UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA: INGENIERÍA ELECTRÓNICA

APLICACIÓN DE LAS TÉCNICAS DE AGRUPAMIENTO PARA LA DISTRIBUCIÓN CUASI–ÓPTIMA DE UNA RED HÍBRIDA WDM–TDM/PON EN CASCADA MULTINIVEL QUE DA SOPORTE A UNA SMART GRID O SMART CITY.

TESIS PREVIA A LA OBTENCIÓN DEL TITULO DE:

INGENIERO ELECTRÓNICO

Autores:

Milton Santiago Amoroso Ordóñez.

Hernán Santiago Ávila Flores.

Director:

Ing. Arturo Geovanny Peralta Sevilla. MSc.

CUENCA, FEBRERO DE 2015

CERTIFICO:

En calidad de DIRECTOR DE LA TESIS ***"APLICACIÓN DE LAS TÉCNICAS DE AGRUPAMIENTO PARA LA DISTRIBUCIÓN CUASI–ÓPTIMA DE UNA RED HÍBRIDA WDM–TDM/PON EN CASCADA MULTINIVEL QUE DA SOPORTE A UNA SMART GRID O SMART CITY"*** elaborada por Milton Amoroso Ordoñez, Santiago Ávila Flores, declaro y certifico la aprobación del presente trabajo de tesis basándose en la supervisión y revisión de su contenido.

Cuenca, Febrero de 2015

Ing. Arturo G. Peralta Sevilla. MSc.

DIRECTOR DE TESIS

DECLARATORIA DE RESPONSABILIDAD

Los autores declaramos que los conceptos desarrollados, análisis realizados y las conclusiones del presente trabajo, son de nuestra exclusiva responsabilidad y autorizamos a la Universidad Politécnica Salesiana el uso de la misma con fines académicos.

A través de la presente declaración cedemos los derechos de propiedad intelectual correspondiente a este trabajo a la Universidad Politécnica Salesiana, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

Cuenca, febrero del 2015

Milton Santiago Amoroso Ordoñez

C.I. 0302083621

Hernán Santiago Ávila Flores

C.I. 0105243703

DEDICATORIA

Dedico esta tesis a mi familia de manera especial a mis padres quienes me apoyaron todo el tiempo. A mi novia María Cristina quien me apoyo y alentó para continuar, cuando parcia que me iba a rendir y a todos los que me apoyaron para escribir mi tesis.

Para ellos es esta dedicatoria de tesis, pues es a ellos a quienes se la debo por su apoyo incondicional.

AGRADECIMIENTOS

Mis agradecimientos van hacia mi familia y demás personas por su apoyo incondicional durante este largo camino, especialmente a mis padres ya que sin su apoyo en todo sentido sería imposible haber cumplido esta meta y a todos quienes de una u otra manera han sido un pilar fundamental para la realización de este documento de manera especial a nuestro director de Tesis el Ing. Arturo Geovanny Peralta Sevilla. MSc por su constante apoyo en el desarrollo del mismo ya que sin él hubiera sido imposible este logro.

Milton Amoroso Ordoñez.

DEDICATORIA

La culminación de este trabajo dedico de manera especial a mis padres que fueron el eje principal para luchar y seguir preparándome a lo largo de todo este trabajo.

AGRADECIMIENTOS

Al haber culminado este proyecto en primer lugar quiero agradecer a DIOS por brindarme la voluntad y conocimiento para seguir en un constante crecimiento personal, a mis Padres y a mi Director de Tesis por ser un apoyo e impulso importante, y también a mi compañero de tesis por su colaboración completa para el desarrollo de este proyecto, y que trabajando como equipo logramos culminar el mismo de una manea enriquecedora y productiva para nuestras futuras vidas personales y profesionales.

Santiago Ávila Flores.

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

PROLOGO

En el siguiente documento se analizara un nuevo concepto en la aplicabilidad del diseño de despliegue de red enfocándose en una red multiservicios bajo una plataforma WDM--TDM.

Partiremos de un estado de arte acerca de Redes PON que estará descrito en el capítulo uno, posteriormente en el capítulo dos se analizara la robustez y estabilidad de las técnicas de agrupamiento para una distribución de Red enfocándonos en tres métodos distintos que solventen estas técnicas. En el capítulo 3 se planteara una breve descripción de los modelos matemáticos que son base analítica de los métodos utilizados y que son un complemento para la nueva solución algorítmica planteada. En el capítulo cuatro se enfocara en la aplicabilidad de esta herramienta, plasmándola un escenario real de la ciudad de cuenca, donde se obtendrán conclusiones técnicas y económicas, que brindaran aplicaciones y proyecciones futuras en los diseños de despliegues de red. Finalmente en el Capítulo 5 se darán las conclusiones y del resultado del trabajo

CAPITULO 1. ESTADO DEL ARTE

Introducción

El constante crecimiento en Telecomunicaciones ha permitido y desarrollado la evolución de nuevos protocolos, métodos y procesos de transmisión de datos. En los últimos años, la sociedad de la Información ha experimentado un rápido desarrollo, debido, en gran parte, a la mayor competitividad impulsada por la regulación del Mercado de las Telecomunicaciones y a la aparición de nuevos servicios de banda ancha. El resultado de estos dos factores se ha traducido en una necesidad de disponer mejores redes de comunicaciones capaces de ofrecer un mayor ancho de banda a un menor costo, siendo en la actualidad la más predominante en Ecuador la tecnología ADSL, y es la que sigue explotando la última milla de abonado en cobre.

Por otro lado, la demanda cada vez mayor de usuarios móviles, fijos, y en adición la inserción de tráfico de datos por parte de servicios que están modernizando las redes eléctricas como son las Smart Grid, o brindando mayor inteligencia a ciertos servicios urbanos como es el manejo de flujo vial y alumbrado público mediante el concepto de Smart Cities, ha hecho replantear a los operadores consolidados y emergentes sus estrategias, comenzando una carrera por el crecimiento de la velocidad en sus backbones, Sin embargo ADSL cuenta con una limitación técnica importante, y es el ancho de banda que puede ofrecer, el mismo que disminuye drásticamente a medida que el usuario se aleja del punto de acceso.

En este sentido, la tecnología de la fibra óptica se presenta como una firme solución al problema gracias a la robustez, a su potencial ancho de banda amplio y a la continua mejora de sus características técnicas y descenso de los costes asociados a los dispositivos involucrados en tales redes.

Si consideramos que hoy (nuevos sectores de vivienda , centros comerciales) ya integran cableado estructurado de fibra óptica por su bajo coste marginal en el proyecto, estamos hablando de un escenario muy apto para poder desplegar soluciones de conectividad en fibra óptica, que directamente lleguen hasta el usuario donde cada vez es más fuerte la necesidad de obtener un **multiservicios** que adapte

estas nuevas plataformas de tal manera que solvente las distintas y escalables necesidades de los mismos.

Desde su aparición la fibra óptica ha sido un medio muy sólido para encaminar las comunicaciones y transmisiones de datos en distintas aplicaciones, ***PON (Optical Network Pasive)*** es una red que representa muy significativamente las ventajas de utilizar fibra y permite opcionar una solución al dimensionar una estructura de red multiservicios deseada, además que el crecimiento en los costos y consumo energético conlleva la configuración de los sistemas eléctricos de potencia actuales, que dirigirá la evolución del sistema de distribución de energía eléctrica hacia *Smart Grid y Smart City* , brindando mayor inteligencia al sistema (infraestructura de medición avanzada), integrando las tecnologías de la información y la comunicación (modelos de arquitecturas de telecomunicaciones), ofreciendo un mayor aprovechamiento de las fuentes de energía alternativa (generación distribuida – micro redes), una más rápida respuesta a los fallos y una adaptación de cara a la evolución de *Smart Grid y Smart City* en los países de la Región Andina [1]

El proyecto que se plantea, se enfoca en la utilización de una **RED MULTISERVICIO,** basada en una topología hibrida WDM/TDM soportada en un sistema computacional que facilite su despliegue y solvente de mejor manera las aplicaciones actuales y futuristas de Smart Grid y Smart City, no obstante, para ello es necesario mencionar algunas consideraciones básicas dentro de este formato que serán útiles para entender desde su base la temática planteada.

Redes PON

Una red PON es una Red Pasiva Óptica de última generación, a pesar de que hoy en día ya se cuenta con despliegues de redes ópticas pasivas con capacidad de Gigabit (Gigabit–capable Passive Optical Network, GPON) dentro de la ciudad, es importante referirnos a PON como base de red a analizar.

Los principales componentes Ópticos Pasivos en una PON, de acuerdo a [2] son:

- Acoplador 1×2 Multiplexador de Longitud de Onda Amplia (Wide Wavelength Division Multiplexer, WWDM)

- Splitter 1xN.
- Cables de fibra óptica (alimentadores, distribuidores y acometida).
- Conectores y cables de ensamble.
- Sistemas de administración de fibra / cajas de empalmes.

A los compoentes mencionados hay que sumar la central de trasmisión denominada Terminal de Línea Óptica (Optical Line Terminal, OLT) y su destino final de llegada denominado Unidad de Red Óptica (Optical Network Unit, ONU), indispensables en la red, sin embargo, los dos componentes críticos son el acoplador y el splitter [2].

El acoplador 1×2 WWDM usado en una PON es bidireccional y típicamente tiene una pérdida de inserción de 0.7 a 1.0 dB. Este componente es usado de manera controlada dentro de la central y de manera no controlado dentro del ONU y tiene por responsabilidad las siguientes funciones:

- Acoplar la señal de alta potencia de vídeo con la señal de voz/datos del OLT para la transmisión de bajada en la red.
- Acoplar las señales de video con las señales de voz en el OLT.
- Acoplar la señal voz/dato para la transmisión de subida y bajada.
- Transmitir las señales de subida de voz/dato hacia el receptor en el OLT

El Splitter en cambio es un elemento divisor óptico usado en las PON, el mismo tiene una sola entrada y múltiples salidas. La señal de entrada óptica se divide en una cantidad de puertos de salida, permitiendo que distintos usuarios compartan una sola fibra y en consecuencia compartan una longitud de onda. Estos elementos son pasivos por no depender de ninguna fuente de energía externa más que la señal óptica de entrada. Además son independientes de la longitud de onda y solamente incorporan una atenuación debido a que se divide la potencia de entrada. Esta pérdida es llamada radio de acoplamiento o perdida de inserción, expresada en (dBs) y va de la mano con el número de puertos o relación de Split tal como se muestra en la TABLA 1.

TABLA 1. ATENUACIÓN DE SPLITTER [2]

Relación de Split	Pérdida de inserción (dB)
1:2	3,6
1:4	7,2
1:8	11
1:16	14
1:32	17,5

1.2.1 Arquitectura de una Red PON

Una red PON está formada como se mencionó anteriormente por el OLT situado en la Oficina Central (Central Office, CO), uno o más separadores/acopladores ópticos distribuidos a lo largo de la red de acceso (optical splitter/coupler) y un conjunto de terminales de red ópticos (OLT/ONU) conectados a un separador óptico mediante fibra. De estos elementos, los separadores ópticos son pasivos, no requieren alimentación y están ubicados fuera de la planta, siendo estos, donde proviene el nombre de la tecnología PON.

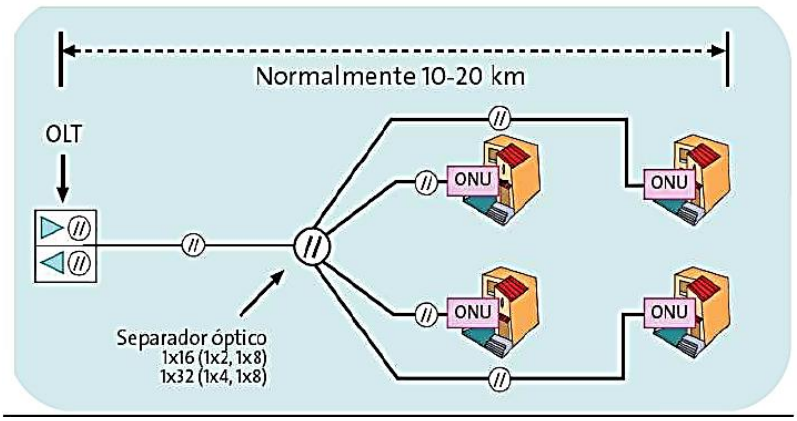

Figura 1.Arquitectura de una Red PON

Con los elementos citados, está red viene a representar una tecnología tipo punto–multipunto cubriendo distancias de 10 a 20 Km. Todas las transmisiones en una red PON se realizan entre OLT. Habitualmente la unidad OLT se interconecta con una red de transporte que recoge los flujos procedentes de varias OLTs y los encamina a la cabecera de la red. La unidad ONU, se ubica en el domicilio de usuario,

configurando esquemas como son fibra hasta el usuario (Fiber To The Home, FTTH), fibra hasta las edificaciones (Fiber To The Building) y fibra hasta las cabinas o armarios de distribución (Fiber To The Cabinet, Fiber To The Curb, FTTC), que conforman despliegues conocidos como FTTx.

Para poder soportar toda la demanda de servicio requerido es necesario que la Red PON funcione bajo ciertos protocolos que optimizaran al máximo el uso de la misma, y es así que en [3] plantea que la tecnología sugerida para la infraestructura de telecomunicaciones de una red de acceso multiservicio es **TDM–PON** (Time Division Multiplexing–Passive Optical Network) integrada o complementada con tecnologías inalámbricas como 4G (Long Term Evolution, LTE y WiMAX), el constante crecimiento del ancho de banda fijo y móvil requerirá el uso de tecnologías ópticas pasivas de próxima generación **WDM (Wave length Division Multiplexing)–PON**, así mismo una alternativa intermedia entre TDM–PON y WDM–PON es la arquitectura híbrida **WDM/TDM– PON** que facilitara la actualización de la red a medida que los usuarios requieren más capacidad.

1.2.2 Características y Clasificación de Redes PON

Haciendo mención a como se ha presentado la evolución en el trayecto del tiempo de las redes PON, esta arquitectura nace en los laboratorios Telecom en el año de 1990 en Nueva York, tras buscar tecnologías que amplíen los anchos de banda y además se pueda usar como medio de transmisión la fibra óptica, y tomando como pila de protocolos los establecidos para Ethernet algunas características importantes se las puede encontrar en [4], citamos algunas de ellas:

- Aumento hasta 20 Km
- Mayor ancho de banda
- Debido a la inmunidad ante ruidos electros magnéticos mejora la calidad del servicio y la simplificación de la red.
- Minimiza el despliegue de fibra óptica gracias a su topología
- Abarata equipos y costos debido a su simplificación de red

- Mucho más efectiva que una red punto a punto

No obstante, este fue el augurio para el despliegue de sub redes nacientes de PON tomando funcionalidades de la misma y mejorándolas o enfatizándose en la versatilidad, es así que en **2001**, se presenta **Broadband PON (BPON)** como una nueva alternativa acoplando una longitud de onda para trasportar video bajo un protocolo TDM.

En [4], **Ethernet PON (EPON)** estaba definida en **2004** por el grupo EFM (Ethernet First Mile) del IEEE como la técnica PON de nueva generación que, influenciada por la tecnología Gigabit Ethernet existente residía bajo un protocolo TDM y permitía a los suministradores de equipos lanzar rápidamente al mercado equipos de mayores anchos de banda a precios más competitivos. No obstante, EPON carecía de muchas funcionalidades necesarias para el transporte de otros servicios con calidad de operador que daban lugar a soluciones propietarias. En el **2010** tras aplicaciones con fines militares en USA se crea **GPON** como una clasificación de PON muy robusta, con una menor sobrecarga de codificación y tiempos de guarda menores, el ancho de banda neto de GPON es mucho mayor que el de EPON salvo por su implementación de red, representa un poco más compleja.

En la TABLA 2 se expone en un resumen la derivación e evolución de redes PON contemplando sus factores principales que las caracterizan.

TABLA 2. CLASIFICACIÓN DE REDES PON [5]

Características	BPON	EPON	GPON
Fecha de inicio	2001	2010	2004
Tasa de bits(Mbps)	Down 1.244;622;15 Up 622;155	Down 1.250 Up 1.250	Down: 2.488;1.2 Up:2.48;1.244;622;155
Codificación de línea	NRZ+(Scranbling)	8b/10b	NRZ+(Scranbling)
Radio de división máxima	1:32	1:32	1:32 1:64 en la practica
Alcance máximo	20 km	20km	60km
Estándares	Serie ITU-TG.983.x	IEEE 802.3 ah	Serie ITU.G.984.x
Soporte TDM	TDM sobre ATM	TDM sobre paquetes	TDM nativo
Soporte de Video RF	No	No	Si
Eficiencia Típica	83% downstream 80% upstream	61% downstream 73% upstream	93% downstream 84% upstream

Como mencionamos anteriormente una red PON ha sido una secuencia de evolución que adapta en su estructura protocolos de transmisión y es este factor lo que permite brindar diferentes topologías, las cuales a través de estos protocolos ya sea TDM o WDM canalizan de la manera más idónea y necesaria el servicio al usuario.

1.2.3 Topología de una Red PON

Existen varios tipos de topologías adecuadas para el acceso a red, incluyendo topologías en anillo (no muy habituales), árbol, árbol–rama y bus óptico lineal. Cada una de las bifurcaciones se consiguen encadenando divisores ópticos 1x2 o bien divisores 1xN donde N es siempre un múltiplo de 2.

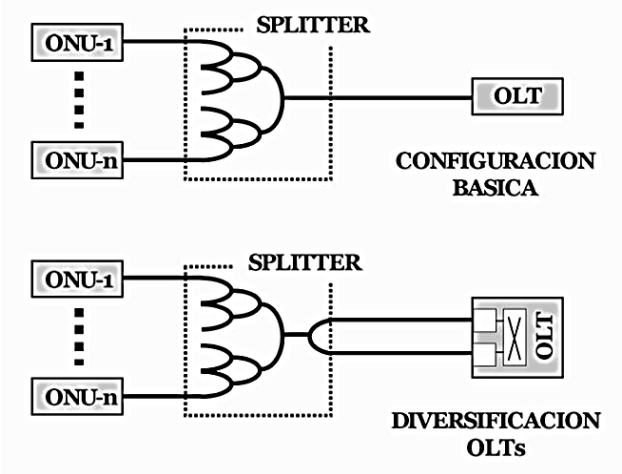

Figura 2.Topología de una Red PON [4]

Todas las topologías PON utilizan monofibra para el despliegue un ejemplo de ello se lo puede ver en la Figura 2.

En canal descendente una PON es una red punto multipunto. El equipo OLT maneja la totalidad del ancho de banda que se reparte a los usuarios en intervalos temporales. En canal ascendente la PON es una red punto a punto, donde múltiples ONUs transmiten a un único OLT.

Adecuando como medio de transmisión monofibra, la manera de optimizar las transmisiones de los sentidos descendente y ascendente sin entremezclarse consiste en trabajar sobre longitudes de onda diferentes, utilizando técnicas WDM (Wavelength Division Multiplexing).

Cabe mencionar que la mayoría de topologías de red PON están compuestas por dos o tres splitters en cascada y estos a su vez son colocados de manera que la transición de datos es secuencial por etapas bajo un protocolo de envió donde si es TDM se hará en secuencias de tiempo y si es WDM en intervalos de frecuencias, pero de ambas formas cualquiera que esta sea la intención es obtener una topología segura y versátil, diferenciándose obviamente por características de cada protocolo [4].

1.3 Redes TDM/PON

El protocolo TDM (Time División Multiplexing) representa un soporte muy útil para una red PON y ha venido siendo muy utilizado en los despliegues de red para estos últimos años aprovechando el ancho de banda del medio de transmisión.

Este protocolo funciona sobre una red de fibra enviando los datos en determinados intervalos de tiempo, multiplexando cada despliegue de bits desde la OLT o CO hacia el splitter donde posteriormente se destinara cada usuario (ONU).

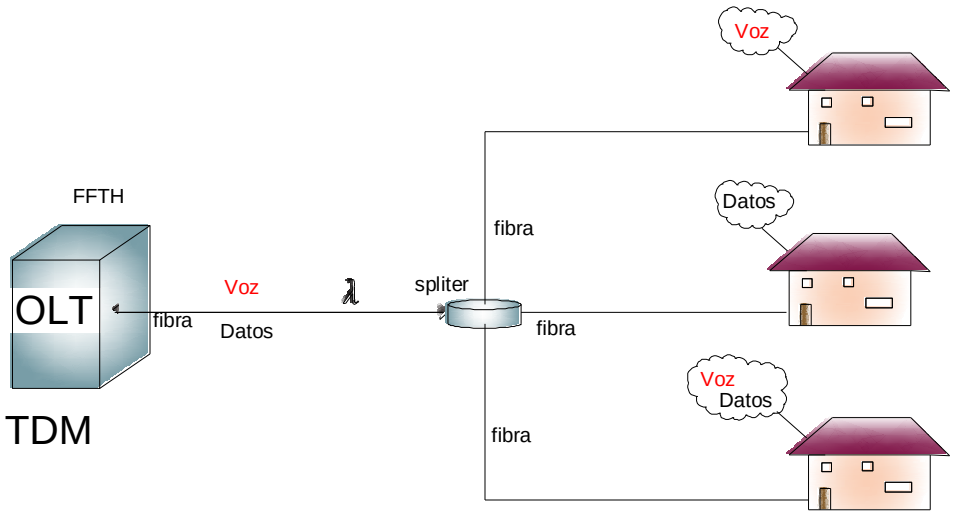

Figura 3. Red TDM-PON

Este proceso de transmisión sobre un sistema óptico de fibra y al manejar técnicas de multiplexación puede ofertar algunas ventajas:

- Brindar todo ancho de banda a un solo canal en un intervalo de tiempo.
- Aumento en velocidad dependiendo de la cantidad de ONUs.
- Control accesible de la Red.
- Fácil implementación y despliegue de red.

1.3.1 Arquitectura y Características de una Red TDM/PON

Al entender lo antes mencionado se podría decir que TDM oferta ventajosas pero limitadas características dependiendo de los servicios y la cantidad de los mismos que sean demandados por los usuarios, pese a ello este protocolo ha sido utilizado desde las redes de cobre hasta hoy en día en despliegues de fibra, convirtiéndose en una arquitectura **punto–multipunto,** permitiendo la utilización de servicios Ethernet e IP, con una básica arquitectura a entender representado en la Figura 4.

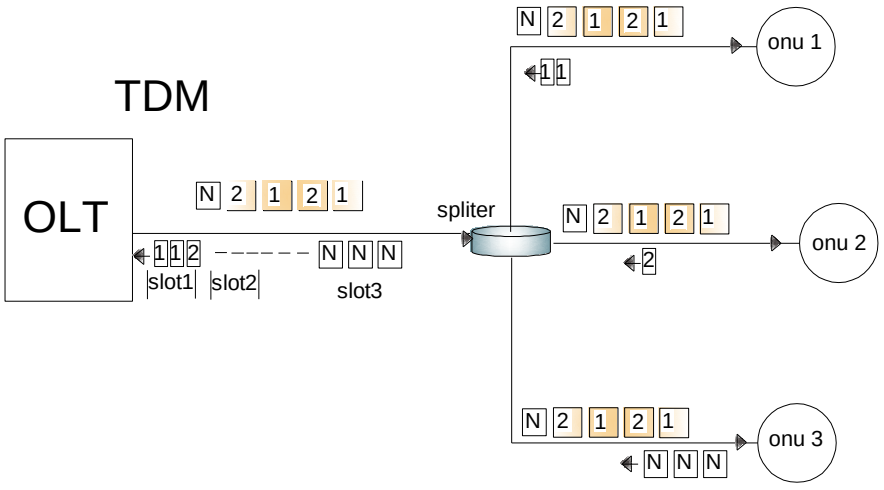

Figura 4. Arquitectura TDM–PON

Como se puede apreciar en la figura anterior, esta arquitectura permite la transmisión a través de un solo canal (dado por la longitud de onda λ), el mismo que a través de

un elemento óptico (splitter), subdistribuye este canal, eso significa que el ancho de banda total del canal, la velocidad y la distancia de cobertura estarán divididos para todos los abonados que se encuentren activos, ahora dependiendo de los equipos utilizados esta distribución será mayor o menor, aunque por lo general vienen para ratios de división entre 1:32 o 1:64, ya que en algún momento todo el ancho banda será completo o dividido para cada usuario, de allí que TDM brinda **ventajosas pero limitadas** características, sin embrago no deja de ser una estructura de red solida fiable y robusta para aplicaciones de Redes Ópticas (PON). [6]

1.3.2 Revisión de la propuesta hacia Smart Grid y Smart City

Basándose en lo que se ha estado desarrollando anteriormente, se puede ya empezar a relacionar el contexto de TDM–PON con la utilidad hacia Smart City que obviamente engloba Smart Grid. El manejo de estos dos términos ha ido en crecimiento a la par con el uso de la fibra y es que hoy en día el termino **red inteligente** es sinónimo de evolución y solución tecnológica, pues la necesidad de nuevos servicios y la manera de controlarlos ha dado origen una nueva era de redes de comunicación.

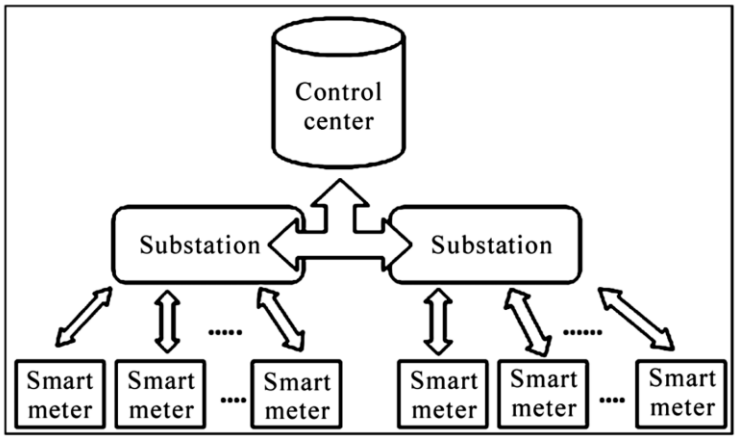

Figura 5.Diagrama de Smart Grid (control) [7]

La propuesta de utilizar un protocolo TDM para una red inteligente es plasmada por el crecimiento de las redes pasivas ópticas que han estado en constante ampliación y es así que, redes GPON, EPON desde ya algún tiempo son topologías fundamentales

en despliegues urbanísticos modernos donde cada vez el servicio de voz, datos y procesos vía IP son más requeridos en los usuarios, en adición se han soportado inicialmente sobre estas redes servicios de Redes Inteligentes como son: Lectura de Medidores (luz, agua, gas) Automatizada *(Automated Meter Reading, AMR)* y en estos últimos años ha evolucionado hacia el concepto de Infraestructura de Medición Avanzada *(Advanced Metering Infrastructure, AMI)*, Automatización en Alimentadores Primarios de la red de energía *(Automation Feeders, AF)* en colaboración con Redes Inalámbricas de Sensores *(Wireless Sensor Networks, WSN)*, Generación Distribuida mediante el concepto de mircrorredes y subestaciones virtuales en el despacho de energía, y cargas adicionales de energía y datos como son los Vehículos Híbridos y los Vehículos Eléctricos Conectados *(Plug–In Electric Vehicle, PEV)*, estos conceptos se muestran en la Figura 6.

En [7] podemos ver que el objetivo de una Smart Grid y Smart City se basa en el control y administración de red a tal punto que esta red pueda funcionar con una alta fiabilidad, donde TDM realmente es una alternativa útil en Smart Grid, ya que servicios como voz y datos, a través de procesos multeplexibles en el tiempo que son cubiertos de una manera muy sólida, sin embargo, Smart Grid es el preámbulo para una ciudad inteligente donde todos los servicios amplían su necesidad en **ancho de banda** y estándares de trasmisión como es comunicación **IP,** sin contar el obvio crecimiento de usuarios, ante este escenario TDM–PON lastimosamente está limitado debido a su arquitectura y características de red especialmente al ancho de banda existente, no obstante, en la actualidad ya contamos en algunas ciudades del país con redes TDM–PON; pero la directriz objetiva que es Smart City, requerirá nuevos protocolos como WDM o la combinación de los dos TDM–WDM (también conocido como TWDM), los mismos que se mencionaran en el posterior desarrollo de este documento.

