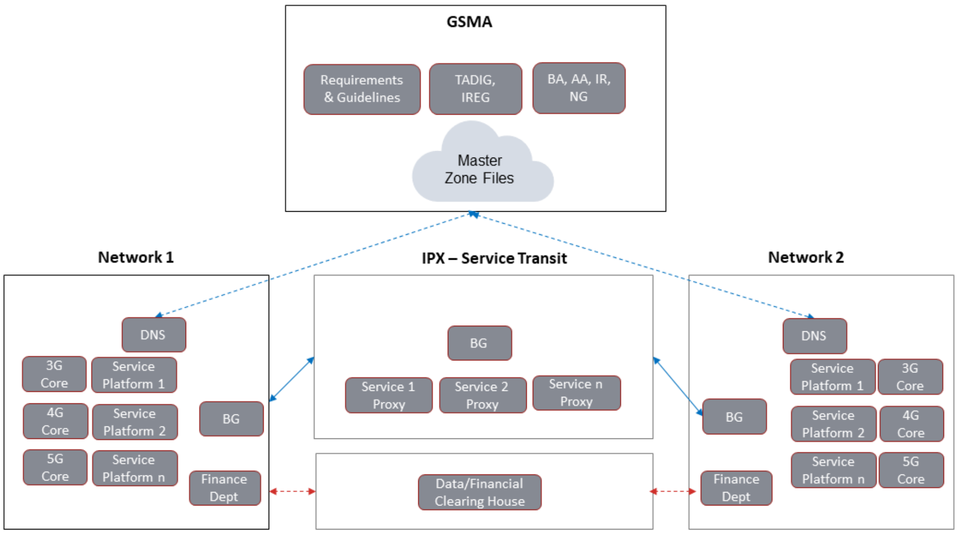


Figure 10 - IPX service transit model

Some of the roaming and interconnect services that IPX providers enable for their customers are DNS lookup, Diameter/SCCP signaling routing, number lookup, SMS/MMS/RCS and IP voice interconnect, and roaming. More information can be found in GSMA PRD AA.51.

The multilateral service connectivity option (IPX hub) (Figure 11) enables multilateral connection based on service awareness. Traffic is exchanged between the service provider and multiple participating service providers through the IPX hub. In this model, a service provider can get access to multiple IPX destinations on a per-service basis. On top of the connectivity and SLA for each service, the service provider is provided with cascaded billing and settlement.

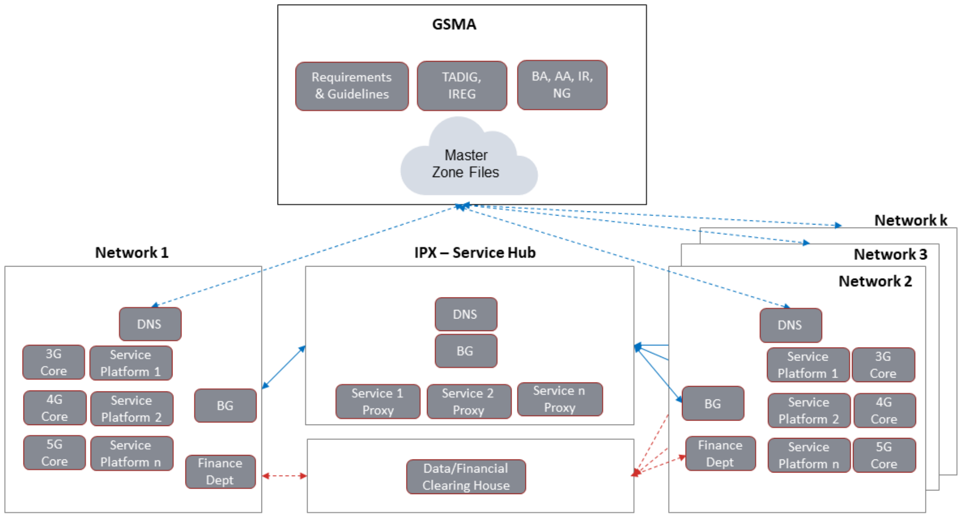


Figure 11 - IPX service hub model and roaming hubs

The hub model can be viewed as a complement to the bilateral roaming model. Depending on business needs and operational capabilities, operators can choose to maintain some relationships as bilateral agreements and migrate others to a multilateral agreement via a roaming hub.

Roaming hubs are intermediaries who support the above-described multilateral service hub model by offering roaming interconnect services across multiple service providers through a single contract, thus allowing a service provider to expand its coverage without the operational overhead of managing a multitude of bilateral agreements with a large number of domestic and global mobile operators. Some of the services offered by a roaming hub include roaming agreement management, signaling routing and associated management, interoperability and interworking testing, monitoring, data analytics, reporting, troubleshooting, isolating faults, and resolution of faults from an end-to-end perspective. A roaming hub can also provide optional complementary services such as data clearing, fraud prevention, settlement and financial clearing, Steering of Roaming (SoR), and welcome SMS.

Service providers have flexibility to pick and choose their IPX and roaming hub providers. GSMA members usually use AA.12 to establish bilateral roaming and AA.73 to establish multilateral roaming via a roaming hub. Note that GSMA only specifies the interfaces and the formats to facilitate interconnection and exchange UDRs.

More recently, to streamline the data clearinghouse capabilities for variety of roaming services (voice, data, IoT, etc.), GSMA has developed the BCE framework. It has developed the concept of detailed data records (DDRs), usage data reports (UDRs), usage summary reports (USRs), billing statement reports (BSs), and charging documents (CDs). The call data records (CDRs) are converted into DDRs. The UDRs are exchanged in lieu of DDRs for usage validation/reconciliation. The USRs are used to reconcile both usage and charges for an invoicing period in order to resolve disputes prior to invoice production. The BSRs are prepared by the serving party in support of the corresponding CDs and are used to reconcile settlement amounts. The CD is the document sent to the served party and can be in the form of an invoice, a credit note, or a debit note.

Under the GSMA's BCE framework and process, both parties agree on a service name identifier for data roaming service. A service can be identified by a combination of the IMSI ranges, APN, radio access technology (RAT), and/or the bearer. The served party is not required to create an H-DDR, H-UDR, or H-USR format for reconciliation and validation. However, it is their responsibility to accept and validate the respective serving party reports. The V-UDR exchange is mandatory. The served party must be able to validate the syntax and the aggregation content of each UDR received from the serving party.

The solution supports processes to allow the home to notify the visited network that information received is not accepted (rejected or disputed). When a report/document is received by the served party, it may be validated and reconciled with a receiving party's own data. Reject and dispute timescales will be based on the specific data exchange process. Reject and dispute reporting formats are defined in GSMA PRD TD.204.

Like GSMA, WBA has also specified a set of specifications to not only enable interconnection but also facilitate data and financial clearing on a bilateral and multilateral basis through a transit/hub network. Figure 12 depicts the high-level architecture of the transit/ hub network. OpenRoaming makes use of RCOIs and defines terms and conditions for entities that wish to join the OpenRoaming federation; however, this is not a requirement for using the WRIX framework.

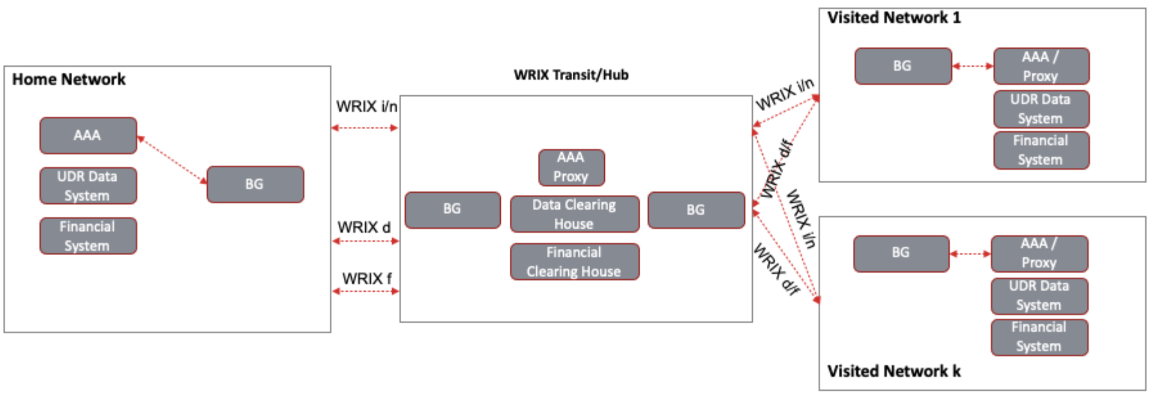


Figure 12 - WBA-specified WRIX transit/hub for Wi-Fi roaming

To facilitate roaming and identify the companies involved, both the access network provider (ANP) and the IdP are identified during WRIX interactions through the use of an identification code. For companies using the WRIX framework for the interaction, both parties will either have a WBAID assigned or use a mutually agreed-upon compatible identification code, such as a GSMA TADIG code.

3. MSO eXchange

MSO eXchange (MSOX) is a roaming hub to facilitate and enable inter-MSO roaming across mobile (4G/5G) and Wi-Fi access networks. It transforms regionalized wireless hotspot deployments of many MSOs into a connectivity network, spanning a much bigger geographic footprint for MSOs' mobile subscribers, and helps MSOs reduce the usage of MNO networks. This is relevant from the perspective of regional MSOs within North America that are contemplating targeted mobile deployments. Unlike traditional roaming, MSOX can also enable connectivity to facilitate inter-PLMN (public land mobile network) handovers with adjacent footprints, and it can host the roaming 4G/5G functions on behalf of MSOs.

At the basic level, MSOX provides a common platform for our members to enable roaming and handover across their mobile and Wi-Fi coverage footprints. In its most elemental form, MSOX provides basic IP routing and the necessary QoS for the roaming traffic. This is similar to the basic roaming service that the IPX and GRX providers have been offering to enable data roaming across 2G/3G/4G operators. This allows MSOs the flexibility to form their own roaming agreements with their own terms and conditions.

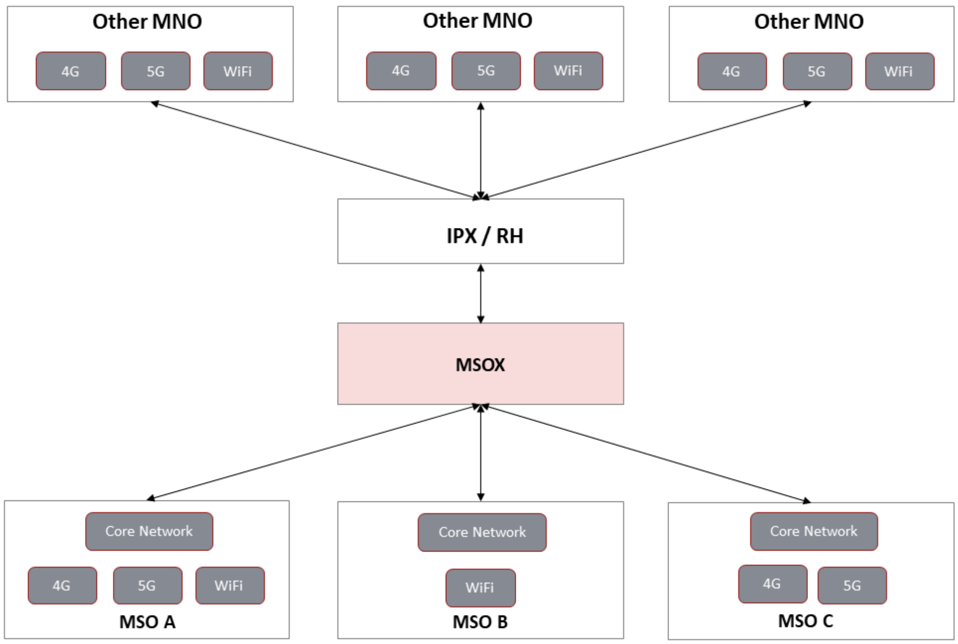


Figure 13 – MSOX Framework

MSOX facilitates both bilateral and multilateral roaming arrangements. For multilateral agreements, MSOX serves as a roaming hub, and for bilateral roaming, MSOX acts as a transport provider between MSOs. Bilateral agreements could be attractive to larger MSOs, whereas the roaming hub agreement could be attractive to smaller MSOs, which may not have a roaming department/team of their own.

MSOX acts as an ecosystem partner/broker to enable and support OpenRoaming capabilities and associated WRIX interfaces for networking, clearing, and settlement. For operators not supporting OpenRoaming, it serves as the transit hub for WRIX-enabled Wi-Fi network operators. It could optionally enable an online signup portal on the behalf of the visited network operator to allow subscribers from MSOs lacking roaming relationships to access the untrusted Wi-Fi network, after which the device will discover the N3IWF and connect to the home network operator's 5G core network to enable home-routed data routing to facilitate enhanced capabilities, like ATSSS (access traffic steering, switching, and splitting).

MSOX can also connect with other roaming hub providers to potentially enable low-cost international data roaming packages for MSO customers. It facilitates interworking using 4G Diameter-based signaling plane protocol and provides value-added roaming services, like Steering of Roaming (SoR) and SMS welcome message services, to further reduce the cost and improve the customer experience for outbound roamers.

In addition, MSOX can provide value-added services for MSOs, such as hosting a common voice/messaging core on behalf of the MSOs. Alternatively, it can provide the necessary connectivity to a third-party voice/messaging service provider with the requisite QoS for voice traffic.

Finally, MSOX can incorporate a shared core that can be used in the event handovers between the MNO and partner MSO networks become desirable and/or to offer a hosted core option for MSOs who prefer managed core solutions.

3.1. Key Benefits

The primary benefit of MSOX is the ability to implement roaming functions and serve the commercial and operational needs of roaming on behalf of MSOs that are also MVNOs (which currently rely on their agreements with MNOs), enabling a more economical mobile service. Some MSOs are contemplating building out their own mobile deployments and developing the necessary expertise; MSOX allows them to take advantage of non-overlapping buildouts of their mobile networks without having to develop in-house roaming skill sets or get into expensive commercial roaming agreements. MSOX assists MSOs in reducing time to leverage each other's wireless infrastructure for roaming.

The traditional roaming hubs are not focused on enabling inter-domain seamless handovers, as service continuity across borders is not a critical need. In the United States, several large metro areas have different MSOs providing services in adjacent areas. To facilitate a seamless data experience across adjacent mobile deployments by MSOs in a large metro area, MSOX can enable seamless inter-PLMN handovers. One mechanism would be to facilitate inter-domain mobility interfaces (typically not specified as part of the GSMA roaming specs); another could be to have a shared core deployed within MSOX. MSOX not only increases the network coverage footprint for the MSOs by enabling outbound roaming, but it also allows the MSOs to monetize and create new revenue streams from their mobile and Wi-Fi networks by allowing inbound roaming from other operators (e.g., rural operators).

Additionally, many MSOs, especially in the United States, have large Wi-Fi deployments. MSOX enables seamless utilization of partner Wi-Fi connectivity where available and further reduces the MVNO costs. In fact, by enabling a common anchor for the user plane through the implementation of the home routing roaming architecture,[1] MSOX can facilitate 3GPP's ATSSS across the MNO's mobile and roaming partners' Wi-Fi networks. By being a transit/roaming hub for standalone Wi-Fi network operators, MSOX can also facilitate ATSSS between the participating standalone Wi-Fi and the MNO's mobile networks.

MSOX can interwork with other roaming hubs to facilitate international outbound roaming for its mobile subscribers. Today, MVNOs rely on MNOs' international mobile subscriber identifier (IMSI) range and roaming partnerships to provide roaming services for their customers and have no ability to negotiate preferable terms of their own. Having their own data roaming agreements with international mobile providers associated with their own SIM could allow MSOs to independently negotiate data roaming rates as part of the international roaming service package for outbound roamers. Connectivity with other roaming hub providers could also facilitate domestic roaming (e.g., to rural operators). For inbound roamers, MSOs can offer an attractive rate plan to international operators through a data-centric eSIM that the home operator can configure to prioritize MSOs' 5G/Wi-Fi data coverage, when available.

