



HAL
open science

Autonomic maintenance of high programmable optical access network

Nejm Eddine Frigui

► **To cite this version:**

Nejm Eddine Frigui. Autonomic maintenance of high programmable optical access network. Networking and Internet Architecture [cs.NI]. Ecole nationale supérieure Mines-Télécom Atlantique, 2019. English. NNT : 2019IMTA0127 . tel-02176863

HAL Id: tel-02176863

<https://theses.hal.science/tel-02176863>

Submitted on 8 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE
COMUE UNIVERSITE BRETAGNE LOIRE

Ecole Doctorale N°601
*Mathématique et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*
Par

« **Nejm Eddine FRIGUI** »

« **Maintenance autonome du réseau programmable d'accès optique de très haut débit** »

Thèse présentée et soutenue à RENNES, le 21 Janvier 2019
Unité de recherche : **IRISA Lab Rennes & Orange Labs Lannion**
Thèse N° : 2019IMTA0127

Rapporteurs avant soutenance :

Abderrezak RACHEDI Professeur des Universités, Université Paris-Est Marne-la-Vallée (UPEM), Paris
Damien MAGONI Professeur des Universités, Université de Bordeaux, Bordeaux

Composition du jury :

Président :	Damien MAGONI	
Examineurs :	Abderrezak RACHEDI	Professeur des Universités, Université Paris-Est Marne-la-Vallée (UPEM), Paris
	Damien MAGONI	Professeur des Universités, Université de Bordeaux, Bordeaux
	Ada DIACONESCU	Maître de Conférence, Telecom ParisTech, Paris
	Benoit RADIER	Docteur Ingénieur de Recherche Expert, Orange Labs, Lannion
Dir. de thèse :	Jean-Marie BONNIN	Professeur des Universités, IMT Atlantique Bretagne, Rennes
Co-dir. de thèse :	Tayeb LEMLOUMA	Maître de Conférence HDR, Université de Rennes 1, Lannion

Invité(s) :

Renaud LE MEUR Ingénieur Chef de Projet, Orange Labs, Lannion
Stéphane GOSSELIN Ingénieur Responsable de Projet de Recherche, Orange Labs, Lannion

ACKNOWLEDGEMENT

Foremost, I would like to thank my PhD examination committee members for the honor that they have given me by accepting to judge this work. Special thanks go to Pr. Abderrezak Rachedi, Pr. Damien Magoni, and Dr. Ada Diaconescu for accepting to review this work in order to give me valuable and detailed reports to improve the thesis manuscript.

I sincerely would like to thank my academic supervisors Pr. Tayeb Lemlouma and Pr. Jean-Marie Bonnin as well as my Orange Labs Lannion supervisors Renaud Le Meur, Benoit Radier and Stéphane Gosselin for their guidance and their help during the whole thesis. My sincere gratitude is expressed to them for their continuous support in my PhD researches, their patience, motivation, knowledge, the time and effort that they dedicated in reviewing my scientific articles, their comments and remarks on the thesis manuscript, and their advices throughout the three years.

My sincere thanks are conveyed to Hubert Sauvage, René Chapelain, Nicolas Neyret and Anne-Marie Blanchard, who helped me to evaluate my approach within a real testing platform and intervened to solve any faced technical problem.

Special thanks to my Orange Labs Lannion managers Emmanuelle Cressan, Jacques Thiebaut, and Michel Beuvelot for their assistance and support to succeed this thesis experience

I am also grateful to all my colleagues in Orange Labs with whom I shared great and special moments.

I am especially thankful to my parents Souad and Najeh, and my brother Ilyes for the unconditional support, their love, the efforts that they have made to ensure my arrival at this level of education, without them, I will never be here.

I would also like to express my gratitude to my brother friends Hafedh and Hamza with whom I spend a lot of time and enjoy funny moments. Thank you for the role you played to encourage me and support me all the time.

Au cours des dernières années, les réseaux de télécommunication ont considérablement évolué permettant la croissance et l'hétérogénéité du nombre des terminaux de communication ainsi que le lancement d'une nouvelle gamme de services gourmands en termes de débit tels que la vidéo à la demande, la visioconférence, les services de vidéo à trois dimensions, etc. A l'horizon 2020 et avec l'arrivée prévue de la 5G, l'internet des objets, et les services à haute définition, nous assisterons à une forte croissance du trafic dans les réseaux fixes et mobiles. Ceci suggère une anticipation de la part des opérateurs afin d'optimiser l'utilisation des ressources dans leurs réseaux pour adapter d'une manière dynamique les services proposés aux besoins des clients.

Le réseau Fiber To The Home (FTTH) représente actuellement l'infrastructure la plus évoluée et la plus fiable des réseaux d'accès fixe et il est recommandé comme la solution la plus adaptée pour bâtir les villes et les maisons intelligentes. Les réseaux optiques passifs (Passive Optical Networks (PONs)) représentent l'une des solutions les plus performantes du réseau d'accès FTTH. Ils ont été largement déployés par les opérateurs grâce à leur capacité d'offrir des services de très haut débit et de satisfaire les besoins de la majorité des clients en termes de qualité de service.