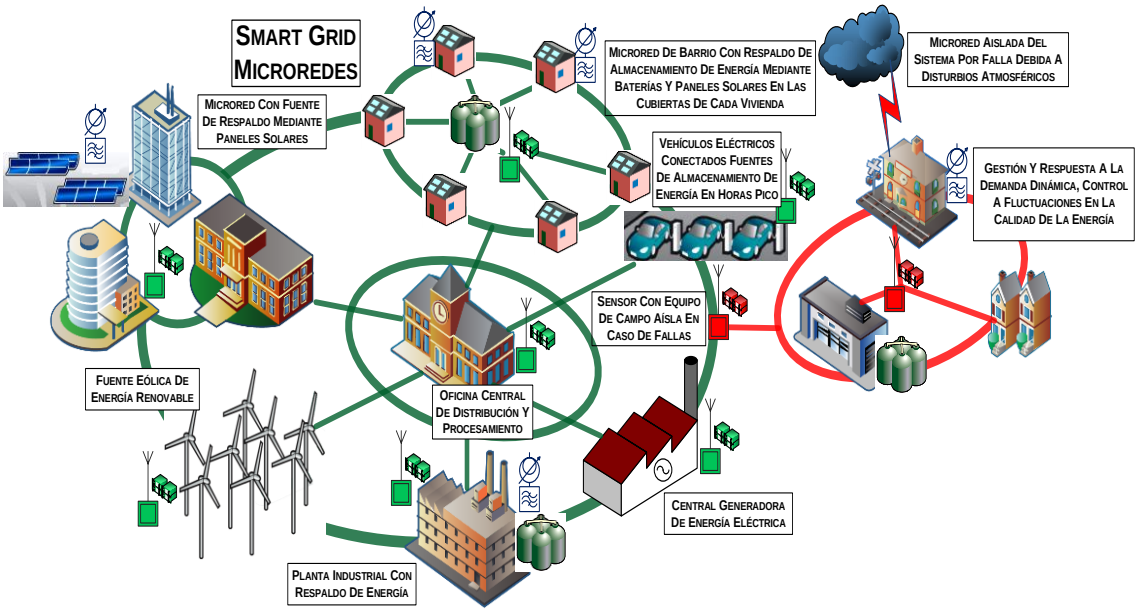

Figura 6. Representación de una Smart Grid y sus componentes [1].

1.4 Redes WDM–PON

WDM–PON es de propósito general y una tecnología extremadamente eficiente del futuro óptico para uso de redes de transporte de acceso y redes Metro. Esta permite altamente el uso eficiente de una planta de fibra externa para proveer una conexión óptica punto a punto a múltiples localidades remotas a través de un único alimentador de fibra.

La necesidad de un ancho de banda no compartido, y la aparición de nuevos servicios y protocolos IP como se mencionó anteriormente hacen que WDM–PON se establezca como nueva plataforma de red en los años venideros.

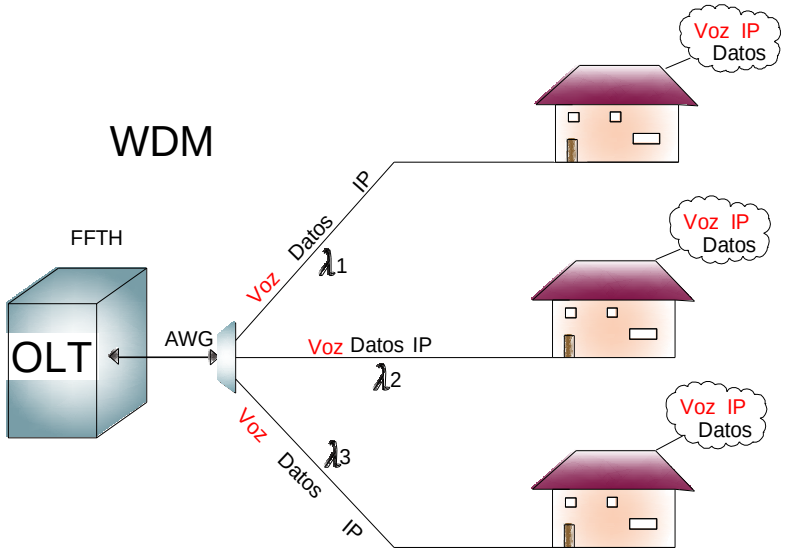

Figura 7. Red WDM/PON

La funcionalidad básica de WDM se ilustra en la Figura 7. Una conectividad óptica punto a punto dedicada a "n" sitios remotos requiere "n" transmisores *(canal dado por longitudes de onda* ***λs)*** tanto como para oficina central y para la ONU remota. En una arquitectura punto a punto convencional, esta funcionalidad es con frecuencia aprovechada usando "2n" alimentadores de fibra como se muestra en la Figura 4.

En la arquitectura WDM PON esos "2n" transmisores son conectados por un solo alimentador óptico por medio del uso de multiplexación y demultiplexación de DWDM.

En realidad, la solución WDM ha estado en el mercado desde el 2005; pero no es sino hasta el 2011 donde la ITU a través de su referencia TG.698.3, da a conocer su aplicabilidad total como una arquitectura sostenible apoyada en su ancho de banda y escalabilidad contemplando aplicaciones de video en alta definición y operaciones de redes inteligentes [8]. Esta arquitectura se plasma de la siguiente manera.

1.4.1 Arquitectura y Clasificación de una Red WDM-PON

En la Figura 8, se indica la arquitectura en una Red **WDM–PON** y el tráfico de paquetes de información:

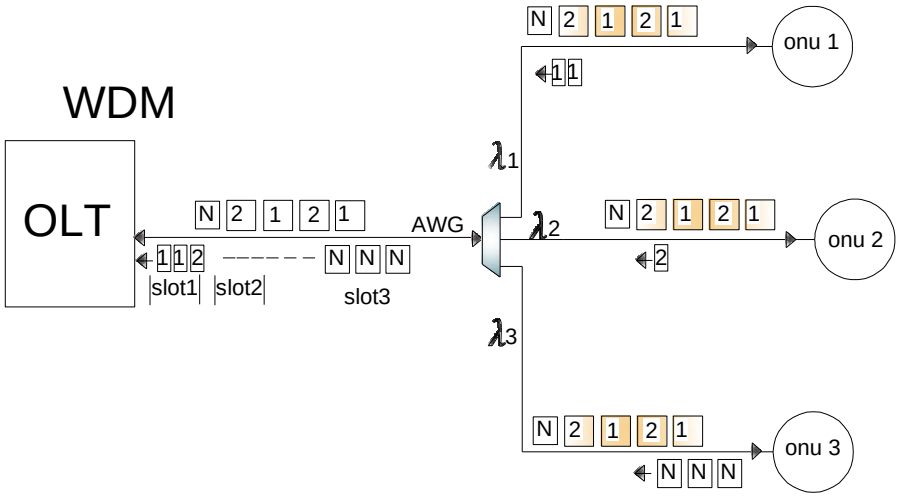

Figura 8. Arquitectura WDM/PON

En [9], se menciona la arquitectura de la Figura 8 como elemental para WDM donde el enrutador óptico de longitudes de onda conocido como *(Arrayed Waveguide Grating, AWG)* posee guías de onda formadas en el substrato que se conectan con fibras para permitir la interconexión del AWG a otros dispositivos y equipos ópticos. Básicamente la función de este quipo es separar las longitudes de Onda en la entrada y distribuirlas en los diferentes puertos de salida, pudiendo emplearse en sentido contrario, es decir, puede actuar como multiplexor o demultiplexor.

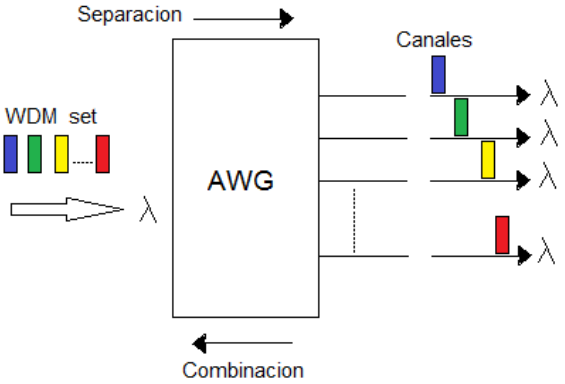

Figura 9. AWG [7]

En una red PON el AWG, se encarga que por cada canal utilizado, se genere una longitud de onda de transmisión λ, la misma que tiene una ancho de banda

individual, no obstante es posible la interconexión a un Splitter si se desearía una subdistribución. Esta arquitectura es muy versátil y permite la aplicación de múltiples servicios o combinación de ellos que demanden requerimientos con un bajo jitter, bajo retardo y data rate, con un gran ancho de banda; una ventaja en los actuales tiempos es que cada vez están disminuyendo los costos de los equipos activos en capa física como son los láseres, fotodiodos, moduladores y demoduladores.

Al discernir la funcionalidad de este protocolo y refiriéndonos a [6], [7], [8] podemos plantear algunas características que catalogan a WDM/PON como una herramienta amplia para redes de fibra:

- Ancho de banda independiente.
- Mayores distancias de cobertura.
- Aplicación de servicios de alta resolución.
- Control y administración de red más precisa.
- Alta adaptabilidad en redes.
- Mayor nivel de escalabilidad.
- Aplicaciones satisfactorias en redes inteligentes.

Como vemos las características de esta red son buenas, sin embargo, el coste tanto en despliegue como equipos WDM retardan la decisión por elegir este protocolo (todavía no se tiene un estándar a nivel mundial para fabricantes de plataformas WDM, DWDM o UDWDM), no obstante y en la actualidad grandes capitales utilizan ya esta arquitectura con la obvia proyección hacia redes inteligentes. En el Ecuador existen ya proyectos basados en WDM/PON y aunque son netamente teóricos se intentan enlazar con las redes TDM/GPON ya establecidas buscando generar una alternativa para la inevitable expansión de servicios.

1.4.2 Revisión de la propuesta hacia Smart Grid y Smart City

El soporte que brinda WDM como hemos visto en las secciones anteriores permite que el enfoque de aplicabilidad hacia Smart Grid y Smart Cities sea idóneo a diferencia de TDM, es decir que una red inteligente, conviene el uso de arquitecturas soportadas sobre protocolos por multiplexación mediante longitudes de onda, en [9] se plantea la optimización en el uso de este protocolo mediante la intercalación de AWG, donde se realiza la agregación o separación de longitudes de onda con anchos de banda desde la OLT hacia las ONUs, estableciendo de esta forma aplicaciones de alta definición en voz, video y datos.

En [10] se mira una comparativa entre Smart Grid con el Internet, la diferencia radica en que el Internet no está disponible en todas las casas, y la intención de Smart Grid es estar siempre y para todos con alta fiabilidad a tal punto de poder controlar la congestión y demanda de tráfico en todos sus servicios adjuntos, esto se esquematiza en la Figura 10.

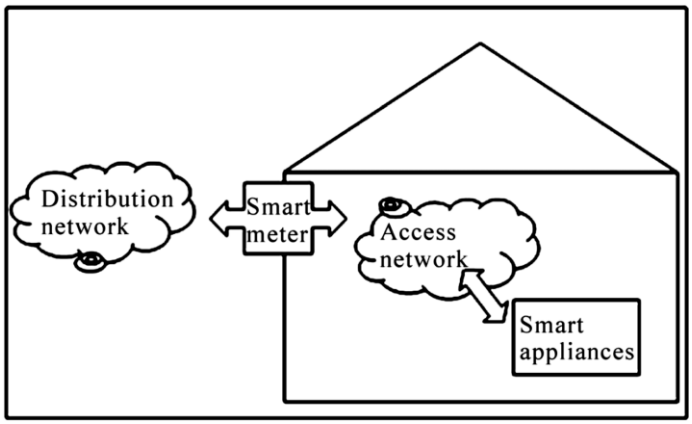

Figura 10. Aplicación Smart Grid [8]

Smart Grid y Smart Cities son redes que demandan muchos servicios y mucha calidad, WDM colaboraría con estas necesidades entregando bandas de trasmisión pasivamente, si bien es cierto la red estaría dada por topología WDM desde la OLT, en algunos casos dependiendo del servicio, la distancia y la cantidad de usuarios sería conveniente la utilización de un divisor óptico u splitter de tal manera que al dividir la potencia se obtiene cuotas en el tiempo lo que crearía una red mucho más

administrable e inteligente [9], al realizar esto estamos ya planteando una red hibrida que permite ya por completo gestionar aplicaciones de Smart City direccionándonos a un modelo de red altamente sostenible y futurista.

1.5 Redes Hibridas en Cascada Multinivel WDM–TDM/PON

Como hemos estado revisando, TDM y WDM por separado funcionan para una RED PON, sin embargo un sistema completo y más global genera la necesidad de fusionar las mismas obteniendo así una red multiservicio, lo que conlleva al planteamiento inicial de este proyecto, teniendo la visión de dar una alternativa de migración escalable en el tiempo que de paso a la transición de redes TDM a redes WDM o DWDM.

Considerando que una red de acceso multiservicio debe integrar diferentes tecnologías y dispositivos de usuario final incluidos los sensores y actuadores de las redes inteligentes que satisfacen aplicaciones como AMR, AMI, y Sistemas de Gestión de la Energía (Energy Management Systems, EMS).

La Figura 11 expresa gráficamente lo mencionado, la intención es llegar a tener una red que permita brindar servicios de alta calidad, optimizando el despliegue de fibra y la funcionalidad óptima de los equipos para cada destino de usuario.

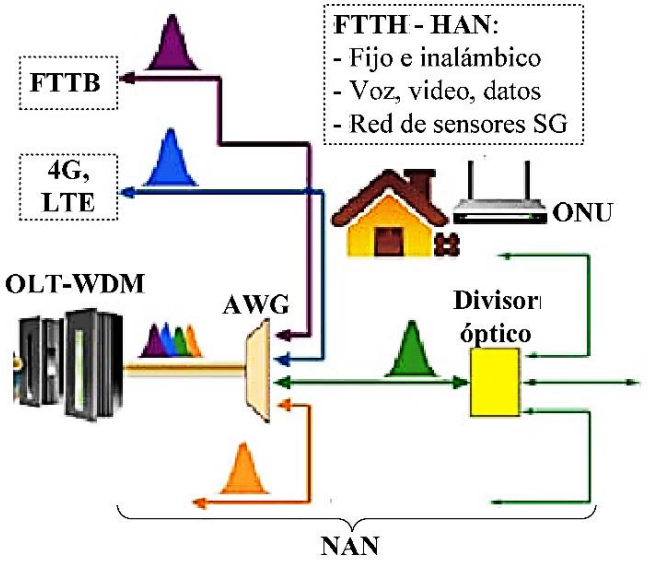

Figura 11. Red hibrida WDM–TDM/PON [3].

La topología de una red de acceso multiservicio basada en la arquitectura híbrida WDM/TDM–PON espera incrementar los niveles de seguridad, retardo, escalabilidad y disponibilidad en las redes de acceso multiservicio de próxima generación. Ésta arquitectura facilita la actualización insertando longitudes de onda a medida que los usuarios requieren más capacidad, manteniendo las ONUs de TDM–PON [3].

En la Figura 11 se identifican varios segmentos como son HAN (Home Area Network), NAN (Neighborhood Area Network) y WAN (Wide Area Network). El segmento HAN incluye los dispositivos al interior del hogar. El segmento NAN se asocia a la red de acceso multiservicio que agrega la información proveniente de los/las redes de sensores y actuadores dentro de cada HAN. La conexión del OLT–WDM hacia la red metro o WAN hace parte del segmento WAN [11].

1.5.1 Arquitectura y Características de una Red Hibrida WDM–TDM/PON

Para describir la funcionalidad de esta arquitectura puede citar a [3], donde plantea que en WDM–PON el OLT (Optical Line Terminal), ubicado en las instalaciones del proveedor del servicio, genera en el canal descendente varias longitudes de onda empleando WDM. Esto a diferencia de lo que ocurre en TDM–PON, en donde el OLT genera una longitud de onda en el canal descendente que es compartida en el tiempo entre varias ONUs ubicadas en los predios de los usuarios., Mientras TDM–PON está estandarizada, aun no se define un estándar para WDM–PON. En TDM–PON se tienen dos principales alternativas estandarizadas Gigabit Ethernet–PON (GE–PON) y Gigabit–PON (G–PON), ambas con hasta 20 km de longitud máxima de fibra óptica, con 64 ONUs por OLT y con tasas de bits por ONU entre 10–1000 Mbps. La Figura 12, puede describir una arquitectura **WDM/TDM PON,** con todos sus componentes importantes.

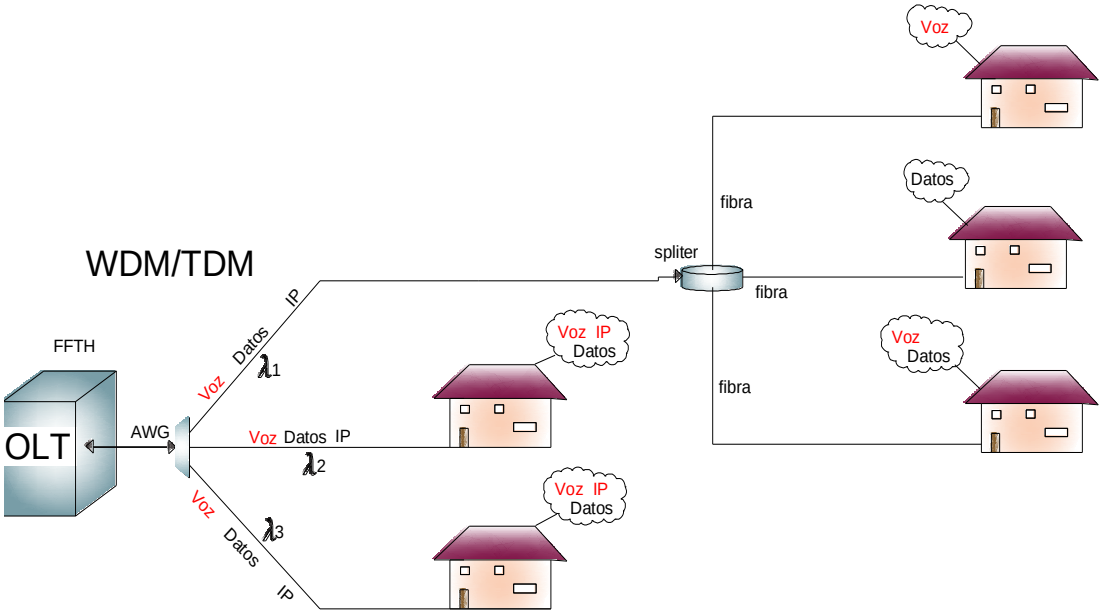

Figura 12. Arquitectura Red Hibrida WDM–TDM/PON

El AWG (Arrayed–Wave guide Grating) separa las longitudes de onda dirigidas hacia los ONUs en el canal descendente y agrega las longitudes de onda provenientes de los ONUs en el canal ascendente. En WDM–PON la señal del AWG va directamente al ONU. Sin embargo, una arquitectura híbrida ofrece la posibilidad de compartir en el tiempo una longitud de onda entre varios ONUs empleando un divisor óptico.

La solución presentada permite llegar con una conexión de fibra al hogar (FiberToThe Home, FTTH), al edificio (Fiber to the Building, FTTB) y servir como backhauling a tecnologías inalámbricas móviles como 3G, 4G, y futuramente a 5G (Ondas Milimétricas). Para el caso de FTTH, la ONU ofrece servicios a dispositivos fijos, móviles y a la red de dispositivos y sensores de Smart Grid.

1.5.2 Revisión de la propuesta hacia Smart Grid y Smart City

El enfoque de implementar una red hibrida, soporta de excelente manera las aplicaciones de Smart Grid y Smart City, esto se debe a la completa aplicabilidad que tendría la red. Una red inteligente se enfatiza en el control y seguridad de la red a tal punto que la transmisión sea óptima desde su inicio hasta su destino, es decir que su

ancho de banda, velocidad, retardo y precisión sean iguales en toda la red, esto asegura al usuario un servicio con calidad y confiabilidad ,es muy común encontrar y darse cuenta que en las redes actuales la claridad de la comunicación es mejor en algunos sectores a diferencia de otros, por ejemplo un especial énfasis se presenta con el internet donde se utiliza una banda compartida. Al proponer un modelo hibrido TWDM corregirá ese factor al entregar una sola longitud de onda a cada sector así mismo TDM entregara un solo slot de tiempo a cada usuario, es decir se corrige y optimiza la utilización adecuada de los recursos de red lo que en contextos económicos representa un gran ahorro en despliegue de fibra canalizaciones y equipos pasivos, este tema los analizaremos con más detalle en el tercer capítulo.

Smart City que es el futuro, demanda una buena administración de una Smart Grid, todos los servicios como internet, aplicaciones de video, sonido, voz y datos deberán ser óptimos en todo el proceso de comunicación, una red hibrida brinda un entorno seguro y fiable de optimidad, en un futuro podríamos disfrutar de videos en ultra alta definición en tiempo real, video bajo demanda, voz mediante formatos IP, aplicaciones incluso vía Smartphone y todo bajo un único sistema de banda propio para cada usuario es decir tendríamos velocidad, capacidad y aplicabilidad sin perjudicar a los demás.

Pero las ventajas no son solo desde el usuario así mismo las empresas de telecomunicaciones facilitarían sus procesos de control y ampliarían sus procesos de aplicación, todo dirigiéndose a un resultado óptimo versátil y ahorrativo que van de la mano con el crecimiento tecnológico se establecerá como una tecnología sólida y solvente para el campo de telecomunicaciones, siendo el escenario inicial para posteriores metodologías, modelamientos de optimización matemáticos, técnicas de agrupamiento y heurísticas que reemplazaran tediosos y obsoletos procesos en despliegues de red.

CAPITULO 2. ANÁLISIS DE LOS MÉTODOS DE AGRUPAMIENTO

2.1 Definición de Clúster

Dar una definición universal de "clúster" no es posible debido a que cada interpretación dependerá de la aplicación a la cual se le asigna, sin embargo, el agrupamiento (clúster) como método puede ser interpretado de diferentes maneras y de acuerdo al caso de aplicación, por lo tanto, el método de agrupamiento puede ser definido según nuestro caso de interés como el proceso de agrupar individuos en una población, con el objetivo de descubrir la estructura de los datos, de manera que los individuos que están dentro de un mismo grupo (población) sean más similares entre sí que con individuos que pertenecen a otra población.

De acuerdo a la definición dada en [12] , el término clúster puede ser expresado como:

Dado un con junto de datos representado por *X*, teniendo, $X = [x1, x2, \dots x_N]$, definiendo una m–agrupación de *X*, $\mathbb{R}$, a una partición de *X* en *m* conjuntos o clústeres, *C1*, *C2*, *C3*.....*Cm*, de modo que se cumplan las siguientes condiciones:

- $Ci \neq \phi, i = 1, \dots, m$ (1)
- $U_{i-1}^{m} Ci = X$ (2)
- $Ci \cap Cj = \phi, i \neq j, i, j = 1 \dots, m$ (3)

El conjunto de vectores que contiene *Ci* son "más similares" entre si y "menos similares" a las características de los vectores del otro clúster, *Cj*. La cuantificación de los términos similar–diferente dependerá de los tipos de agrupamientos que prevé sean la base de la estructura de *X*.

Según lo dicho anteriormente, cada objeto es asignado a una única clase, este es el caso de un clúster de tipo duro (Hard Clustering).

También tenemos que cada punto u objeto puede pertenecer a todos los clústeres hasta cierto grado, este es el caso de la fuzzy clustering que matemáticamente se puede definir de la siguiente manera.

Una agrupación de tipo fuzzy de *X* dentro de *m* está caracterizada por m funciones u_j. En donde tenemos que:

$$u_j : \underline{x} \to [0,1], \quad j = 1,2,...,m \qquad (4)$$

Además

$$\sum_{j=1}^{m} u_j(\underline{x}_i) = 1, \quad i = 1,2,...,N \qquad (5)$$

Cumpliendo la condición:

$$0 < \sum_{i=1}^{N} u_j(\underline{x}_i) < N, \quad j = 1,2,...,m \qquad (6)$$

Estas son conocidas como funciones de pertenencia en donde cada $\underline{x}_i$ pertenece a algún agrupamiento hasta cierto grado, dependiendo de

$u_j(\underline{x}_i), \quad j = 1,2,...,m$

En donde se puede decir lo siguiente:

Si $u_j(\underline{x}_i)$ se acerca a $1 \Rightarrow$ tenemos un alto grado de pertenecia de $\underline{x}_i$ hacia el agrupamiento j.

Si $u_j(\underline{x}_i)$ se acerca a $0 \Rightarrow$ tenemos un bajo grado de pertenencia hacia el cluste.

2.1.1 Proceso para el Agrupamiento

En nuestro caso como se trata de un problema concreto, se debe tener en cuenta algunas fases para obtener un resultado deseado, que de acuerdo al autor de [12], se pueden señalar los siguientes:

- Selección de características. Este caso consiste en discernir y elegir los subconjuntos de datos con los que posean una mayor información útil para la predicción y separar de aquella no es de ayuda para nuestro fin, es decir

tomar los datos que más nos sirven y eliminar aquellos que sean irrelevantes, es decir, se busca la mínima redundancia de información de las características, el pre–procesamiento de características puede ser necesario antes de su utilización en las etapas subsiguientes.

- Medidas de proximidad. Es la cuantificación de los términos similar o diferente, es decir, esta sirve para saber cuánto se asemejan o difieren dos objetos, su objetivo es garantizar que todas las características seleccionadas contribuyen por igual al cálculo de la medida de proximidad y no hay características que predominen sobre otras. En un análisis de agrupación de individuos, la proximidad suele venir expresada en términos de distancias, mientras que el Análisis de Clúster por variables involucra generalmente medidas del tipo coeficiente de correlación [13].

- *Criterio de agrupamiento:* Este depende de la interpretación sensata por parte del experto, el tipo de agrupaciones que se utilice estará en función de la base del conjunto de datos, que consiste en una función costo o un conjunto de reglas.
- *Algoritmo de agrupamiento:* Después de haber adoptado una medida de proximidad y un criterio de agrupamiento, este paso se refiere a la elección de un esquema algorítmico específico que describa la estructura de la agrupación del conjunto de datos.
- *Validación de los resultados:* Una vez que se han obtenido los resultados del algoritmo de agrupamiento, tenemos que verificar si estos son correctos mediante la realización de pruebas.
- *Interpretación de los resultados:* En muchos casos, los expertos en el campo de aplicación deben comparar los resultados obtenidos con otros experimentos y analizarlos con el fin de sacar las conclusiones correctas.

Medidas de proximidad

Entre las definiciones relativas a las medidas entre vectores tenemos las siguientes consideraciones:

De acuerdo a lo señalado en [14]. Una medida de disimilitud *d*, dentro de un conjunto de datos

X, está representada como la función:

$$d: \ X \times X \longrightarrow \Re$$

Dándose el cumplimiento de las siguientes propiedades:

- $\exists d_0 \in \Re: \ -\infty < d_0 \leq d(\underline{x}, \underline{y}) < +\infty, \ \forall \underline{x}, \underline{y} \in X$ (7)
- $d(\underline{x}, \underline{x}) = d_0, \ \forall \underline{x} \in X$ (8)
- $d(\underline{x}, \underline{y}) = d(\underline{y}, \underline{x}), \ \forall \underline{x}, \underline{y} \in X$ (9)

También tenemos:

- $d(\underline{x}, \underline{y}) = d_0 \ Si \ y \ solo \ si \ \underline{x} = \underline{y}$ (10)
- $d(\underline{x}, \underline{z}) \leq d(\underline{x}, \underline{y}) + d(\underline{y}, \underline{z}), \ \forall \underline{x}, \underline{y}, \underline{z} \in X$ (11)

Teniendo una desigualdad triangular, en donde *d* se conoce como medida de disimilitud métrica.

Medida de similitud o distancia

Existen diferentes técnicas y de cual se utilice dependerá de cómo queramos formar los grupos, por lo cual, será de suma importancia utilizar la técnica adecuada ya que de esta dependerá el éxito en los resultados,

La medida de similitud entre vectores de *X* está dada por la siguiente función:

$$s: \ X \times X \longrightarrow \Re$$

Con las siguientes propiedades:

- $\exists s_0 \in \Re: \ -\infty < s(\underline{x}, \underline{y}) \leq s_0 < +\infty, \ \forall \underline{x}, \underline{y} \in X$ (12)
- $s(\underline{x}, \underline{x}) = s_0, \ \forall \underline{x} \in X$ (13)
- $s(\underline{x}, \underline{y}) = s(\underline{y}, \underline{x}), \ \forall \underline{x}, \underline{y} \in X$ (14)

Con lo cual tenemos:

- $d(\underline{x}, \underline{y}) = d_0 \ Si \ y \ solo \ si \ \underline{x} = \underline{y}$

- $d(\underline{x},\underline{z}) \le d(\underline{x},\underline{y}) + d(\underline{y},\underline{z}),\ \ \forall \underline{x},\underline{y},\underline{z} \in X$ (15)

En donde s es conocida como una medida de similitud métrica.

Como se mencionó, existen varias medidas de similitud o distancia, no obstante, aquí no se analizaran todas sino únicamente las más utilizadas y que posteriormente serán utilizadas, en te sentido vamos a mencionar algunas medidas de asociación entre individuos.