[1] "Evolved MVNO Architectures for Converged Wireless Deployments," white paper, June 2021, CableLabs—Omkar Dharmadhikari, John Kim, Ojas Choksi

SCTE STANDARDS JOURNAL
SCTE a subsidiary of CableLabs®

In the long term, MSOX could host a variety of service functions spanning location (LMF), voice (IMS + TAS), and SMS (SMSF, SMSC) on a shared basis as MSOs acquire more mobile subscribers, at which time usage-based fees for voice and SMS being charged by the MNO may no longer be economical. It could also host, operate, and manage a centralized mobile core on behalf of some of the operators and reduce the operational cost through sharing among MSOs.

Another benefit of MSOX is that it enables MSOs to outsource operationally intensive activities like interoperability and interworking testing, collecting, and analyzing inter-operator traffic analytics and troubleshooting, as well as management of bilateral and multilateral commercial agreements with other mobile operators. It also prevents duplication of efforts across the MSOs and reduces the overall cost.

Finally, MSOX is envisioned to leverage MSOs' backbone infrastructures through SD-WAN-like APIs. The same APIs can be used to provide connectivity solutions to private networks/enterprises spanning across footprints of multiple MSOs.

4. MSOX Functional Description

Figure 14 depicts the functional (logical) architecture of MSOX, comprising three main functional domains: interconnect (light gray shading), value-added services (dark gray shading), and clearinghouse functions (yellow shading).

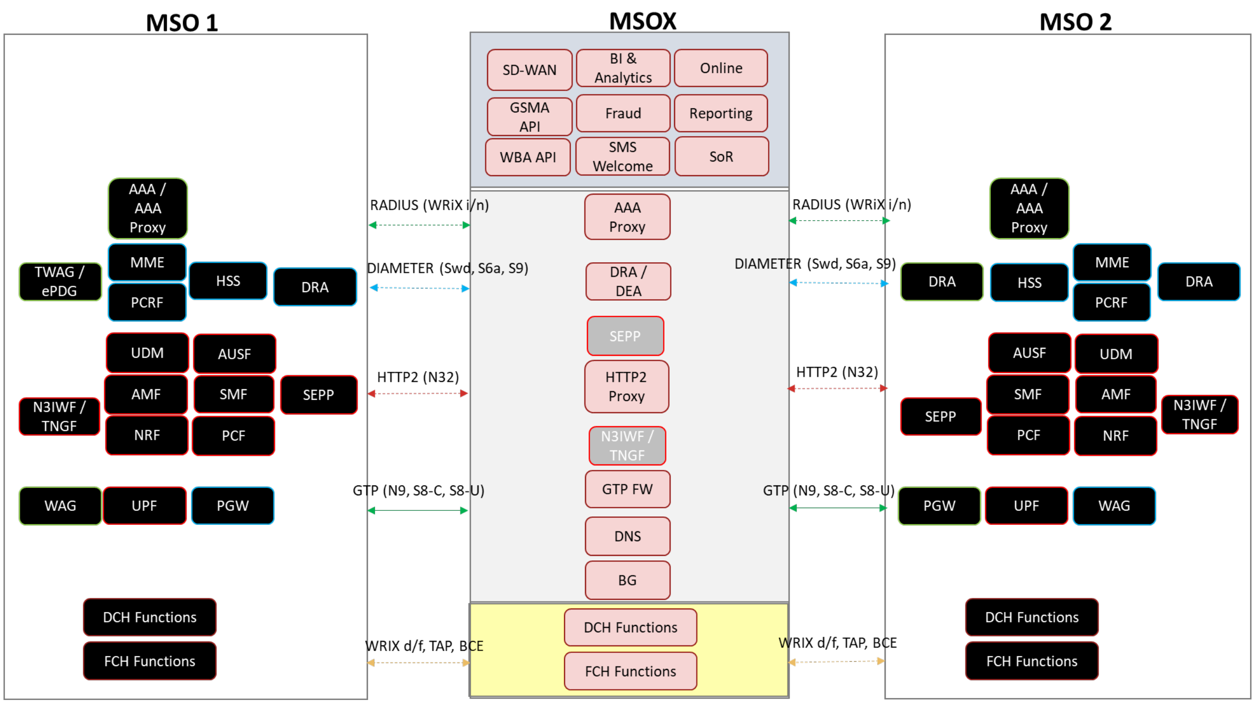


Figure 14 – MSOX functional architecture

4.1. Interconnect Functions

The interconnect functions consist of the following:

- 4G and non-3GPP-related control plane functions (DRA, AAA proxy) to route control plane traffic between the visited and home networks,
- SEPP HTTP2 proxy and GTP firewall (GTP FW) to secure the 5G signaling and 4G/5G data traffic between the visited and home operator networks,
- Repositories (DNS) to look up destination IP addresses of the targeted control plane and services functions, and
- Border gateways (BG) to advertise IP routes and facilitate routing of user traffic.

The MSOX platform may also host certain network functions such as SEPP, DRA/DEA, and/or N3IWF on behalf of the MSOs to reduce their capital and operational costs, especially for smaller operators.

The following interfaces specified by 3GPP and WBA are utilized to interconnect the visited and home networks via MSOX.

- N32 interface - To facilitate greater security during roaming, 3GPP has specified a new function, SEPP, and an interface, N32, to securely transfer control plane signaling messages between the visited and home networks. The N32 interface comprises two components: N32-c and N32-f. The N32-c is used by the SEPPs in the visited and home network to securely exchange the credentials for the N32-f interface. N32-c is secured using TLS and/or PRINS (application layer security) depending on the architecture configuration preferred by the MSOs. The N32-f is used by the operators to exchange messages related to various services interfaces (identified in Table 1). The confidentiality and the integrity protection of the message is enabled using the credentials exchanged via the N32-c.

Table 1 - Services-Based Interfaces Transported Over N32-F Interface

INTERFACE	NETWORK FUNCTIONS
N8	vAMF - hUDM
N12	vAMF - hAUSF
N16	vSMF - hSMF
N10	vSMF - hUDM
N24	vPCF - hPCF
N27	vNRF - hNRF
N31	vNSSF - hNSSF
N21	vSMSF - hUDM

As shown in Figure 15, while the N32-c is end-to-end between the visited and home operators, N32-f can be hop-to-hop, thereby allowing the MSOX the visibility of information embedded in the control plane messages between the visited and home network functions. This visibility can be used by MSOX to deliver value-added services (e.g., fraud management and welcome SMS) to both visited and home network operators.

Depending on the agreement, each operator may send certain IEs in cleartext, thereby allowing the MSOX to view and, if required, modify them before forwarding the message to the target network. Each MSO agrees to a modification policy with MSOX and provides it to its roaming partner prior to the establishment of the N32 interface.

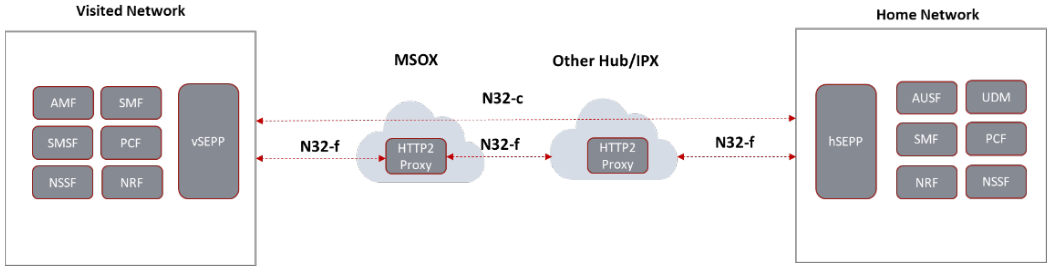


Figure 15 – Comparison of end points for N32-c vs. N32-f

- WRIX-i and WRIX-n interfaces - These interfaces are used to authenticate and authorize roaming Wi-Fi users. They use IETF specified RADIUS Protocol to exchange information. These interfaces are also used to periodically deliver charging details by the visited network to the home network.
- SWd, S6a, and S9 interfaces - These are 3GPP-specified interfaces that use Diameter protocol. S6a is used by the visited network to authenticate the roaming user via the home HSS. S9 is used by the visited network Policy and Charging Rules Function (PCRF) to retrieve the policies for the roaming user's data traffic from the home PCRF. SWd is used by the visited network AAA proxy to authenticate the roaming user via the home AAA server.
- S8-C interface - This is a 3GPP-specified interface that uses GTP to exchange signaling messages to set up/modify/release sessions between the visited S-GW and the home P-GW. It uses UDP to transport GTP messages.
- N9 and S8-U interfaces - These interfaces are specified by 3GPP to tunnel user traffic between the visited and home networks using GTP encapsulation.

4.2. Value-added Service Functions

MSOX supports the following value-added service functions.

- WBA API function - This function handles interactions with WBA to retrieve authorized WBAIDs/subIDs. MSOX issues sub WBAIDs and manages the assignment of WBAIDs to OpenRoaming Wi-Fi network operators while acting as a WBA broker.
- GSMA API function - GSMA provides an online repository, stored in an XML format, of operators' IR.21 and IR.85. This function will interface with the GSMA IR database repository and upload/download necessary IR.21 and IR.85 documents, extract the network information, and configure them in its internal routing tables.
- Steering of roaming (SoR) service function - This function enables steering of roaming users to a prioritized 4G roaming partner when the home network is unavailable. It procures the prioritized list from the home MSO and provisions it locally within the MSOX. It monitors the control plane signaling and steers the roaming user to a preferred roaming partner of the home MSO. If steering is expected to be subscriber specific, it will interface with the home MSO's subscription database to determine the prioritized list of roaming partners for that subscriber. SoR function assists the home operator to minimize its cost and/or maximize user experience when their subscribers are roaming onto a 4G network. 5G has specified explicit control plane signaling as part of the NAS registration procedure to deliver the prioritized list of roaming partners from the home network to the roaming device, thereby eliminating the need for such a service.

- Fraud management service function - This function identifies and minimizes the fraudulent use of the network through real time monitoring and correlation of control plane signaling with the user plane traffic. MSOX will use the data captured via signaling interfaces and usage reports to detect fraudulent and high-usage activities on a real-time basis. This function will interface with the business intelligence and reporting functions to alert and deliver reports to the MSOs on a real-time basis based on various criteria set by the MSOs. Some of the potential fraud-related incidents that will be reported are high-usage activities, frequent registrations into roaming networks with services usage to avoid fraud detections (velocity check), and multiple registrations from different networks that are geographically apart from one another.
- Business intelligence (BI) and reporting functions - These functions analyze the roaming usage pattern and provide actionable reports to the MSOs to optimize roaming packages. They generate custom reports analyzing roaming usage patterns for the MSOs. The analytics engine of the BI and reporting functions utilizes the roaming usage data collected from various nodes within the MSOX to create customized reports for the MSOs. They are then stored and made available via an online portal function.
- SD-WAN function - This function is used to set up the transport connections between different points of presence of the MSOX deployment within MSOs' transport networks. It interfaces with each MSO's service orchestrator to set up the connection paths with the requisite bandwidth and QoS.
- Online function - This function serves as a portal for the MSOs to view the reports in graphical format and download them. It facilitates delivery of GSMA- and WBA-specific documents related to any network configuration updates. The portal also allows MSOs to configure various criteria to receive individualized alerts and reports. MSOs can also download the datasets associated with the reports for further integration into their own BI tools and reporting functions.

4.3. Clearinghouse Functions

The clearinghouse functions can be categorized into data clearing functions (DCH) and financial clearing functions (FCH).

Figure 16 provides a consolidated view of DCH and FCH functions of MSOX implementing the BCE and WRIX-d/f agents to facilitate data collection, reconciliation, and settlement.

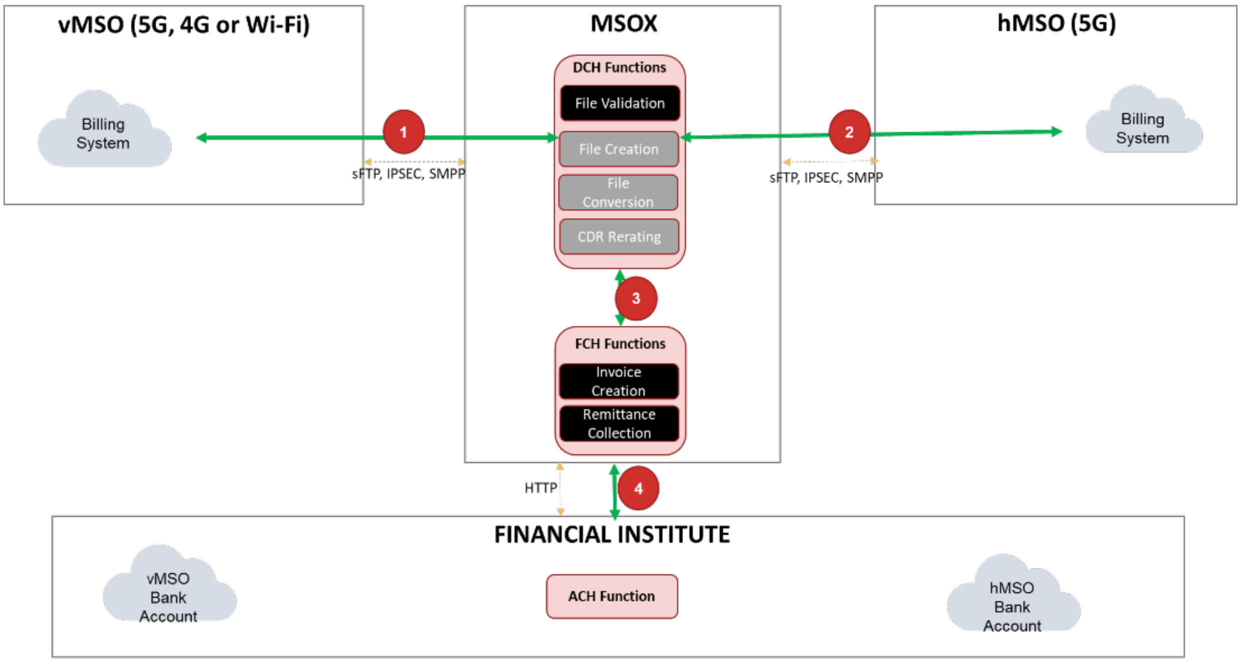


Figure 16 – DCH and FCH functionality of MSOX

The DCH function processes the roaming usage files based on GSMA-BCE and WBA WRIX-d specifications. It retrieves the usage data records from the MSOs via a secured FTP or VPN connection (steps 1/2 in Figure 16), in formats specified by BCE and WRIX-d. The DCH function then validates the file records (based on formats specified by BCE/WRIX-d) and generates any error record files and sends it back to the originating MSO. After validation, it reconciles the usage data records from the visited and home operators, as well as those collected within MSOX, via its control and user plane functions (e.g., GTP FW, IPUPS, DRA, and HTTP2 proxy). As part of the reconciliation, the DCH function also validates that the usage records within these files are charged as per the wholesale rates agreed within the roaming agreements, including any discounts that may apply.

MSOs will have access to all roaming usage CDRs and files processed by the DCH, as well as any processing done on these files, and records via MSOX's business intelligence functionality in various customizable reporting formats.

After the reconciliation of the usage data, the DCH function makes available the necessary files to the FCH function, responsible for providing financial clearing and settlement services to the MSOs. Based on the reports received from the DCH, the FCH generates and distributes invoices to the MSOs in a format specified by BCE and WRIX-f specifications. Invoice distribution can be performed via secure FTP/SMPP and, if requested by the MSO, via a courier. The FCH also computes all relevant taxes, such as sales tax or VAT, and includes them in the invoice.

The FCH supports a variety of settlement models, including: bill and keep (where each MSO bills their subscribers for roaming, but they do not pay any amount to their roaming partner); netting (where the FCH takes all the amount the MSO owes and all the amount the MSO will be credited from its roaming partner and then nets the total amount owed versus the total amount billed); or a simple financial settlement based on what is owed and what is billed separately. The mode of settlement is typically specified as part of the roaming agreement. The FCH will interface with the MSOs' internal systems via secured FTP and/or email to deliver periodic reports (step 1/2 in Figure 16).