Cependant, en raison de la dynamique du trafic des différents clients, les réseaux PON doivent s'appuyer sur un mécanisme efficace pour l'allocation de ressources, plus particulièrement dans le sens montant. En effet, des débits modérés peuvent entraîner des conséquences insoutenables pour l'opérateur. Des clients insatisfaits et des pertes économiques importantes peuvent être causés si la qualité de service est dégradée. Ainsi, le mécanisme d'allocation de ressources doit être aussi fiable que possible pour éviter toute circonstance négative, notamment que nous sommes dans un environnement où l'opinion et l'expérience de l'utilisateur ont un impact direct sur la réputation de l'opérateur. Dans les réseaux PON actuels, l'allocation de la bande passante

dans le sens montant est assurée par le mécanisme Dynamic Bandwidth Allocation (DBA) qui réside dans l'Optical Line Terminal (OLT). Ce mécanisme définit pour chaque client le temps de transmission et la taille de données à envoyer en utilisant la technologie d'accès multiple à répartition dans le temps (TDMA) [99]. Ainsi, pour un intervalle de temps bien déterminé, seul un client est autorisé à transmettre ses données.

Etant donné que le DBA affecte directement les performances du réseau, de nombreux travaux ont été proposés dans la littérature dans le but d'améliorer son fonctionnement. Plusieurs critères ont été pris en compte tels que le temps de décision, la distribution de la bande passante par classe de service (distribuée ou centralisée), la capacité à respecter les contraintes liées aux débits des clients (ex., débit garantie et débit maximum définis comme des paramètres Service Level Agreement (SLA)), etc. Malgré que la majorité des travaux examinés contribuent à l'optimisation du l'algorithme DBA, un critère a été négligé. En effet, le DBA est un protocole de contrôle intégré dans l'OLT qui n'est pas sous le contrôle direct du l'opérateur. Ainsi, pour chaque modification dans cet algorithme, un recours au fournisseur d'équipement est exigé pour assurer la tâche de la mise à jour dans tous les OLTs. Ceci représente un obstacle pour l'opérateur qui veut optimiser son réseau d'une manière indépendante et de la façon la plus efficace possible.

L'objectif de cette thèse est de proposer une nouvelle architecture pour optimiser l'allocation de ressources dans les réseaux PON tout en agissant uniquement sur les paramètres SLA, désignés comme des paramètres gérables par l'opérateur. Comme ces paramètres sont actuellement de nature statique, nous avons proposé de les réajuster dynamiquement de manière à améliorer l'exploitation de la bande passante disponible dans le réseau sans modifier l'algorithme DBA lui-même dans l'OLT. L'architecture proposée apporte deux contributions majeures. Tout d'abord, elle peut être directement intégrée dans le système de gestion du réseau PON sans aucun besoin à modifier l'algorithme au niveau de l'équipement. Deuxièmement, comme elle agit seulement sur des paramètres gérables, l'architecture proposée peut se reposer sur le concept autonome/cognitif pour enrichir le réseau avec des éléments d'auto-décision qui confient la tâche de reconfiguration dynamique des paramètres SLA au réseau lui-même avec le minimum d'intervention humaine directe.

L'ajustement dynamique des paramètres SLA est assuré sur la base de l'estimation des modèles de trafic de chaque client. Des techniques de classification basées sur l'apprentissage automatique sont utilisées pour analyser le comportement des différents utilisateurs et déterminer leurs tendances de trafic (ex., intensif, faible, ou flexible). Ensuite, la bande passante qui a été estimée comme inutilisée dans le réseau est réattribuée aux utilisateurs qui ont besoin d'un surplus de débit sous la forme d'un supplément dans leurs paramètres SLA. Ce qui permettrait

au final de maximiser la satisfaction globale des clients vis-à-vis du réseau.

Pour la partie classification, on s'est référé dans cette thèse à deux types de méthodes appartenant à la famille de l'apprentissage non supervisé: méthodes de partitionnement et méthodes basées sur la densité des points. Le but est de classer les clients sur la base de leurs historiques de débits sans hypothèse préalable.

Les méthodes de partitionnement représentent le moyen le plus simple et le plus fondamental pour regrouper des points dans des clusters. Elles prennent un ensemble de n points et essayent de construire k clusters de telle façon que chaque point appartienne exactement à un seul cluster. La métrique utilisée pour différencier les clusters dans la plupart des méthodes de partitionnement est la distance entre les points. Par conséquent, les points appartenant au même cluster sont probablement proches. Les méthodes basées sur la densité représentent les approches de regroupement qui divisent un ensemble de points en régions denses séparées par des régions dispersées. La capacité d'un cluster donné continue à augmenter jusqu'à que le nombre de points, c'est-à-dire la densité, atteigne un certain seuil. Deux algorithmes bien connus ont été évalués, respectivement K-means et DBSCAN, afin de choisir la meilleure solution en termes de distribution équilibrée.