$s:\ \ X \times X \longrightarrow \Re$

Teniendo las medidas de disimilitud (DMs) y partiendo de la ecuación para encontrar la distancia de Minkowski, que tienen la propiedad de que las características con varianzas y valores más grandes serán predominantes sobre las otras características, teniendo :

$$d_p(\underline{x},\underline{y}) = (\sum_{i=1}^{l} w_i \mid x_i - y_i \mid^p)^{1/p}$$

En la cual al definir el valor de la variable *p,* tendremos otras medidas de distancias tales como:

- Para *p=1* , tenemos medida de distancia Manhattan descrita por la ecuación (16).

$$d_{man}(\underline{x},\underline{y}) = (\sum_{i=1}^{l} w_i \mid x_i - y_i \mid) \qquad (16)$$

Cuando empleamos este método los datos se tienden a agrupar en forma de grupos hiperestésicos.

- Si *p=2* , tenemos como resultado la distancia Euclidea descrita por la ecuación (17);

$$d_{eu}(\underline{x},\underline{y}) = (\sum_{i=1}^{l} w_i \mid x_i - y_i \mid^2)^{1/2} \qquad (17)$$

Esta es la distancia más aplicada ya que es invariante a translaciones y rotaciones lineales y tiende a formar grupos de forma esférica.

2.1.2 Clasificación de los Métodos de Agrupamiento

Como se mencionó existen diferentes tipos de clasificaciones de los algoritmos de agrupamiento; pero en este caso, vamos a dividirlos de la misma manera como se dio una definición del mismo, es decir, de acuerdo a cada caso de aplicación, que se pude dar de dichos algoritmos, aquí se dará una breve clasificación de acuerdo al enfoque de los problemas de agrupamiento que se puedan dar y particularmente con perspectiva al problema presentado por este documento, se tiene:

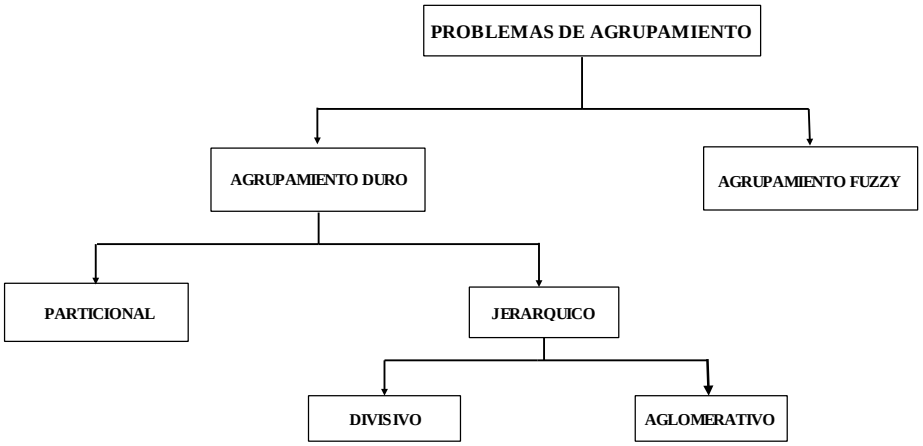

Figura 13. Clasificación de los algoritmos de agrupamiento de acuerdo al tipo de problema.

Los algoritmos de agrupación convencionales se pueden clasificar en dos categorías: algoritmos jerárquicos y algoritmos particionales. Teniendo dos tipos de algoritmos jerárquicos: algoritmos jerárquicos divisivos y algoritmos jerárquicos aglomerativos. En un algoritmo jerárquico divisivo, el procedimiento va desde la parte superior a la parte inferior, es decir, el algoritmo comienza con un clúster grande que contiene todos los puntos de datos en el conjunto de datos y continúa dividiendo en grupos; en un algoritmo jerárquico de aglomeración, el algoritmo procede de la parte inferior a la parte superior, es decir, el algoritmo comienza con las agrupaciones que contienen cada uno de los puntos de datos y continúa la conformación de los grupos. Un algoritmo particional divide un conjunto de datos en una única partición, mientras que un algoritmo jerárquico divide un conjunto de datos en una secuencia de particiones anidadas.

Algoritmo de agrupamiento duro

En este tipo de agrupamiento se tiene que cada vector pertenece exclusivamente a un único clúster, el motivo de análisis de este algoritmo es debido a que son los más utilizados y posteriormente en este documento serán analizados.

El punto de partida, es la suposición de que los coeficientes de pertenencia u_{ij} , pueden ser un ‘’1’’ o un’’ 0’’, además, tenemos un ‘’1’’ para un agrupamiento C_j, y un cero para todos los demás, C_k, $k \neq j$, lo que representa:

$$u_{ij} \in \{0, 1\}, \quad j=1,\ldots,m \tag{18}$$

A demás de

$$\sum_{j=1}^{m} u_{ij} = 1 \tag{19}$$

De acuerdo a las ecuaciones (18), (19); se puede observar como un caso especial extremo de los esquemas algorítmicos fuzzy, sin embargo, la función coste está definida como:

$$J(\underline{\theta}, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij} d(\underline{x}_i, \underline{\theta}_j) \tag{20}$$

No es diferenciable con respecto a $\underline{\theta} j$. No obstante, el marco general de los esquemas algorítmicos fuzzy generalizados, puede adoptarse para el caso especial de la agrupación de tipo dura. Estos esquemas se han usado ampliamente en la práctica.

Si se fija $\underline{\theta}_j$, $j=1.....m$. Dado que para cada vector $\underline{x}_i$ solamente un u_{ij} tiene valor ‘’1’’ y el resto son ‘’0’’, si se asigna cada uno de los $\underline{x}_i$ a su correspondiente grupo que se encuentre más cercano teniendo:

$$u_{ij} = \begin{cases} 1, & if \;\; d(\underline{x}_i, \underline{\theta}_j) = \min_{k=1,\ldots,m} d(\underline{x}_i, \underline{\theta}_k) \\ 0, & otro \end{cases}, \quad i = 1,\ldots, N \tag{21}$$

Algoritmos particionales

Es uno de los algoritmos de agrupamiento más conocidos y más empleado, es aquel que a partir de un conjunto de datos se obtiene una única partición, a diferencia de los algoritmos jerárquicos, los cuales dividen un conjunto de datos en una secuencia

de particiones anidada, es decir, realiza varios niveles de agrupamientos, es utilizada en los casos donde los datos de entrada son de gran dimensión.

Este tipo de algoritmo emplea una función criterio, que estará orientada a la búsqueda aproximada de la mejor partición, mediante el escaneo de forma iterativa del conjunto de datos.

Estos algoritmos asumen un conocimiento a priori del número de ‘’***k***’’ clústeres en que debe ser dividido el conjunto de datos, llegan a una división en clases que optimiza un criterio predefinido o función objetivo [15].

El problema en torno a la implementación de los algoritmos particionales, de acuerdo a [16], consiste en lo siguiente:

Dado ***n*** patrones en un espacio métrico ***d***–*dimensional*, determinar una partición de los patrones en grupos ***k***, o grupos, de tal manera que los patrones en un grupo sean más similares entre sí que a los patrones en los otros grupos.

El procedimiento para realizar este tipo de algoritmos y la solución teórica consiste en los siguientes pasos mencionados en [17]:

a) Seleccionar ***k*** puntos representantes (cada punto representa un grupo de la solución).
b) Asignar cada elemento al grupo del representante más cercano, de forma de optimizar un determinado criterio.
c) Actualizar los ***k*** puntos representantes de acuerdo a la composición de cada grupo.
d) Volver al punto b).

Este ciclo se repite hasta que no sea posible mejorar el criterio de optimización.

Otro inconveniente en este método es la cantidad gigantesca de particiones, incluso teniendo un pequeño número de patrones es un obstáculo formar los grupos, las soluciones a este problema serán dada más adelante de acuerdo al caso y tipo de algoritmo analizado.

Existen varios algoritmos de este tipo, entre ellos está el más común de todos, como es el método de ***K–Means***, del cual se basan varios algoritmos más como son el método ***K–Medoids***, el ***Mean–Shift***, etc. Los cuáles serán estudiados a continuación con más detenimiento, ya que son los métodos empleados para este proyecto.

2.2 Algoritmo K–Means

Es uno de los algoritmos más simples y conocidos, no solo dentro de los de tipo particional sino de los algoritmos en general, sigue una forma fácil y sencilla para dividir un conjunto de datos en ***k*** grupos o Clústers conocidos a priori, cada clúster tiene asociado un centroide (centro geométrico del Clúster).

La forma más simple del algoritmo K–Means consiste en la alternancia de dos procedimientos: La primera consiste en la asignación de objetos a grupos. Un objeto usualmente es asignado a cuya media está más cercano en el sentido euclidiano, esto es, los vectores de medias de las variables medidas sobre los individuos del Clúster. El segundo procedimiento consiste en el cálculo de las nuevas medias del grupo basándose en las asignaciones. El proceso termina cuando no haya más movimientos de un objeto hacia otro grupo.

El objetivo de este procedimiento es disminuir la suma de errores cuadráticos o función objetivo llamada (Sum Squares Error, SSE), es decir, cuando ya no se pueda disminuir más dicha función será el agrupamiento asignado, este tema se tratara con detalle más adelante.

El problema del empleo de estos esquemas radica en que fallan cuando los puntos de un grupo están muy cerca del centroide de otro grupo, también cuando los grupos tienen diferentes tamaños y formas [15].

2.2.1 Procedimiento del Algoritmo

La idea principal del algoritmo K–Means consiste en los siguientes pasos:

Inicialización

Primeramente, se tiene que definir ***k*** centroides, que pueden ser ***k*** elementos seleccionados al azar de la base de datos, o los obtenidos mediante aplicación de alguna técnica de inicialización, posteriormente se toma cada uno de los elementos del conjunto de datos y se los asignan al grupo con el centroide más próximo.

Proceso iterativo

Mientras los centroides no cambien, se procede a calcular la distancia del centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano. El proceso se repite hasta que ya no hay cambio en los grupos, es decir, los ***k*** centroides no cambian luego de una iteración, lo cual es equivalente a decir que el valor de la función utilizada como criterio de optimización no varía [18].

Condición de convergencia: Existen varias condiciones de convergencia, sin embargo de acuerdo a [19], se puede decir que:

Es posible converger cuando alcanza un número de iteraciones dado, converger cuando no existe un intercambio de objetos entre los grupos, o converger cuando la diferencia entre los centroides de dos iteraciones consecutivas es más pequeña que un umbral dado. Si la condición de convergencia no se satisface, se repiten los pasos anteriores del algoritmo.

La complejidad del algoritmo K–Means está dada por *O(n*k*I*d)* en donde:

- *n* = número de puntos,
- *k* = número de Clústers,
- *I* = número de iteraciones,
- *d* = número de atributos

Se tiene un problema de tipo No Polinomial **(NP)** si ***k*** no se fija.

Los diferentes tipos de medidas que pueden ser utilizados en este tipo de algoritmo son: la distancia Manhattan, la distancia Euclidea y similitud coseno; pero la medida de distancia euclidea es la más empleada. Se pueden obtener diferentes agrupamientos para diferentes valores de ***K*** y medidas de proximidad. La función objetivo empleada por K–Means se denomina suma de errores cuadráticos denotada por SSE o también conocida como sumas residuales de cuadrados (RRS).

Cuando se utiliza la distancia euclidea, SSE se minimiza usando la media aritmética (por cada atributo o variable). Cuando se emplea la distancia de Manhattan, SSE se minimiza usando la mediana.

Dado el conjunto de datos $X = [x1, x2, \dots x_N]$ consta de **N** puntos, se denota la agrupación obtenida, después de aplicar K–Means, es decir:

$C = [C1, C2, \dots\dots, C_K]$

El SSE para este agrupamiento está definido por:

$$SSE = \sum_{K=1}^{K} \sum_{xi \in Ck} \| x_i - c_k \|^2 \qquad (22)$$

En donde c_k es el centroide del agrupamiento. El C_k dado por la ecuación (22), donde el objetivo es encontrar la agrupación que minimice el valor de SSE.

$$c_k = \frac{\sum_{xi \in Ck} x_i}{|C_k|} \qquad (23)$$

El propósito del proceso iterativo y los pasos de actualización del algoritmo K–Means, tienen como objetivo el minimizar el valor de SSE para el conjunto dado de centroides.

Minimización de la suma de errores cuadráticos

Como se ha mencionado el agrupamiento K–Means es esencialmente un proceso de optimización por medio de la minimización de la función objetivo conocida como SSE. En [20], se explica este procedimiento de la siguiente manera.

Se puede probar matemáticamente el motivo de la elección de las medias de los puntos, de un conjunto de datos en un agrupamiento, como un prototipo representativo para una agrupación en el algoritmo K–Means.

Es posible denotar $\boldsymbol{C_k}$ como el *k- ésimo* agrupamiento, x_i es un punto en $\boldsymbol{C_k}$, y $\boldsymbol{c_k}$ es la media del *k- ésimo* agrupamiento. Se puede resolver para la característica de C_j que minimiza los SSE para diferenciar el SSE con respecto a $\boldsymbol{C_j}$ y se establece igual a cero, con lo cual la ecuación (22) se escribe como:

Dado

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in Ck} (c_k - x_i)^2 \qquad (24)$$

Y aplicando la diferencial a ambos lados de la ecuación

$$\frac{\partial}{\partial c_j} SSE = \frac{\partial}{\partial c_j} \sum_{k=1}^{K} \sum_{x_i \in Ck} (c_k - x_i)^2 \qquad (25)$$

De donde resulta

$$\frac{\partial}{\partial c_j} SSE = \sum_{x_i \in c_j} (c_j - x_i) \tag{26}$$

De lo que se puede obtener

$$\sum_{x_i \in c_k} 2 * (c_j - x_i) = 0 \Rightarrow |C_j| \cdot c_j = \sum_{x_i \in c_j} x_i \Rightarrow c_j = \frac{\sum_{x_i \in c_j} x_i}{|C_j|} \tag{27}$$

Por lo tanto se puede concluir que la mejor manera representativa de la minimización del SSE de un agrupamiento es la media de puntos de un agrupamiento, el SSE automáticamente decrece con cada iteración, convergiendo monótonamente a un mínimo local [20].

El algoritmo de las K–Means no garantiza que los centroides finales obtenidos sean los que minimizan globalmente la función objetivo SSE.

2.2.2 Desarrollo del Algoritmo

En la mayoría de los casos prácticos, los algoritmos K–Means convergen rápidamente, por lo que en [21], se propone implementar un procedimiento convergente, que se llamara método de las K-Means convergentes:

- Comenzar con una partición inicial de los individuos en clústeres. Si se desea, la partición puede ser construida usando cualquier método de los establecidos para particiones iniciales puede ser empleado.
- Tomar cada caso sucesivamente y calcular las distancias a todos los centroides de los clusters; si el centroide más próximo no es el del clúster padre de dicho caso, reasignar dicho caso al clúster con centroide más próximo y recalcular los centroides de los clústeres afectados en el proceso.
- Repetir el paso segundo hasta que se obtenga la convergencia, o sea, continuar hasta que un ciclo completo a través de todos los casos no proporcione ningún cambio en los miembros de los clústeres.

Con lo dicho antes es posible plantear el pseudocódigo para poder resolver el algoritmo K–Means, como se tiene a continuación:

Algoritmo 1

Algoritmo de Agrupamiento K–Means

1: Inicializar aleatoriamente las K particiones. Encontrar $\boldsymbol{C} = [\boldsymbol{C1}, \boldsymbol{C2}, \ldots\ldots, \boldsymbol{C_K}]$**;**

2: **while** no exista cambios en $\boldsymbol{C}$ **do**

3: Clasificar los objetos de acuerdo al c_j más cercano;

4: Recalcular $\boldsymbol{C}$ de acuerdo al actual;

5: **end while**

Out: $\boldsymbol{C}$

Una desventaja de este algoritmo consiste en que el resultado obtenido es dependiente de la selección inicial de los centroides de los clústeres y puede converger a óptimos locales [22]. Por lo tanto, la selección de los centroides iniciales afecta directamente el proceso principal de K–Means y el resultado final del proceso.

Existen múltiples métodos para la inicialización, en este caso se aplicara el método propuesto por MacQueen que busca de una manera sencilla de manera aleatoria las *k* semillas, este es el método más popular y más sencillo de emplear, sin embrago, en [20], se encuentran algunos métodos que son útiles para la inicialización según el caso, así como estimar el número del agrupamientos para optimizar el algoritmo; pero en este caso es necesario dar el número de clúster a priori ya que es un dato del cual se tiene que basar.

2.3 Agrupamiento K–Medoids

Este es una variación del algoritmo K–Means, por lo tanto es muy similar al mismo, el objetivo del algoritmo K–Medoids, está en encontrar una solución de agrupamiento que minimice una función objetivo predefinida, en la que los clúster que agrupa el conjunto de $\boldsymbol{N}$ objetos en $\boldsymbol{k}$ grupos conocidos a priori.

El algoritmo K–Medoids es más robusto al ruido y a valores grandes de los datos en comparación con el K–Means, ya que minimiza la suma de diferencias por parejas en lugar de la suma de los cuadrados de las distancias euclidianas.

El algoritmo K–Medoids tiene como objetivo minimizar el criterio de error absoluto en lugar del SSE, como en el caso del algoritmo K–Means, sin embargo, procede interactivamente hasta que cada objeto representativo sea en realidad el medoide de la agrupación.

Este algoritmo utiliza los medoides en lugar de los centroides como el caso de los algoritmos K–Means. K–Medoids se basa en el objeto más céntrico de un clúster haciéndolo menos sensible a los valores atípicos en comparación con la agrupación K–Means [23].

Un medoide, puede ser definido como un objeto de un grupo, cuyo promedio de disimilitud a todos los objetos en grupo es mínima, es decir, que es un punto más céntrico del clúster [24].

2.3.1 Procedimiento del Algoritmo

El procedimiento del algoritmo empieza por la inicialización de los agrupamientos mediante la selección aleatoria de lo ***k*** nodos semillas. Todos los nodos se asignan al clúster del nodo de la semilla más cercana (medido en distancia geodésica), esta distancia es un número entero de los pasos entre nodos, muchos nodos semillas pueden estar a la misma distancia de un nodo dado ***v***, en este caso ***v*** es asignado aleatoriamente a uno de los nodos semilla más cercanos. El siguiente paso es calcular la proximidad central para cada nodo en cada agrupamiento. La proximidad central de un clúster, es una medida de la distancia de un nodo a los otros nodos en el agrupamiento. Si el nodo ***i*** está en la agrupación ***k*** y denotamos la distancia geodésica entre un nodo ***i*** y un nodo ***j*** por $d_g(i,j)$, donde la cercanía del nodo ***i*** hacia otro nodo en esta agrupación es definida como el reciproco de la suma de las distancias geodésicas, con lo cual se da la siguiente relación:

$$c_l(i) = \frac{1}{\sum_{j \in cluster_k} d_g(i, j)} \qquad (28)$$

Los nuevos nodos centrales son los que tienen la proximidad central más pequeña en cada grupo. El procedimiento de iteración se da hasta que se obtenga una solución estable, no necesariamente la convergencia, ya que la naturaleza aleatoria de la asignación de nodos a las agrupaciones en el caso de distancias geodésicas iguales puede resultar en cierta inestabilidad. Por lo tanto, el algoritmo puede ser considerado estable cuando hay un cambio de menos el x% en el número de medoides en el agrupamiento, según [25] sugiere un valor de x de entre 1 y 3.

En el algoritmo de agrupamiento K–Medoids, se considera un punto $\boldsymbol{x_i}$ aleatorio que es usado para remplazar un objeto representativo. A continuación el siguiente paso es comprobar el cambio de la composición de los puntos que originalmente pertenecían a $\boldsymbol{m}$, el cambio en la composición de estos puntos pueden ocurrir en una de las siguientes maneras.

Estos puntos pueden ahora estar más cerca de $\boldsymbol{x_i}$, este es el nuevo punto representativo, o a todos los otros conjuntos de puntos representativos.

El costo de intercambio, se calcula como el criterio de error absoluto para K–Medoids. Cada operación de reasignación de este costo de intercambio es calculado y esto contribuye a la función general de costos, el costo es considerado como la distancia geodésicas antes descrita.

2.3.2 Desarrollo del Algoritmo

Los pasos para realizar este algoritmo según [26] pueden ser los siguientes:

1. Inicializar un conjunto de objetos como semillas para los centros de grupo.
2. Se asignan los nodos a la agrupación centro de la agrupación más cercana (usando la distancia geodésica – los lazos son resueltos aleatoriamente).
3. Para cada grupo, se calcula un nuevo conjunto de centros de grupo. Estos son los nodos de cada grupo con el valor más bajo de cercanía con la ecuación (27).
4. Si el algoritmo se ha estabilizado, salir, de lo contrario ir al paso 2.

Algoritmo 2

Algoritmo K–Medoids

1: **Choose** estimaciones iniciales arbitrarias $\theta j(0)$ para $\theta j'\,s,\ j=1,\ldots,m.$

2: Repeat

3: **For** $i=1$ to N

4: Determinar el representante más cercano, expresar θj, for xi

4: **Set** $b(i)=j.$

End {For}

5: **For** $j=1$ to m

6: Actualizar Parámetros: Determinar θj como la media de los vectores $xi \in X$ con $b(i)=j.$

End *{For}*

7: Hasta que exista cambio en $\theta j'\,s$ que se produce entre dos iteraciones sucesivas.

2.4 Agrupamiento Mean–Shift

Es una técnica no paramétrica para el análisis de un conjunto de puntos dados en un espacio de características que obtiene un máximo local de la función de densidad estimada para la muestra, este método generalmente es empleado en la segmentación de imágenes.

Una técnica no paramétrica es aquella que se enfoca en el reconocimiento de patrones sin suponer nada acerca de la distribución de los valores en el espacio de características, es decir, no se requiere de información a priori, como es el número de grupos y permite la forma arbitraria de los datos.

Los métodos que necesitan información sobre el número de clústeres o la forma de éstos, en general no son aplicables en situaciones reales y ésta es la razón por la que los métodos no paramétricos son estudiados para su uso. La razón para utilizar una técnica general no paramétrica de estimación de la densidad se debe a que el espacio de características se puede considerar como la función de densidad de probabilidad del parámetro representado.

Las regiones con gran cantidad de datos del conjunto de datos corresponderán a los máximos locales de la función de densidad.

La trayectoria que sigue el método de Mean Shift es una trayectoria suave, esto significa que el ángulo entre dos vectores Mean Shift siempre es menor que 90 grados [27].

2.4.1 Procedimiento del Algoritmo

Este método consiste en un proceso iterativo de punto fijo, el cual converge a un máximo local de la densidad, obteniendo en cada paso una estimación del gradiente de dicha función de densidad a partir de la muestra, es decir, se aplica el método iterativamente con el fin de encontrar el máximo local o punto fijo de la función de densidad más cercano a un punto P del conjunto de datos. Al aplicar este método varias veces desde diferentes puntos de partida, se puede encontrar las modas de densidad, es decir, las zonas en las que se tiene mayor densidad de puntos. Una vez que se haya encontrado la moda para cada punto, se asociaran en un grupo común a todas aquellas que se encuentren en un contorno determinado.

Dicha función de densidad es encontrada mediante un método de estimación basado en Kernels o también conocidos como Ventanas de Parzen:

Dado un conjunto de datos $x_i \ , i = 1, \dots, n$ de un espacio d–dimencional, teniendo:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (29)$$

En dónde EL Kernel $\boldsymbol{K}$ y el ancho de banda $\boldsymbol{h}$, corresponden a la ventana y la longitud de uno de sus lados respectivamente. Existen varios tipos de Kernels; pero uno de los más utilizados es el Kernel Normal o Gauissiano, definido como:

$$K_N(x) = (2\pi)^{\frac{-d}{2}} exp\left(-\frac{1}{2}\|x\|^2\right) \qquad (30)$$

Siendo este un Kernel radialmente simétrico tenemos como su superficie:

$$K_N(x) = \exp\left(-\frac{1}{2}x\right) \quad \text{x} \geq 0 \qquad (31)$$

Como se mencionó antes, el objetivo es encontrar las modas de un espacio con densidad *f(x)*, dichas modas son encontradas cuando $\widehat{\nabla} f(x) = 0$, es decir, cuando la función de densidad es nula.

Por lo expuesto anteriormente se define al gradiente que estimara la densidad por medio de la siguiente expresión:

$$\widehat{\nabla} f_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} g\left(\left\|\frac{xi - x}{h}\right\|^2\right) \left[\frac{\sum_{i=1}^{n} xig\left(\left\|\frac{xi - x}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{xi - x}{h}\right\|^2\right)} - x\right] \quad (32)$$

En donde el término de la derecha es el conocido como el vector de desplazamiento medio, el cual apunta siempre a la dirección de máximo incremento de densidad y está definido como:

$$m_h(x) = \left[\frac{\sum_{i=1}^{n} xig\left(\left\|\frac{xi - x}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{xi - x}{h}\right\|^2\right)} - x\right] \quad (33)$$

Es decir este vector es una estimación normalizada del gradiente de la función de densidad de la muestra.

En el caso de utilizar un Kernel Gaussiano se tendría:

$$m_h(x) = \frac{\sum_{i-1}^{n} xi\ exp^{-\frac{1}{2}\left\|\frac{xi-x}{h}\right\|^2}}{\sum_{i-1}^{n} xi\ exp^{-\frac{1}{2}\left\|\frac{xi-x}{h}\right\|^2}} - x \quad (34)$$

Teniendo ya definidos estos parámetros se puede dar una definición más clara de lo que es el algoritmo Mean Shift [27]:

Se denomina Mean Shift, al desplazamiento desde un punto inicial ***x*** en el espacio a otro que resulta del promedio de los pesos de los datos dentro de una vecindad determinada por una región ***S*** centrada en ***x***. Los pesos asociados a los datos dentro de la hiperesfera ***S*** están directamente relacionados con sus distancias respecto al origen y sus valores quedan determinados por la función kernel ***K***.

Con esta definición se puede tener una mejor idea de cómo funciona el algoritmo Mean Shift mediante los siguientes apartados.

2.4.2 Desarrollo del Algoritmo

Con lo antes expuesto se puede desarrollar el al algoritmo Mean–Shift, mediante la ejecución de los siguientes pasos:

- Teniendo un punto ***x*** en el espacio de características, se procede a encontrar el vector de desplazamiento hacia la media $m_h(x^t)$ a partir de los datos dentro de la región ***S***.
- Cuando se tenga el nuevo punto, desplazar el centro de la ventana hacia este punto.
- Repetir los anteriores hasta que el desplazamiento sea menor a un épsilon dado.

Teniendo la siguiente iteración de punto fijo definida como:

$$x^{t+1} = x^t + m_h(x^t) \qquad (35)$$

Este método converge cuando el gradiente de función de densidad es nulo. Mean Shift alcanzará un máximo en tanto se tenga la certeza de que en cada iteración exista un único punto de mayor densidad de datos, dentro del entorno determinado por el ancho del kernel utilizado.

Algoritmo 3

Mean Shift

Entrada: Conjunto de datos ***X*** donde $xi \in \mathbb{R}^d$ *; i = 1……n* y ancho de banda ***h.***

1: **for all** $xi \in X$ ***do***

2: $xt \leftarrow xi$

3: **while** *mh(xt)* ! = *umbral* **do**

4: calcular *mh(t);*

5: $xt \leftarrow xt + mh(t)$

6: **end while**

7: *ClustCent$_i$= {x ϵ D | dist (x,x$_i$)* es mínima}

8: **if** *ClustCent$_i$* no existe **then**

9*:* *ClustCent$_i$=nuevo Cluster Id;*

10: **end if**

11: **end for**

Salida: *Vector cluster* $\in \mathbb{R}^d$

2.5 Experimentos y Resultados

Los experimentos son realizados con los algoritmos antes descritos y aplicados a tres diferentes escenarios con distintos número de datos en cada uno, siendo el algoritmo K–Means analizado con el escenario de los datos representados en la Figura 14, el cual consta de 84 puntos. El algoritmo K–Medoids con el segundo escenario que se muestra en la Figura 15, con 142 puntos y finalmente la técnica de Mean–Shift en el tercer escenario el mismo que está conformado por 394 puntos.

En la Figura 14, se pueden observar un conjunto de datos correspondientes a las coordenadas de los nodos de un sector de la ciudad de Cuenca, los mismos que son agrupados por medio de la técnica de K–Means.