4.4. Scenarios for MSOX Use (with Depiction of Data Flow)

This section describes the scenarios supported by MSOX. First, mobile roaming scenarios are described. Thereafter, different Wi-Fi roaming scenarios are described. For each scenario, high-level signaling and data flows are depicted to highlight the interfaces and functions of MSOX that will be involved.

Also, even though each scenario depicts only the visited and home operators, the relationship between the two roaming partners could be established either via a multilateral or bilateral agreement facilitated by MSOX. The figures below do not distinguish between bilateral and multilateral roaming, as the functions performed by MSOX are similar. For each scenario, MSOX could be facilitating a bilateral or a multilateral agreement.

In the case of bilateral roaming, where the roaming agreement is between the MSOs, MSOX simply uses the IR.21 document supplied by the MSOs to enable routing of signaling (N32) and user (N9) plane interfaces for home-routed traffic between the MSO networks. MSOX provides the necessary physical and logical links between the MSOs that meet the QoS requirements and ensures that these links are redundant and secure.

In the case of multilateral roaming, MSOX acts as the roaming hub between MSOs. In this scenario, the MSO's IR.21 is converted to IR.85 by MSOX and shared with other MSOs or the MSO's IPX provider.

4.4.1. Mobile Roaming Scenarios

There are three mobile roaming scenarios that can be supported by MSOX.

1. 5G-to-5G roaming—This scenario is applicable when MSOs have only deployed 5G SA networks (e.g., CBRS deployments by MSOs in the United States).
2. Inter-operator 5G-to-5G handovers—This scenario is applicable when MSOs that are neighboring franchisees within a large metro region have negotiated inter-operator N14 connectivity between themselves.
3. 4G-to-5G roaming—This scenario is applicable when MSO subscribers of 5G SA-only deployment roam onto 4G networks (e.g., rural 4G deployments, international data roaming).

MSOX can also support 4G-to-4G roaming; however, it is not described here, as it is identical to the 4G-to-5G roaming scenario from an implementation perspective.

The dotted lines generally labeled 1 depicted in the below figures show the control plane, whereas the solid lines generally labeled 2 depicted in the below figures show the data plane. The solid lines depicted in green labeled 3 and 4 show the flow of clearing data flow.

SCTE STANDARDS JOURNAL

4.4.1.1. 5G-to-5G Roaming

The scenario in Figure 17 depicts the roaming enabled by MSOX between MSOs with regional/hotspot footprints of their own 5G mobile networks (H-MVNOs[2]) and where they want to reduce the cost of utilizing the MNO network through existing MVNO arrangements. It is assumed that each device is provisioned with a dual SIM—one to access the MNO network (MNO SIM) and another to access its own regional network (H-MVNO SIM). To support this roaming scenario, the H-MVNO SIM will have to be configured with PLMN IDs of all its data roaming partners, so that whenever the user roams outside the home network, it searches for the partner MSO's 5G network and, if available, re-registers as a data-centric device via the partner's 5G core network. The device remains connected to the MNO network via the MNO SIM for data service when outside MSO coverage.

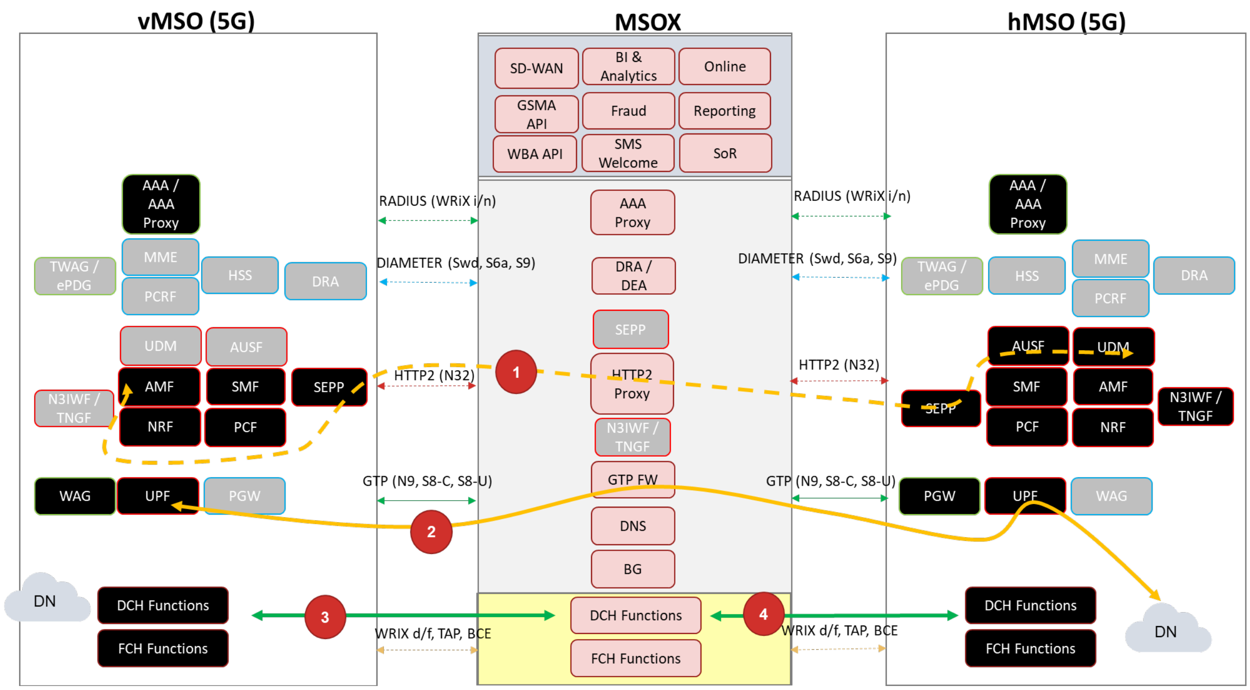


Figure 17 – 5G-to-5G mobile roaming between MSOs (locally deployed SEPP)

The 5G control plane signaling between the H-MVNO and its domestic MSO roaming partners will be via the N32 interface. Figure 17 depicts the scenario where SEPP is deployed by individual MSOs. However, as shown in Figure 18, it can also be hosted within MSOX on behalf of one or more MSOs.

Irrespective of whether the MSOs deploy their own SEPPs or host them in MSOX, the N32-f will need to be set up on a hop-by-hop basis to enable viewing/modification of cleartext for facilitating MSOX's value-added services, such as fraud management, business analytics, and detailed roaming traffic

- [2] Hybrid Mobile virtual Network Operator (H-MVNO) – An operator that has deployed its own mobile network and has an MVNO relationship with a nationwide mobile operator. When available, the localized mobile coverage from its own network is prioritized for data usage over the MNO network.

reporting. If the N32-f is enabled directly between the two MSOs and if the SEPPs are deployed by the MSOs, MSOX will have limited ability to view and provide value-added services to the MSOs when their users are roaming.

If the SEPP is hosted within MSOX on behalf of an MSO, it will act either as a V-SEPP or an H-SEPP. It will act as a V-SEPP for the home MSO and an H-SEPP toward the visited MSO. MSOX will manage the necessary certificates and configuration of the hosted SEPP. The MSO and MSOX will have an IPSec tunnel between their networks to enable connectivity between the hosted SEPP and the MSO's network elements. When new MSO members are added to MSOX, the N32-f context can be modified or renegotiated. HTTP2 headers could be added to provide robust attribution for each message originating from the visited MSO and traversing through MSOX so that the home MSO can attribute it to a particular visited MSO.

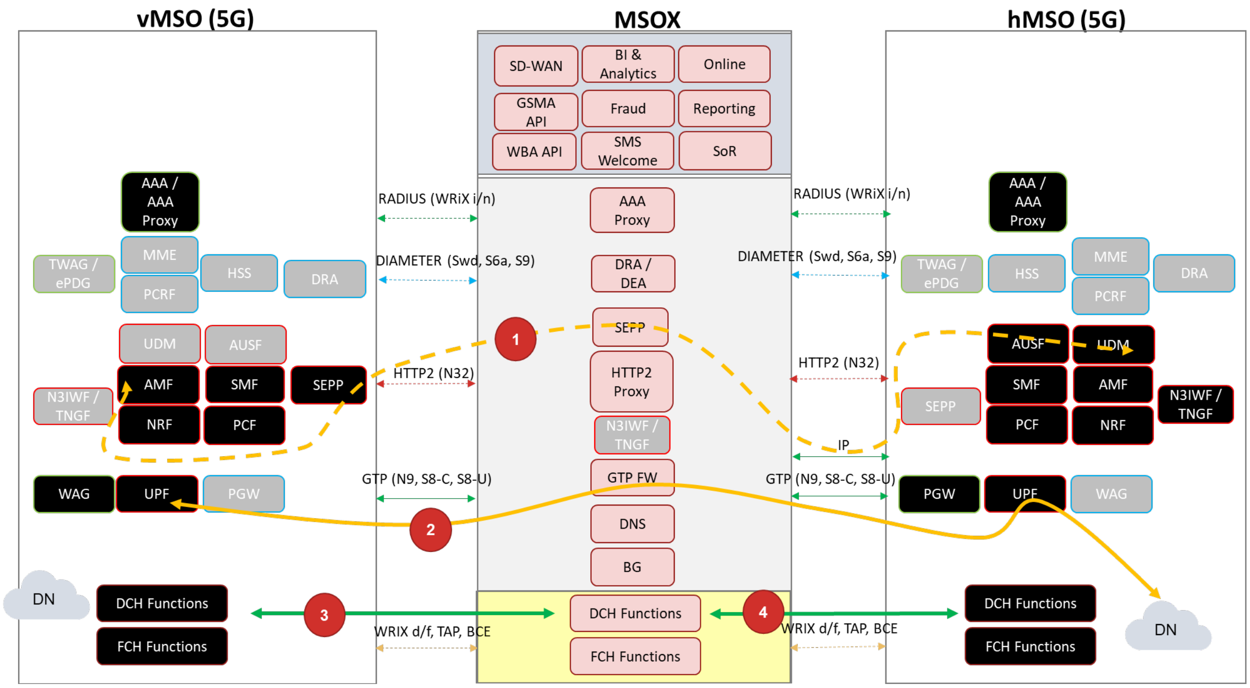


Figure 18 – 5G-to-5G mobile roaming between MSOs (SEPP hosted for home operator in MSOX)

The user plane GTP traffic over the N9 interface will be transferred via the GTP FW to protect the MSOs' core networks from denial-of-service attacks. MSOX can correlate the F-TEID exchanged via the HTTP2/JSON messages over the N32-f interface, with the F-TEID included in the GTP tunnel header, to ensure only authorized user plane traffic is transported between the two MSOs. Alternatively, or additionally, each MSO can enable IPUPS in its network to ensure passage of only authorized GTP-encapsulated user traffic.

Finally, the newly specified BCE process will be used to facilitate data and financial clearing between the MSOs. As described in Section 3.3, the DCH/FCH functions in the MSOX will provide the clearing and settlement process, indicated as steps 3 and 4 in above figures.

4.4.1.2. Seamless Handovers Between 5G Networks

This scenario applies to MSOs that have deployed their 5G networks in adjacent areas of large metro regions and require the ability to facilitate seamless handover of their customers' data sessions across their networks.

Figure 19 depicts the architecture to enable handovers across two operators with roaming arrangements. It is the same roaming architecture described in Figure 2, with one notable addition—N14 is also enabled via the N32 interface between the visited and home operators to facilitate handovers and relocation of UE context between the two networks. N14 enables the interconnection of AMF in the visited network to the AMF in the home network.

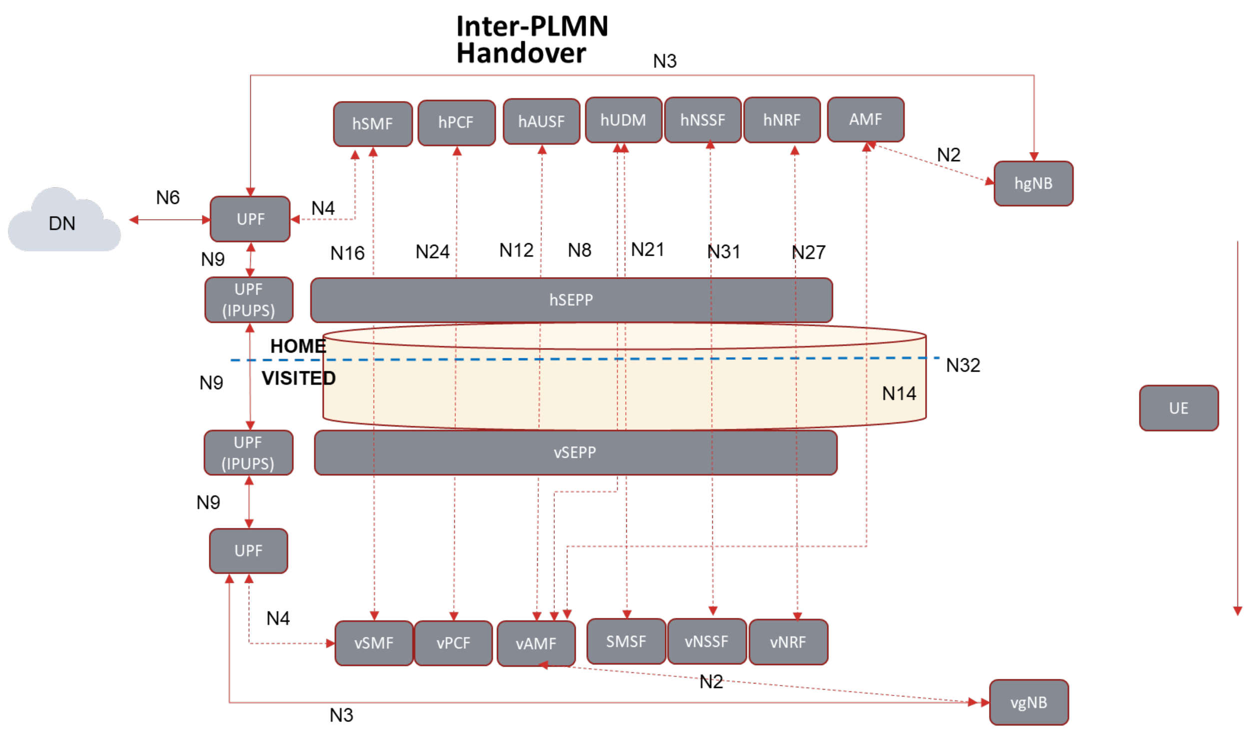


Figure 19 – Architecture for inter-PLMN handovers across operators

In this scenario, in addition to the interfaces described in Section 3.4.1, MSOX also allows the transport of N14 interface related messages over the N32 interface. MSOX will prioritize N14 traffic to ensure low latency handovers. The rest of the functionality offered by MSOX remains the same as in Section 3.4.1.

4.4.1.3. 5G Roamers into 4G

Figure 19 depicts the outbound roaming of MSO subscribers onto a 4G network. This scenario occurs when a subscriber of a 5G-only MSO roams internationally or in a rural area where the preferred roaming partner has only deployed 4G. It is assumed that the home network has deployed a combined HSS+UDM, PCF+PCRF, SMF+PGW-C, and PGW-U+UPF and configured the 4G credentials within the HSS+UDM for the H-MVNO SIM.

The signaling messages will be routed via the DRA/DEA[3] within MSOX to the home HSS and PCRF. The GTP traffic will continue to be routed via the GTP FW in MSOX.