Même s'il avait été proposé il y a longtemps, K-means reste l'un des algorithmes de partitionnement les plus utilisés par les chercheurs grâce à sa simplicité et à sa facilité de mise en œuvre. Sur la base d'un dataset donné, K-means sélectionne arbitrairement k points parmi les points de dataset appelés centroïdes de cluster. Ensuite, chaque point restant sera affecté à l'un des k groupes de manière à minimiser la distance euclidienne entre les points et les centroïdes.

DBSCAN représente l'un des algorithmes les plus courants et les plus cités de classification basée sur la densité. Proposé pour la première fois en 1996, l'idée de DBSCAN est de trouver des points qui ont des voisinages denses afin de les connecter et de former les clusters correspondants. Contrairement aux algorithmes de partitionnement, DBSCAN ne nécessite aucune spécification concernant le nombre de clusters. Seuls deux paramètres doivent être pris en compte: le rayon maximal de chaque cluster et le plus petit nombre de points d'un cluster, qui représente le seuil de densité d'une région dense.

Les résultats produits ont montré que K-means surclasse DBSCAN selon les critères fixés. La classification proposée est effectuée par intervalle de temps (ex. 5 min). Pour pouvoir avoir une vision globale sur la tendance de trafic de chaque client, un algorithme de calcul d'index a été proposé. Cet algorithme permet sur la base des résultats de la phase d'apprentissage non

supervisé de donner sur une échelle de 1, la probabilité qu'un client ait une tendance faible, intensive ou flexible par rapport à sa consommation.

Bien que l'approche proposée se caractérise par sa grande précision au niveau de l'affectation d'un utilisateur PON à une certaine classe de trafic, la phase d'évaluation des modules de classification proposés a montré que la distribution des utilisateurs n'est pas équilibrée dans la manière où la majorité des clients ont été assignés à la classe flexible. Cela peut être un obstacle à l'objectif global qui consiste à maximiser autant que possible la satisfaction de la majorité des clients. Il est donc préférable d'avoir un nombre important d'utilisateurs à forte consommation et d'utilisateurs légers afin de maximiser l'efficacité de l'allocation de la bande passante dans le réseau.

Pour résoudre ce problème et faire face au comportement de trafic inattendu de certains utilisateurs du réseau PON tout en en apprenant plus sur la tendance de trafic des utilisateurs flexibles, nous avons supposé que le futur comportement d'un utilisateur PON dépend de son comportement antérieur. Ainsi, les approches de prédiction pouvant traiter l'analyse et la prévision de l'attitude du trafic des utilisateurs du réseau ont été fortement recommandées.

Dans ce contexte, nous avons étendu l'approche proposée par un module prédictif que nous l'avons considéré comme un complément de la première phase de classification. Ce module permet d'améliorer la phase d'apprentissage en anticipant le comportement futur possible d'un client flexible et sa tendance du trafic. Nous avons fait référence aux techniques de prédiction sur la base des séries temporelles (time series forecasting). Grâce à la collecte et à l'analyse d'observations passées, les méthodes de prédiction sur la base des séries temporelles tentent d'extraire la relation entre les différents points afin de proposer une extrapolation à l'avenir de la série temporelle étudiée. Dans la littérature, les modèles classiques et les plus simples en termes de compréhension, d'évaluation et de mise en œuvre restent les modèles linéaires qui ont été principalement utilisés au cours des dernières décennies. Ces modèles reposent sur l'utilisation de fonctions linéaires pour prédire les valeurs futures en fonction des valeurs passées.

Lorsqu'il s'agit de séries temporelles ayant un motif non linéaire, un modèle non linéaire est préféré. Dans ce cas, les réseaux de neurones artificiels (ANN) sont les meilleurs candidats à sélectionner en raison de leur flexibilité en termes de modélisation et d'identification de motifs. Comme il est habituel que les modèles linéaires ainsi que les réseaux de neurones ont généralement besoin d'un nombre considérable de données pour effectuer une bonne précision de prédiction, une autre théorie de prédiction est investiguée s'intitulant la théorie de Grey. Cette dernière a été conçue pour traiter des systèmes caractérisés par des informations incomplètes ou

des ensembles de données limités.

Dans notre thèse, une étude évaluative a été menée sur la base des deux approches ARIMA, et Grey (les réseaux de neurones n'ont pas été évalués à cause du dataset limité en termes de nombre de points). Le modèle ARIMA est l'un des modèles de prédiction linéaires les plus efficaces. Il est construit sur l'hypothèse que la série temporelle est stationnaire alors que la stationnarité peut être définie comme le maintien des propriétés statistiques de la série temporelle (c'est-à-dire que les propriétés statistiques telles que la moyenne ou la variance restent constantes) dans le temps. Ainsi, si une série temporelle suit un comportement particulier au cours d'une période donnée, il est hautement probable qu'elle continue à suivre le même comportement au cours de la période suivante.

Concernant le modèle de Grey, grâce à sa capacité à estimer le comportement possible des données en se basant uniquement sur quelques échantillons d'informations, il est devenu l'un des modèles de prédiction les plus utilisés dans le monde de la recherche. Il est très avantageux lorsqu'il s'agit de systèmes caractérisés par des informations incomplètes (partiellement connues) et/ou pauvres (minimum de 4 points de données) et n'impose aucune hypothèse relative à la distribution statistique des données. Le concept du modèle de Grey repose sur l'identification de la relation mathématique entre les différents points pour en savoir plus sur le comportement de l'ensemble de données et pour prendre la bonne décision concernant la tendance future.