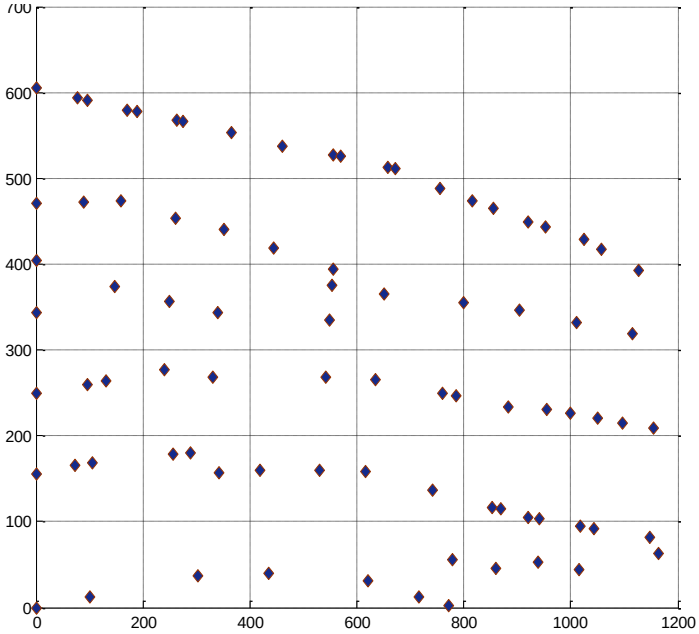

Figura 14. Escenario 1 de 84 puntos pertenecientes a la Ciudad de Cuenca

El segundo escenario corresponde al sector del estadio de la ciudad de cuenca constituido por 142 Puntos o nodos como se muestra en la Figura 15:

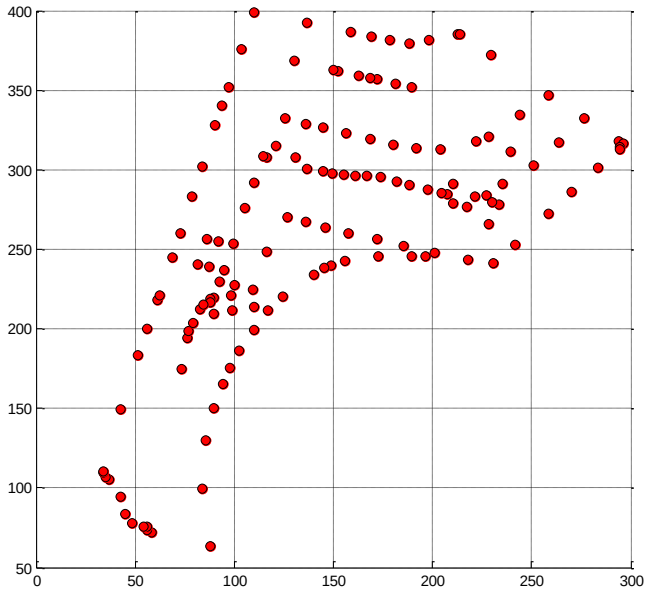

Figura 15. Escenario 2 de 142 puntos pertenecientes a la cuidad de Cuenca

El tercer escenario en análisis es el que comprende parte del sector de la "Avenida Remigio Crespo Toral" de la ciudad de Cuenca con 394 nodos como se ve en la Figura 16.

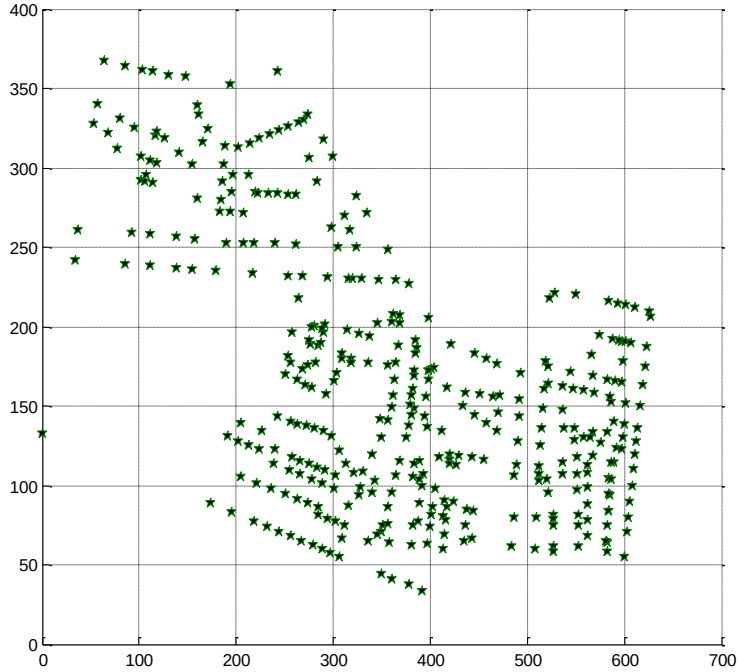

Figura 16. Escenario 3 de 394 puntos pertenecientes a la Ciudad de Cuenca

Los puntos de color azul, rojo y verde representan los nodos dados por las coordenadas en metros de los nodos (esquinas) de un sector de la ciudad de Cuenca, estos datos serán agrupados en diferentes particiones mediante las técnicas de agrupamiento K–Means, K–Medoids y Mean–Shift respectivamente, luego se procederá a realizar una comparativa y análisis de los datos obtenidos con cada algoritmo, para posteriormente aplicarlos en el despliegue de la red de fibra óptica.

2.5.1 Iteraciones versus Tiempo

Esta grafica muestra el tiempo que se demora en ejecutarse el algoritmo durante cada iteración en este caso se ha realizado el análisis mediante 10 intentos aplicando los métodos y escenarios de acuerdo a lo que se mencionó antes.

Las pruebas fueron realizadas con agrupaciones de 2, 4, 6 y 8 del conjunto de datos, aplicando los tres métodos en estudio como son: K–Means, K–Medoids y Mean–Shift con 10 iteraciones para cada agrupamiento de manera consecutiva.

La Figura 17, muestra las curvas de cada caso aplicando el algoritmo K-Means a los datos de la Figura 14, representadas mediante el número de iteraciones en el eje de

abscisas y el tiempo en segundos en el que tardo en ejecutarse cada proceso en el eje de las ordenadas.

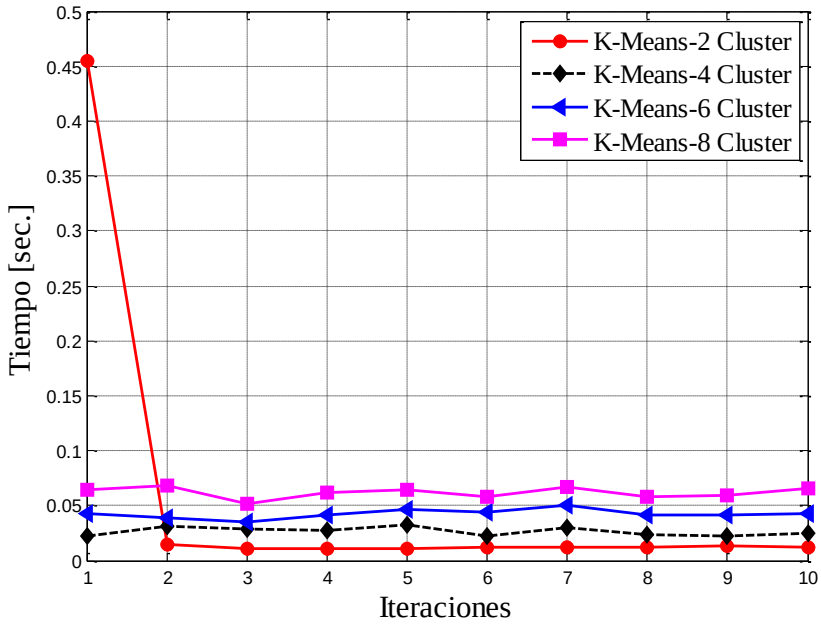

Figura 17. Curvas de Iteraciones versus Tiempo del escenario 1 mediante k-means

La curva de color rojo representa el tiempo transcurrido en agrupar todos los datos en dos clúster en 10 iteraciones, de igual manera la curva de color negro en cuatro, la curva de azul en seis y la de color magenta en 8 clúster respectivamente.

En la Tabla 3**,** se detallan los valores de los tiempos de ejecución de la Figura 17, para cado caso propuesto, los tiempos están dados en segundos.

TABLA 3. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 17

Iteración	Tiempo en segundos clúster 2 grupos	Tiempo en segundos clúster 4 grupos	Tiempo en segundos clúster 6 grupos	Tiempo en segundos clúster 8 grupos
1	0,454240453	0,022907111	0,042166865	0,064437072
2	0,014322077	0,031253361	0,038748473	0,06857717
3	0,010609557	0,028839197	0,03451554	0,051748093
4	0,010331944	0,027064168	0,041638864	0,062204562
5	0,010236877	0,032277228	0,046801935	0,064015743
6	0,011527645	0,022007769	0,044618967	0,057445414
7	0,012340847	0,029703726	0,050335479	0,067197583
8	0,012407349	0,024172884	0,041629492	0,058541138
9	0,012805916	0,022108192	0,041378211	0,059282928
10	0,011884258	0,024403634	0,042786809	0,066111678

Como se puede observar con esta técnica en el proceso de la primera iteración el tiempo es considerablemente superior a las iteraciones posteriores, indiferente del número de aglomerados, a partir de la segunda el tiempo es menor y muy estable para las siguientes iteraciones, esto se debe a que este algoritmo es una función interna de Matlab y al momento de correrla una vez se queda guardada en la memoria del programa, hasta que esta se vacíe. Los tiempos de ejecución son estables y bajos en todos los casos.

Al emplear el algoritmo llamado K–Medoids para agrupar los datos que contiene la Figura 15 y al igual que en el caso anterior se grafican las curvas de las 10 iteraciones y su tiempo de ejecución todo ello se ilustra en la Figura 18.

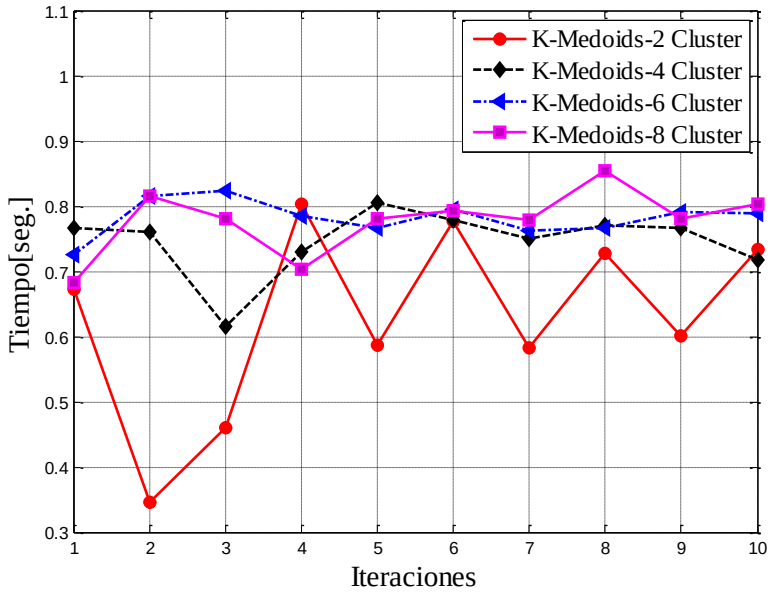

Figura 18. Curvas de Iteraciones versus Tiempo del escenario 2 mediante K–Medoids

En la gráfica se ve claramente la irregularidad y mayor retardo de los tiempos que en el caso anterior, sin embargo, cabe recalcar que para este caso se tomaron un escenario con un mayor número de datos y de diferente topología los mismos que se aprecian en la Figura 15, obteniendo los resultados representados en la Tabla 4.

TABLA 4. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 18

Iteración	Tiempo en segundos clúster 2 grupos	Tiempo en segundos clúster 4 grupos	Tiempo en segundos clúster 6 grupos	Tiempo en segundos clúster 8 grupos
1	0,64950657	0,36918145	0,45677602	0,5293884
2	0,42655054	0,6974618	0,73603978	0,76941229
3	0,46265365	0,54227867	0,78264266	0,8042532
4	0,46234212	0,72510887	0,57601984	0,82153663
5	0,35832507	0,67657966	0,7680028	0,80568054
6	0,34940664	0,74743442	0,77925908	0,79175257
7	0,39555069	0,34685881	0,69295795	0,79634702
8	0,66992676	0,74233874	0,60102512	0,75254036
9	0,44693771	0,7473215	0,75830776	0,78148668
10	0,55368848	0,61343737	0,77405717	0,77329664

Se puede apreciar que los tiempos no son constantes en ninguna iteración, pudiendo en una iteración ser superior, en otra ser inferior y en la siguiente ser superior nuevamente y así en todos los casos de agrupamiento, es decir, cuando se tiene 2, 4, 6 y 8 grupos de datos, en adición los resultados de la Figura 18 revelan un mínimo en la segunda iteración cuando se tienen 2 clústeres, y un máximo en la octava iteración cuando se tienen 8 clústeres.

De igual manera se analiza los puntos de la Figura 16, pero esta vez aplicando el algoritmo Mean–Shift para ser repartidos en grupos de 2 hasta tener 8 clústeres que es el mayor número de aglomerados que se utiliza en este estudio teniendo los tiempos de cada intento como se muestra en la Figura 19.

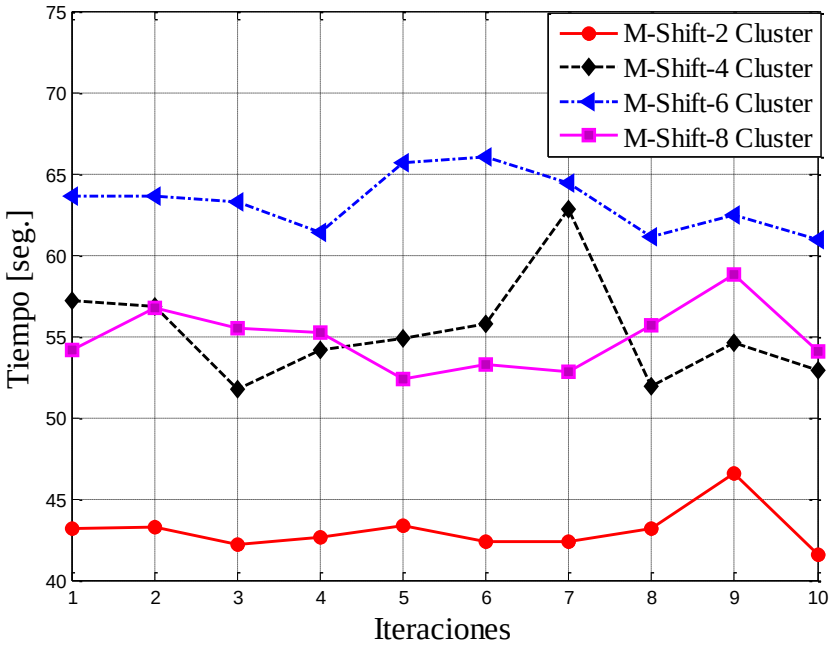

Figura 19. Curvas de Iteraciones versus Tiempo del escenario 3 mediante Mean–Shift

En la primera curva de color rojo es decir, cuando tenemos dos grupos se nota que se tiene más continuidad y muestra menores tiempos en relación con el resto de curvas de este algoritmo, teniendo el caso más irregular al tener 4 agrupamientos y el caso con un mayor retardo en todas las iteraciones en la curva que representa al Mean–Shift de 6 clústeres, teniendo en el de 8 aglomerados la mayoría de tiempos no tan elevados como se creería, solo siendo superiores al tener 2 clústeres, sin embargo, no siendo tan irregular. Estos valores de acuerdo a cada iteración y que se puedan interpretar de una mejor manera se puede apreciar en la Tabla 5.

TABLA 5. VALORES DE LOS TIEMPOS OBTENIDOS DE LA Figura 19

Iteración	Tiempo en segundos clúster 2 grupos	Tiempo en segundos clúster 4 grupos	Tiempo en segundos clúster 6 grupos	Tiempo en segundos clúster 8 grupos
1	43,182594910	57,190384184	63,600784551	54,116851602
2	43,270598259	56,819272962	63,596678802	56,749872462
3	42,190699735	51,727641879	63,287150242	55,471365761
4	42,626589756	54,109649246	61,377366027	55,200295289
5	43,341509571	54,908086839	65,713060423	52,322791124
6	42,406483861	55,753962144	66,020627381	53,233202426
7	42,388778572	62,812155855	64,408791810	52,776916221
8	43,139700346	51,928167026	61,160040740	55,650409705
9	46,594517248	54,637470280	62,481688824	58,801224003
10	41,598883651	52,888376656	60,978043014	58,801224003

Como ya se había dicho los tiempos más bajos y repetitivos son los del clúster de 2 grupos dándose el menor tiempo de ejecución en la iteración 10 del primer caso con un valor de 41,598883651 segundos y el más elevado de 66.020627381 segundos en la sexta iteración al tener 6 grupos.

Los tiempos son muy elevados en comparación con los otros algoritmos esto se debe a la misma forma del algoritmo, pues el Mean–Shift realiza una iteración para ubicación de cada punto.

En las figuras anteriores se mostraron los resultados de los diferentes métodos para los escenarios planteados, por otro lado, es de gran importancia conocer los resultados de los tres escenarios para cada uno de los métodos, en este sentido se han considerado los tiempos promedios de las 10 iteraciones para cada caso en todos los escenarios.

En la Figura 20 se muestran las curvas de los tiempos promedios del algoritmo K–Means para los tres escenarios, teniendo como resultado los tiempos promedios para cada clúster es decir con 2, 4, 6 y 8 agrupamientos.

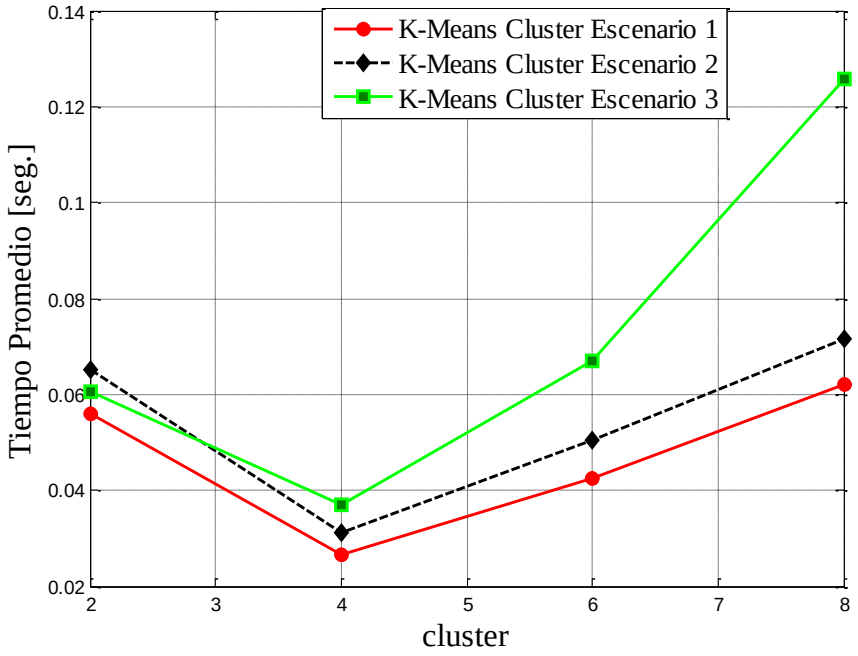

Figura 20. Curvas de los tiempos promedio de los 3 escenarios para el agrupamiento K–Means

La línea de color rojo representa los tiempos promedios para cada clúster al ser distribuidos mediante el método de k–means del escenario 1, la de color negro con el escenario 2 y la de color verde con el escenario 3. En la Tabla 6 se detallan estos valores.

TABLA 6. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 20

Escenarios	Tiempo promedio en segundos clúster 2 grupos	Tiempo promedio en segundos clúster 4 grupos	Tiempo promedio en segundos clúster 6 grupos	Tiempo promedio en segundos clúster 8 grupos
1	0,056070692	0,026473727	0,042462064	0,061956138
2	0,059906933	0,031092194	0,050391314	0,071727678
3	0,060526608	0,036420628	0,066862394	0,613319188

Se puede observar que los tiempos de ejecución de este método son sumamente cortos haciéndolo un algoritmo muy rápido, a pesar de que el tercer escenario es más numeroso no se observa un incremento sustancial en relación con los dos restantes, a

pesar de que en la primera iteración hay un mayor retardo en relación con las demás se tienen tiempos muy cortos haciéndolo muy ágil.

Con la misma metodología, en la Figura 21 , se muestran los resultados de los tres escenarios con el método de K–Medoids.

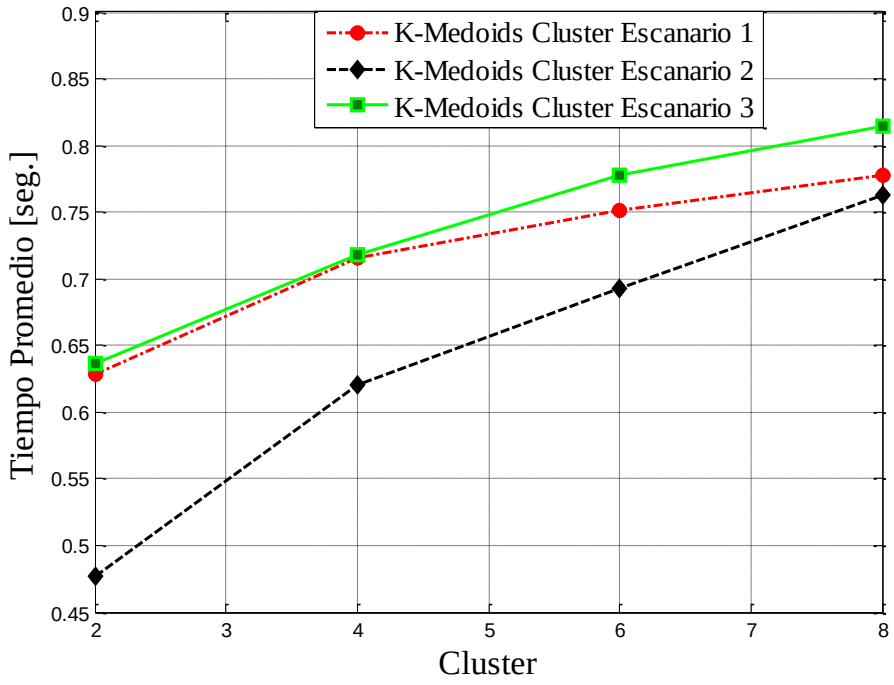

Figura 21.Curvas de los tiempos promedio de los 3 escenarios para el agrupamiento k-medoids

En la figura se ve claramente que los tiempos de agrupamiento del escenario 1, que es el que tiene menor cantidad de puntos, sin embargo, no son los más bajos, siendo los de menor retardo los del escenario 2, en todos los casos el tiempo promedio se va incrementado conforme mayor cantidad de grupos se obtenga, estos valores se muestran en la Tabla 7.

TABLA 7. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 21

Escenarios	Tiempo promedio en segundos clúster 2 grupos	Tiempo promedio en segundos clúster 4 grupos	Tiempo promedio en segundos clúster 6 grupos	Tiempo promedio en segundos clúster 8 grupos
1	0,62881398	0,716075638	0,751798662	0,777205497
2	0,477488823	0,62080013	0,692508816	0,762569431
3	0,684630089	0,743263231	0,782283264	0,782283264

La diferencia entre los casos no es considerable entre un número de grupos y otro, si comparamos con el K-Means este presenta mayor tiempo de procesamiento pero aun así puede ser considerado este algoritmo como ágil.

Los resultados de los tiempos promedios al agrupar los datos de los escenarios mediante Mean–Shift se muestra en la Figura 22.

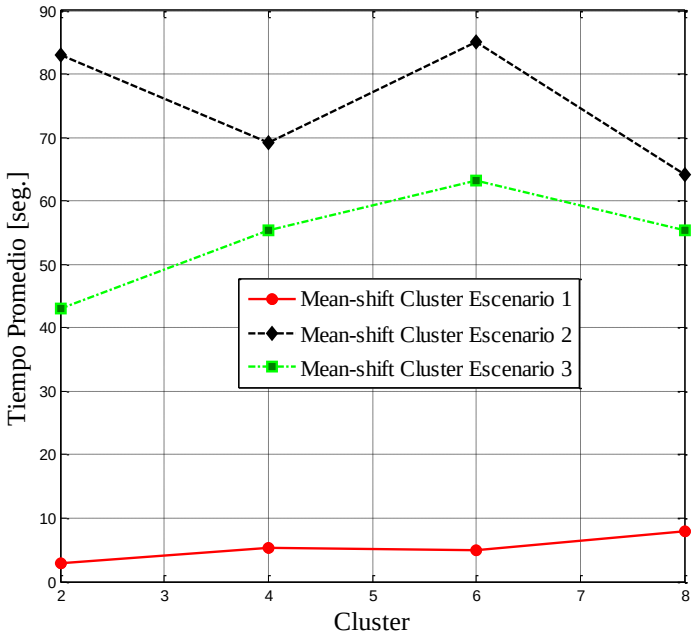

Figura 22. Curvas de los tiempos promedio de los 3 escenarios para el agrupamiento Mean Shift

Los tiempos son considerables con relación de los otros algoritmos ya que como se mencionó la naturaleza de este método lo hace muy lento pero muy predecible a la vez. El escenario 1 muestra retardos muy inferiores si los comparamos con los dos restantes, además es el más estable es decir en todos los casos se ejecutó casi en tiempos similares, por otro lado el segundo escenario es el que mayores pérdidas en tiempo presenta, mismos que se detallan en la Tabla 8.

TABLA 8. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 22

Escenarios	Tiempo promedio en segundos clúster 2 grupos	Tiempo promedio en segundos clúster 4 grupos	Tiempo promedio en segundos clúster 6 grupos	Tiempo promedio en segundos clúster 8 grupos
1	2,83321643	5,24611689	4,99383661	7,987611452
2	83,05376585	69,188864	84,99326718	64,07616501
3	43,074035591	55,277516707	63,262423181	55,312415260

El menor tiempo se presenta en el caso de tener dos clústeres con los datos del escenario 1 con un valor de 2,8 segundos y el máximo en el arreglo de seis grupos con el segundo escenario con 85 segundos es decir, el resto de tiempos es también considerable siendo este factor un aspecto importante ah tener en cuenta cuando utilizamos estos algoritmos en aplicaciones de ingeniería como es el caso de redes de telecomunicaciones, estos tiempos influencian en los tiempos totales de procesamiento.

Cabe recalcar que este tiempo depende de la velocidad de procesamiento de la PC en la cual se está corriendo en este caso fue realizado en una laptop DELL AMD Athlon II Dual–Core (2.30GHz) con una RAM de 4 GB.

2.5.2 Intervalos de Confianza

Otro aspecto importante a ser analizado y que nos dará una idea más clara de que tan efectivo son estos algoritmos es el intervalo de confianza de la cantidad de miembros que contiene cada grupo en las diferentes iteraciones, es decir, la métrica está orientada a mirar el número máximo y mínimo de puntos con respecto a un promedio sostenido durante las 10 iteraciones en los diferentes casos como son clústeres de 2, 4,6 y 8.

Manteniendo la misma secuencia anterior se estudiaran los tres métodos uno para cada escenario.

La Figura 23 muestra las curvas del algoritmo K–means con los datos del escenario1 y sus intervalos de confianza con el número de miembros que contiene cada grupo.

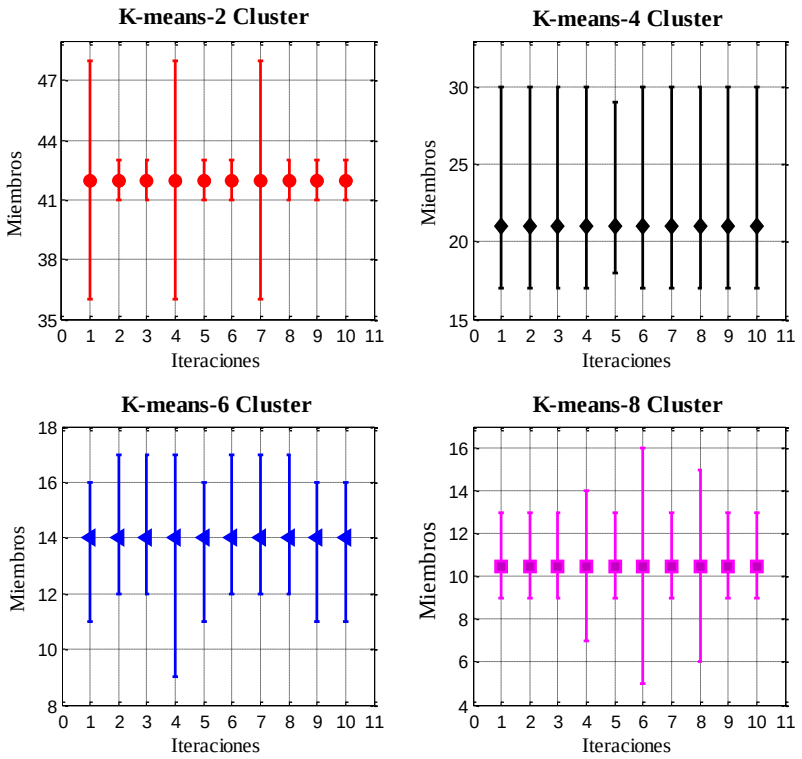

Figura 23. Intervalos de confianza de miembros del escenario 1 en cada grupo mediante k-means

Lo expuesto en la Figura 23 revela que en el caso de tener dos agrupaciones, se encuentra entre un máximo de 48 y un mínimo de 36 miembros por clúster en la iteración 1,4 y 7, mientras que en el intento 2, 3, 5, 6, 8, 9 y 10 un máximo y mínimo 43 y 41 puntos respectivamente, teniendo como referencia un promedio de 42 para todos los casos.