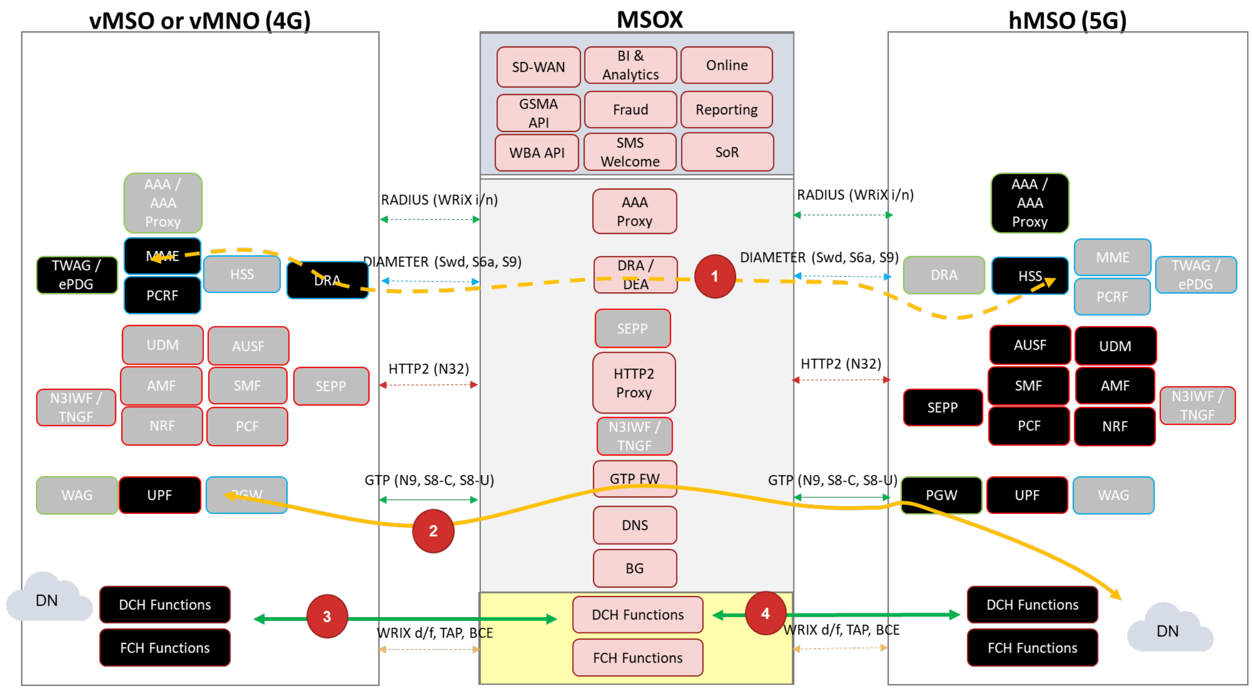


Figure 20 – 4G-to-5G mobile roaming between MSOs

If the 4G operator only supports TAP-based data clearing and settlement, MSOX will consolidate and perform the settlement using GSMA's new BCE process toward the home network.

4.4.2. Wi-Fi Roaming

Depending on whether Wi-Fi is integrated into the 5G core in the visited and home networks or not, there are several different Wi-Fi roaming scenarios supported via MSOX. 3GPP-integrated Wi-Fi in the visited network is considered first; thereafter, roaming into a standalone Wi-Fi deployment (either by the MSO or a non-MSO) is considered.

4.4.2.1. 3GPP Integrated

Figure 21 depicts the scenario where the visited network operator has Wi-Fi integrated into its 5G core. The home operator also supports non-3GPP access integrated into its 5G core.

[3] Given that Diameter signaling is not required in a 5G core, except for roaming to 4G, there will be very little need for a Diameter routing agent within a 5G SA network. In such a scenario, to reduce operational overhead, the MSOX can host the necessary DRA functionality on behalf of the MSO.

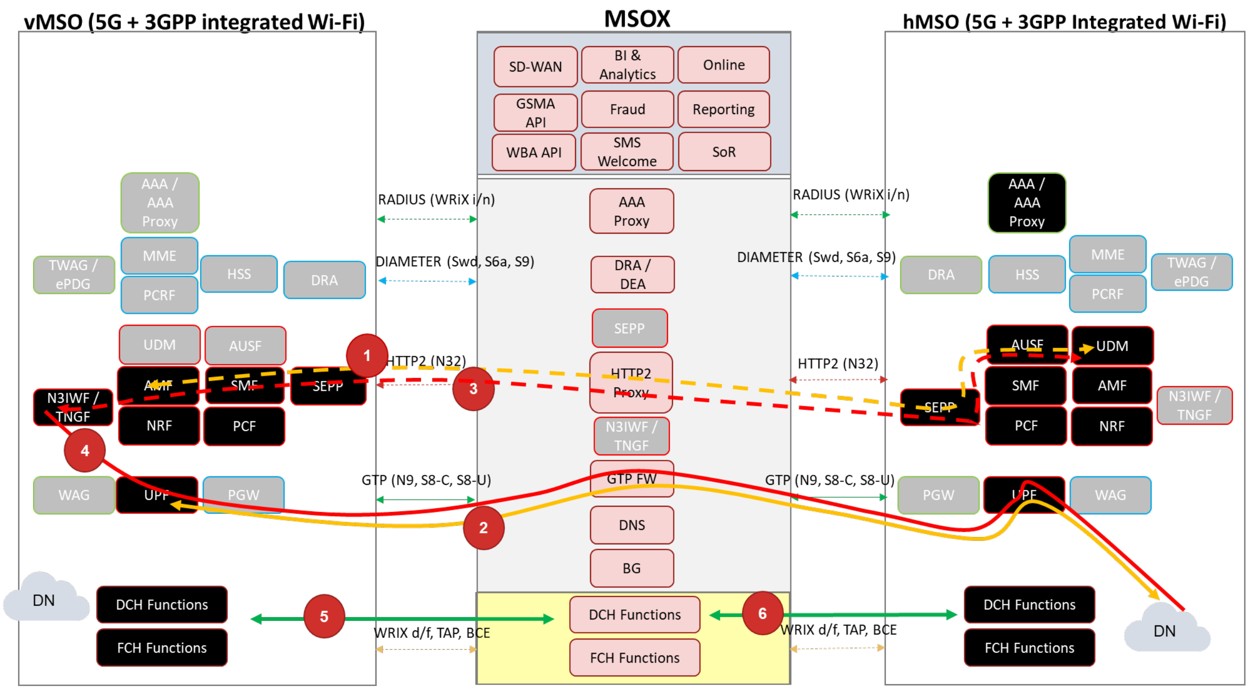


Figure 21 – Wi-Fi roaming - Wi-Fi integrated into 5G core

The mobile roaming operation is the same as shown in Figure 17 and Figure 18. MSOX will continue to provide HTTP2 proxy service (if operators decide to host their own SEPPs) for the signaling path and the GTP firewall and routing for user plane data traffic with the required QoS.

Irrespective of whether the visited operator's integrated Wi-Fi network is trusted or untrusted by the 5G core, the signaling and data paths will remain unchanged. If the Wi-Fi network is treated as untrusted, the visited operator will have deployed N3IWF to integrate the Wi-Fi network into the 5G core. Otherwise, it would have deployed TNGF. The signaling will be authenticated by the visited AMF via MSOX, and the data path will be same as that of mobile authentication. The home network will assign the same packet session anchor assigned for the mobile operation in the home network. This will facilitate ATSSS operation and improve the quality of experience for the roaming subscribers by leveraging both Wi-Fi and mobile access networks based on application requirements.

The clearing and settlement will be done within the MSOX in the same way as it is done for mobile roaming; i.e., using the GSMA BCE process and interface.

4.4.2.2. Non-3GPP Integrated (Standalone)

Figure 22 depicts the scenario where the visited operator has not integrated its Wi-Fi deployment as part of its mobile deployment. The roaming over the visited Wi-Fi network could be enabled either through a bilateral/multilateral agreement using WRIX or OpenRoaming via MSOX.

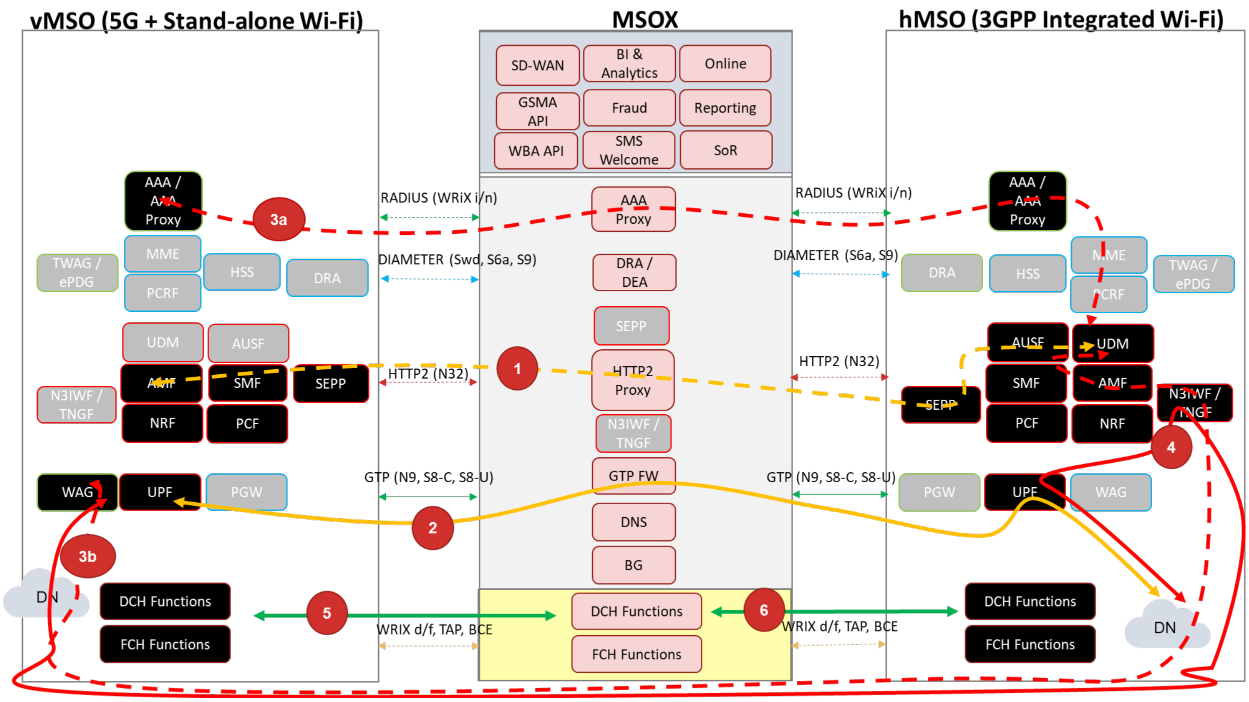


Figure 22 – Wi-Fi roaming - standalone Wi-Fi deployment using WRIX/OpenRoaming (N3IWF deployed by home operator)

In this scenario, the authentication/authorization signaling will be proxied via the MSOX AAA proxy using the RADIUS interface between the visited and home networks. After the access to Wi-Fi is authenticated, the device will discover the public IP address of the home N3IWF using DNS (if not preconfigured within the device) and establish an IPSec tunnel to perform 3GPP authentication using NAS signaling with the home AMF. Upon successful authentication/authorization, the home network will assign the same packet session anchor (SMF and UPF) as the mobile session. Like the previous scenario, this will ensure that ATSSS is available to the roaming users while roaming into the partner's Wi-Fi and mobile networks.

For standalone Wi-Fi networks that do not support OpenRoaming or even Passpoint authentication, roaming can be enabled by MSOX by providing online signup service on behalf of the Wi-Fi network or the visited MSO network provider, and the home MSO's subscribers will be able to sign up and register for the service on the fly. Here MSOX provides all the relevant infrastructure, such as AAA proxy/server, PKI RADSEC, DNS, and roaming management in the form of custom data and financial settlement.

The clearing and settlement for Wi-Fi access will be done using the WBA process and using the associated WRIX d/f processes and interfaces, whereas for mobile access, it will be via the GSMA BCE process. MSOX will consolidate across disparate accesses to streamline the settlement.

In the above scenario, N3IWF was deployed and operated by the home operator. However, if the home operator does not have a widespread deployment of its own Wi-Fi network, instead of deploying and managing its own N3IWF, the home operator could host the N3IWF in MSOX. The signaling and data flow will be identical to that in Figure 22, except that the connectivity between the N3IWF in MSOX and the home operator's 5G core will be over a secured interface. Figure 23 depicts such a scenario.

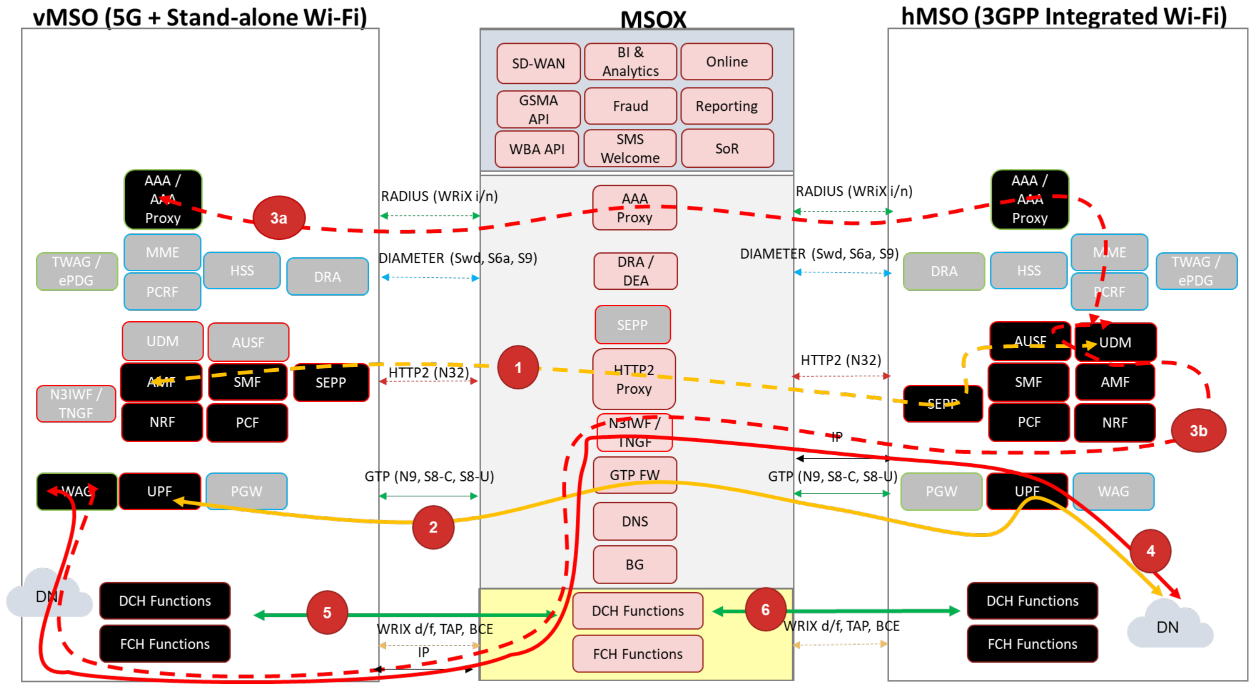


Figure 23 – Wi-Fi roaming - standalone Wi-Fi deployment using WRIX/OpenRoaming (N3IWF hosted and managed in MSOX on behalf of the home operator)

4.4.3. Data Clearing House for Evolved DSDS Interconnect Options

Several MVNO interconnect options were identified in the technical paper prepared for SCTE by CableLabs1 to facilitate better user experience and control over the data usage of H-MVNO's mobile subscribers. These options rely on interconnection options between the MNO and the MSO core to facilitate a common packet session anchor point within an MSO's mobile core infrastructure. This requires that the S8 and, optionally, S6a and S9 interfaces be set up between the MNO and the MSO networks. MSOs can optionally leverage MSOX to secure the connection and host the DRA. Additionally, MSOX's DCH functionality could be leveraged to consolidate, rate, and reconcile the data usage over the MNO network. Figure 24 depicts the use of MSOX for interconnecting the MSO's core with that of its respective MNO partner and using MSOX's DCH capability to perform data settlement and reconciliation. The existing financial clearing and settlement capabilities can continue to be used, or, optionally, MSOX's FCH can be leveraged (not shown in Figure 24).