L'étude évaluative a montré des performances élevées du modèle de Grey par rapport au modèle ARIMA sur la base de plusieurs métriques tels que la précision, l'erreur quadratique moyenne, etc. Par la suite, le modèle de Grey a été utilisé pour générer la nouvelle distribution des utilisateurs. Une amélioration au niveau de la classification des clients a été distinguée avec moins de clients flexibles contre une augmentation des utilisateurs des autres classes. Ce qui justifie la validité de l'ajout du module prédictif.

Nous avons utilisé par la suite le concept autonome/cognitif pour assurer l'accomplissement de la reconfiguration dynamique des paramètres SLA par le réseau lui-même avec un minimum d'intervention humaine. En effet, les deux versions proposées de notre approche doivent d'une part, effectuer la mise à jour des paramètres SLA au niveau des équipements et d'autre part, informer l'opérateur des nouvelles modifications. Étant donné que la gestion actuelle des paramètres SLA est sous la responsabilité des équipes opérationnelles, le processus de réglage effectué à différentes périodes de la journée peut entraîner des réactions lentes et inefficaces à mesure que le système de gestion devient de plus en plus complexe. En général, dans un contexte d'optimisation de réseau, il convient d'éviter autant que possible la gestion manuelle afin de

fournir des services de réseau efficaces et efficients. Ainsi, la tâche de reconfiguration dynamique des paramètres SLA doit être accomplie par le réseau lui-même, avec le minimum d'intervention humaine directe.

Une étude comparative a été effectuée entre plusieurs modèles autonomiques sur la base de critères bien définis, tels que le type d'approche (centralisée, décentralisée, hybride), la coordination, la résolution des conflits, la traduction des commandes, etc.

Suite à cette étude, le modèle GANA a été choisi pour implémenter la boucle décisionnelle de notre approche proposée. L'avantage majeur de GANA par rapport aux autres modèles est qu'il est normalisé par l'ETSI, ce qui facilite la mise en œuvre du modèle par les opérateurs. La conception architecturale de ses boucles de contrôle hiérarchiques facilite la mise en œuvre d'éléments opérationnels orientés vers la décision dans différents équipements de réseau. GANA offre aux opérateurs la possibilité de gérer leur réseau via des éléments décisionnels qui permettent la gouvernance de réseau en fournissant des interfaces appropriées avec un opérateur humain. Ces éléments décisionnels sont utilisés pour assurer la gestion autonome du processus d'optimisation. L'ensemble de l'architecture est contrôlé par une interface de gouvernance qui permet d'injecter des politiques, de modifier des paramètres, de contrôler et d'arrêter, si nécessaire, le processus de self-optimisation.

GANa propose quatre niveaux principaux de boucles de contrôle, respectivement le niveau de protocole, le niveau fonction, le niveau nœud pour les boucles de contrôle rapides et le niveau réseau pour les boucles de contrôle lentes. Les différents éléments décisionnels sont orchestrés dans ces quatre niveaux de manière à ce que chaque élément décisionnel gère un ou plusieurs éléments décisionnels de niveau inférieur qui lui sont associés. Ainsi, les éléments décisionnels de niveau inférieur sont considérés comme des entités gérées des éléments supérieurs. Pour définir ce que le réseau est censé faire et pour offrir des services et des règles à différents équipements de réseau, GANA introduit un plan de connaissances au niveau réseau qui peut également être considéré comme le niveau cognitif supérieur de GANA.

Pour évaluer l'approche proposée et montrer son efficacité en termes d'amélioration de la qualité de service, une étude expérimentale a été conduite. Deux cas d'utilisation ont été considérés pour démontrer l'applicabilité de la méthodologie proposée: (i) un cas d'utilisation de démonstration basé sur un scénario de réseau réel inspiré d'une infrastructure PON réelle utilisant les traces de trafic collectées comme données historiques. (ii) Un cas d'utilisation du trafic selon lequel différents scénarios d'utilisation de la bande passante ont été considérés (par exemple, faible utilisation de la bande passante, utilisation moyenne de la bande passante, utilisation élevée de la bande passante).

Pour évaluer ces deux scénarios expérimentaux avant et après l'application de l'approche d'optimisation, nous avons effectué nos tests au sein d'un réseau d'accès GPON utilisé par Orange Labs France à des fins de test. Conformément aux spécifications de la norme GPON, nous avons considéré des capacités de bande passante en amont de 1,25 Gbit/s et de 2,5 Gbit/s en aval. Pour la génération de trafic, nous avons injecté des charges de trafic dans le réseau PON à l'aide des outils de génération de trafic fournis par la plate-forme de banc d'essai Spirent Test Center. Pour avoir un raisonnement réaliste, ces charges de trafic étaient basées sur le profil de trafic de chaque ONU. Selon l'étude d'apprentissage réalisée, les utilisateurs légers et flexibles ne demandent pas une transmission de données en amont au-delà de leurs paramètres SLA. Pour montrer l'impact de notre approche d'optimisation, nous avons supposé que les gros utilisateurs exigent davantage de transmission de données que ce qu'ils sont autorisés à envoyer. Leur trafic est supposé dépasser leurs PIR. Les charges de trafic sont modifiées toutes les 5 minutes pour toutes les ONUs.