Cuando se tiene 4 grupos tenemos un máximo 30 y un mínimo de 17 para todas las iteraciones excepto para la quinta en la cual tenemos un intervalo de confianza entre 15 a 29 con un promedio de 21 miembros, es decir, los máximos y mínimos son más constantes que en el caso anterior no obstante es menos equitativo al distribuir los datos en cada grupo.

A medida que se requiera de un mayor número de aglomerados la confianza se vuelve menor al momento de aplicar la técnica. Como es notorio el en el K–Means–6 Clústeres y más aún en el K–Means–8 Clústeres, este presenta en las iteraciones 1, 2, 3, 4, 7, 9 y 10 un máximo 13 y un mínimo de 9 miembros, en la cuarta iteración se

observa un intervalo de 7 a 14, mientras que en la sexta iteración de 5 a 16, en la octava de 6 a 13 con un promedio de 10,5 miembros.

En la Figura 24 se muestran los intervalos de confianza pertenecientes al algoritmo K–Medoids con el escenario 2.

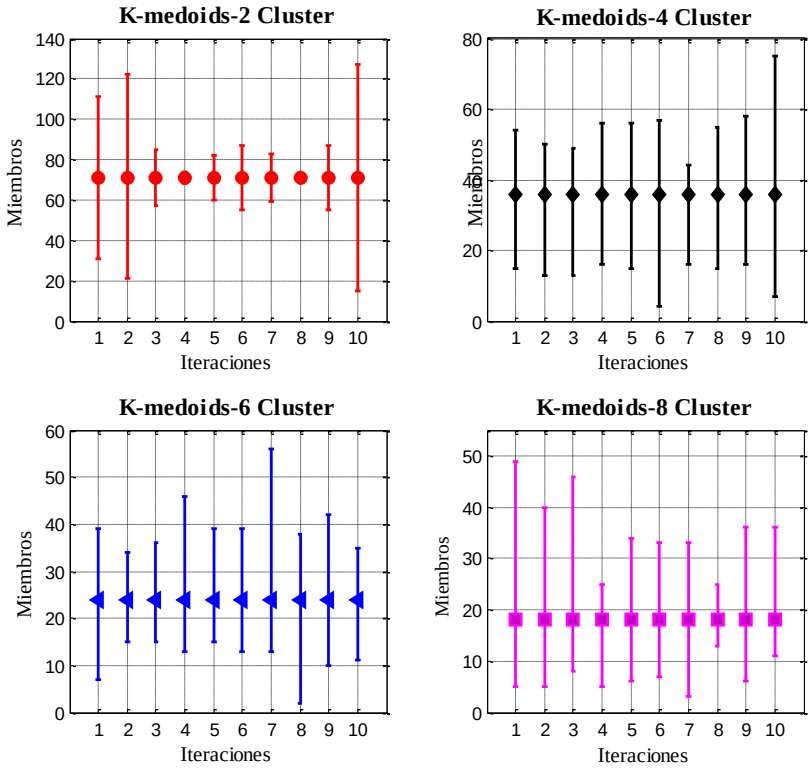

Figura 24. Intervalos de confianza de miembros del escenario 2 en cada grupo mediante k-medoids

La curva de color rojo muestra un intervalo de confianza de un valor máximo de integrantes dentro de un clúster de 127 y un valor mínimo de 15 miembros en el peor de los casos el cual se da en la iteración número 10 tomando como referencia un promedio de 71, el mejor arreglo es en la octava corrida con un arreglo perfecto, es decir, con 71 en los dos grupos, en la iteración 1 se tiene una diferencia considerable de igual manera en la segunda a diferencia de las restantes: 3, 4, 5, 6, 7, 9 en las cuales se tiene un resultado aceptable.

En la segunda grafica al tener un arreglo de cuatro clúster tomando como promedio un valor de 35.5 puntos, el mejor resultado se da en la iteración siete con un máximo de 44 y un mínimo de 16 y el peor en la corrida 10 como en el caso anterior, cuyos valores son 75 y 7 como máximo y mínimo respectivamente, con en el resto de las

corridas no se tienen buenos resultados en relación a la distribución equitativa de los miembros.

Considerando la tercera gráfica, al tener 6 grupos el mejor resultado se da en la iteración 2 obteniendo un valor máximo de 35 y mínimo de 15 miembros, tomando como referencia para cada uno un valor promedio de 24 miembros.

Por ultimo en la gráfica de color magenta, 8 clústeres, el mejor resultado se tiene en el intento 8 con un máximo de 25 y un mínimo de 13 con un promedio referencia de 18, es decir, una diferencia de 5 miembros en el mínimo y de 7 en el valor máximo y el resultado más pobre en cuanto a este análisis es la primera iteración con un máximo de 49 y un mínimo de 5.

Los intervalos de confianza para el tercer escenario con Mean–Shift se muestran en la Figura 25 y se analiza para la misma cantidad de agrupaciones de los casos anteriores.

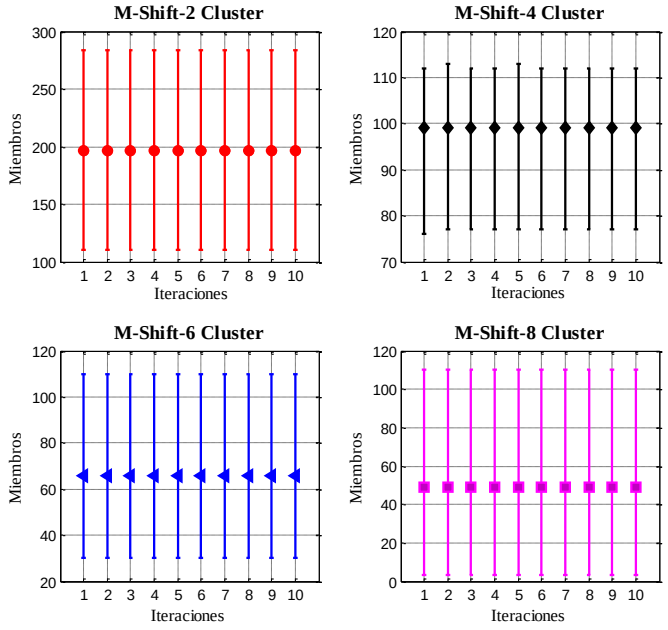

Figura 25. Intervalos de confianza de miembros del escenario 3 en cada grupo mediante Mean Shift

En este caso en todas las iteraciones se dan los mismo resultados ya que es un algoritmo predecible, es decir, el arreglo encontrado es el mismo indiferente del número de corridas ya que esta es una propiedad del mismo pudiendo tener la certeza en los resultados ya que los mismos dependerán de los anchos de banda elegidos no

como en los casos anteriores que se tenían diferentes resultados parea cada iteración, depende del valor de este ancho de ancho de banda para encontrar el mejor resultado, es decir, el valor de este deberá ser escogido con suma cautela y de acuerdo al caso.

El valor máximo en la primera grafica es de 284 y el mínimo de 110 miembros con un valor de referencia de 197, es decir, el arreglo perfecto cuando todos los grupos tengan el mismo número de miembros.

Al pretender tener cuatro agrupamientos tenemos un intervalo de confianza con valores de 112 como máximo y un mínimo de 77 miembros con un promedio óptimo de 99 en caso de darse una distribución perfecta.

En la tercera grafica correspondiente a 6 clústeres, se obtuvo un máximo de 110 y un mínimo de 30, teniendo ya una diferencia considerable entre estos, pudiendo decirse un mal resultado.

En el caso de dividir los datos en 8 grupos muestra que el grupo que más llega a tener es de 110 miembros y un mínimo de 3 miembros con una referencia de 49 miembros.

En todas las gráficas el valor de 110 miembros se repite independientemente del número de grupos, es decir, este algoritmo considera que ese número de miembros debe pertenecer al mismo grupo en algún momento.

En el experimento se ha considerado como valores óptimos, si el grupo distribuye los datos en grupos con el mismo número de miembros si es posible, para cada uno, o muy cercanos entre sí, no obstante; cada algoritmo decide a que grupo pertenece uno u otro miembro, de acuerdo a sus condiciones y reglas, algunos consideran la media como es el caso del K–Means, otros consideran las medias entre los datos para ubicar los centroides , otros mediante los medoides y otros una serie de aspectos estadísticos como es el caso del Mean–Shift, por lo tanto, los grupos no deben ser equitativos sino de acuerdo a las condiciones y a donde se pretenda utilizar estas herramientas de agrupamiento.

2.5.3 Número de Miembros promedio de acuerdo al número de Grupos

En este apartado se analiza el número de agrupamientos y el valor promedio de los miembros durante todas las 10 iteraciones para cada caso de clúster.

En este sentido el primer caso de análisis cuando se tiene el algoritmo K–Means agrupando los datos del escenario 1 se muestra en la Figura 26.

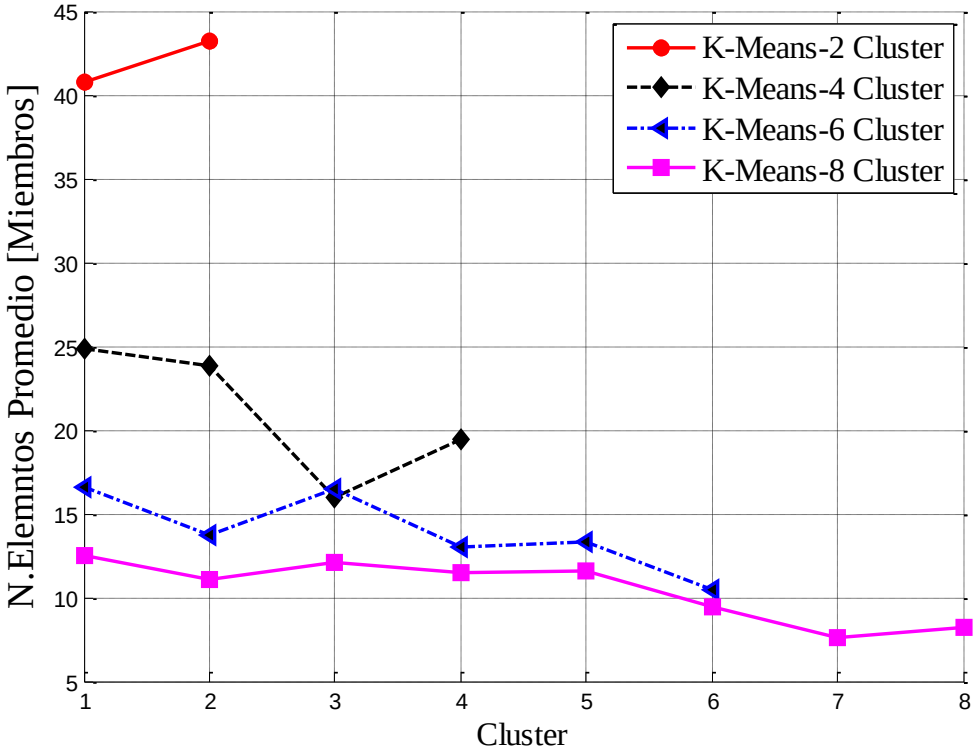

Figura 26. Número de miembros promedio del escenario 1 en cada grupo mediante K–Means

De lo observado en la Figura 26 en promedio este algoritmo distribuye a los miembros de una manera casi simétrica entre grupos teniendo el mejor resultado en el arreglo de ocho grupos, los datos de esta figura se muestran en la Tabla 9.

TABLA 9. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 26

Casos	Clúster 1	Clúster 1	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
2	42,9	41,1						
4	20,8	21,6	21,4	20,2				
6	13,4	14	13,9	14,3	15	13,4		
8	10,4	10,2	10,9	10,1	11,3	9,9	11,2	10

En la primera curva en la que se tiene a todos los datos en dos grupos, el primer clúster alberga a 43 puntos y el segundo a 42, se tiene una distribución casi perfecta. Cuando se tiene 4 grupos se observa: 29, 22, 21, y 20 miembros para los clústeres 1, 2, 3 y 4 respectivamente teniendo un resultado aceptable. Al agrupar los datos en 6

grupos tenemos 13, 14, 14, 14, 15 y 13 para los clústeres 1, 3, 4, 5 y 6, finalmente en el caso de agrupar los datos en un número mayor de conglomerados en este caso de 8 clústeres como se observa se tiene una gráfica casi lineal, es decir, este algoritmo está distribuyendo los datos de una manera equitativa y conveniente para el caso.

Este método arrojo resultados excelentes para este caso puesto que tenemos pocos datos, no obstante, como se observara más adelante a medida que se aumenta la cantidad de datos el algoritmo tiende a bajar su efectividad, sin embargo, aún se tiene resultados aceptables con un tiempo mínimo de ejecución, factor de suma importancia en este caso de estudio.

La Figura 27 contiene las curvas del método K–Medoids con los datos del escenario 2.

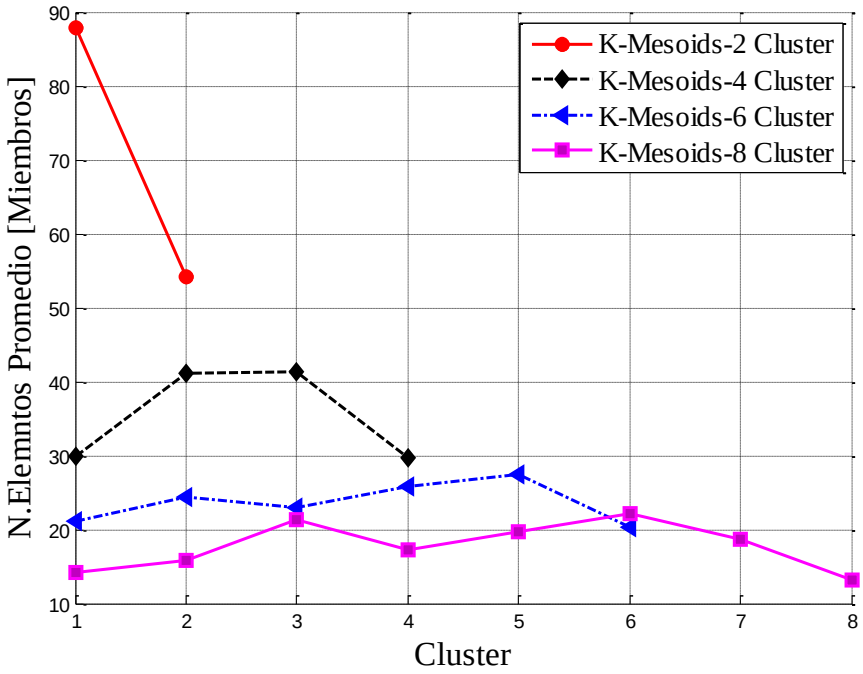

Figura 27. Número de miembros promedio del escenario 2 en cada grupo mediante K-medoids

Como es notorio en el caso de la primera curva K–Medoids de dos agrupamientos los datos agrupados en uno y otro grupo no son uniformes, dándose en el segundo caso uniformidad de dos en dos, es decir, el primer grupo con el cuarto y el segundo con el tercero, en el último caso cuando se tiene un arreglo de 8 conglomerados se presenta una curva aceptable, estos datos están detallados en la Tabla 10.

TABLA 10. VALORES DE LOS TIEMPOS OBTENIDOS DE LA Figura 27

Casos	Clúster 1	Clúster 1	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
2	87,9	54,1						
4	29,8	41,1	41,4	29,7				
6	21,2	24,3	23	25,8	27,4	20,3		
8	15,9	21,3	17,2	19,6	22,1	18,6	13,1	15,9

El experimento muestra resultados favorables en el caso de 6 y 8 clústeres, en comparación con los arreglos de 2 y 4 clústeres.

Al realizar este análisis con el algoritmo de Mean–Shift con los datos del escenario 3 se tienen las curvas representadas en la Figura 28.

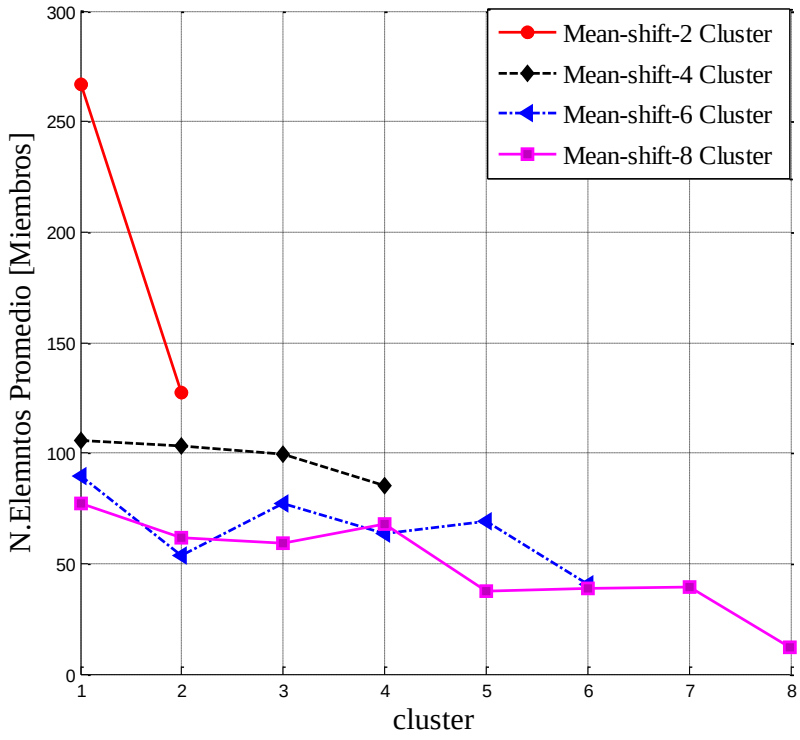

Figura 28. Número de miembros promedio del Escenario 3 en cada grupo mediante Mean–Shift

En la primera curva de la Figura 28, se muestra una amplia diferencia en el caso de obtener dos clústeres, los resultados de 4 arreglos obtiene una curva casi uniforme, mostrándose el peor de los caso en la curva de color violeta que corresponde al arreglo de 8 clústeres especialmente en el último grupo como se ve en la Tabla 11.

TABLA 11. VALORES DE LOS TIEMPOS OBTENIDOS DE LA FIGURA 28

Casos	Clúster 1	Clúster 1	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
2	266,6	127,4						
4	105,7	103,3	99,5	85,5				
6	89,7	53,7	77,2	63,6	69,1	40,7		
8	77,2	61,9	59,1	68,1	37,4	38,7	39,3	12,2

La diferencia del grupo dos en el primer caso es más del doble es decir, 267 en comparación con 127 de otro grupo, el segundo caso muestra que el grupo 1 y el 2 son casi iguales así como el clúster 3 con el cuarto, es decir, 106, 103 y 99, 86 respectivamente, en el tercer caso el grupo que más miembros tiene es primero y el clúster 8 el que tiene menor cantidad con una diferencia de 65 puntos entre estos.

2.6 Análisis comparativo de los algoritmos K–Means, K–Medoids y Mean–Shift

Complejidad: El algoritmo K–Means normalmente requiere sólo un número $\boldsymbol{O(kN)}$, de operaciones, donde ***k*** es el número de agrupaciones y ***N*** el número de puntos, por lo tanto el algoritmo K–Means se puede aplicar con facilidad a un gran conjunto de datos. Fácil su interpretación e implementación.

En el algoritmo K–Medoids la complejidad aumenta con $\boldsymbol{O(ik(n-k)^2)}$, donde *i* es el número total de iteraciones, ***k*** es el número total de agrupaciones, y ***n*** es el número total de objetos.

El algoritmo de cálculo Mean–Shift es muy intensivo, pues requieren de $\boldsymbol{O(kN^2)}$ operaciones que son principalmente los cálculos de la distancia euclidiana,) donde **N** es el número de puntos de los datos y ***k*** es el número de pasos de iteración promedio para cada punto de datos, por lo cual Mean Shift tarda mucho más tiempo en ejecutarse a medida que se incrementa el número de puntos y, por tanto, es menor en escalabilidad que el K–Means.

Parámetros de ingreso: Para poder emplear el algoritmo K–Means así como el método de K–medoids, es necesario tener predeterminado el número ***K*** de grupos.

Mean–Shift solo necesita un parámetro de entrada que es el ancho de banda de la ventana de desplazamiento. Este algoritmo presenta una ventaja sobre los otros

algoritmos, y es el hecho de que no necesitan asumir el número de grupos, sino que son capaces de identificar los grupos existentes sin necesidad de este parámetro.

Efectividad y forma de agrupar: Como se evidencio durante lo experimentación, la técnica de K–Means es la que muestra una mayor uniformidad al momento de distribuir los datos dentro de los grupos, este método se realiza a través de la media del conjuntos de datos y así asociándolos a cada una de ellas para cada grupo, en cambio, con K–Medoids no se mostraron mucha uniformidad en la distribución en algunos casos, esta técnica se realiza a través de los medoides.

Mean–Shift no reconoce el grupo, sino que divide el grupo en distintos subgrupos. Esto sucede, a causa de que el algoritmo Mean–Shift ubica los grupos asociando cada uno con el punto en la zona de mayor densidad. Es decir, supone que los grupos tienen una estructura con una parte más densa y otra menos densa, y cada punto del grupo es asociado con el centro de la zona de mayor densidad.

Mean–Shift obtiene los grupos asociando cada punto a un pico o moda que se encuentra en la región de máxima densidad. Mean–Shift es capaz de descubrir grupos de formas arbitrarias.

Forma de los Datos: Tanto la técnica de K–Means así como K–Medoids no es adecuado para resolver las agrupaciones con forma no convexa, o grupos de muy diferente tamaño. Mean–Shift es capaz de descubrir grupos de formas arbitrarias.

Sensibilidad al ruido y outliers: El método de K–Means utiliza un centroide para representar al grupo y este es sensible a los valores outliers o atípicos. Esto significa, que un conjunto de datos con un valor extremadamente grande puede perturbar la distribución de estos. Ello conlleva a que este algoritmo sea incapaz de manejar datos ruidosos y valores atípicos.

El método K–Medoids supera este problema mediante el uso de medoides para representar el clúster en lugar de centroide. Un medoide son los datos más céntricos de un clúster. Aquí, se seleccionan los ***k*** conjuntos de datos aleatoriamente como medoides para representar ***k*** clúster y permaneciendo todos los objetos de datos se colocan en un clúster que tiene el medoide más cercano (o la mayor parte similar) para ese grupo de datos. Después de procesar todos los objetos de datos, un nuevo medoide es determinado que puede representar a un clúster de mejor manera y todo el proceso se repite. Una vez más todos los conjuntos de datos se unen a los grupos

basados en los nuevos medoides. En cada iteración los medoides se mueven. Este proceso continúa hasta que no exista cualquier otro movimiento en los medoides. Dando como resultado, los ***k*** grupos encontrados y representados en un conjunto ***n*** integrado de ***k*** subconjuntos de datos.

K–Medoids es más robusto que el K–Means en la presencia de ruido y valores atípicos; porque un medoide es menos influenciado por los valores atípicos u otros valores extremos que una media.

Mean–Shift clustering no es capaz de reconocer ruido pero, como se muestra en [12], donde se dice que se puede hacer una aproximación de los grupos encontrados cuyo número de puntos que sean inferiores a un cierto umbral y estos serán clasificados como puntos de ruido.

Costos de Tiempos y Escalabilidad: Como se observó en el apartado anterior de experimentos, el que menor tiempo de ejecución consume es el del algoritmo K–Means, así como en la escalabilidad ya que en todos los casos mostraron tiempos aceptables es decir, a medida que se incrementaban la cantidad de puntos los tiempos no aumentan considerablemente. El resultado y el tiempo de ejecución total del algoritmo K–Means como K–Medoids dependen de la partición inicial, en el caso del segundo se muestra un mayor retardo con respecto a K–Means así como menor escalabilidad. Mean–Shift tarda mucho más tiempo en ejecutarse a medida que se incrementa el número de puntos, por tanto, es el peor en escalabilidad y el que mayor tiempo consume.

CAPITULO 3. BREVE DESCRIPCIÓN DE LOS CRITERIOS Y CONSIDERACIONES SOBRE EL MODELO MATEMÁTICO EMPLEADO EN LA PLANEACIÓN DE REDES PON.

En un despliegue de red el objetivo final consiste en generar un proyecto óptimo y económico, y como se ha revisado anteriormente un modelo hibrido PON WDM/TDM promete sólidas y ventajosas aplicaciones en la transmisión de datos, no obstante el éxito de los resultados se acentúa en la manera de gestionar este modelo, lo que se logra a través de los métodos de agrupamiento, los mismos que basan su funcionalidad sobre modelos matemáticos que describen condiciones, complejidades y técnicas propias de cada red,

3.1 Consideraciones topológicas y técnicas de las redes PON.

Como ya se ha establecido, una red PON al ser una Red Óptica pasiva es necesario conocer sus características técnicas y funcionalidades para poder plantear un modelo algorítmico sobre ella.

Consideraciones

Es necesario establecer parámetros tanto técnicos como físicos propios de cada tipo de red para poder generar un modelo, cada vez tiene mayor importancia el tiempo y la calidad de comunicación con la que se puede transmitir, y es justo aquí donde las redes PON acentúa sus criterios, pues este tipo de RED se enfatiza en el ancho de BANDA, la versatilidad y la adaptabilidad que tiene y podrá tener la implementación de esta estructura de RED. La alternativa de poder establecer un método que permita un análisis más exacto, más completo y más concluyente al momento de realizar un estudio de red, desvía por completo los métodos tradicionales que se han venido utilizando a lo largo del tiempo, y busca la utilización de un nuevo sistema que simplifique y optimice este proceso.

Una red PON es una tecnología punto-multipunto donde todas las transmisiones se realizan entre la terminal Óptica de línea (OLT) y la unidad óptica de usuario (ONU) configurando un esquema FTTx.

Pueden emplearse varios tipos de topologías para el acceso de red, (árbol, anillo, cascada, multinivel), y PON no hace la excepción ya que puede adaptarse a cualquiera de estos tipos bastando con la encadenación de divisores ópticos, pero dependiendo de ciertos factores la red de acceso OLT puede requerir protección.

Algunas consideraciones topológicas y técnicas de una red PON son:

- Toda topología PON utiliza monofibra para su despliegue
- En canal descendente PON es una red punto-multipunto donde la OLT maneja la totalidad del ancho de banda que se reparte a los usuarios.
- En canal ascendente PON es una red punto-punto donde múltiples ONUs transmiten a una única OLT.
- Una red PON maneja multiplexacion TDMA y WDMA, con el fin de contener al máximo el coste de la optoelectrónica.

Sin embargo en términos físicos los aspectos a considerar al momento de estimar un modelo de red son:

- *Geografía del sector*: Que describe el perímetro o área de cobertura donde se pretende plasmar la red, por lo general estará indicada por coordenadas georreferenciadas.
- *Cantidad de ONUs*: Hace referencia a los usuarios activos a los que se pretende brindar servicio.
- *Cantidad de enlaces*: Representa todas las posibles interconexiones entre todos los diferentes puntos georreferenciados que constituyen el sector de cobertura.

Considerando los puntos anteriores, se puede sumar un aspecto técnico que se debe de plantear en el estudio de tecnologías PON, que es el de la potencia de transmisión, ya que como bien se conoce en redes de telecomunicación, la distancia entre central y abonados muchas veces ha estado limitada por la potencia de transmisión, donde se consideraba evitar saturar el fotodiodo transmisor lo que obliga a que el diseño y estructura de red varíe en base a la distancia entre usuarios y central, es decir que

mientras más lejos se encuentre el usuario mayor será el despliegue de fibra y mayor será el servicio exigido lo que representa un aspecto importante al momento de decidir la posición de la central y acopladores ópticos, en adición también se debe tomar en cuenta los efectos No Lineales y de Dispersión Cromática como lo descrito en [28], donde se plantea como importantes estos dos factores ya que limitan la máxima tasa de bits y alcance en la transmisión de información qua viaja a través de la fibra óptica. Los efectos No Lineales y de Dispersión son efectos presentes en la funcionalidad de red que están específicamente en los canales de fibra durante todo el tiempo de transmisión.

PON considera todos estos aspectos, incluyéndolos en sus transceptores ópticos brindando una notable simplificación a la electrónica anteriormente establecida necesaria para actuar sobre un control externo del transceptor, lo que representa una nueva óptica que miniaturiza, integra y simplifica el trabajo con ráfagas de diferente nivel de potencia.

Para poder plantear un método nuevo y simplifacativo en el diseño e implementación de red es necesario las consideraciones anteriores acerca de la tecnología PON, pero también hay que tomar en cuenta otros aspectos propios y externos de la red. Ambas consideraciones trabajan a la par, pues como características propias se tiene todo el estudio matemático que define, plantea y hace funcional a una red PON, la misma que brindara su máxima aplicabilidad gracias al modelo más compacto en el que se acentúe y considere de mejor manera todos los requerimientos de funcionalidad , como son distancias entre usuarios, topología del terreno, ubicaciones georreferenciadas de escenarios reales y herramientas computacionales que procesen esos aspectos, donde el trabajo armónico en conjunto generará el método óptimo de funcionalidad.