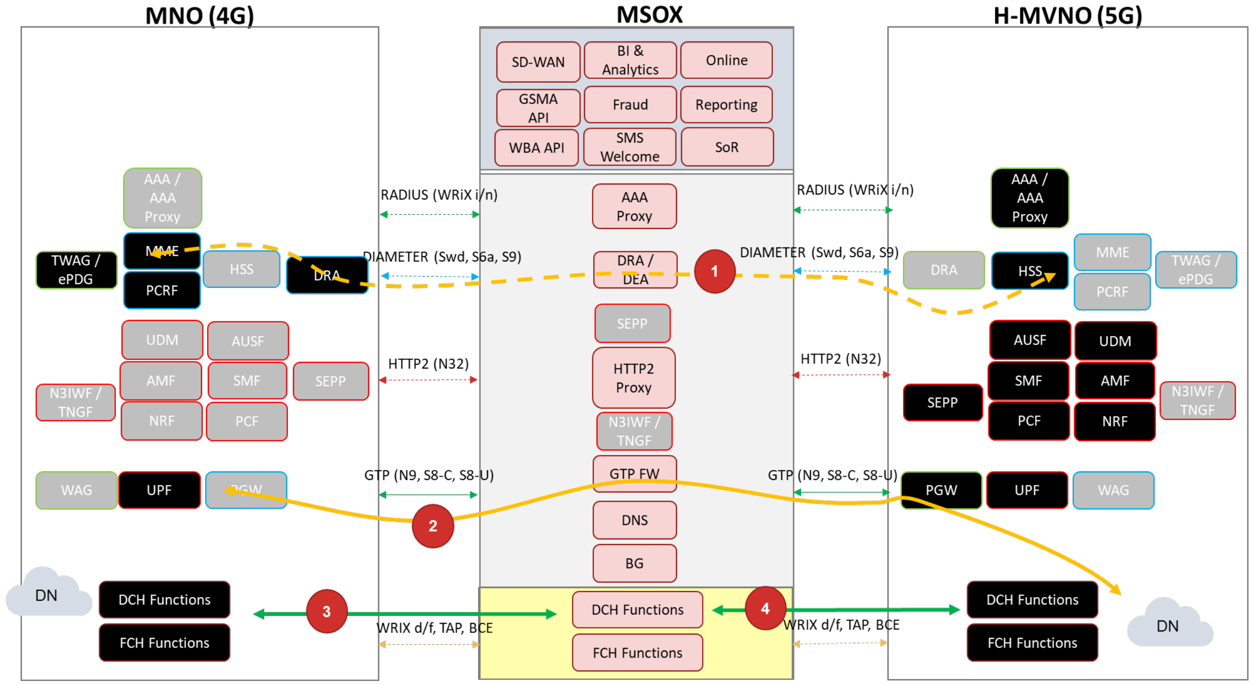


Figure 24 – Leveraging MSOX for evolved DSDS H-MVNO interconnect options

4.4.4. ATSSS Using Partner’s Wi-Fi and MNO’s 4G/5G

ATSSS refers to the capability by which the operator can dynamically steer, switch, or split traffic between different mobile and Wi-Fi accesses based on the availability and/or quality of the connection to each access. This capability requires feature support in the visited and home 5G core networks. Even though ATSSS is a 5G capability, it is transparent to the 4G and, therefore, can be supported when connected to the partner’s 4G network. As alluded to in previous sections, enabling ATSSS requires either N3IWF or a TNGF to be deployed either in the visited or home networks.

Deploying N3IWF as part of the 5G core deployment or hosting it within MSOX will enable ATSSS between the MNO partner with whom the MSO has an H-MVNO relationship and the Wi-Fi deployment of its MSO roaming partner. This will be made possible by implementing one of the evolved DSDS architectural solutions described in Section 4.4.3. Figure 25 depicts such a possibility with a hosted N3IWF that facilitates prioritizing Wi-Fi access (provided the quality of the connection is adequate) over the MNO connection, thereby reducing the data traffic transmitted over the MNO network in areas inside of the MSO roaming partner’s Wi-Fi footprint. This in turn will translate into further reduction in costs incurred in terms of payments to the MNO.

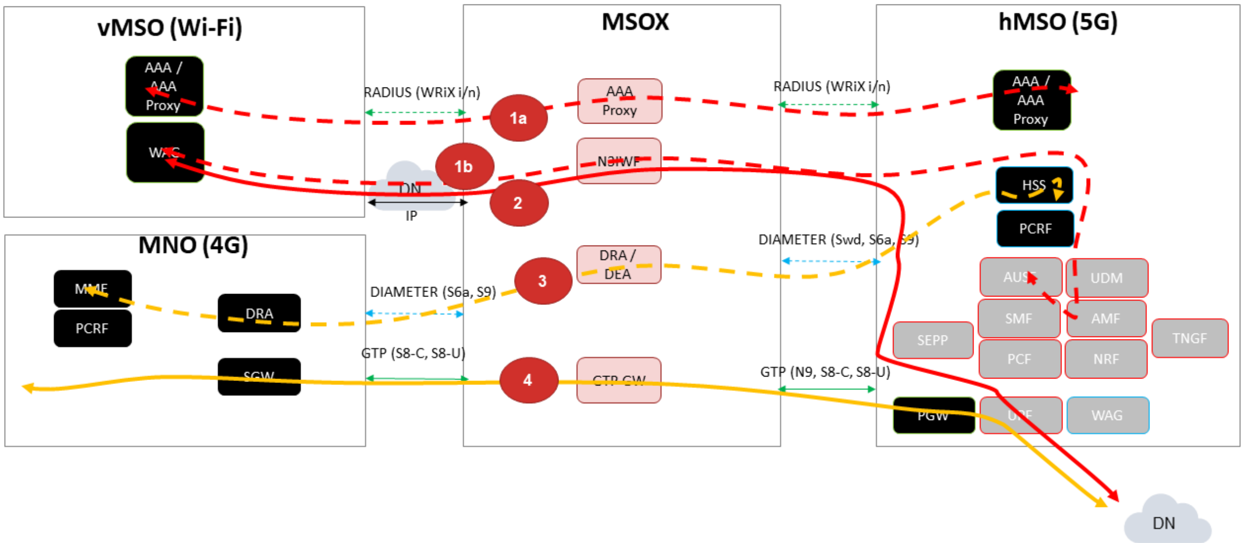


Figure 25 – ATSSS between H-MVNO's MNO partner and vMSO's Wi-Fi outside hMSO's coverage footprint

4.4.5. Inbound Roamers (Other Operators)—Setting up a Common PLMN ID Across All Operators

Operators prefer to steer their roaming subscribers to a network that minimizes roaming costs. With proliferation of multi-SIM devices, international operators could enable new roaming opportunities by enabling a data-only roaming service using a secondary eSIM that prioritizes MSOs' mobile deployments[4] in the United States. However, this will require international operators to configure several PLMN IDs (one for each MSO deployment) in the roaming list for the eSIM. One way to minimize this operational overhead and facilitate seamless roaming across MSO networks will be to configure a common PLMN ID that can be broadcast by all MSOs as a second PLMN ID. The international operators will configure this common PLMN ID as part of the roaming PLMN ID list for the secondary eSIM. Data access will be via the secondary eSIM wherever coverage from MSO deployment is available. All other areas, the primary SIM, and the existing roaming relationship with the tier-1 operator can be leveraged.

Optionally, MSOs could implement an MOCN (Multi-Operator Core Network)-based dedicated core deployment whereby the dedicated core associated with this common PLMN ID is hosted in MSOX. One potential advantage will be the operational consistency for the inbound roamers, with a single enforcement point within MSOX. Another will be the lowering of operational overhead for the MSOs associated with configuring and enforcing agreements for a large number of roaming partners.

[4] It is assumed that MSOs are able to offer lower roaming rates than Tier-1 mobile operators across their mobile and Wi-Fi deployments.

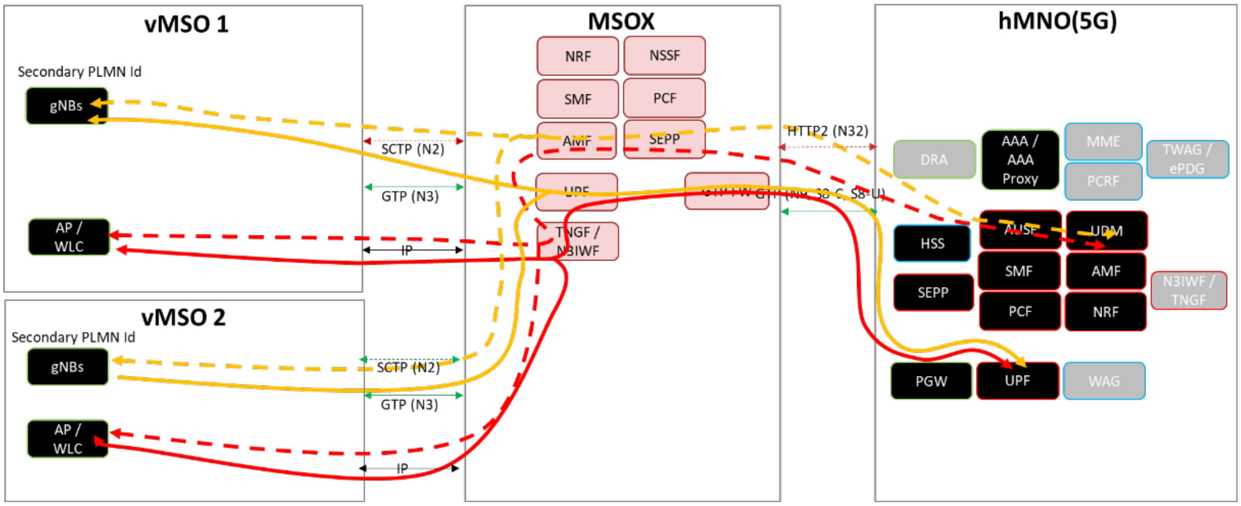


Figure 26 – Dedicated core for shared PLMN ID in MSOX

5. Conclusions

As outlined above, the MSOX platform will assist MSOs to quickly and cost effectively increase the service area footprints of their wireless networks by allowing subscribers to roam onto each other's networks. As 5G deployment accelerates, MSOs may deploy their own 5G networks and may look to utilize other MSOs' networks via roaming agreements. The MSOX platform is flexible to allow MSOs to utilize all functions and services provided by MSOX or by selecting specific functions and services, which allows MSOs to maintain varying degrees of control over their roaming businesses. The platform is flexible enough to allow hosted or outsourced functions and services.

This paper provided an overview of 5G, 4G, and Wi-Fi roaming; operationalization of roaming; and IPX network as described by 3GPP, GSMA, and WBA. The paper also outlined the key benefits of the MSOX platform and provided a high-level functional description of the MSOX platform and potential value-added services that could be provided by MSOX. It identified various scenarios where MSOX would be beneficial. Finally, it provided a high-level architecture of the MSOX platform. This paper can be considered as a blueprint for laying down the groundwork for enabling inter-MSO convergence.

6. Bibliography and References

3GPP TS 23.060, "General Packet Radio Service (GPRS); Service description; Stage 2" (Release 16), v16.0.0, March 2019

3GPP TS 23.261, "IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2" (Release 16), v16.0.0, July 2020

3GPP TS 23.401, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access" (Release 17), v17.2.0, September 2021

3GPP TS 23.402, "Architecture enhancements for non-3GPP accesses" (Release 17), v17.0.0, March 2021

3GPP TS.23.501, "System architecture for the 5G System (5GS)" (Release 17), v17.2.0, September 2021

3GPP TS 23.502, "Procedures for the 5G System (5GS)" (Release 17), v17.2.1, September 2021

3GPP TS 23.682, "Architecture enhancements to facilitate communications with packet data networks and applications" (Release 17), v17.1.0, September 2021

3GPP TS 24.312, "Access Network Discovery and Selection Function (ANDSF) Management Object (MO)" (Release 16), v16.0.0, July 2020

3GPP TS 25.304, "User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode" (Release 16), v16.0.0, July 2020

3GPP TS 29.272, "Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol" (Release 17), v17.1.0, September 2021

3GPP TS 29.274, "3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3" (Release 17), v17.3.0, September 2021

3GPP TS 29.281, "General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)" (Release 17), v17.1.0, September 2021

3GPP TS 29.303, "Domain Name System Procedures; Stage 3" (Release 17), v17.1.0, March 2021

3GPP TS 29.500, "5G System; Technical Realization of Service Based Architecture" (Release 17), v17.4.0, September 2021

3GPP TS 29.502, "5G System; Session Management Services; Stage 3" (Release 17), v17.2.0, September 2021

3GPP TS 29.503, "5G System; Unified Data Management Services; Stage 3" (Release 17), v17.4.0, September 2021

3GPP TS 29.509, "5G System; Authentication Server Services; Stage 3" (Release 17), v17.3.0, September 2021

3GPP TS 29.510, "5G System; Network function repository services; Stage 3" (Release 17), v17.3.0, September 2021

3GPP TS 29.512, "5G System; Session Management Policy Control Service; Stage 3" (Release 17), v17.4.0, September 2021

3GPP TS 29.513, "5G System; Policy and Charging Control signaling flows and QoS parameter mapping; Stage 3" (Release 17), v17.4.0, September 2021

3GPP TS 29.518, "5G System; Access and Mobility Management Services; Stage 3 (Release 17), v17.3.0, September 2021

3GPP TS 29.531, "5G System; Network Slice Selection Services; Stage 3" (Release 17), v17.2.0, September 2021

3GPP TS 29.573, "5G System; Public Land Mobile Network (PLMN) Interconnection: Stage 3" (Release 17), v17.2.0, September 2021

3GPP TS 33.117, "Catalogue of general security assurance requirements" (Release 17), v17.0.0, June 2021

3GPP TS 33.501, "Security architecture and procedures for 5G System" (Release 17), v17.3.0, September 2021

3GPP TS 36.304, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode" (Release 16), v16.5.0, September 2021

GSMA PRD AA.12, "International Roaming Agreement," v16.7, August 2019

GSMA PRD AA.73, "Roaming Hubbing Client to Provider Agreement," v8.3, April 2020

GSMA PRD BA.19, "RAEX Op Data Business Requirements," v10.0, November 2019

GSMA PRD BA.27, "Charging Principles," v49.1, September 2021

GSMA PRD BA.29, "RAEX IOT Business Requirements and IOT Rules and Procedures," v4.2, July 2019

GSMA PRD BA.30, "Steering of Roaming," v4.3, April 2019

GSMA PRD BA.40, "Roaming Guide," v7.4, July 2019

GSMA PRD BA.60, "Roaming Hub Handbook," v2.5, July 2020

GSMA PRD BA.62, "Roaming Hubbing Business Requirements Commercial Model," v4.4, July 2020

GSMA PRD BA.65, "LTE, VoLTE and 5G NSA Roaming Implementation Handbook," v7.0, November 2021

GSMA PRD FS.19, "NRTRDE Commercial Implementation Handbook," v11.2, March 2017

GSMA PRD FS.20, "Normal Business Fraud and Crime Threats," v1.1, December 2014

GSMA PRD FF.21, "Fraud Manual," v18.0, February 2021

GSMA PRD IR.21, "GSM Association Roaming Database, Structure and Updating Procedures," v15.0, August 2021

GSMA PRD IR.33, "GPRS Roaming Guidelines," v10.0, July 2017

GSMA PRD IR.34, “Guidelines for IPX Provider networks (Previously Inter-Service Provider IP Backbone Guidelines),” v17.0, May 2021

GSMA PRD IR.61, “Wi-Fi Roaming Guidelines,” v14.0, May 2021

GSMA PRD IR.67, “DNS Guidelines for Service Providers and GRX and IPX Providers,” v18.0, May 2021

GSMA PRD IR.77, “InterOperator IP Backbone Security Req. For Service and Inter-operator IP backbone Providers,” v5.0, October 2019

GSMA PRD IR.85, “Roaming Hubbing Provider Data, Structure and Updating Procedures,” v2.1, March 2015

GSMA PRD IR.88, “EPS Roaming Guidelines,” v24.0, May 2021

GSMA PRD TD.36, “Data Clearing Procedures,” v10.0, May 2019

GSMA PRD TD.201, “Common Billing and Charging Processes,” v5.4, September 2021

GSMA PRD TD.204, “Reject Dispute Reporting,” v1.1, May 2021

GSMA PRD TD.206, “Detail Data Records,” v1.4, June 2021

GSMA PRD TD.207, “Billing and Charging Evolution (BCE) Testing,” v1.1, January 2021

GSMA PRD TD.91, “RTDR Format Specification,” v5.1, May 2019

GSMA PRD TD.92, “Roaming Hubbing RAEX, TAP, RAP and NRTRDE Flow,” v4.2, November 2012

WBA, “WRIX Umbrella Doc,” v2.0.0, March 2020

WBA WRiX-n, “Network and AAA Focus,” v2.0.0, May 2020

WBA WRiX-i, “Interconnection,” v2.0.0, May 2020

WBA WRiX-d&f, “Data Clearing and Financial Settlement,” v2.0.0, March 2020

WBA WRiX-L, “Location Data Exchange,” v3ss.0.0, March 2020

SCTE, CableLabs: Evolved MVNO Architectures for Converged Wireless Deployments, October 2021

SCTE STANDARDS JOURNAL

Workforce Shortage? Technology to the Rescue:

Built-in Intelligence and Thoughtful UI Design
Simplify Fiber Test, Avoid Errors and Reduce Test Time

A Technical Paper prepared for SCTE by

Michael Scholten, Sr. Product Line Manager, AFL
2 Technology Park Dr.
Westford, MA 01886
Michael.scholten@AFLglobal.com
+1 978.455.9213

Table of Contents

List of Figures

List of Tables

1. Introduction

Fiber deployment is booming driven by a perfect storm of fiber-to-the-home (FTTH) broadband access, 5G transport, hyperscale datacenters, and edge computing. At the same time, the "Great Resignation" exacerbates the growing demand for experienced fiber installers, creating a crippling shortage in the fiber installation workforce.