Les résultats obtenus ont montré que le mécanisme d'optimisation est très performant lorsque la distribution des clients tend à avoir une majorité de clients avec une faible utilisation de trafic. Avec un nombre important d'utilisateurs à forte utilisation de trafic et lorsque nous nous rapprochons de la capacité totale en amont du réseau PON, le mécanisme d'optimisation devient moins efficace à mesure que le supplément de bande passante estimé diminue. En conclusion, l'approche d'optimisation est très bénéfique lorsqu'il existe une quantité importante de bande passante en amont théoriquement inexploitée par le DBA et gaspillée en raison des contraintes SLA. Ainsi, plus la bande passante inutilisée par les utilisateurs légers est élevée, plus le mécanisme d'optimisation est efficace.

Comme perspectives, nous comptons améliorer l'approche proposée par l'ajout d'un module d'apprentissage par renforcement pour optimiser la sélection des différents paramètres dans la phase de classification. En effet, nous pensons que la phase de sélection des paramètres comme les intervalles de temps, les inputs de clustering, les différents seuils utilisés, etc., devrait être davantage étudiée pour converger vers la classification d'utilisateurs la plus efficace. Pour cela, l'ajout d'un module d'apprentissage par renforcement au mécanisme d'optimisation pourrait permettre d'améliorer la sélection des paramètres grâce à des expériences itératives, des essais et des retours d'actions et d'expériences précédentes. L'apprentissage par renforcement est une technique d'apprentissage automatique qui utilise des agents informatiques qui interagissent avec l'environnement, apprennent et prennent les décisions appropriées en fonction des actions passées.

Nous prévoyons aussi d'ajouter un module de sensibilité au contexte qui permet de consolider la phase d'apprentissage par des données contextuelles tels que les conditions météorologiques, les évènements quotidiens, hebdomadaires ou mensuels, les débuts et fin de vacances, etc. L'approche proposée a été destinée au trafic montant. Néanmoins, nous essayerons de l'adapter dans des prochains travaux pour qu'elle soit applicable aussi dans le sens descendant.

The introduction of new bandwidth-hungry services as well as the growth in the number of communication terminals in the last years has led to an exponential growth of data traffic in both fixed and mobile networks. Passive Optical Networks (PONs) representing one of the most attractive Fiber To The Home (FTTH) access network solutions, have been widely deployed for several years thanks to their ability to offer high-speed services and satisfy the needs of the majority of customers in terms of QoS. However, due to the dynamicity of users traffic patterns, PONs need to rely on an efficient upstream bandwidth allocation mechanism to define for each customer the amount of data that needs to be transmitted at a specific time. This mechanism is currently limited by the static nature of Service Level Agreement (SLA) parameters which can lead to an inefficient bandwidth allocation in the network.

The objective of this thesis is to propose a new management architecture for optimizing the upstream bandwidth allocation in PON while acting only on manageable parameters to allow the involvement of self-decision elements into the network. To achieve this, classification techniques based on machine learning approaches are used to analyze the behavior of PON users and specify their upstream data transmission tendency. A dynamic adjustment of some SLA parameters is then performed to maximize the overall customers' satisfaction with the network. The proposed architecture comes with two major contributions. First, it can be directly and easily integrated in the PON management system without a need to modify the resources allocation mechanism itself in the equipment. Second, as it focuses only on manageable parameters, the proposed approach gives us the opportunity to apply the autonomic and cognitive paradigm in order to enrich the network with self-decision capabilities that leave the task of the dynamic reconfiguration of the SLA parameters to the network itself with the minimum of direct human intervention.

The proposed architecture is extended by adding a forecasting module as a complement to the first classification phase. Some well-known forecasting techniques are used in this context to learn more about the traffic trend of users and improve their assignment. Moreover, autonomic and cognitive networking fundamentals are used to ensure the accomplishment of the task of dynamic reconfiguration of SLA parameters by the network itself with the minimum degree of human intervention using specific self-decision elements. Lastly, an experimental study is conducted to evaluate the proposed architecture in terms of QoS using some testing scenarios.