3.2 Descripción del problema en la Planeación de redes PON.

Para ilustrar la realidad del problema de una manera más robusta en el momento de analizar un estudio de despliegue de red, se describirá dos trabajos realizados por: GU Rentao, LIU Xiaoxu, LI Hui, BAI Lin [29] y Attila MITCSENKOV, Géza PAKSY, Tibor CINKLER [30], las dos contribuciones parten de diferentes enfoques pero engloban muy significativamente las condiciones y restricciones al momento de

estudiar un red, describiendo de una manera amplia y matemáticamente la problemática en la planeación de redes PON.

Partiendo de [29], en un proyecto es necesaria la planeación de Red más allá de haber definido qué tipo de red o aplicabilidad de la misma, un aspecto de mayor importancia es el saber la ubicación y distribución de los elementos ópticos pasivos más precisa para dar mayor y mejor cobertura a un menor costo en despliegue de red, y para ello se utiliza un modelo que describa el sector, las condiciones y características donde se va a desplegar la red.

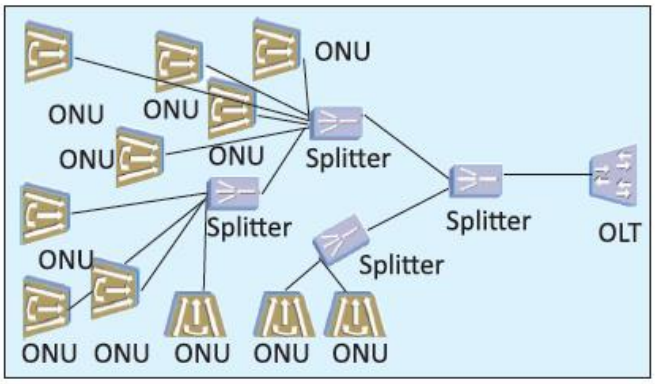

Figura 29. Red Óptica Pasiva en cascada [29]

En la Figura 29 se puede observar una red óptica pasiva multinivel (PON), donde en base a la ubicación de las ONUs se debe encontrar la mejor configuración para poder conectarlas hacia los splitters y hacia la OLT. Buscando la mejor alternativa ante ello se podría decir que la "mejor configuración" y más económica seria conectar cada usuario (ONU) a un splitter y este a su vez a través de un canal activo conllevarlo a su central (OLT), a esto hay que sumar el costo de implementación que estaría dado por un costo de la OLT, el cable de fibra, elementos pasivos, entre otros y muy aparte el costo laboral, materiales de obra civil, licencias etc.

Por otro lado en el momento de realizar la planeación de red pasiva óptica existen consideraciones técnicas que pueden aumentar o no el coste del proyecto tales como:

- El presupuesto de energía por potencia de transmisión.
- Distancia de transmisión lo que relaciona un costo con el proveedor de servicio.

- El diferencial de distancia entre ONUs y OLT, que representa un ancho de banda.
- Elementos pasivos de red en base a las condiciones geográficas del proyecto.

Todos estos aspectos sumados pueden definir un proyecto de planeación de red, pero la verdadera interrogante nace en saber por dónde y cómo empezar a planear, conllevado obviamente de una estimación económica para solventar el proyecto así como de la mejor decisión para optimizar el mismo, dando como punto de partida este problema [29].

3.3 Planteamiento del problema en la Planeación de redes PON.

3.3.1 Formulación del problema y análisis de restricciones del modelo matemático descrito.

En [29] describe que al existir un determinado sector o área donde se desea brindar servicio de cobertura a través de una red PON, primero es necesario establecer los parámetros indicados, sean estos individuales o en conjunto, en redes PON los parámetros se acentúan en la ubicación de los elementos de red que se establecerá como:

- El sitio de la OLT o conjunto de OLT
- La ubicación de los splitters
- La ubicación de las ONUs

A todos estos conjuntos de ubicaciones el autor las representa por (Ω), así mismo al conjunto de posibles candidatos para ubicar splitter lo denomina Ω_{CS}, al de posible ubicación de ONUs Ω_{ONU}

Es importante recordar que todas las posiciones son candidatos posibles de ubicación por lo que al conjunto de posibles ubicaciones del splitter será cambiante en torno a la cartografía y dimensiones del espacio predeterminado [29].

Una vez que se establecen siglas y denominaciones, el autor plantea la necesidad de los puntos de ubicación o coordenadas de cada nodo, estos a su vez estarán presentes

en el plano o en el sector donde se plantea el estudio, a estos puntos se los representara por (Vs), que contiene la ubicación de cada nodo, posterior a esto se puede considerar los enlaces que representaran el manzano o conjunto de nodos unidos que dan forma al perfil geográfico, a estos enlaces se los representa como (Es). También necesario preestablecer la forma de agrupar los nodos y eso se manifiesta atreves de (T) que describe la unión de nodos que puede ser (1–4, 1–8, 1–16, 1–32, 1–64, 1–128).

Una vez que se tiene los parámetros identificados empieza el proceso matemático que establece las bases para el modelo de optimización.

Rentao, plantea el uso de matrices que calculan la distancia y centroides para cada figura que se representa a través de sus nodos y enlaces.

La matriz que calcula las conexiones entre OLT, SPLITTERS, y ONUs está dada por:

$$D_{OC} = \left(d_{1i}^{OC}\right)1xm \tag{36}$$

Donde de d_{1i}^{OC} indica la distancia de un posible candidato hasta la OLT

$$D_{OC} = \left(d_{1i}^{OC}\right)mxm \tag{37}$$

De la ecuación anterior si $d_{li} = 1$ significa que existe una conexión directa desde ese candidato existente dentro del conjunto (m x m) hacia la OLT. Otra ecuación necesaria para el autor, es la que describa una conexión entre los posibles candidatos existentes o nodos y para ello nombra un nodo (i) y otro nodo (j).

$$D_{CC} = \left(d_{ji}^{CC}\right)mxm \tag{38}$$

Donde si $d_{ji} = 1$, significa que existe una conexión directa entre el nodo (i) al nodo (j), convirtiéndose en una conexión en cascada entre los splitter.

Ahora si bien es cierto estas matrices representan solo ubicaciones o posibles candidatas, el autor plantea otras consideraciones para proseguir con el modelo matemático.

Entre estas consideraciones tenemos:

- **Posición de los parámetros.** Hace referencia a la ubicación (X_0, Y_0) o coordenadas donde se encontrara cada OLT , asi mismo la posición de un sitio candidato estará representada por $\left(X_i^S, Y_i^S\right)$, mientras que la posición de la ONU estará dada por $\left(X_j^O, Y_j^O\right)$.
- **Potencia de los parámetros**. Aquí tenemos a la potencia de atenuación de fibra por kilómetro(km) dada por (A), $\frac{dB}{km}$, así mismo también se considera la potencia de transmisión (P_B), la potencia de perdida en elementos pasivos (P_S).
- **Trafico de parámetros.** Se considera en esta etapa, canales de subida (N_{up}), canales de bajada (N_{down}), y el ancho de banda.
- **Costo de parámetros.**

  Costo de tipo de splitter $\mathrm{C_t^S}$.

  Costo por cable de fibra $\mathrm{C_b^f}$.

- **Parámetros de distancia.** Aquí se considera la máxima distancia de transmisión l_{max}^{dis} entre la OLT y la ONU, además se refiere cada tramo de un nodo (i) a otro nodo (j) este destinado a la ONU, lo que denomina distancia total de nodo l_{ij}^{total} , naciendo una primera restricción i > j, lo que significa que para que exista un enlace debe existir una distancia entre un nodo y una ONU, además de un factor de costo por cada enlace existente que se representa por η_{ij}.

Una vez que se tiene todos estos parámetros, lo que se ilustra en [29] es minimizar todas las ecuaciones a sumatorias de todos los datos con los parámetros predefinidos, representada por un modelo de programación lineal entera (ILP):

(39)

$$\sum_{i\in\Omega_{CS}}(S_{it}C_t^S)+\sum_{i\in\Omega_{CS}}(C_0^f+C_b^f)d_{li}^{OC}l_{1i}\eta_{1i}+\sum_{i,j\in\Omega_{CS}}(C_0^f+C_b^f)d_{ji}^{CC}l_{ji}\eta_{ji}$$
$$+\sum_{j\in\Omega_{CS};k\in\Omega_{ONU}}(C_0^f+C_b^f)d_{jk}^{CO}l_{jk}\eta_{jk}$$

Reduciendo a:

$$\sum_{i\in\Omega_{CS}}d_{li}^{OC}=1 \tag{40}$$

$$d_{ji}^{CC}+d_{ij}^{CC}<2,\qquad i,j\in\Omega_{CS} \tag{41}$$

$$\sum_{j,k}d_{jk}^{CO}=1(\forall j\in\Omega_{ONU}) \tag{42}$$

$$\sum_{K\in\Omega_{ONU}}d_{jk}^{CO}+\sum_{i\in\Omega_{CS}}d_{ji}^{CC}+d_{1j}^{OC}\le\sum_{t\in T}S_{jt}\,T_t+1,\qquad(\forall j\in\Omega_{CS}) \tag{43}$$

$$\sum_{K\in\Omega_{ONU}}d_{jk}^{CO}+\sum_{i\in\Omega_{CS}}d_{ji}^{CC}+d_{1j}^{OC}\ge\sum_{t\in T}(S_{jt}+1)S_{jt}\ ,(\forall j\in\Omega_{CS}) \tag{44}$$

$$\sum_{j\in\Omega_{CS}}d_{ji}^{CC}+d_{1i}^{OC}=\sum_{t\in T}S_{it},\quad(\forall i\in\Omega_{CS}) \tag{45}$$

$$\sum_{t\in T}S_{it}\le 1,\qquad(\forall i\in\Omega_{CS}) \tag{46}$$

$$l_{1_i}^{total}=\sum_{e_{jk}\in E_{1i}}l_{jk} \tag{47}$$

$$P_{1_i}^{total} = \sum_{e_{jk} \in E_{1i}} l_{jk} A_0^f + \sum_{V_j \in V_{l1} t \in T} A_t^S S_{jt} \quad (48)$$

$$l_{1_i}^{total} \leq l_{max}^{dis}, \quad (\forall i \in \Omega_{ONU}) \quad (49)$$

$$P_{1_i}^{total} \leq P_{B,} \ (\forall i \in \Omega_{ONU}) \quad (50)$$

$$\sum_{1 \leq i \leq n} B_{0,i} \leq B_{down} \quad (51)$$

$$\sum_{1 \leq i \leq n} B_{0,i} \leq B_{up} \quad (52)$$

Con las ecuaciones anteriores se tiene la longitud total de fibra para el despliegue, el mismo que deberá representar un costo por ello, además se puede ver que el autor plantea restricciones que garantizan la topología de red como acíclica y conectada. La restricción (40) garantiza que sólo hay un divisor conectado con la OLT . La restricción (41) tiene como objetivo evitar un bucle local entre un par de nodos, la restricción (42) plantea que cada ONU conecta a uno y sólo un divisor, así mismo las restricciones (43-44) garantizan el grado del divisor, que sea adecuado para la correspondiente relación de split . La restricción (44) también indica que un divisor debe tener un sucesor, y a través de la restricción (45) describe que cada divisor tiene un precursor . La restricción (46) significa que no se asigna más de un tipo a un sitio candidato, es decir , habrá como máximo un divisor en cada candidato o sitio, dado que es preferible colocar un divisor con relación divisor más grande en lugar de dos más pequeñas, la restricción (46) es consistente con el práctico funcionamiento de la red real, en cambio las restricciones (40-45) son un conjunto de restricciones para mantener la topología acíclica y asegurarse de que todas las unidades ONU , divisores y OLT se lleguen a conectar . Las restricciones (49-50) forman las limitaciones de distancia y limitaciones presupuestarias que en combinación con las restricciones (40-45) determinan si un ruta establecida satisface el requisito de la distancia y el presupuesto de alimentación, mientras que

las restricciones (51-52) garantizan que el ancho de banda tanto de subida como de bajada que exijan la OLT y las ONUs sea satisfactorio.

3.3.2 Complejidad del Modelo Descrito

Hay que recordar que en el momento de analizar el despliegue de red, se plantea como candidatos de soluciones a todos los puntos que se encuentran en el área, los mismos que pueden ser cambiantes a lo que el autor los llama código binario pues puede ser un "1" y activarse como candidato, o ser un "0" y no ser candidato, la complejidad del modelo radica en el aumento de nodos y usuarios, ya que aumentan los candidatos posibles a lo que se denomina un problema NP completo, debido a que la solución crece de manera exponencial, y matemáticamente se imposibilita la solución por la cantidad de expresiones a resolver, a lo que una nueva solución computacional enfrentaría de mejor manera la problemática existente, mediante el uso de heurísticas que llevan el tamaño del problema a convergencia polinomial.

3.4 Descripción del problema en la planeación de redes PON (segundo modelo)

En un segundo documento [31] , el autor analiza un enfoque en el que manifiesta que las tecnologías de red de acceso están desplegando altos requerimientos de ancho de banda y crecientes necesidades de tráfico en tiempo real y fiabilidad donde las operadores de red pueden disminuir la administración, mantenimiento y gestión de gastos, por ofrecer un servicio de triple play en una sola red, dando el potencial de reemplazar las redes separadas para voz, datos y el tráfico de vídeo, y el uso de una única red convergente con gestión de los recursos de control común, lo requerirá gran ancho de banda y bajo retardo simultáneamente.

Citando en [30] el mismo autor enfatiza como importante el diseño de topología de red, plasmándolo como clave en cuanto a negocios, donde para su análisis hay que considerar las diferentes arquitecturas, y la topología que tiene que ser ajustada a la seleccionada área de servicio, por no hablar de la elección fundamental de la correcta

tecnología. El modelo pretende englobar toda la información relevante para el proceso de diseño de la topología: los datos geográficos y de infraestructura de la zona donde la red se desplegará, las limitaciones específicas de la tecnología y los parámetros de costos, además los enlaces de red no se pueden construir donde se desee sino que debe seguir algunos caminos de cableado existentes, el sistema de calles y otras infraestructuras.

Con esta descripción del problema, la topología deseada puede interpretarse como un (costo mínimo), que conecta los usuarios del CO en un punto–multipunto (o punto a punto) de la arquitectura a través de un subconjunto de la red de enlaces, utilizando varias unidades de distribución cuando así lo permita y sea necesario, y que cumpla con las limitaciones específicas de la tecnología [30].

3.5 Planteamiento del problema en la Planeación de Redes PON (segundo modelo)

3.5.1 Formulación del problema y análisis de restricciones del modelo matemático descrito.

Para formular el problema el autor en [31] , plantea la existencia de una red G=(V,E) donde a E le asigna los posibles enlaces o interconexiones que pueda tener la infraestructura, mientras que V representa el CO, los suscriptores o usuarios y los lugares permitidos para las unidades de distribución (splitter). Aquí se manifiesta que todos los bordes E, que llegarían a ser las calles o líneas que describan el sector no deben de tener una longitud negativa, así mismo le asigna a cada uno un costo de cable de despliegue $C_o(e)$ y costo de fibra representado por $C_v(e)$, estos costos son muy comunes para analizar un despliegue de red, pero no necesariamente se preestablecen al enlace ya que dependen mucho de la tecnología e infraestructura a analizarse.

Para poder minimizar el costo de despliegue el autor plantea las siguientes ecuaciones

$$Costo\ minimo\ C = C_{equipamiento} + C_{cable} \quad (53)$$

De la ecuación (53) se puede citar en un documento posterior del mismo autor [30], donde se desglosa la obtención de la misma:

$$C_{equipamiento} = \sum_{PONs} (C_{OLT} + C_{SPLITTER}) + \sum_{USUARIOS} (C_{ONUs}) \quad (54)$$

$$C_{cable} = C_{fibra} + C_{distribucion} \quad (55)$$

Dónde:

$$C_{fibra} = \sum_{e \in fibra} (C_o^{fbra} + \sum C_v^{fibra}) \quad (56)$$

$$C_{distribucion} = \sum_{e \in distri} (C_o^{distri} + \sum C_v^{distri} \quad (57)$$

Donde minimizando se tiene

$$C = \sum_{GRUPOS} C_{DU} + \sum_{e:F(e)>0} (C_o(e) + \sum_{F(e)} C_v(e)) \quad (59)$$

En las ecuaciones (58) y (59) la función F(e) representa la cantidad de conexiones paralelas que se pueden dar entre los diferentes puntos del escenario. En cuanto a los equipos, el costo de abonado o usuario por unidades (CSU) no depende de la

topología (sólo de la cantidad de abonados); estos pueden ser retirados de la ecuación como constantes. El costo de la distribución de unidades (CDU) representa la cantidad de abonados o grupos activos dependientes de un servicio. El costo de la oficina central (CCO) depende normalmente de la cantidad de abonados y la cantidad de grupos de abonados en este último caso, puede añadirse al coste de las unidades de distribución (CDU). Por lo tanto, el costo CO también se elimina de la ecuación, y el problema se reduce a la ecuación (59).

Las restricciones específicas de este modelo son diferentes para cada tecnología y, por desgracia, requiere una gran de cantidad de variables integrales, debido a la estructura de punto-multipunto y diversidad de trayectorias limitantes, sin embargo en un anterior documento de mismo autor [31], se plantea estas restricciones

$$|V| \cdot I_e \geq X_e \tag{60}$$

$$\sum_{e \in V\rightarrow} X_e - \sum_{e \in \rightarrow V} X_e = \begin{cases} N & if v = CO \\ -SP_V & if v \neq CO \end{cases} \tag{61}$$

$$N = \sum_{V} SP_V \tag{62}$$

$$\sum_{e \in V\rightarrow} Y_e - \sum_{e \in \rightarrow V} Y_e = \begin{cases} -1 & if v = S \\ \leq SP_V \cdot D_{max} & if v \neq S \end{cases} \tag{63}$$

En las ecuaciones anteriores el autor asigna variables (X_e, Y_e), que describen el tráfico total de la red para cada enlace representado por X_e para enlaces y Y_e para los usuarios ya que la interacción entra ambos generara el flujo de red.

La representación I_e es una variable que indica los enlaces utilizados y es así que la ecuación (60) indica que puntos fueron ya utilizados como enlaces. Junto con la ecuación (61) denotan el flujo de utilización de los puntos como enlaces a través de sumatorias donde (v) representa al nodo, SPv representa el número de divisores o splitter en el nodo y N representa el número total de splitter en la red, por lo que a

través de esta expresión se interpreta que el número total de enlaces va a estar restringido por el número total de nodos y el número total de splitter, que se obtiene la ecuación (62).

Una última restricción que plantea es en la ecuación (63), donde la mejor distribución de flujo de red donde se pueda conectar todos los usuarios a un óptimo nodo o punto de acceso estará restringida por el número de splitter y el costo que representa cada uno de ellos, denotado por $SP_V \cdot D_{max}$, cumpliendo asi un modelo matemático y aproximado a un solución, eso sí dependiendo de cada espacio o lugar geográfico donde se pretenda desplegar la red. [31].

3.5.2 Complejidad del Modelo Descrito

El problema de optimización que se plantea en este modelo es que tiene que ser escalable y solvente, incluso para problemas a gran escala existentes en la vida real, donde se habla de grandes áreas de servicio, lo que resulta en miles de nodos y enlaces. El objetivo de diseñar un sistema de despliegue de red es el poder solventar de la mejor manera las necesidades de usuarios existentes en un sector, sin violar leyes ni reglamentos técnicos y sobre todo optimizando los costos en fibra y equipos.

Para poder cumplir ese objetivo es necesario procesar, analizar y plasmar la mejor ubicación de la central y los elementos de la red, y la verdadera complejidad del modelo radica aquí, pues si se tienen los puntos que forman el área de cobertura, lo óptimo es ubicar en menor tiempo y mayor precesión dichos elementos de red, pero al encasillarse a un proceso netamente matemático es necesario que el modelo brinde un elevado tiempo polinomial de resolución, lo que no lo hace pues a mayor cantidad de datos las expresiones matemáticas generan un problema NP–duro, es decir que existe una mayor complejidad no polinomial en resolverse, con lo que ante esto se pone en duda la decisión por utilizar un método tradicional y abre la inquietud y necesidad de hallar un método algorítmico robusto y solvente para satisfacer necesidades amplias en despliegues de red [30].

CAPITULO 4. DESCRIPCIÓN DEL ESCENARIO EN BASE A LOS MÉTODOS UTILIZADOS.

4.1 Propuesta para el despliegue de la Red hibrida WDM–TDM/PON

El modelo matemático plantea una heurística cercana a una posible solución, pero al tener una gran cantidad de nodos y a su vez estos formen escenarios muchos más complejos el cálculo matemático aumenta su trazabilidad y sobre todo empieza a formar limitaciones en el mismo.

Todo sector o ubicación es un posible escenario a describir, pero para poder analizar un determinado sector en base a un modelo utilizado se puede mencionar tres aspectos a tomar en cuenta:

- **El número de nodos.** Como se mencionó al aumentar el número de nodos aumenta el tamaño de las matrices, lo que representa un cálculo mucho largo y tedioso e incluso en algunas ocasiones incalculable, por lo que una restricción seria el tamaño o forma de lugar donde se realice el estudio de Red.

- **Distancias y formas de los enlaces.** esto se ve como una restricción al modelo, ya que el objetivo del modelo es llegar a buscar la mejor ubicación de la OLT, y al existir una forma compleja de ubicación o distancias muy abstractas e irregulares, el modelo busca tomar en cuenta cada espacio o forma que se presente lo que conlleva a soluciones lejanas y confusas, creando así un problema para su análisis matemático, sin embargo un modelo algorítmico solventa de la manera más óptima esta restricción, solo que hay considerar los puntos georreferenciados de la manera más exacta y clara posible.

- **Parámetros técnicos.** Al analizar un estudio de Redes PON, y ser una red óptica, sus elementos presentan condiciones de funcionalidad técnicas necesarias para su óptimo desarrollo, el ancho de banda, potencia de atenuación, potencia de transmisión son factores que intervienen en el modelo, pero representan en números y costos que al concluir el análisis son representativos para el estudio.

Para poder describir un ejemplo de un escenario se tiene la Figura 30.

Figura 30. Escenario de la ciudad de Cuenca para el despliegue de una red PON

En la Figura 30, podemos observar dos escenarios dentro de la ciudad de cuenca, la línea color purpura está montada sobre el mapa cubriendo cada sector, se puede observar que cada esquina de manzano o cada cuadra forman enlaces entre ellas obteniendo un área de despliegue de red, el escenario de la izquierda está formado por 120 puntos o esquinas.

El algoritmo de modelamiento por métodos de agrupación que se plantea, toma la idea de modelos matemáticos, y lo que hace es:

- Leer los datos de ubicación (x,y) de un punto o llámese nodo y ubicarlos en una matriz.
- Luego lee los datos de las uniones llamados enlaces entre los nodos.
- Posterior a través de un método de agrupamiento seleccionado, este revisa y en base a las longitudes de los enlaces y la forma que representan estos, calcula pequeños centroides que finalmente, en otro proceso algorítmico de selección ubica de manera idónea los componentes ópticos (OLT, AWG, Splitters), a tal punto que de cómo resultado una longitud, una ubicación y un costo aproximado de despliegue de fibra, así como la mejor ubicación de los componentes de red para obtener la mejor cobertura.

Para poder describir la complejidad que conlleva el uso de estos modelos matemáticos como métodos de solución de despliegues de red, se va a utilizar un

pequeño escenario real para poder dimensionar el problema que abarca el uso de modelos matemáticos.

En la siguiente figura se observa un pequeño ambiente donde ya se han ubicado la OLT, y los splitters para poder llegar a cada nodo de enlace.

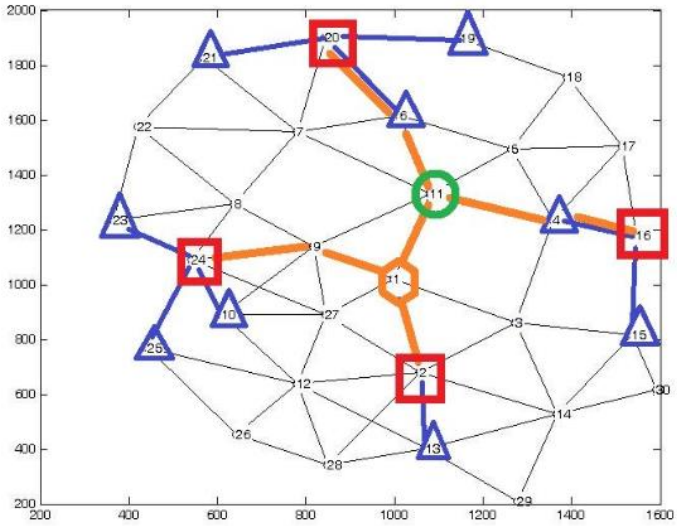

Figura 31. Escenario explicativo para ejemplo de agrupamiento

4.2 Comparación de las Topologías según los Algoritmos empleados para el despliegue de Red.

En base a todo lo analizado antes en este documento, y de acuerdo a los requerimientos para el estudio de los métodos descritos en el capítulo 2 de los tres escenarios antes analizados nos enfocaremos en el escenario 1 de la figura, así mismo los posibles casos de despliegue de su red serán analizados mediante las técnicas de agrupamiento k–means y k–medoids.

4.2.1 Despliegue de la Red Mediante la técnica de Agrupamiento K–means

Por las características propias de la red se analizan su distribución mediante tres condiciones o escenarios, el primero despliega la red con dos AWG mediante la técnica de k–means, los resultados obtenidos se observan en la Figura 32.

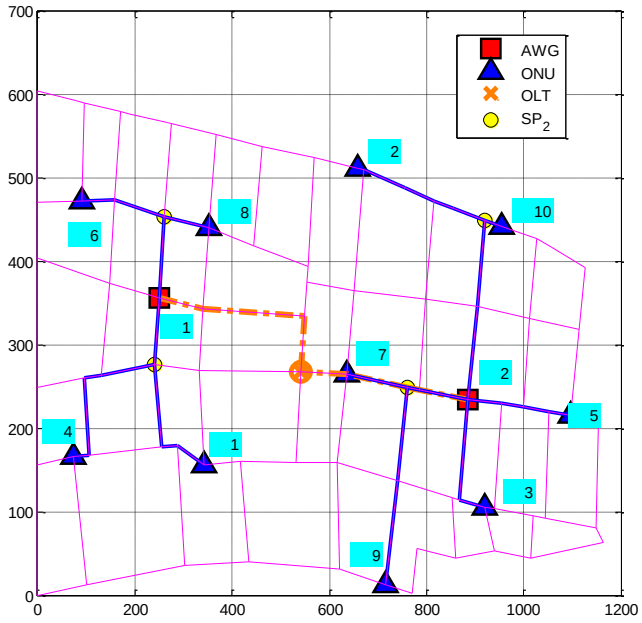

Figura 32. Despliegue de la red PON con 2 AWG mediante k-means

Se observa la red primaria compuesta por una OLT conectada a una AWG 1 y a una AWG 2, y la red secundaria compuesta por 10 ONUs, conectada mediante fibra óptica respectivamente como se observa en la Figura 32, los resultados de esta grafica se detallan a continuación:

Pérdidas por atenuación en paths de la red Primaria

- De la OLT a la AWG 1: 3.073704 dB.
- De la OLT a la AWG 2: 3.068682 dB.

Pérdidas por atenuación en paths de la red Secundaria

- De la AWG 1 a la ONU 1: 3.054324 dB.
- De la AWG 1 a la ONU 4: 3.070156 dB.
- De la AWG 1 a la ONU 6: 3.054057 dB.
- De la AWG 1 a la ONU 8: 3.070156 dB.
- De la AWG 2 a la ONU 2: 3.097532 dB.
- De la AWG 2 a la ONU 3: 3.482561e-02 dB.
- De la AWG 2 a la ONU 5: 4.272081e-02 dB.
- De la AWG 2 a la ONU 7: 3.050101 dB.
- De la AWG 2 a la ONU 9: 3.072956 dB.
- De la AWG 2 a la ONU 10: 3.050404 dB.

Como es notorio por los datos obtenidos son bastante aceptables ya que él valor de atenuación que tiene un AWG es de 3 dB y aquí hemos obtenidos cercanos a ese valor e incluso valores menores en algunos casos.

Las longitudes de onda designadas para cada zona, las mismas que han sido representados por un AWG respectivamente es decir, en este caso tenemos dos zonas las cuales tienen un total de 2 longitudes de onda para la zona 1, es decir, AWG 1. Para el caso de la zona, AWG 2, se tienen 4 longitudes de onda, así mismo se tiene 2 y 4 puertos en uso respectivamente.