Expanding the skilled workforce cannot happen overnight. However, cable, component, and equipment manufacturers are introducing products and providing training to simplify fiber splicing and termination. They are doing this by further automating single- and multi-fiber connector inspection as well as by integrating intelligence into installation test and diagnostic equipment so less skilled users can install, verify, troubleshoot, and repair optical networks more efficiently.

This article summarizes recent improvements in cabling, components, and fiber installation, verification and troubleshooting equipment, and then identifying how service providers and contract installers may utilize those improvements to successfully complete more work with less skilled workers.

2. Why a Critical Shortage of Skilled Fiber Installers?

Telecommunications industry news is replete with articles regarding the critical shortage of skilled workers to deploy and maintain rapidly expanding fiber optic networks [1, 2]. The causes for both the surge in demand and the shortage of skilled workers are commonly understood:

Table 1 - Surging demand is only one cause of skilled worker shortage.

Surge in Demand	Shortage of Workers
• Telco deployment of FTTH is booming…	• Demand for workers is growing...
• Applying pressure to cable operators to increase subscriber bandwidth including more upstream bandwidth…	• Even while the most experienced workers are aging out of workforce…
• Causing operators to split nodes and accelerate deployment of fiber deep into the network …	• Or simply not returning from COVID shutdowns or vaccine mandates…
• Compounded by 5G rollouts requiring fiber backhaul for more closely spaced macro-cells, micro-cells and DAA.	• As newer, less skilled workers begin to fill the gap, coax technicians must be retrained to install and maintain fiber.

In response, vendors, government, and industry trade associations -- including the Society of Cable Telecommunications Engineers (SCTE), the Fiber Broadband Association (FBA), and the Fiber Optic Association (FOA) – have been expanding their fiber optic skills training programs [3, 4, 5].

Even as training expands the pool of semi-skilled workers, what else can be done to simplify and accelerate fiber optic network installation, verification, and maintenance to enable a still limited supply of workers to accomplish more in less time? Here's where technology can come to the rescue.

3. Technology Can Help Overcome Skilled Installer Shortage

Installing, verifying, and maintaining fiber networks requires a fair amount of sophisticated equipment including fusion splicers, fiber termination (aka connectorization) equipment, laser sources and power meters, connector end-face inspection and cleaning products, non-intrusive optical fiber identifiers, optical time domain reflectometers (OTDRs), and live passive optical network (PON) troubleshooters.

Coincident with the broadband-fueled surge in fiber deployment which began several years ago, equipment manufacturers began integrating increased intelligence into fiber installation, verification, and maintenance equipment. At the same time, technology advancements enabled equipment users to reduce the time required to:

- Splice fibers;
- Inspect and clean connectors;
- Verify network insertion and return loss; and
- Characterize the location, loss and reflectance of installed components including connectors, splices, and PON splitters or taps.

Technology advancements associated with personal computers, tablets and phones have also enabled equipment vendors to incorporate larger, high visibility, higher-resolution displays, gesture-recognition touchscreen controls, and powerful processing and communications capabilities into smaller, lighter, more rugged, battery-operated packages.

Let us review how technology improvements simplify and reduce the time required to join, terminate, verify, and maintain optical networks.

3.1. Fiber Splicing

Due to the length of outside plant networks, obstacles encountered during installation, and the need to branch networks at various locations, fiber cables must be installed in sections, joined, or terminated in splice closures or cabinets, and verified post installation.

Fibers can be joined in two ways: by thermally fusing glass fibers together or by mating cleaned and cleaved fibers in a mechanical splice. Fusion splicing always reduces optical power loss and reflectance, and with an installed splice protector forms a more robust joint. However, fusion splicing traditionally required expensive and possibly delicate splicing equipment, a skilled operator to strip, clean, cleave, splice, and then install a heat-shrink protector one fiber at a time. With today's cables carrying hundreds or thousands of fibers, fusion splicing one cable to another could take days. Mechanical splicing was not any faster; you still had to strip, clean, cleave, and splice one fiber at a time. While it does not require a fusion splicer, the material cost per splice is higher and splice performance is reduced.

With technology advances, multiple cameras integrated into the fusion splicer, and processor-controlled motor driven chucks that optimally align fiber cores, today's single fiber fusion splicers provide higher quality splices in less time. Ribbon fiber splicers are also available to align and simultaneously splice twelve fibers in flat or rollable ribbon cable, dramatically reducing splice time and the staff required to splice multi-fiber cables.

Bad splices are typically the result of poor fiber preparation, inadequate cleaning, or poor mechanical cleaves. Splicer-integrated cameras evaluate splice quality and alert installers of poor splices, usually requiring the technician to re-splice and increasing the time to complete an installation. Once again, technology provides solutions. Cleavers that track the number of cleaves since the last blade rotation, communicate with the fusion splicer via Bluetooth and provide motorized blade rotation to prevent dull edges from producing poor cleaves. Real-time arc control in the fusion splicer improves splicing performance when fibers are cleaved with less than ideal cleave angles. This results in fewer bad splices and shorter time to job completion.

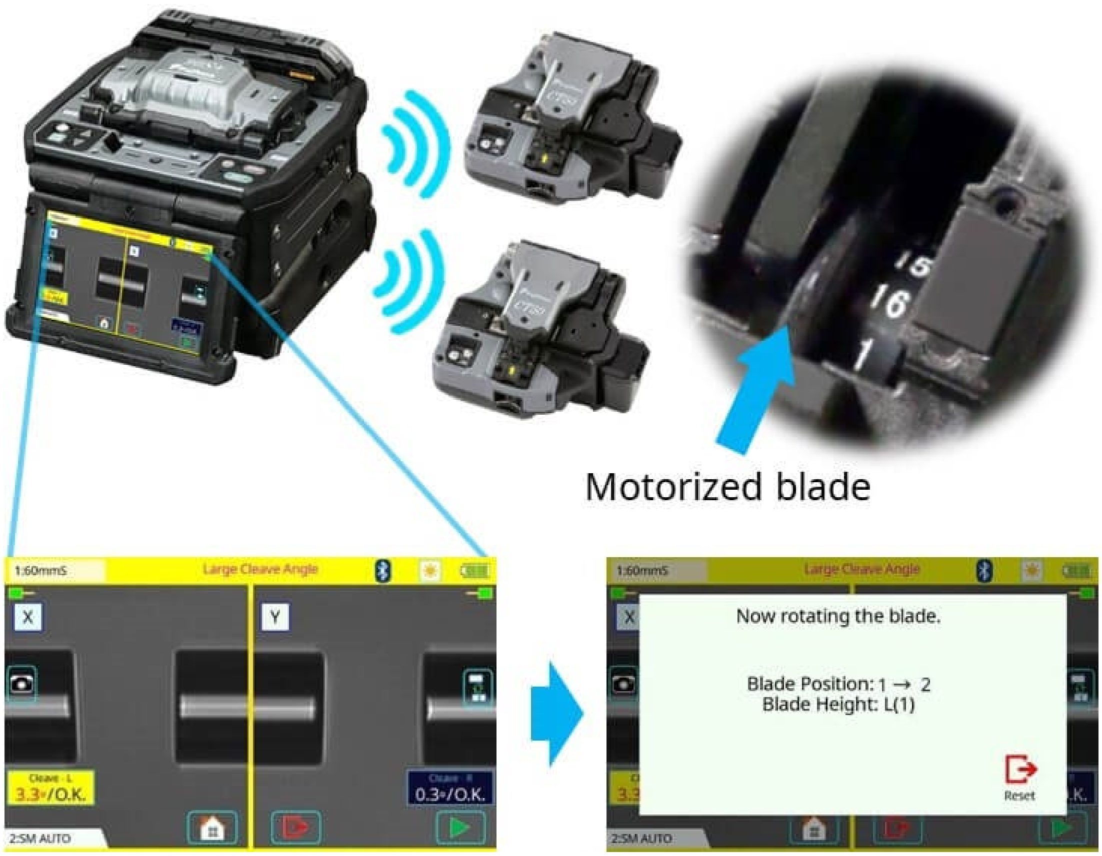


Figure 1- Automated fiber alignment, Bluetooth communications and real-time arc control reduce the skill and time required to produce high quality splices.

3.2. Fiber Termination

Terminating fiber requires installing a fiber connector. Terminations allow networks to be reconfigured without having to break and resplice the fiber. Single, duplex, and multi-fiber connectors are all commercially available. Field termination previously required the installer to strip, clean and cleave the fiber, epoxy it into the ceramic ferrule, then polish the connector end to remove any excess epoxy, producing a smooth, convex-polished end, without over-polishing which would create an air gap at a mated connection. Air gaps are a source of unacceptable insertion loss and/or reflection.

Once again, technology has provided a solution which simplifies connector installation, reduces the skill required to terminate, reduces termination time, and provides consistent, high-quality connections. The solution? Splice-on or mechanically spliced connectors containing a pre-polished fiber stub. Available in both fuse-on and mechanically spliced form factors, installation utilizes the familiar strip, clean, and cleave process followed by either fusion splicing or mechanical alignment. When fully assembled, the connector body combined with strain relief protect the fusion or mechanical splice inside the connector body.

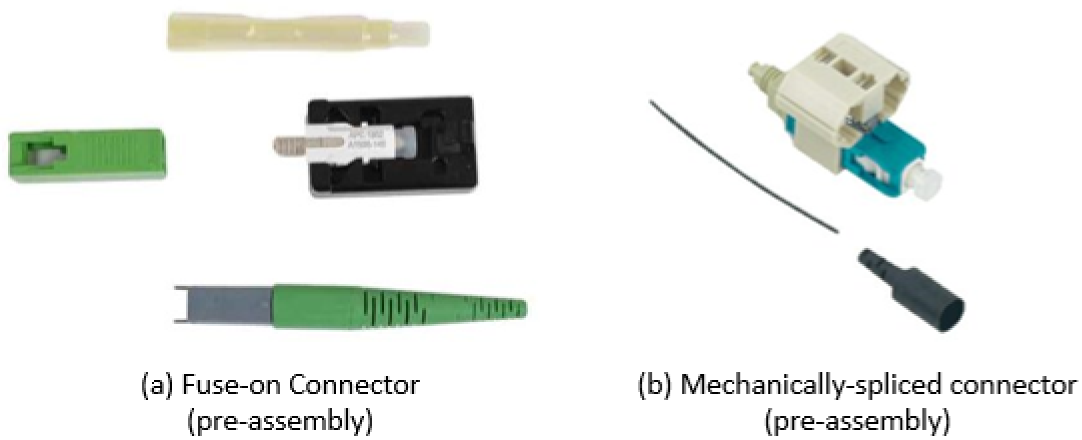


Figure 2 - Pre-terminated fuse-on or mechanical splice connectors simplify connector termination.

3.3. Connector End-Face Inspection and Cleaning

A survey of fiber installers by NTT [6] identified dirty and damaged connectors as the single most frequent cause of fiber network problems. Dirty connector end-faces reduce or prevent light transmission through a mated connection. Damaged connectors can create an air gap at a mated connection resulting in excess loss and reflection at the connection, and increased error rate on optical signals transmitted through the connection. Mating a dirty connector to a clean connector may damage both connectors as spring force in the mated connection grinds dirt into both end-faces.

Service providers all preach cleaning to their technicians. Unfortunately, the more inconvenient cleaning is, the less likely technicians are to follow good cleaning practices. Technology's solution: make cleaning as simple and fast as possible. Today's "one-click" cleaners can rapidly clean the end-faces of exposed ferrules on patchcords as well as ferrules mounted behind bulkhead adapters in patch panels. While some one-click cleaners require the user to install or remove a ferrule alignment cap when switching between cleaning patch cord and bulkhead-mounted connectors, the newest one-click cleaner for common SC connectors and field-hardened optical connectors cleans both connector types without cap swapping (see Figure 3). While this may seem a simple improvement, if it means the difference between an installer cleaning or not, it also can help prevent a poor connection or having to replace one or both connectors, often at a later time after the error has been identified.

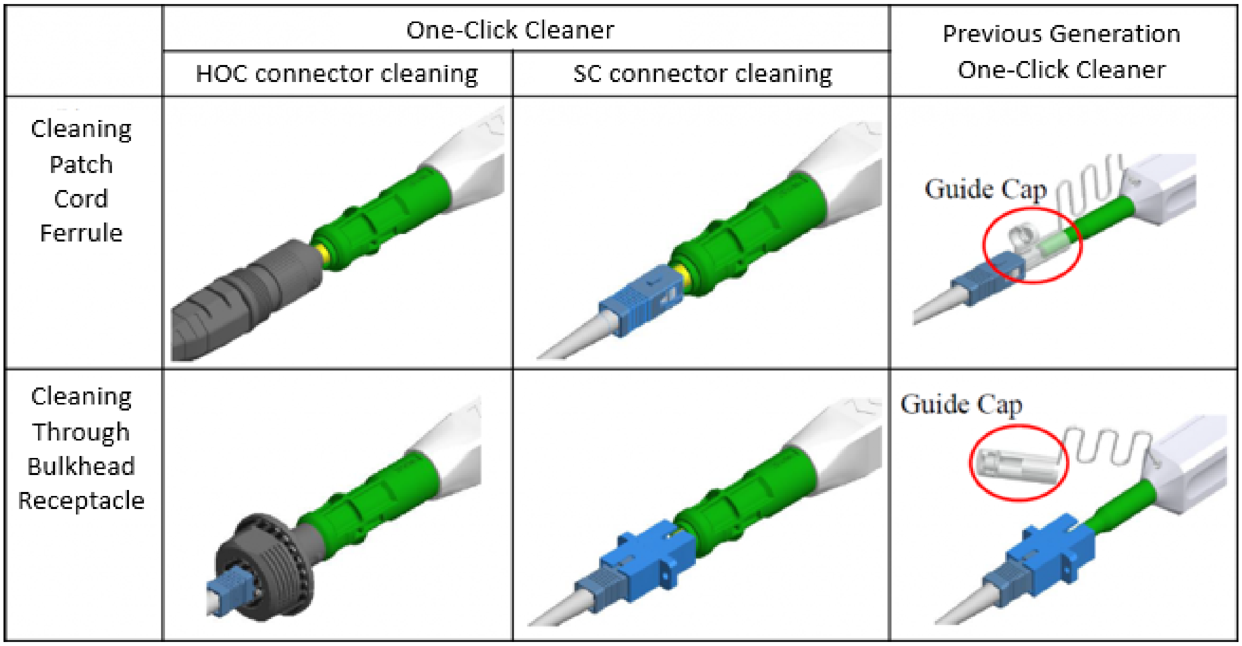
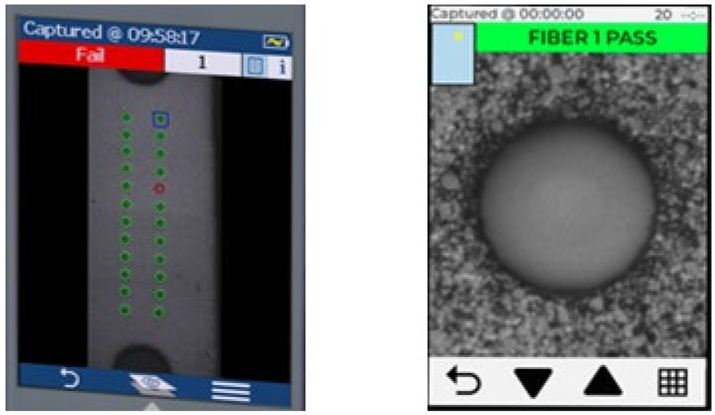


Figure 3 - Latest generation one-click cleaners clean both SC and HOC connectors without swapping cap to test patch cords or through bulkhead adapters.