TABLE OF CONTENTS

1	Introduction	31
1.1	Context & Motivation	31
1.2	Contributions	34
1.3	Thesis Organization	35
2	Context	37
2.1	Introduction	37
2.2	FTTH Access Network Overview	38
2.2.1	FTTH Network Architectures	38
2.2.1.1	Point-to-Point (PtP) Architecture	38
2.2.1.2	Point to Multi-Points (PtMP) Architecture	39
2.2.2	PON Overview	40
2.2.2.1	PON Architecture	40
2.2.2.2	PON Standards	41
2.2.2.3	PON Management System	42
2.3	Traditional PON Maintenance	44
2.4	Towards Autonomic Proactive Maintenance	47
2.5	Conclusion	51
3	Management Architecture for Optimizing PON Upstream Resources Allocation	53
3.1	Introduction	53
3.2	Related Work	54
3.2.1	Upstream Bandwidth Allocation	54
3.2.1.1	Static Bandwidth Allocation (SBA)	55

TABLE OF CONTENTS

3.2.1.2	Dynamic Bandwidth Allocation (DBA)	55
3.2.2	Dynamic Bandwidth Allocation Optimization Approaches	58
3.2.3	Limits & Motivations	60
3.3	Management Architecture for Optimizing PON Upstream Resources	61
3.3.1	Overall Architecture	61
3.3.2	Detailed Architecture	63
3.3.2.1	Monitored Data Collector	63
3.3.2.2	Users Classification	63
3.3.2.2.1	Clustering Module	63
3.3.2.2.2	Assignment Index Calculator	64
3.3.2.3	Reallocation of SLA Parameters	67
3.4	Users Classification	68
3.4.1	Clustering Tendency	68
3.4.2	Clustering Algorithms	69
3.4.2.1	Partitioning Methods	69
3.4.2.2	Density-Based Methods	70
3.4.3	Experimental Evaluation	70
3.4.3.1	Environment Settings	70
3.4.3.2	Results & Analysis	71
3.4.3.3	Discussion	77
3.5	Conclusion	78
4	Forecasting PON Users Traffic Behavior	79
4.1	Introduction	79
4.2	Related Work	80
4.2.1	Auto-Regressive Integrated Moving Average (ARIMA)	81
4.2.2	Artificial Neural Networks (ANNs)	83
4.2.3	Grey Theory	85
4.3	Forecasting PON Users Traffic Behavior	87
4.4	Evaluation	89
4.4.1	Environment Settings	89
4.4.2	Results & Analysis	90
4.4.3	Discussion	100
4.5	Conclusion	100
5	A GANA Oriented Architecture for Self-Optimized Passive Optical Network	103
5.1	Introduction	103
5.2	Related Work	104

5.2.1	Autonomic Network Approaches	106
5.2.1.1	Centralized Autonomic Approach	106
5.2.1.2	Decentralized/Distributed Autonomic Approach	106
5.2.1.3	Hybrid Autonomic Approach	107
5.2.2	Autonomic Network Models	107
5.2.3	Cognitive Networks: An Alternative Vision of Autonomic Networks	115
5.3	A Comparative of Autonomic models	116
5.4	GANA-Oriented Architecture for Self-Optimizing Passive Optical Network Up- stream Resources Allocation	119
5.4.1	Proposed Model	119
5.4.2	Runtime Scenario	122
5.5	Conclusions and Future Work	124
6	Use Cases - Evaluation of the Self Optimized Passive Optical Network Archi- tecture	125
6.1	Introduction	125
6.2	Evaluation Environment	126
6.3	Demonstration Use Case	126
6.3.1	Use Case Objective	126
6.3.2	Experimental Scenario	126
6.3.3	QoS Assessment	129
6.4	Traffic Usage Variation Use Case	132
6.4.1	Use Case Objective	132
6.4.2	Experimental Scenario & QoS Assessment	132
6.4.2.1	Low Bandwidth Usage	133
6.4.2.2	Average Bandwidth Usage	135
6.4.2.3	High Bandwidth Usage	137
6.4.3	Discussion	138
6.5	Conclusion	139
7	Conclusion	141
7.1	Summary	141
7.2	Perspectives	143
	Bibliography	143

LIST OF FIGURES

2.1	Passive Optical Network architecture [47]	40
2.2	Model for layering management functions	42
2.3	PON management system	43
2.4	Autonomic control loop	49
3.1	Static bandwidth allocation and dynamic bandwidth allocation mechanisms . . .	55
3.2	Status reporting DBA mode	57
3.3	Management architecture for optimizing PON upstream resources	61
3.4	Distribution of users per mean upstream bitrate slices in a given day period . . .	71
3.5	Hopkins statistic index	72
3.6	K-means results (k=3, metric: ONUs mean upstream bitrates)	73
3.7	K-means results (k=3, metric: decimal logarithm of ONUs mean upstream bitrates) . . .	73
3.8	DBSCAN results ($\epsilon=10$ Mbit/s, $MinPts = 1$)	74
3.9	OLT user distribution for Wednesdays	75
3.10	OLT user distribution for Saturdays	76
3.11	A PON port user distribution for Wednesdays	76
3.12	A PON port user distribution for Saturdays	77
3.13	A PON port user distribution for Wednesdays and Saturdays	77
4.1	A typical MLP ANN structure with one hidden layer and one output	84
4.2	A feed-forward scheme for training process in MLP	84
4.3	Enhanced model using a forecasting module	87
4.4	Proposed method for forecasting flexible users assignment indexes	88
4.5	A PON port user distribution for Wednesdays and Saturdays	92