En el caso de tener una distribución de la red mediante el empleo de 4 AWG se tiene lo mostrado por la Figura 33.

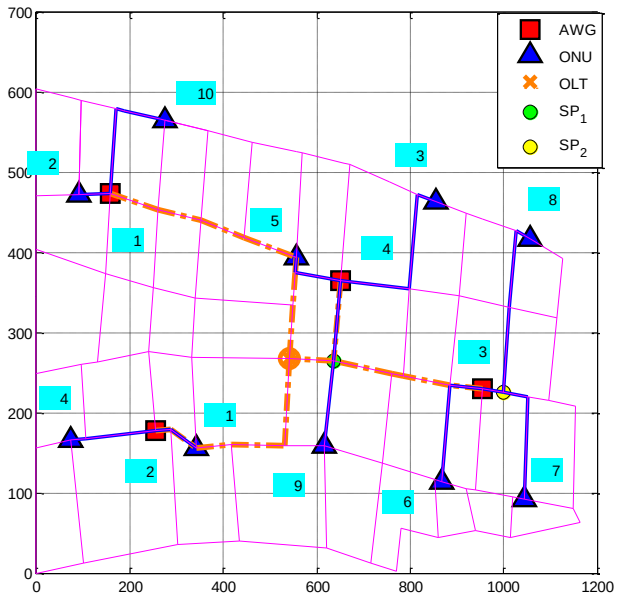

Figura 33. Despliegue de la red PON con 4 AWG mediante k-means

La figura muestra la distribución de los elementos sobre la red tanto primaria como secundaria teniendo los siguientes valores como resultados:

Pérdidas por atenuación en paths de la red Primaria

- De la OLT a la AWG 1: 3.106443 dB.
- De la OLT a la AWG 2: 3.077857 dB.
- De la OLT a la AWG 2: 6.082702 dB.

- De la OLT a la AWG 2: 6.038822 dB.

Pérdidas por atenuación en paths de la red Secundaria

- De la AWG 1 a la ONU 2: 1.369804e-02 dB.
- De la AWG 1 a la ONU 10: 4.225312e-02 dB.
- De la AWG 2 a la ONU 1: 1.829965e-02 dB.
- De la AWG 2 a la ONU 4: 3.654711e-02 dB.
- De la AWG 3 a la ONU 6: 3.818519e-02 dB.
- De la AWG 3 a la ONU 7: 3.045225 dB.
- De la AWG 3 a la ONU 8: 3.056809 dB.
- De la AWG 4 a la ONU 3: 6.151847e-02 dB.
- De la AWG 4 a la ONU 5: 2.323205e-0 dB.
- De la AWG 4 a la ONU 9: 4.183993e-02 dB.

Los AWG, además, tienen en el caso del AWG 1, 2 longitudes de onda y 2 puertos en uso, en el AWG 2, se tiene 2 longitudes de onda con 4 puertos en uso, así mismo en el caso del AWG 3 se tienen 2 longitudes de onda con 2 puertos en uso y finalmente en el AWG 4 con 4 longitudes de onda y 3 puertos en uso, es decir, se tendría un puerto libre.

En la Figura 34 se puede apreciar las curvas de mayor atenuación para cada tramo al tener cada uno de los casos para el escenario, el algoritmo ha sido ejecutado en 10 iteraciones para tenar un panorama más claro de los resultados, debido a su característica propia de algoritmo de agrupamiento particional.

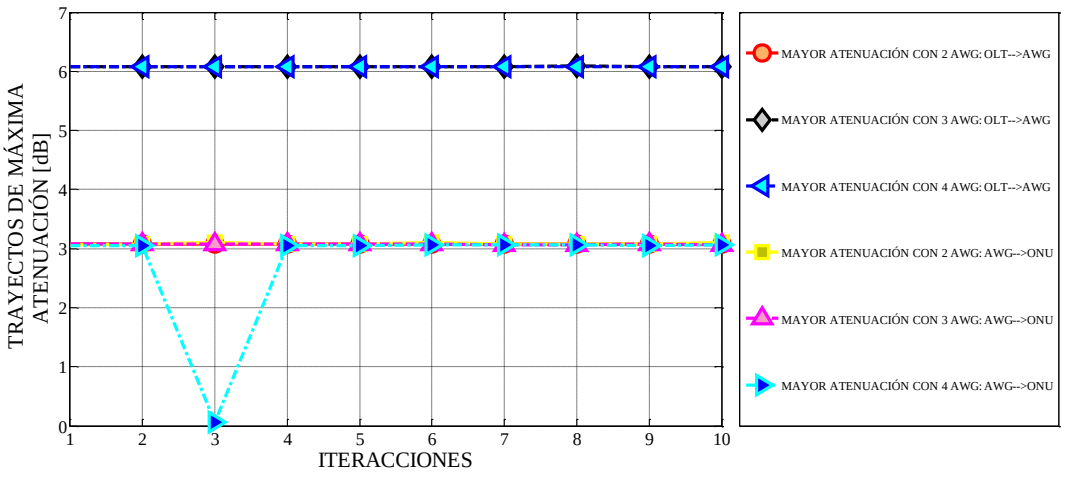

Figura 34. Atenuación de los tramos de la Red al aplicar k-means

Como se puede notar en este caso las atenuaciones son bastante constantes durante cada una de las iteraciones realizadas, teniendo un valor de 3 dB como valor promedio de la red secundaria con un mínimo en la iteración 3 en el caso de tener 4 AWG en su tramo hasta las ONU y un promedio de 6 dB para red primaria.

Otro aspecto importante al diseñar una red son las longitudes de ondas que los AWG están ocupando y cuantas están libres para después ser utilizadas, es decir, tener una proyección de crecimiento y despliegue de la red para así captar más usuarios.

La Figura 35 muestra las curvas de las lambdas que están siendo utilizados por los AWG y los que quedarían libres al aplicar el algoritmo con la técnica de k-means durante 10 iteraciones.

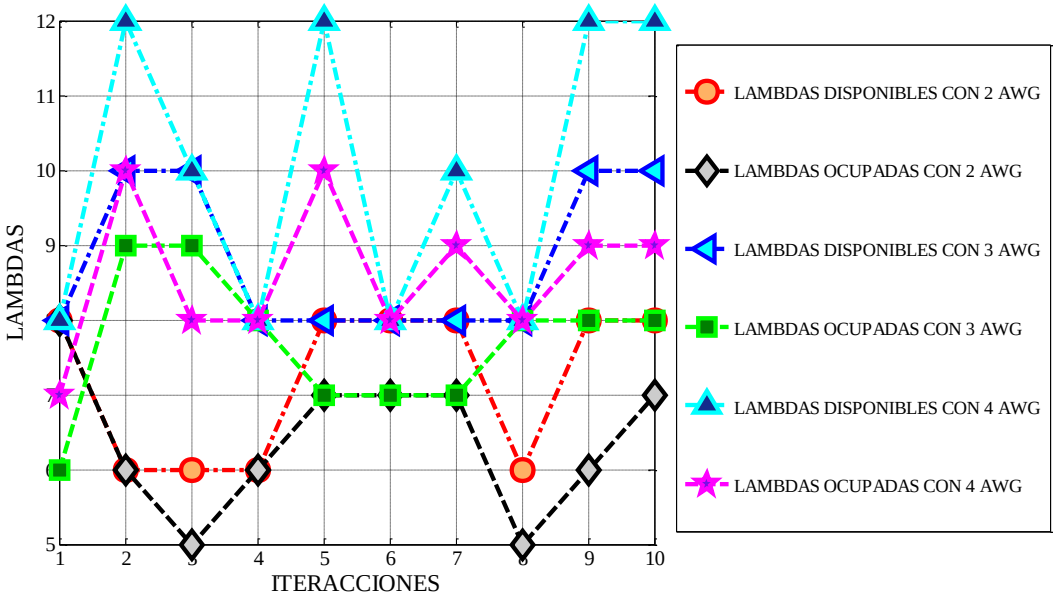

Figura 35. Lambdas de los AWG obtenidos mediante K-means

Con la gráfica se puede apreciar que en el caso de tener 2 AWG se puede tener una proyección de crecimiento en todas las iteraciones a excepción de la iteraciones 1, 2 y 4, en el segundo caso, 3 AWG, las iteraciones que muestran una mayor posibilidad de crecimiento son la 1, 9 y 10 con un total de hasta 2 γ no en uso, finalmente el caso de tener 4 AWG la mayor posibilidad de crecimiento se la en las iteraciones 9 y 10 con 3 γ libres y en su mayoría 2 γ para el resto de iteraciones.

Como se pudo observar en los experimentos realizadas al desplegar la red PON en estudio se alcanzaron muy buenos resultados en cada uno de los casos llegando a tener valores de atenuación aceptables así como buena posibilidad de crecimiento de Red.

4.2.1 Despliegue de la Red Mediante la técnica de Agrupamiento k-medoids

El escenario a ser modelado será el mismo que en caso anterior, esta vez se agrupara la técnica de k-meoids para distribuir los elementos de la red PON, dependiendo de este resultado tanto la cantidad equipos como son ONUs, Splitters y por ende cantidad de fibra óptica para la conexión de estos equipos dicho despliego se observa en la Figura 36.

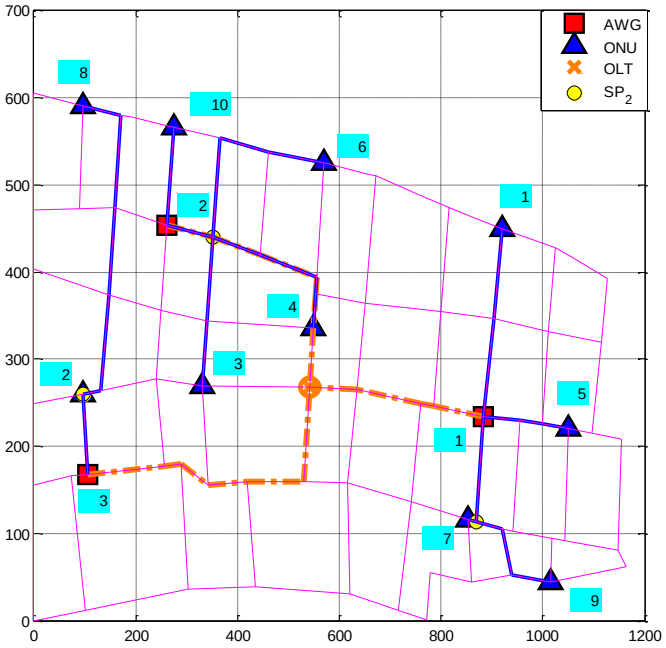

Figura 36. Despliegue de la red PON con 3 AWG mediante k-Medoids

En este caso se ha graficado el caso en donde se tiene 3 AWG y 10 ONUs como en el caso anterior cuando se empleó el método de k-means, dando como resultado lo siguiente.

Pérdidas por atenuación en paths de la red Primaria

- De la OLT a la AWG 1: 3.068682 dB.
- De la OLT a la AWG 2: 3.085559 dB.
- De la OLT a la AWG 3: 3.107774 dB.

Pérdidas por atenuación en paths de la red Secundaria

- De la AWG 1 a la ONU 1: 4.366637e-02 dB.
- De la AWG 1 a la ONU 5: 3.351497e-02 dB.
- De la AWG 1 a la ONU 7: 3.027388 dB.
- De la AWG 1 a la ONU 9: 3.061124 dB.
- De la AWG 2 a la ONU 3: 6.052677 dB.
- De la AWG 2 a la ONU 4: 6.072071 dB.
- De la AWG 2 a la ONU 6: 6.081900 dB.

- De la AWG 2 a la ONU 10: 2.267510e-02 dB.
- De la AWG 3 a la ONU 2: 3.01848 dB.
- De la AWG 3 a la ONU 8: 3.104031 dB.

En cuanto al número de longitudes de onda se tienen 4 en el AWG 1 con 3 puertos en uso, 2 longitudes y 2 puertos ocupados para AWG 2 y para el AWG 3 se tienen 2 longitudes de onda y 1 puerto ocupado.

Otro escenario a ser analizado es el caso de tener un número de 4 AWGs, el resultado se observa en la Figura 37.

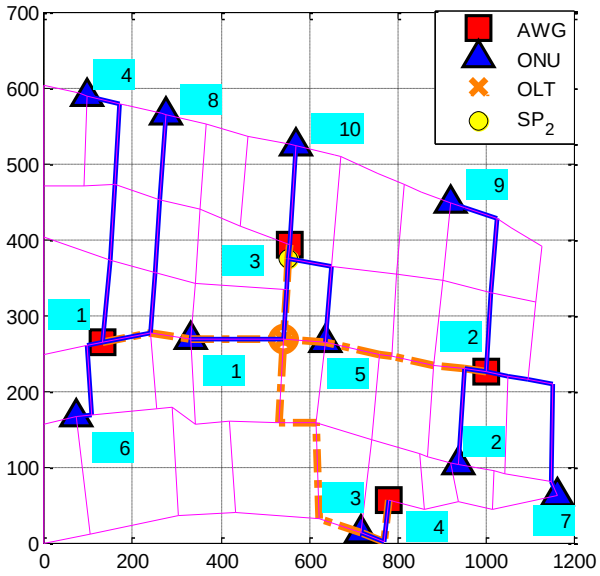

Figura 37. Despliegue de la red PON con 4 AWG mediante k-Medoids

Los elementos distribuidos en esta red PON presentan las siguientes características, que se analizaran con cuidado más adelante.

Pérdidas por atenuación en paths de la red Primaria

- De la OLT a la AWG 1: 3.082656 dB.
- De la OLT a la AWG 2: 3.092003 dB.
- De la OLT a la AWG 3: 3.025310 dB.
- De la OLT a la AWG 4: 3.105911 dB.

Pérdidas por atenuación en paths de la red Secundaria

- De la AWG 1 a la ONU 4: 7.880997e-02 dB.
- De la AWG 1 a la ONU 6: 3.185143e-02 dB.
- De la AWG 1 a la ONU 8: 8.045393e-02 dB.
- De la AWG 2 a la ONU 2: 3.492477e-02 dB.
- De la AWG 2 a la ONU 7: 6.121234e-02 dB.
- De la AWG 2 a la ONU 9: 6.233767e-02 dB.
- De la AWG 3 a la ONU 1: 3.067531 dB.
- De la AWG 3 a la ONU 5: 3.043473 dB.
- De la AWG 3 a la ONU 10: 2.637371e-02 dB.
- De la AWG 4 a la ONU 3: 2.222949e-02 dB.

Las longitudes de onda para la zona 1 es de 4 con 3 puertos ocupados, la zona dos consta de 4γ con 3 puertos usados, en la zona tres se tiene 4γ y 3 puertos, en la zona 3, 2γ y dos puertos ocupados y finalmente para zona 4 dos longitudes de onda y un puerto utilizado.

Como se observa en este caso se tienen atenuaciones aceptables tanto en la red primaria como secundaria, la discusión en este sentido será presentada más adelante.

De la misma manera que en el caso del algoritmo anterior la Figura 38 muestra las curvas de atenuación de los tramos de la red durante cada una de las 10 iteraciones sino esta vez aplicando la técnica de k-medoids.

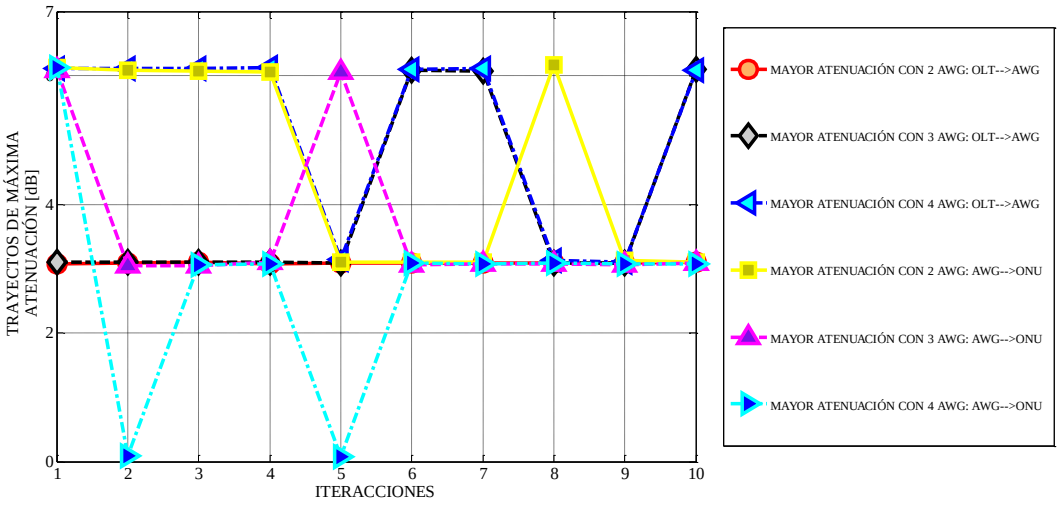

Figura 38. Atenuación de los tramos de la Red al aplicar k-medoids

En este caso se puede notar que se da una atenuación constante durante las iteraciones en el primer caso, es decir, al tener 2 AWG, en el resto de iteraciones se da una atenuación bastante inestable variando de aproximadamente 3 dB a 6 dB, teniendo la menor atenuación en la red secundaria en el caso de 4 AWG hasta la ONU en las iteraciones 2 y 5.

La muestra la posibilidad de despliegue de la Red al ser distribuida mediante la técnica de agrupamiento de k-meoids.

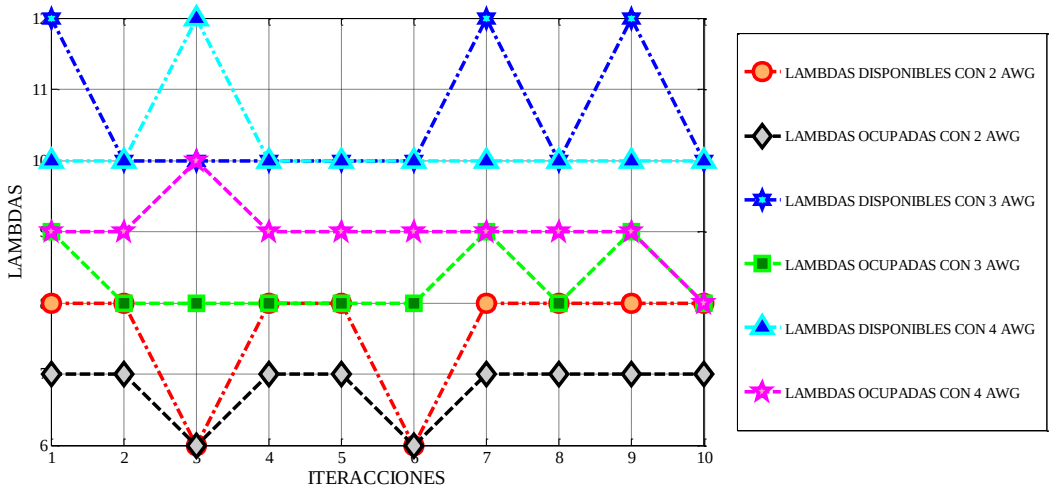

Figura 39. Lambdas de los AWG obtenidos mediante K-medoids

La figura muestra que en el primer caso existe un máximo escenario de crecimiento de hasta 2 longitudes de onda en la mayoría de iteraciones, mientras el segundo caso 3 AWG se disponen de una máximo de 3γ en la iteraciones 1, 7 y 9, finalmente en el último caso, es decir, al tener 4 AWG se tiene un mayor número de puertos disponibles en la iteración 3 y 10 con un máximo de 2γ.

En el caso de aplicar el algoritmo de k-medoids para el despliegue de la red PON hibrida en cuanto a la atenuación de obtuvieron datos aceptables en algunos casos no así en otros por lo que demuestra ser un algoritmo inestable de la misma manera fue en la posibilidad de crecimiento, aunque aquí se mostró mayor posibilidad de crecimiento también se observó que no agrupo de buena manera quedando a veces 1 AWG para una sola ONU, es decir, desperdiciando recurso de la Red.

4.3 Comparación de los Costos de implementación de la Red

Al momento de realizar la planificación de una red, el motivo de analices en la distribución de los equipos y la cantidad de los mismos utilizados para abastecer a todos los usuarios, radica en que estos resultados serán los responsables del costo que dicha Red llegue a tener siendo este el aspecto más importante dentro del despliegue de la Red y razón de varios casos de análisis y motivo principal de este trabajo.

4.3.1 Costos Reales

Este es el aspecto más importante dentro del tema de los costos, ya que se hablan de valores monetarios lo que indica que tan rentable sea este proceso mientras menos dinero se emplee en la elaboración de la Red con un alcance optimo hacia los usuarios y proyecciones de crecimiento mejor serán los resultados en su implementación.

Para el análisis en este documento han sido tomados precios referenciales de los equipos se acuerdo al mercado actual así como a un posible incremento en futuro.

En este sentido los costos de la red de la Figura 33 son los siguientes.

- Costo total de la red Primaria en canalización, tuberías y fibra óptica = $ 10915,2
- Costo total de red secundaria en Canalización, tuberías y fibra óptica = $ 11969,99
- Costo de la OLT = $ 20000
- Costo total de las ONU = 16,000 dólares
- Costo total de las AWG = $ 3,950
- Costo total de las SP = $ 1.600

COSTO TOTAL DE LA RED HIBRIDA WDM / PON - TDM / PON = $ 64435,25

Par el caso de los costos de la Figura 37. Tenemos los siguientes valores:

- Costo total de la red Primaria en canalización, tuberías y fibra óptica = $ 11623,42

- Costo total de red secundaria en Canalización, tuberías y fibra óptica = $ 16424,72
- Costo de la OLT = $ 20000
- Costo total de las ONU = 16,000 dólares
- Costo total de las AWG = $ 4100
- Costo total de las SP = $ 800

COSTO TOTAL DE LA RED HIBRIDA WDM / PON - TDM / PON = $ 68948,14

De acuerdo a costos en desplegar la Red con ambos métodos, el que representa un menor valor en su despliegue es el caso con el método de k-means con una diferencia de $ 4512,89, esta diferencia puede aumentar de acuerdo al número de puntos y tipos de escenarios.

Para poder tener un mejor panorama de la diferencia en cuánto a los costos de la Red se comparan los resultados de los dos métodos en 10 iteraciones

La Figura 40 muestra una comparación de los costos de la Red entre ambos métodos.

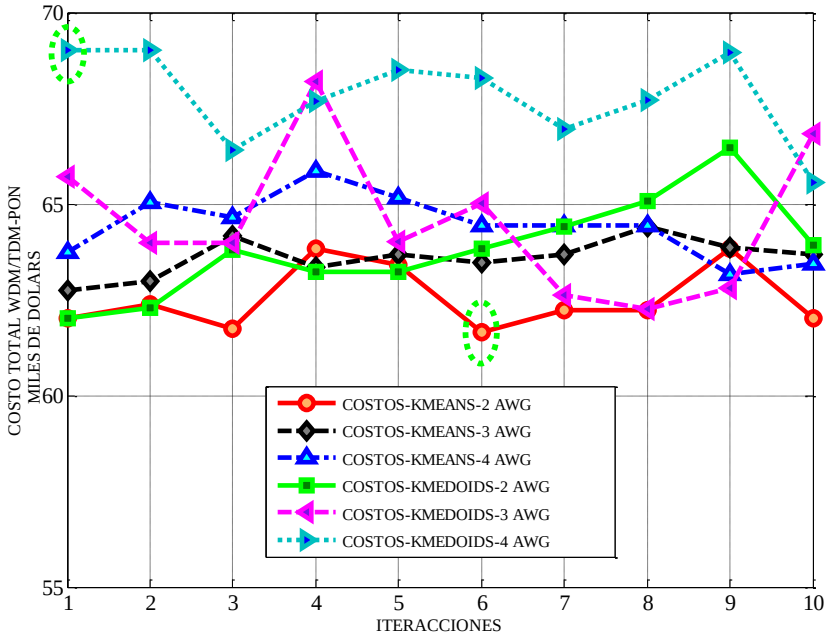

Figura 40. Comparación de los Costos De la Red PON Hibrida mediante k-means y k-medoids

Los costos son más elevados al desplegar la red con el método de k-meodoid que con k-means en cada uno de los casos, sin embargo, el costo más alto de la Red se da en

el caso de tener 4 AWG en la primera iteración por el medo de k-medoids y el más bajo en caso de tener 2 AWG con la técnica de k-means en la sexta iteración.

4.3.2 Costos Computacionales

Este es también un factor muy importante a ser tomado en cuenta ya que aunque en este caso no representa mucha la diferencia al aumentar los nodos, la forma del escenario así como los casos ira también siendo trascendental una optimización en tiempo, por ahora nos enfocaremos en lo que refiere la comparación y conclusión de con cuál de los métodos en estudio se logran los mejor tiempos así como cual es más estable en este sentido, más adelante en la justificación de la decisión tomada se anualizara también la razón de porque utilizar este método y no otros métodos conocidos para el despliegue de una red .

La Figura 41 muestra los tiempos necesarios para la ejecución del algoritmo que hace posible el despliegue de la Red empleando el método de k-means para cada uno de los casos en 10 iteraciones.

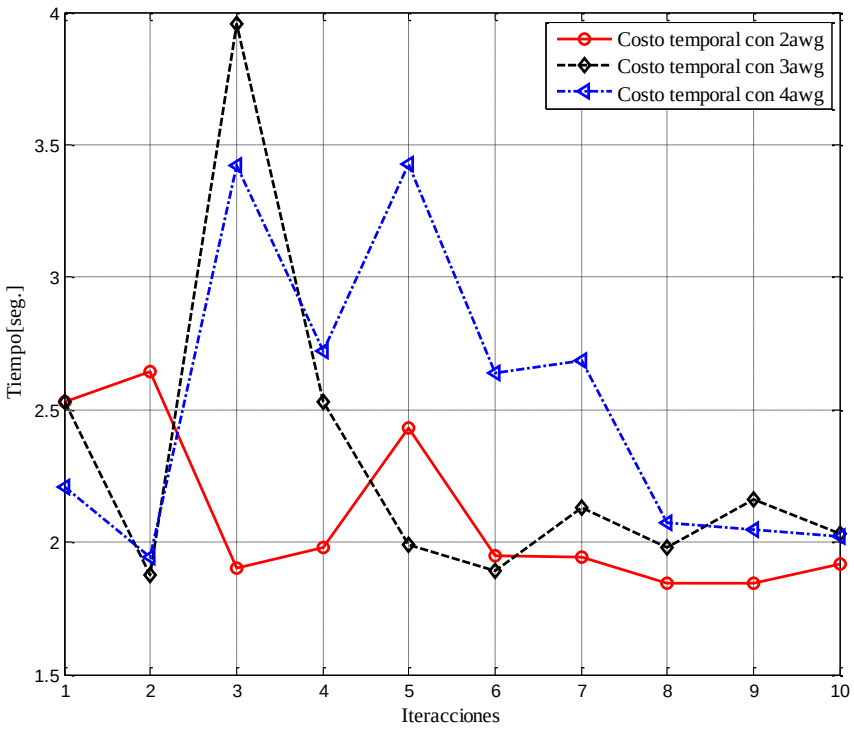

Figura 41. Tiempos de ejecución del Algoritmo mediante k-means

Se muestra un mayor retardo al tener el caso de 3 AWG en la iteración 3, el más bajo en la iteración 9 en el caso 2 AWG, siendo también este el caso de mayor estabilidad en comparación con los otros casos.

En la Figura 42 se observa ahora los retardos al utilizar la técnica de k-medoids.

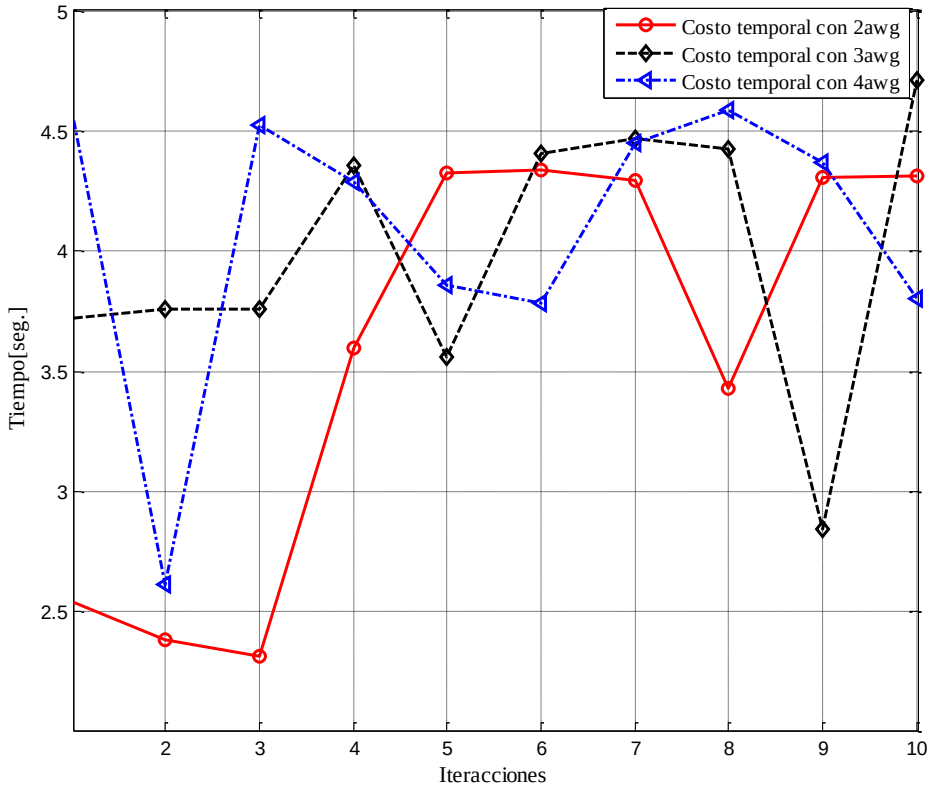

Figura 42. Tiempos de ejecución del Algoritmo mediante k-medoids

La grafica demuestra el menor retardo en la iteración 3 del caso 2 AWG y la más elevada en la iteración 10 del caso 3 AWG, también muestra gran inestabilidad en cuanto a los tiempos de ejecución.