The small diameter of optical fibers (smaller than a human hair) requires a microscope to examine the connector end-face for contamination or damage. Today's preferred solution is a video microscope capable of at least 400x magnification. To objectively evaluate the presence or absence of scratches, contamination or defects, IEC developed a standard [7] to establish the size and number of permissible scratches or defects in the core, cladding, and physical contact region of a connector end-face. Pass/fail limits are a function of fiber type (single-mode or multimode) and core diameter (8-9 um for single-mode, 50-62.5 um for multimode).

Technology incorporated into today's connector inspection probes simplifies and reduces test time by implementing auto-focus, auto-center, capture, analyze and display at the press of a button. Displayed connector end-face and pass/fail results may be automatically saved and/or sent via Bluetooth to a mobile device for remote archiving and/or reporting.

The latest generation of MPO multi-fiber connector inspection probes can auto-focus, capture and analyze the end-faces of all fibers in a single- or multi-row MPO connector at once, further reducing test time for MPO-terminated cables (see Figure 4).



(a) MPO-24 Full End-face (b) Individual Fiber End-face

Figure 4 - MPO inspection probes image entire end-face of a multi-row connector (a), while providing sufficient resolution to evaluate pass/fail for each fiber end-face (b).

3.4. Post-Installation Network Verification

Once optical network cables are installed, spliced, and terminated, the installed optical plant must be verified. At a minimum, network length, insertion loss and optical return loss (ORL) must be tested against expected length, maximum allowed loss, and minimum acceptable return loss limits. More typically, service providers also require verification of each splice and connection against maximum allowed loss and reflectance limits.

While an optical loss test set (OLTS) can verify length, loss and possibly ORL, it cannot locate and evaluate connectors and splices within the network. An OTDR is capable of also measuring length, loss and ORL, while additionally locating, identifying, and measuring the loss and reflectance of connectors, splices, splitters (in FTTH PONs), and defects such as breaks or macro-bends.

Operating like a radar, OTDRs transmit narrow pulses of light down the fiber measuring the time and amplitude of light backscattered or reflected from anomalies back to the OTDR. OTDRs can present a trace of returned backscatter and reflections vs. distance. A skilled operator can interpret this graph to identify splices, connectors, and the fiber end, measure the distance to each, and measure the loss or reflectance at each location. Manual settings allow the user to configure test range, select the pulse width, select test wavelengths, and adjust the averaging time, all of which affect the quality of the obtained trace. Wider pulse widths improve signal to noise ratio (SNR), but closely-spaced events may overlap and be interpreted as a single event. Longer averaging time also reduces noise in the traces, but at the expense of longer test times.

Intelligence integrated into today's OTDRs has significantly reduced the skill required to configure an OTDR and interpret results. Auto-ranging selects optimal test range. Multi-pulse acquisition allows the OTDR to automatically test the network using multiple pulse widths, selecting the best traces to evaluate for each event. Most importantly, signal processing software interprets the trace, locating, identifying, and measuring length, loss, ORL, and the location, event type, loss, and reflectance of each event.

A ”LinkMap” view depicts detected events as a linear sequence of easily understood, color-coded icons, where each icon indicates the type of event and its pass/fail status (see Figure 5). Touchscreen navigation allows the user to pan along the LinkMap, viewing each event and its associated measurement results. Displayed results may be saved internally and uploaded via Bluetooth or Wi-Fi for often mandatory archiving and reporting. Some OTDR models even include integrated print to PDF capabilities for built-in report generation.

Often overlooked in OTDR model design are uncomplicated features which reduce required operator skill. These include auto-saving results, restoring last result and last-used setup on power-up, and auto-off and display auto-dimming to prevent draining the battery when it is left unattended.

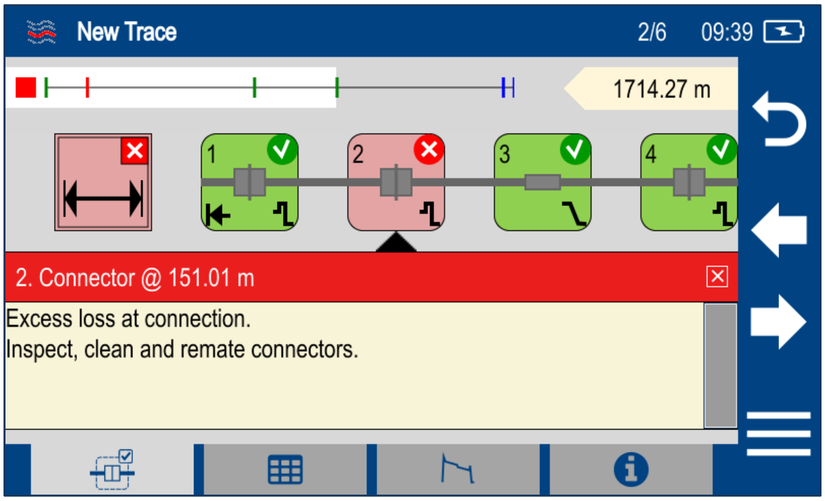


Figure 5 - Color-coded LinkMap icons identify detected events and pass/fail status.

3.5. Live PON Troubleshooting

Unlike point-to-point (P2P) transport networks, FTTH passive optical networks (PONs) utilize a point-to-multi-point architecture (PMP). In this architecture, a single feeder fiber carries both upstream and downstream traffic to/from the headend from/to an optical splitter (typically 1:32 or 1:64 split ratio) located close to subscribers' homes or apartments. Individual distribution and drop fibers can then connect splitter outputs to up to 32 or 64 subscribers.

In this architecture, one or more subscribers may lose service while others remain in service. When this occurs, a maintenance technician will likely need to troubleshoot the network from the subscriber's premise or an access point near the premise. A typical test procedure involves checking to see if the downstream signal from the headend is present at the access point or subscriber's terminal. If the downstream power level is absent or too low, but other customers remain in service, there is likely a fault in the distribution or drop fiber for that subscriber. If downstream power level is present and within expected min/max limits, the problem is likely the connection from the test point to the customer's optical network terminal (ONT), the ONT itself, or in-home connections.

Because PONs are a PMP system and all ONTs transmit upstream using the same wavelength and spectrum, they are allowed to transmit only in timeslots assigned to them by the optical line terminal

(OLT) in the headend. A consequence of this is measurement of an ONT's output power level cannot occur unless an OLT is simultaneously connected.

A through mode PON power meter (TPPM) can monitor the downstream and upstream signals as they pass through the test set. With a TPPM, a user can verify the presence and power level of both the upstream and downstream signals simultaneously. Once again, software intelligence can automatically apply pass/fail limits based on the test location and PON protocol, providing clear pass/fail indications to the user rather than requiring them to interpret power levels.

As PONs migrate from today's widely deployed GPON and EPON technology to 10Gb/s XG/XGS/10GE PON technology, additional downstream and upstream wavelengths using wavelength division multiplexing (WDM) will be utilized. This is intentional as it allows previously deployed GPON/EPON to coexist with emerging XG/XGS/10GE PON. A TPPM which identifies which wavelengths are present, measures each independently, and evaluates them against appropriate pass/fail levels reduces the skill level required to identify if PON signals are present and within acceptable power limits.

Figure 6 - PON power meter evaluates pass/fail of detected PON power levels.

Just as an OLTS cannot identify the location and cause of excess loss or reflection in a P2P network, a PON power meter cannot identify the root cause of absent or low PON power levels in an FTTH PON. An OTDR is well-suited for that task, but a traditional OTDR faces several challenges unique to PONs:

- If the PON is live and multiple customers remain in-service while one or a few are out of service, disconnecting the fiber at the headend or the input to the splitter will disrupt service to all subscribers.
- Additionally, when testing downstream through the splitter, backscatter and reflections from all 32 or 64 fibers connected to the splitter will be combined and overlap in the OTDR's trace. It is not possible to determine the backscatter or reflections contributed from each subscriber's distribution and drop cable, so results beyond the splitter are not useful.

- However, when testing in the upstream direction from the subscriber's premise, the OTDR pulses traverse only a single distribution and drop fiber before reaching the splitter, then continue through the single feeder fiber.

- The PON's PMP architecture means there may be live downstream signals present. These would normally interfere with an OTDR test in the upstream direction. To overcome this, live PON OTDRs (or troubleshooters) integrate a filter to block PON wavelengths and test using an out-of-band wavelength (typically 1625 or 1650 nm).

The newest generation of live PON troubleshooters combine both a downstream PON power meter supporting GPON/EPON, video and XG/XGS/10GE PON wavelengths, along with a live PON OTDR or troubleshooter capable of testing upstream to detect, locate and measure the loss and reflection of splices, connections, splitters, and faults. Once again, to reduce the skill level required of technicians activating and maintaining subscriber connections in a PON, built-in software intelligence compares downstream power levels, event loss and reflectance to pass/fail limits and provides clear pass/fail indications. A live PON troubleshooter may even recommend corrective action when poor splices or connections are detected.

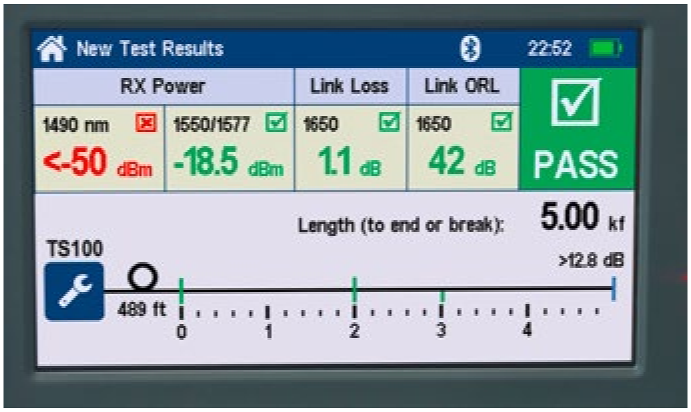


Figure 7 - Live PON troubleshooters provide clear pass/fail indications based on detected PON power levels, link loss, ORL and detected events.

4. Conclusions

The shortage of skilled fiber installation and maintenance technicians is well documented. Causes of the shortage and the service providers' need to meet bandwidth demand for residential, business and mobile backhaul services are also well understood. Numerous training programs have been created to grow the pool of skilled workers. Not as well documented are the technology-based improvements equipment manufacturers have been developing to reduce the skill and the time required to install and maintain fiber networks. Technicians and managers have a growing array of technology-based features to consider when selecting splicing, termination, inspection, cleaning, network verification and troubleshooting equipment for their growing, but less-skilled workforce. Easy-to-use products with built-in intelligence enable novice workers to deliver expert results.

5. Abbreviations and Definitions

5.1. Abbreviations

DAA	distributed antenna array
EPON	ethernet PON
FBA	Fiber Broadband Association
FOA	The Fiber Optic Association
FTTH	fiber to the home
GPON	gigabit PON
MSO	multiple system operator
OLT	optical line terminal
OLTS	optical loss test set
ONT	optical network terminal
ORL	optical return loss
OTDR	optical time domain reflectometer
PMP	point-to-multi-point
PON	passive optical network
P2P	point-to-point
SCTE	Society of Cable Telecommunications Engineers
TPPM	through-mode PON power meter
XG-PON	10Gb/s asymmetric PON
XGS-PON	10Gb/s symmetric PON
5G	fifth generation mobile network
10GEPON	10Gb/s ethernet PON

5.2. Definitions

Downstream	Information flowing from the headend to the user
Upstream	Information flowing from the user to the headend
Headend	Cable operator facility from which voice, data, video services originate
Node	Point at which coax or fiber splitter branches services to multiple subscribers
Patchcord	Fiber jumper used to interconnect equipment
Patch Panel	Connector panel enabling optical signals to be rerouted by reconnecting patchcords

6. Bibliography and References

1. "Labor Shortages in Broadband are Likely to Slow Deployment," Doug Dawson, The Center for Growth and Opportunity at Utah State University, Nov-2021.

2. "Equipment, Fiber Shortage Could Delay Fiber Rollout," Mike Farrell, https://www.nexttv.com/news/fiber-deficiency , Apr-2022.

3. "Overcoming FTTH Deployment Obstacles," Kevin Morgan, Broadband Technology Report, March-2022.

4. "AT&T, Corning Tackle Labor Shortage with New Fiber Training Program," Diana Goovaerts, Fierce Telecom, Apr-2022.

5. “The Broadband Surge is Here. Is Your Team Ready?”, SCTE, https://www.scte.org/education/technology-solutions/network-architecture-and-construction/
6. “The Impact of Optical Connector End-Face Contamination and How to Prevent It,” NTT Advanced Technology, 2017.
7. “Visual Inspection of Fibre Optic Connectors and Fibre-Stub Transceivers,” IEC-61300-3-35, Ed 2.0, Jun-2015.

SCTE STANDARDS JOURNAL

AI Pairing in the Contact Center

How Service Providers Can Increase Revenue and Improve Subscriber Satisfaction

Letter to the Editor prepared for SCTE by

Fanny Heneine, Senior Director, Product Management, Afiniti
Chrysler Building, 405 Lexington Avenue
New York, NY 10174
fanny.heneine@afiniti.com

Theresa Gebert, Director, Product Management, Afiniti
Chrysler Building, 405 Lexington Avenue
New York, NY 10174
theresa.gebert@afiniti.com

Madeleine Lamm, Marketing Manager, Afiniti
Chrysler Building, 405 Lexington Avenue
New York, NY 10174
madeleine.lamm@afiniti.com

Eleanor Mcdonald, Director of Data Strategy, Afiniti
Gateway Building, 1 Macquarie Place
Sydney, NSW 2000
eleanor.mcdonald@afiniti.com

1. Introduction to AI Pairing

Service providers continue to embrace data-driven decision making to improve their businesses. Such advancement often takes the form of simple data analytics used for customer segmentation as part of marketing campaigns. And yet, both inbound and outbound contact centers have lagged behind in employing this technology, perhaps because they are often perceived by organizations as merely “cost centers” rather than an integral part of the customer journey and a massive revenue generating machine in their own right.