4.6	Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using ARIMA(1,0,0)	93
4.7	Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using ARIMA(2,0,1)	94
4.8	Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using GM(1,1)	95
4.9	A PON port user distribution for Wednesdays And Saturdays Using GM(1,1)	97
4.10	Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using ARIMA(1,0,0)	98
4.11	Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using ARIMA(2,0,1)	98
4.12	Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using GM(1,1)	99
5.1	The proposed model for optimizing PON upstream resources: initial version	105
5.2	The proposed model for optimizing PON upstream resources: enhanced version	105
5.3	IBM MAPE-K reference model architecture [60]	108
5.4	FOCALE architecture [102]	109
5.5	GARSON architecture [23]	110
5.6	UMF functional blocks [103] Copyright © 2012, IEEE	111
5.7	GANA architecture [108]	113
5.8	GANA-oriented architecture for self-optimizing passive optical network upstream resources allocation	119
5.9	Runtime scenario	122
5.10	Continuous monitoring phase	123
6.1	Users classification	128
6.2	Allocated upstream bitrates before applying the optimization mechanism	130
6.3	Allocated upstream bitrates after applying the optimization mechanism	131
6.4	Upstream bandwidth usage ratio before and after applying the optimization mechanism in a low bandwidth usage scenario	135
6.5	Upstream bandwidth usage ratio before and after applying the optimization mechanism in an average bandwidth usage scenario	136
6.6	Upstream bandwidth usage ratio before and after applying the optimization mechanism in a high bandwidth usage scenario	138
6.7	Maximum upstream bandwidth usage ratio for each traffic variation scenario before and after applying the optimization mechanism	139

6.8 Average upstream bandwidth usage ratio for each traffic variation scenario before and after applying the optimization mechanism 139

LIST OF TABLES

3.1	T-CONT types	56
3.2	SLA parameters	58
3.3	Clustering explanation	64
3.4	Table of notations	65
4.1	Dataset instantiated for each flexible user	88
4.2	Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using ARIMA(1,0,0)	93
4.3	Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using ARIMA(2,0,1)	94
4.4	Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using GM(1,1)	95
4.5	Comparative table of forecasting models	96
4.6	Comparative table of forecasting models for all weekdays	99
5.1	Comparison of autonomic approaches	107
5.2	Evaluation of autonomic models	118
6.1	ONUs upstream SLA parameters	127
6.2	Estimation of the bandwidth used by light users	128
6.3	Extra bandwidth, η , and α values in a demonstration use case scenario	129
6.4	ONUs new upstream SLA parameters	129
6.5	Average latency measures	131
6.6	Users classification for traffic usage variation use case scenarios	133
6.7	Ranges of injected loads for different traffic classes	133

LIST OF TABLES

6.8 ONUs classification in a low bandwidth usage scenario 134

6.9 Results of bandwidth estimation and sharing in a low bandwidth usage scenario . 134

6.10 Estimation of the bandwidth used by light users in a low bandwidth usage scenario134

6.11 Extra bandwidth, η , and α values in a low bandwidth usage scenario 134

6.12 ONUs classification in an average bandwidth usage scenario 135

6.13 Results of bandwidth estimation and sharing in an average bandwidth usage scenario136

6.14 Estimation of the bandwidth used by light users in an average bandwidth usage
scenario 136

6.15 Extra bandwidth, η , and α values in an average bandwidth usage scenario 136

6.16 ONUs classification in a high bandwidth usage scenario 137

6.17 Estimation of the bandwidth used by light users in a high bandwidth usage scenario137

6.18 Extra bandwidth, η , and α values in a high bandwidth usage scenario 137

Published Work in International Conferences

- N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J.-M. Bonnin, “Dynamic reallocation of SLA parameters in passive optical network based on clustering analysis,” in 2018 21st Conference on Innovation in Clouds, Internet and Networks (ICIN 2018), Paris, France, Feb. 2018.
- N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J. M. Bonnin, “Optimization of the upstream bandwidth allocation in passive optical networks using internet users’ behavior forecast,” in ONDM 2018 - 22nd International Conference on Optical Network Design and Modeling (ONDM 2018), Dublin, Ireland, May 2018.

Unpublished Work/Work Under Submission

- N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J.-M. Bonnin, “Self Optimization of Passive Optical Network Upstream Resources Allocation”, Journal Paper
- N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J.-M. Bonnin, “Cognitive Architecture for Self Optimized Passive Optical Network Using Machine Learning Techniques”, Conference Paper

International Work Presentations

- N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J.-M. Bonnin, “Autonomic Maintenance of FTTH Access Network” in 2017-2nd GDR RSD and ASF Winter School on Distributed Systems and Networks

ACF	Auto Correlation Function.
AI	Assignment Index.
AIC	Akaike's Information Criterion.
AIR	Assured Information Rate.
ANN	Artificial Neural Network.
AON	Active Optical Network.
API	Application Programming Interface.
ARIMA	Auto-Regressive Integrated Moving Average.
BML	Business Management Layer.
BSS	Business Support System.
CIR	Committed Information Rate.
CLI	Command Line Interface.
CO	Central Office.
CoS	Class of Service.
CPI	Cognitive Process Instance.
CRM	Customer Relationship Management.
DAG	Data Acquisition and Generation.
DBA	Dynamic Bandwidth Allocation.
DBRu	Dynamic Bandwidth Report upstream.
DBSCAN	Density Based Spatial Clustering of Applications with Noise.
DE	Decision making Element.