La muestra una comparación de los tiempos promedios de la Figura 41 y la Figura 42, para así tener una idea más clara de la diferencia de tiempo de ejecución entres los métodos.

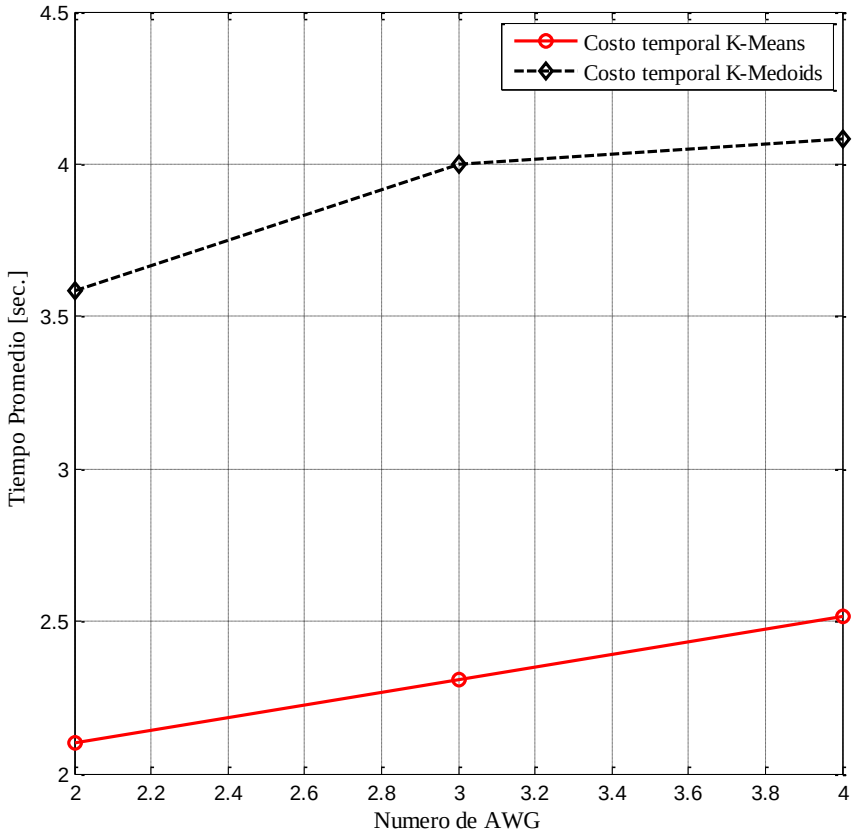

Figura 43. Tiempos promedios en la ejecución del algoritmo con los métodos k-meas y k-meoids

La figura muestra una diferencia de aproximadamente 2 segundos en la ejecución del algoritmo, siendo la técnica de k-means la más rápida de las dos incrementándose la diferencia conforme se aumente los puntos y números de AWG.

4.4 Toma de decisión

En base a los diferentes análisis en todo este documento como son que tan óptimos son los algoritmos en cuanto a el agrupamiento de los miembros y que tan estable son con respecto a este tema, así como la comparación entres métodos en lo que se refiere a la rapidez como se ejecutan primero analizando este tema en el capítulo dos, para luego analizarlos ya con el algoritmo completo para la distribución de la Red en este capítulo así como los costos que representa uno u otro método al formar la Red PON hibrida para el escenario en estudio, además de la posibilidad para

escalabilidad que la red posea para el futuro ser capaces de captar nuevos usuarios con un servicio óptimo para los mismos.

El método de agrupamiento **k-means** fue el más sencillo en implementarse, así como el más rápido, robusto, el que mejor agrupa, el que mejor se estabilizo, también este mostro menor atenuación en los tramos de red y el que menor costo mostro para su implementación con una gran opción de crecimiento por lo que es el método que concluimos que mejor se adapta para este tipo de aplicaciones y obtener un despliegue de una Red de este tipo cuasi optimo con valores muy aceptables.

4.5 Justificación del algoritmo empleado para la distribución de la red hibrida WDM–TDM/PON

En principio se analizaron tres métodos de agrupamiento como son el k-means, k-medoids y Mean-shift, siendo este último desechado posteriormente ya que la característica de este método no suponía el conocimiento a priori del número de clústeres, siendo este un factor indispensable en el caso de estudio de este documento, por lo que solo se analizaron finalmente los métodos restantes, quedándonos con lo0s resultados obtenidos mediante la implementación del algoritmo k-means por todo lo expuesto antes, además, en comparación con otros métodos comunes para el despliegue de una Red.

La idea de poder utilizar un método matemático cuestiona mucho la toma de decisiones ya que quizá para este escenario tras un largo proceso matemático se pueda obtener una solución, pero que sucede cuando existen más nodos o puntos, ante esta situación podemos ya realmente hablar de un gran limitante por no utilizar el término ¨problema¨ puesto que el estudio de implementación y despliegue tendría un elevado factor de imprecisión al momento de ubicar los elementos ópticos-pasivos de la red, y esto a su vez generaría una deficiente cobertura de servicio y por ende un disparado costo de despliegue. Ante esto una solución algorítmica basándose obviamente en una plataforma matemática proyecta una alternativa fiable y óptima para un estudio de Red.

Como se ha revisado anteriormente el uso de un modelo matemático nos brinda una solución muy compacta ante cualquier escenario donde se desee establecer un estudio de red PON (fiber to fiber), pero presenta limitantes, que recae sobre algo muy complejo que es la cantidad de puntos a tratarse, pues como son cálculos

diferenciales, las soluciones son más grandes y tardías acorde a los puntos y enlaces, lo que proyecta a utilizar una herramienta algorítmica de soporte, que obviamente centrara su funcionalidad en una base matemática pero la ejecutara a través de una herramienta computacional adaptándose de mejor manera a la topología de red planteada.

Para poder entender como la complejidad del modelo matemático y la utilidad de un modelo logarítmico, se puede desglosar cada solución por separado para obtener una diferencia de análisis entra la una de la otra

Solución mediante MILP.

Las ecuaciones que describen un modelo matemático donde se llegan a obtener expresiones que representan cada enlace y nodo del escenario partiendo de un origen, para el uso y manejo de estas representación matemáticas es necesario el uso de una herramienta computacional de cálculo y solución, como LPSOLVE, SOLVE, MODELLUS entre otros para este caso las ecuaciones planteadas y soluciones obtenidas se han obtenido de un software llamado **LPSOLVE IDE 5.5,** un programa gratis y disponible en la red.

Una vez que se conoce los puntos y los enlaces del escenario a estudiar, lo siguiente es representarlos mediante las ecuaciones mencionadas anteriormente y transcribirlas en el programa LPS, que para este caso se muestra en la siguiente figura parte del programa con las ecuaciones transcritas.

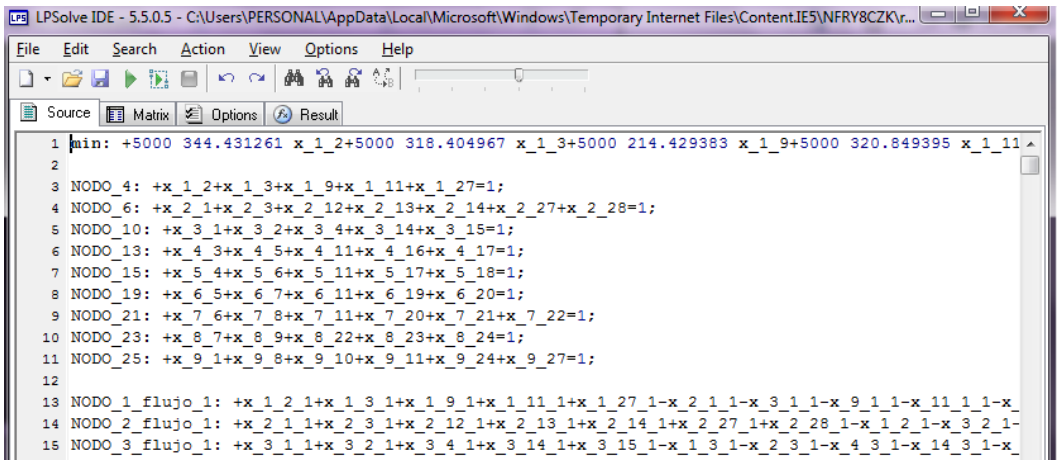

Figura 44. Ecuaciones transcritas en LpS

En la pantalla se observa que cada enlace a cada nodo representa una ecuación, así mismo cada nodo activo representa otra, lo que significa que si aumenta el número de nodos, aumentara el número de enlaces y obviamente aumentara el número de ecuaciones.

Estas ecuaciones serán reescritas para cada flujo, es decir para cada utilización de los nodos, en este caso son ocho por lo que deben existir 8 grupos de ecuaciones para cada nodo, como son 30 nodos en este escenario serian en total solo en esta sección 240 ecuaciones tal como lo indica la siguiente figura.

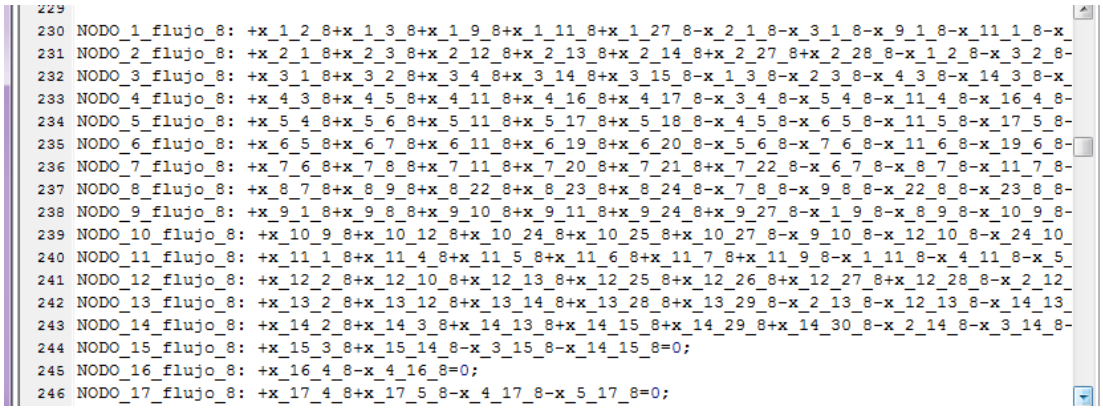

Figura 45. Ecuaciones escritas de nodos en LpS

Una vez que se tiene las ubicaciones de nodos representados en ecuaciones se escriben las ecuaciones de enlaces formadas por todos los nodos o puntos que constituyen el escenario.

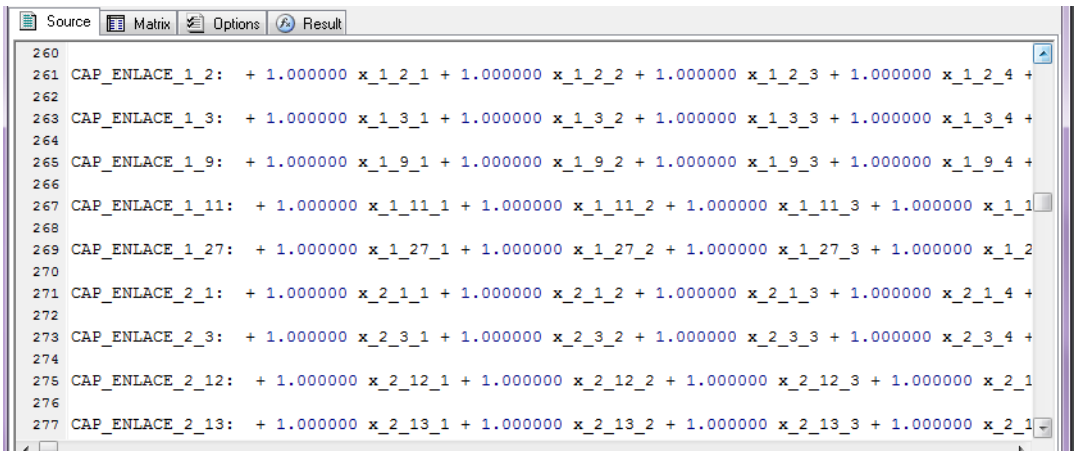

Figura 46. Ecuaciones de enlaces LpS

Las ecuaciones de los enlaces serán las mismas que describan cada unión de los mismos, estas en sus términos computacionales son números pero en términos de despliegue de obra representaran distancias reales de sector analizado, las mimas que llevaran un coste de despliegue de fibra por cada enlace planteado.

Una vez que todas las ecuaciones están plateadas se corre el software de solución para poder obtener algún resultado., en este caso y para este escenario LpS nos da lo siguiente:

```
SUBMITTED
Model size:      240 constraints,     476 variables,          646 non-zeros.
Sets:                                   0 GUB,                   0 SOS.

Using DUAL simplex for phase 1 and PRIMAL simplex for phase 2.
The primal and dual simplex pricing strategy set to 'Devex'.

Relaxed solution       663419.911869 after         29 iter is B&B base.

Feasible solution      672054.493461 after         36 iter,          7 nodes (gap 1.3%)
Improved solution      671027.755746 after         56 iter,         21 nodes (gap 1.1%)
Improved solution      671004.429176 after         57 iter,         22 nodes (gap 1.1%)
Improved solution      670955.238714 after         99 iter,         58 nodes (gap 1.1%)
Improved solution      669973.111945 after        232 iter,        133 nodes (gap 1.0%)
Improved solution      669773.403205 after       1520 iter,        844 nodes (gap 1.0%)
Improved solution      668742.085974 after       2396 iter,       1328 nodes (gap 0.8%)
Improved solution      668698.872671 after       3295 iter,       1811 nodes (gap 0.8%)
Improved solution      668684.693341 after       6211 iter,       3476 nodes (gap 0.8%)
spx_run: Lost feasibility 10 times - iter       7670 and      4321 nodes.
spx_run: Lost feasibility 10 times - iter       7774 and      4384 nodes.
spx_run: Lost feasibility 10 times - iter       8024 and      4531 nodes.
Improved solution      668683.055104 after      10970 iter,       6171 nodes (gap 0.8%)
Improved solution      668608.886352 after      24586 iter,      13168 nodes (gap 0.8%)
Improved solution      667577.569121 after      24963 iter,      13357 nodes (gap 0.6%)

Optimal solution       667577.569121 after      27494 iter,      14661 nodes (gap 0.6%).
Excellent numeric accuracy ||*|| = 2.22045e-016
```

Figura 47. Resultados mostrados en LpS

Como se ve en la figura anterior se obtiene un conjunto de posibles soluciones para cada iteración que realiza el programa, y la solución más óptima la obtiene en la iteración 27494, lo que significa que se podría considerar como una solución real y precisa.

El problema para obtener una solución utilizando un modelo matemático y este a su vez apoyado de alguna herramienta computacional no radica en la iteración que obtuvo el mejor resultado sino en el tiempo en que lo hizo, podemos ver solo para este escenario de 30 puntos LpS demoro 2638 segundos es decir 1,8 días lo que significa que es muy complejo el plantear y solucionar este tipo de ecuaciones, y más aún si el escenario es grande, con lo que se explica la dimensión del problema al tratar de plantear un proyecto utilizando herramientas tradicionales, concluyendo

ante esto se enfatiza en la búsqueda, generación y utilización de una nueva forma de trabajar estos datos que optimice tiempo, dinero y recursos.

Solución mediante Heurísticas.

Tomando el mismo escenario como ejemplo y utilizando modelos de agrupamiento como K-means, k-medoids u otro, se puede observar una diferencia.

Una vez que se tiene los puntos de los nodos georeferenciados que serán los mismos con los que se trabajó en MILP, se los ubica en una matriz utilizando MATLAB, esta matriz es la base para poder crear los enlaces que se compone de todas las uniones entre nodos que describa el escenario y estas se las ubica en otra matriz. Con estas dos fuentes de información se elige el método de agrupamiento de datos que consiste de líneas de comando que a través de sub-tareas tratan y procesan esa información acercándonos una posible solución que toma el nombre de Heurística.

Para este escenario se utiliza métodos de agrupamiento y se obtienen diferentes iteraciones con diferentes resultados como se muestra a continuación:

Iteración 1 :

```
-----------------------------------------------------
TOTAL COST PRIMARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.552188e+04
TOTAL COST SECONDARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.638365e+04
COST OLT = $ 50000
COST TOTAL ONU = $ 58500
COST TOTAL AWG = $ 3950
COST TOTAL SP = $ 800
COST TOTAL HYBRID NETWORK WDM/PON - TDM/PON = $ 1.451555e+05
-----------------------------------------------------
Elapsed time is 1.190310 seconds.
```

Iteración 2

```
-----------------------------------------------------
-----------------------------------------------------
TOTAL COST PRIMARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.552188e+04
TOTAL COST SECONDARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.681090e+04
COST OLT = $ 50000
COST TOTAL ONU = $ 58500
COST TOTAL AWG = $ 3950
COST TOTAL SP = $ 1600
COST TOTAL HYBRID NETWORK WDM/PON - TDM/PON = $ 1.463828e+05
-----------------------------------------------------


Elapsed time is 0.968319 seconds.
```

Iteración 3

```
----------------------------------------------------
TOTAL COST PRIMARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.552188e+04
TOTAL COST SECONDARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.931581e+04
COST OLT = $ 50000
COST TOTAL ONU = $ 58500
COST TOTAL AWG = $ 3950
COST TOTAL SP = $ 1600
COST TOTAL HYBRID NETWORK WDM/PON - TDM/PON = $ 1.488877e+05
----------------------------------------------------

Elapsed time is 1.018136 seconds.
```

Iteración 4

```
----------------------------------------------------
TOTAL COST PRIMARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.629525e+04
TOTAL COST SECONDARY NETWORK OF TRENCHING, PIPELINES AND OPTICAL FIBER = $ 1.928090e+04
COST OLT = $ 50000
COST TOTAL ONU = $ 58500
COST TOTAL AWG = $ 3950
COST TOTAL SP = $ 0
COST TOTAL HYBRID NETWORK WDM/PON - TDM/PON = $ 1.480262e+05
----------------------------------------------------

Elapsed time is 0.967755 seconds.
```

Como se ve en las diferentes iteraciones, nos da como resultado el costo total del despliegue de red para ese escenario, pero parte de ello, es el tiempo que demora en darnos una solución el mismo que promedia entre 1 a 1,5 segundos como máximo, optimizando por completo una herramienta computacional y abriendo múltiples opciones de proyectos ya que permite más nodos, más datos y más enlaces, lo que se concluye como una herramienta muy versátil a gran dimensiones en menor tiempo, rompiendo por completo tradicionales métodos y estableciendo, uno nuevo y óptimo.

La red diseñada debe cumplir con un presupuesto de potencia determinado por los equipos de transmisión y recepción empleados en el diseño. Si se cumplen las restricciones, el dimensionamiento ha finalizado con éxito obteniendo una solución sub–óptima. Si no se cumplen las restricciones se procede a incrementar el número de zonas en el área geográfica definida. En este caso el algoritmo se ejecuta nuevamente hasta satisfacer las restricciones de diseño.

Por todo lo expuesto se realizó este documento es decir, de utilizar métodos de agrupamiento con heurísticas para la el despliegue de una red hibrida PON, de los cuales el que se decide es el mejor es el algoritmo **k-means** por todos los resultados obtenidos.

CAPITULO 5. CONCLUSIONES Y RECOMENDACIONES

Los servicios de transmisión de datos hoy en día son más globales y cada vez necesitan de plataformas y estructuras de Red más versátiles y compactas que soporten de mejor manera las aplicaciones requeridas, es así que una notable evolución del cobre hasta la fibra nos lleva a una primera conclusión del uso necesario de una red PON para despliegues de red actuales ya que esta ofrece mayor velocidad, mayor cobertura y sobre todo su funcionalidad se encuentra abierta a múltiples protocolos de aplicabilidad como TDM , WDM y un concepto hibrido WDM-TDM/PON.

Un protocolo hibrido WDM-TDM/PON solventa de optima manera el consumo, la necesidad y la aplicabilidad del ancho de banda que actualmente se requiere, si bien es cierto hoy en día ya contamos con despliegues de red GPON, este nuevo concepto hibrido planteado junto con el constante crecimiento de abonados y la tecnología requerida, impulsan como se ha visto ya, a el manejo de nuevos conceptos futuristas como son Smart Grid y Smart City, llevándonos a una segunda conclusión donde el análisis, el mecanismo y aplicación de un método algorítmico que nos brinde una solución fiable y cercana, es importante en el estudio de un despliegue de red para estas futuras aplicaciones.

Otra conclusión final y muy importante es que si en un determinado sector deseamos hacer un estudio de despliegue de red, la ubicación de los equipos y costo del proyecto va a depender de método que se utilice para ello, si nos orientamos por un método tradicional donde se define por pura matemática y un poco al azar, aparte de que resulta muy tedioso el resultado final no va ser muy optimo debido a la complejidad de las ecuaciones por lo que como se ha visto en el capítulo 4 ya en un escenario real el método algorítmico facilita mucho la obtención en una solución tanto técnica como económica solventado y proyectando como mejor alternativa esta herramienta para estudios de despliegue de red.

TRABAJOS FUTUROS

El campo que hemos analizado en este documento es muy extenso y de mucha aplicabilidad en la realidad, como futuras investigaciones se puede incluir trabajar flujos en los enlaces y migrar hacia redes WDM o DWDM se piensa también considerar proyección de usuarios en el tiempo mediante métodos estocásticos.

REFERENCIAS

[1] A. G. Peralta-Sevilla y F. Amaya-Fernandez, «Evolución de las redes eléctricas hacia Smart Grid en países de la Región Andina,» Revista Educación en Ingeniería, vol. 8, nº 15, pp. 48-61, 2013.

[2] M. Abreu, «Características generales de una red de fibra óptica al hogar (FTTH),» Memoria de trabajos de difusión cientIfica y tecnica, nº 7, pp. 38-46, 2009.

[3] A. G. Peralta-Sevilla, F. Amaya-Fernandez y R. Hincapie, «Multiservice hybrid WDM/TDM-PON dimensioning using a heuristic method,» de Communications and Computing (COLCOM), 2014 IEEE Colombian Conference on, 2014.

[4] A. Eira, J. Pedro y J. Pires, «Optimized design of multistage passive optical networks,» Journal of Optical Communications and Networking, vol. 4, nº 5, pp. 402-411, 2012.

[5] F. J. Effenberger, K. McCammon y V. O'Byrne, «Passive optical network deployment in North America [Invited],» Journal of Optical Networking, vol. 6, nº 7, pp. 808-818, 2007.

[6] P. M. C. C. R. M. B. a. F. J. Salgado, «Evolution of access networks from GPON toXGPON to WDM-PON from operator point-of-view,» in 15th EuropeanConf. on Networks and Optical Communications (NOC),Faro, Portugal, 2010..

[7] M. S. Ahsan, M. S. Lee, S. Newaz y S. M. Asif, «Migration to the next generation optical access networks using hybrid WDM/TDM-PON,» Journal of Networks, vol. 6, nº 1, pp. 18-25, 2011.

[8] J. B. Rosolem, R. S. Penze, E. W. Bezerra, F. R. Pereira, B. C. de Camargo Angeli, E. Mobilon, J. C. Said y A. D. Coral, «Arquiteturas baseadas em WDM para as próximas redes PON,» Cad. CPqD Tecnologia, vol. 6, nº 1, pp. 65-76, 2010.

[9] A. Gomez-Martinez, F. Amaya-Fernandez, R. Hincapie, J. Sierra y I. Tafur Monroy, «Optical access multiservice architecture with support to smart grid,» de Communications and Computing (COLCOM), 2013 IEEE Colombian Conference on, 2013.

[10] A. K. Aggarwal y P. S Verma, «A proposed communications infrastructure for the smart grid,» 2010.

[11] M. Liserre, T. Sauter y J. Y. Hung, «Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics,» Industrial Electronics Magazine, IEEE, vol. 4, nº 1, pp. 18-37, 2010.

[12] M. Abreu, «Características generales de una red de fibra óptica al hogar (FTTH),» Memoria de trabajos de difusión cientIfica y tecnica, nº 7, pp. 38-46, 2009.

[13] M. Gallardo Campos, «Aplicacion de tecnicas de agrupación Para La Mejora del Aprendizaje,» 2009.

[14] J. F. V. Vera y J. A. G. San Salvador, «ANALISIS DE DATOS,» 2013/14.

[15] I. Amon y C. Jimenez, «Funciones de similitud sobre cadenas de texto: una comparacion basada en la naturaleza de los datos,» de CONF-IRM Proceedings, 2010.

[16] D. P. F. &. S. S. Pascual, «Algoritmos de agrupamiento,» Métodos Informáticos Avanzados, 2007.

[17] A. K. Jain, R. P. W. Duin y J. Mao, «Statistical pattern recognition: A review,» Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, nº 1, pp. 4-37, 2000.

[18] E. B. P. P. G. &. G.-M. R. Yolis, «Algoritmos genéticos aplicados a la categorización automática de documentos,» Revista Eletrônica de Sistemas de Informação ISSN 1677-3071 doi: 10.5329/RESI, 2(2), vol. 2, nº 2, 2003.

[19] R. C. Gonzalez y R. E. Woods, Digital image processing, Prentice hall Upper Saddle River, NJ:, 2002.

[20] J. H. M. F. P. R. C. L. R. G. S. J. \. M. A. Pérez, «Mejora al algoritmo de agrupamiento K-means mediante un nuevo criterio de convergencia y su aplicación a bases de datos poblacionales de cáncer,» II Taller Latino Iberoamericano de Investigación de Operaciones, 2007.

[21] C. C. Aggarwal y C. K. Reddy, Data Clustering: Algorithms and Applications, CRC Press, 2013.

[22] B. F. H. F.-A. T. Gómez y J. F. de la Mora, «Desigualdades territoriales en el bienestar. Clasificación jerarquizada de las provincias españolas mediante un análisis Cluster con semillas predeterminadas,» de Anales de estudios económicos y empresariales, 1996.

[23] A. Villagra, A. Guzman, D. Pandolfi y G. Leguizamon, «Análisis de medidas no-supervisadas de calidad en clusters obtenidos por K-means y Particle Swarm Optimization,» 2009.

[24] H.-S. Park y C.-H. Jun, «A simple and fast algorithm for K-medoids clustering,» Expert Systems with Applications, vol. 36, nº 2, pp. 3336-3341, 2009.

[25] Q. Zhang y I. Couloigner, «A new and efficient k-medoid algorithm for spatial clustering,» de Computational Science and Its Applications--ICCSA 2005, Springer, 2005, pp. 181-189.

[26] M. J. Rattigan, M. Maier y D. Jensen, «Graph clustering with network structure indices,» de Proceedings of the 24th international conference on Machine learning, 2007.

[27] A. R. Webb, Statistical pattern recognition, John Wiley \& Sons, 2003.

[28] N. M. Ignacio, «Un estudio basado en la Técnica de Mean Shift para Agrupamiento y Seguimiento en video,» Departamento de Computación, FCEyN,Universidad de Buenos Aires, 2011.

[29] A. Peralta-Sevilla, M. Tipan-Simbaña y F. Amaya-Fernandez, «Análisis de los efectos dispersivos y no lineales en un canal óptico empleando métodos numéricos».

[30] G. Rentao, L. Xiaoxu, L. Hui y B. Lin, «Evolutional algorithm based cascade long reach passive optical networks planning,» Communications, China, vol. 10, nº 4, pp. 59-69, 2013.

[31] A. Mitcsenkov, G. Paksy y T. Cinkler, «Geography-and infrastructure-aware topology design methodology for broadband access networks (FTTx),» Photonic Network Communications, vol. 21, nº 3, pp. 253-266, 2011.

[32] A. Mitcsenko, G. Paksy y T. Cinkler, «Topology design and capex estimation for passive optical networks,» de Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009. Sixth International Conference on, 2009.