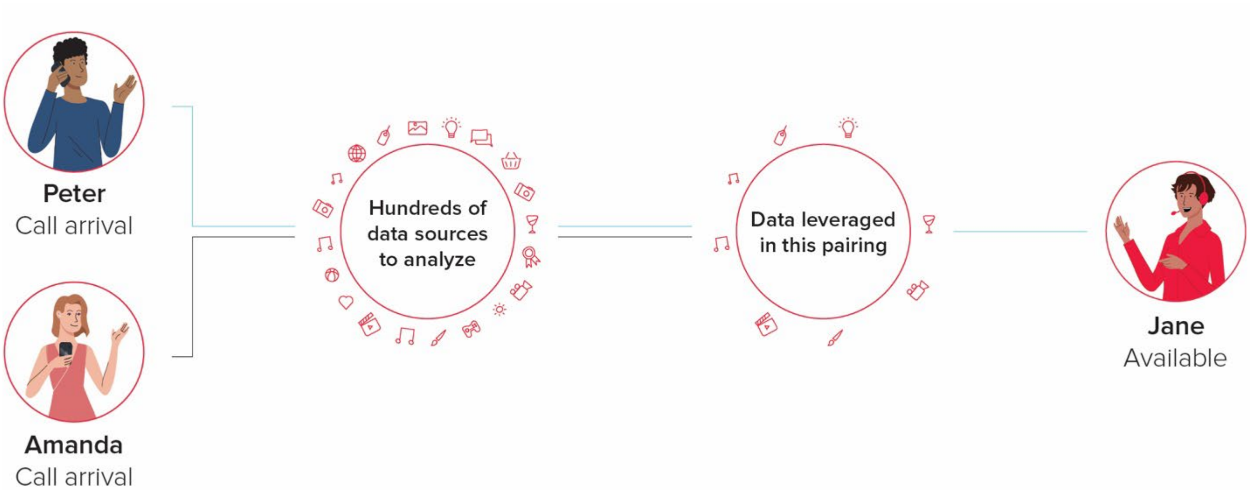
Everyone has an example of poor customer service. For me, it was with an internet provider. After receiving years of offer letters from a competitor, I finally decided to check them out after my current provider let me down. Imagine my disappointment, when after calling the competitor, I was offered different packages and bundles at a higher price than my offer letters, as well as additional services that were not relevant to me. Both the agent and I were confused and frustrated. If instead, I was correctly identified, and offered a consistent omni-channel experience, the competitor would have gained a new customer and additional revenue. The contact center was the final piece of the puzzle to convert me as the customer on whom they had spent years of marketing dollars, and it didn’t fit into place.

Using data and artificial intelligence (AI) to enhance experience during customer interactions has the potential to provide real and tangible benefits to providers’ bottom lines. But after constantly hearing “AI” in buzzword headlines and flashy catchphrases, what does it really mean? Does it mean interacting with robots that communicate in code? Or is it automating every process so that you never speak to a customer again? No; in reality, it’s neither. Ultimately, the real-world AI in existence now is a method of recognizing statistical patterns in rich, complex data. This allows companies to deeply understand their customers, their staff, their products, and the interactions between these elements in ways that were previously impossible with traditional data analytics. When this technical capability is combined with real-world human to human interactions, the possibilities of success are endless.

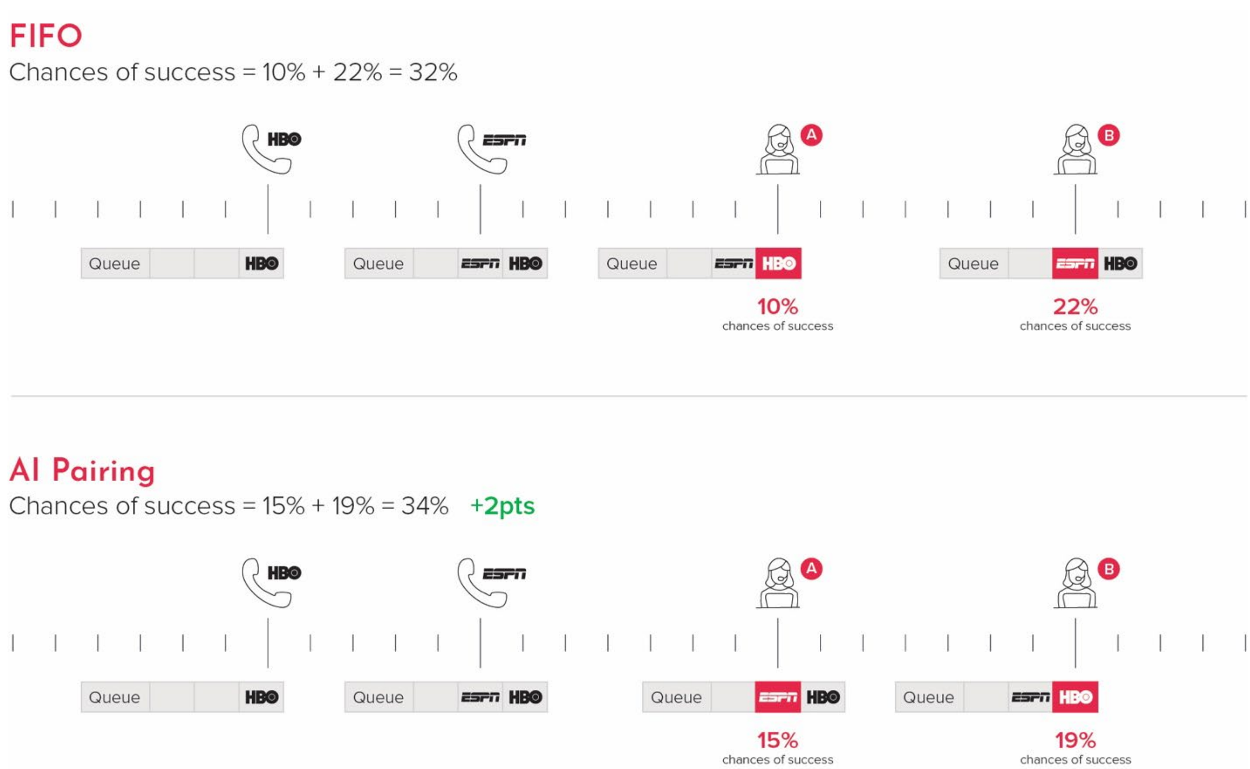
2. AI Pairing To Improve Outcomes

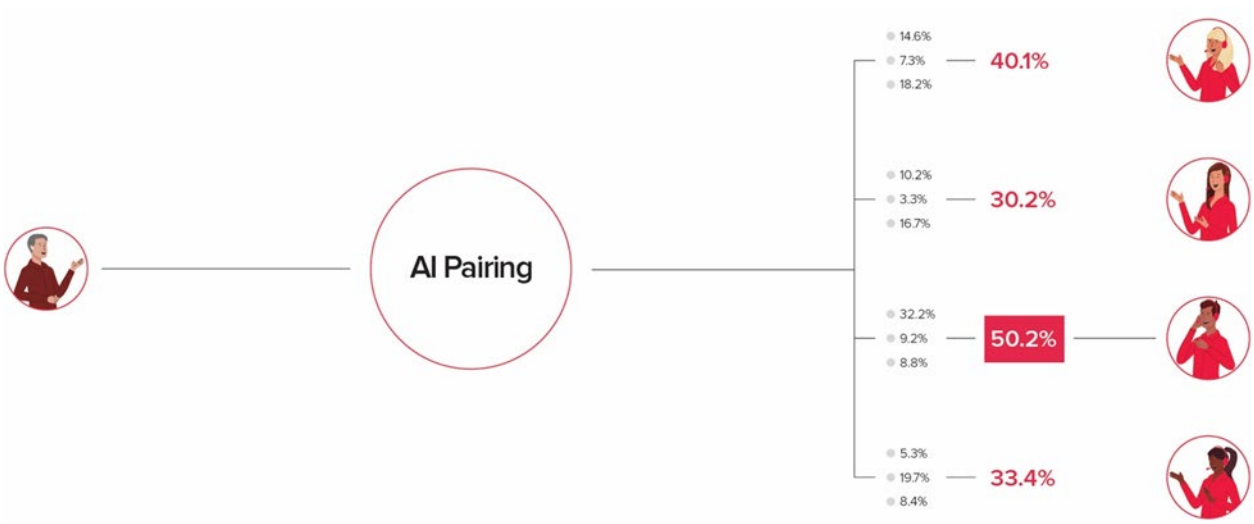
In the contact center specifically, AI is most applicable and beneficial in “AI pairing”. With traditional first-in, first-out (FIFO) call routing, customers and agents are paired solely based on their place in queue. Some contact centers have implemented performance based routing (PBR), where routing is based on the historical performance of agents. PBR can cause unintended operational effects whereby few agents are prioritized and routed too many calls, and it also only optimally functions in agent surplus environments. Unlike FIFO and PBR, AI pairing connects agents and callers as determined by best predicted fit, and works in both agent and caller surplus scenarios. AI pairing utilizes customer and agent data to find statistical patterns of behavior; models are trained and validated to identify predictors and patterns of behavior which can then be implemented in real-time. The patterns enable the algorithms to estimate the probability of success of various potential agent-caller pairs and then select the pair that maximizes the likelihood of a successful interaction. With AI pairing, success is not based on classifying agents by their recent performance. Instead, the success metric is defined as a specific, measurable outcome chosen by their organizations. Examples of success metrics optimized by the pairing algorithm might be customer lifetime value (CLV), satisfaction, revenue, or retention. It is important to note that AI pairing exists as an intelligent layer on top of a contact center’s existing routing structure and does not change the skills or volume of agents.

The application of behavioral pairing becomes clearest at a customer level. In real-time, several callers come into the contact center queue. When an agent becomes free, AI pairing technology uses the patterns identified in historical data to calculate the agent-caller pair that maximizes the probability of success with the pairing possibilities available, and the corresponding caller is routed or directed to the paired agent. This is all done in milliseconds to ensure the smooth and continuous routing of calls even in high volume environments. Consider the example of Peter, who is calling with the intention to disconnect from his cable subscription. The AI models can identify that Peter's usage data shows a medium customer tenure of two years and a drastic recent drop in viewership, as well as two technical faults in the past month, and can identify that customers with similar characteristics exhibit low customer satisfaction and high customer churn propensity. Thus, AI pairing can link him to available agent Jane, who has strong retention outcomes when speaking with frustrated customers and historically resolves retention calls with lower discounts given. Intelligently pairing Peter and Jane maximizes the chances of outcome success, in this case a successful retention, and Peter remains a customer, with a minimum one-time credit offer.



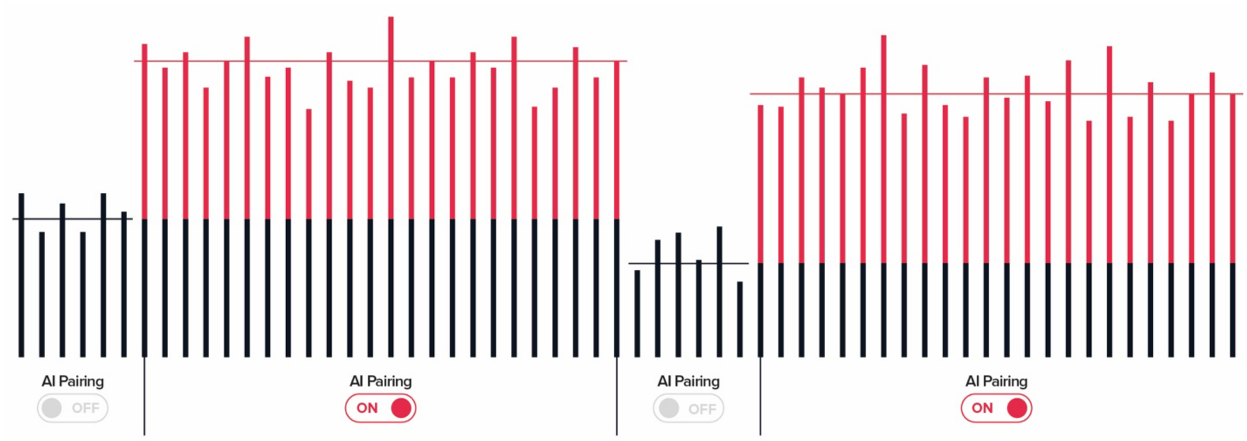
If we expand this example to consider the entire contact center environment, we can illustrate the benefits of maximizing the chance of success in a given interaction through the concept of comparative advantage. For example, say 50% of agents are drama fans and 50% are sports fans. Also assume 50% of customers are HBO customers and 50% are ESPN customers. Drama fan agents are 9% more successful at converting a sale with HBO customers than sports fan agents, and 7% more successful at converting a sale with ESPN customers than sports fan agents. While a PBR approach would route as many calls as possible to the top-performing drama fan agents, it is both impossible and unbalanced to pair the most successful agents with all customers all the time. Instead, we want to choose the best combination of pairings that maximizes overall success. In this example, pairing drama fan agents with HBO customers and sports fan agents with ESPN customers results in a 2% absolute comparative advantage overall. At scale, this number can represent millions of dollars of increased revenue.





3. Measuring The Impact

While the approach of AI pairing can be shown to generate benefits in theory, many contact center leaders wonder how these translate into defined, successful outcomes that ensure a return on their investment in the technology in reality. The answer is surprisingly straightforward: by using benchmarking methodology, the causal impact of AI pairing can be precisely measured. This is achieved by cycling the AI pairing algorithm on and off in short, alternating periods throughout each day. These short cycles control for any exogenous factors that could affect the integrity of the measurement, including day of the week, agent shifts, high and low call volume times, market changes, advertising campaigns, holidays, and even the impacts of weather. In statistical literature, this is known as a randomized controlled experiment or randomized controlled trial (RCT). By comparing the results from when AI pairing algorithms are on to results from when the intelligent pairing is switched off, clients can precisely measure the impact of the technology.



The unique benchmarking measurement approach results in what is known as a triple-blind experiment. This means that neither agents nor callers know whether a pairing algorithm is being used or not at any time, and also that the treatment of ON or OFF being applied at any point in time is decided ahead of time, prior to any outcome of a call. This allows for a clear and unbiased comparison over time, without the ability to cherry-pick certain calls for special handling into either ON or OFF or to categorize results after outcomes are final. Put simply, this means that the additional incremental gain on a chosen KPI and resulting impact to the bottom line solely attributable to implementing AI pairing technology in the contact center can be precisely measured any time, and over time.

4. Multidimensional AI Pairing

The benefits of AI pairing can extend beyond agent-customer pairing by improving the channel or platform the customer uses in the first place. For instance, if a customer is trying to upgrade their plan via your company's web chat tool, it may better to direct them to a live voice call if they are more likely to successfully upgrade via phone than chat. Or perhaps a customer is trying to get support via phone that is actually resolved faster by logging into your app. Pairing multidimensionally in this way delivers a more customized customer journey experience that ultimately drives higher revenue and customer satisfaction. When customers reach out, contextual and behavioral data allows AI to intelligently pair them not only with the agent best-suited to helping them, but also with the retail, digital, or voice-based channel that have the greatest chance of succeeding with that agent.

This optimization can be enhanced further by recommending the most optimal offers as well. AI pairing can be applied to recommending the right offers with the right people at the right time, without changing offer prices or details. While traditional next best offer engines might suggest an offer for a particular customer, multidimensional pairing includes the choice of agent in that selection. Bringing both the call data and agent variables into the picture, offers are even more targeted for greater success. In my initial personal example, multidimensional pairing could have recommended an offer option on the agent's screen that aligned the offer emails I had received previously with the real-time best offer for both myself and the agent to whom I spoke, and thus converted the conversation into a sale by gaining me as a customer. Moreover, intelligently selecting the best channel through which to serve a customer has the potential to further increase the ability to have the most successful interaction.

5. Conclusion

Put simply, intelligent and multidimensional pairing creates a more holistic experience for consumers and minimizes disconnection between the many puzzle pieces of sales and retention for providers. AI pairing also transforms the contact center from what may have been considered a sunken cost with large operational expenses into a profit-driver with high return on investment. Implementing AI pairing technology is a win-win for all parties involved: customers have better experiences, agents have better interactions, service providers increase customer value and revenue while improving customer journey, and contact center operations continue seamlessly.

SCTE

a subsidiary of CableLabs®

140 Philips Road, Exton, PA 19341-1318 | 800.542.5040 | **scte.org**