DEN-ng	Directory Enabled Network-next generation model.
EIR	Excess Information Rate.
EML	Element Management Layer.
EMS	Element Management System.
eTOM	extended Telecommunications Operations Map.
FA	Forecast Accuracy.
FB	Functional Block.
FE	Forecast Error.
FIR	Fixed Information Rate.
FOCALE	Foundation, Observation, Comparison, Action, Learn, rEason.
FSMs	Finite State Machines.
FTTH	Fiber To The Home.
FTTH/P	Fiber to the Home/Premise.
GANA	Generic Autonomic Networking Architecture.
GARSON	Generic Architecture of Self-Organized Networks.
GEM	GPON Encapsulation Method.
GM(1,1)	Grey Model First Order One Variable.
GPON	Gigabit-capable Passive Optical Network.
GUI	Graphical User Interface.
IEEE	Institute of Electrical and Electronics Engineers.
IoT	Internet of Things.
IPTV	Internet Protocol television.
IT	Information Technology.
ITU-T	Telecommunication Standardization Sector of the International Telecommunication Union.
MAD	Mean Absolute Deviation.

MAPE-K	Monitor-Analyze-Plan-Execute over a shared Knowledge.
MBTL	Model Based Translation Layer.
MBTS	Model Based Translation Service.
MEs	Managed Entities.
MIB	Management Information Base.
MLP	Multilayer Perceptron Model.
MR	Mean Residuals.
NE	Network Element.
NEL	Network Element Layer.
NETCONF	Network Configuration Protocol.
NG-PON1	Next-Generation Passive Optical Network 1.
NG-PON2	Next-Generation Passive Optical Network 2.
NML	Network Management Layer.
NMS	Network Management System.
NSR	Non-Status Reporting.
ODN	Optical Distribution Network.
OEO	Optical-Electrical-Optical.
OID	Object Identifier.
OLT	Optical Line Termination.
ONIX	Overlay Network Information Exchange.
ONU	Optical Network Unit.
OPEX	Operational Expenditure.
OSS	Operational Support System.
PACF	Partial Auto Correlation Function.
PIR	Peak Information Rate.
PON	Passive Optical Network.
PtMP	Point to Multi-Points.
PtP	Point-to-Point.
QoE	Quality of Experience.
QoS	Quality of Service.
SBA	Static Bandwidth Allocation.

SLA	Service Level Agreement.
SML	Service Management Layer.
SNMP	Simple Network Management Protocol.
SOAP	Simple Object Access Protocol.
SR	Status Reporting.
T-CONT	Transmission Container.
TDMA	Time-Division Multiple Access.
TS	Tracking Signal.
TWDM	Time and Wavelength Division Multiplexing.
UHD	Ultra High Definition.
UMF	Unified Management Framework.
US-BWmap	Upstream Bandwidth Map.
VDSL	Very-high-bit-rate Digital Subscriber Line.
VoD	Video on Demand.
VoIP	Voice over Internet Protocol.
XG-PON	10 Gigabit-Passive Optical Network.

Contents

1.1 Context & Motivation	31
1.2 Contributions	34
1.3 Thesis Organization	35

1.1 Context & Motivation

Over the recent years, telecommunication networks have evolved significantly enabling the introduction of new services such as Video on Demand (VoD), Internet Protocol television (IPTV), video conferencing, 3D video services, etc., and the growth and the heterogeneity of communication terminals. With the expected arrival of the 5G, Internet of Things (IoT) and Ultra High Definition (UHD) services at the horizon 2020, we will witness a traffic explosion in both fixed and mobile networks. This suggests that network operators should think about cost-effective ways to handle the dynamicity of their customers' traffic patterns and maximize the efficiency of the bandwidth usage in their networks to adapt in a dynamic fashion the offered services to customers requirements.

Fiber To The Home (FTTH) networks represent currently the most reliable fixed access network infrastructure and they are considered as the most suitable solution for building smart cities and smart homes. Passive Optical Networks (PONs), representing one of the most attractive FTTH access network solutions, have been widely deployed for several years thanks to their ability to offer high-speed services and satisfy the needs of the majority of customers in terms of Quality of Service (QoS).

For reasons of innovation and control of technological developments, operators should keep their own PON infrastructure and management system effective for a long time. Indeed, a network failure or a low throughput means unsustainable consequences for the network operator. Dissatisfied customers and significant economic losses may be caused either by the failures themselves or by the time required to fix them without forgetting that the pressure to quickly restore operations may result in human configuration errors which may lead to a future network failure [81]. Thus, a maintenance system becomes an increasingly important need and should be as reliable as possible to avoid any negative circumstances, especially that we are in an environment where the opinion and the experience of the user have a direct impact on the reputation of the network operator. Thus, the most important thing after the PON deployment phase is how to improve its maintenance system and make it work in the most efficient way.

The current PON maintenance system requires a continuous intervention to provide the operational environment required for services to run smoothly, recover from faults, and maximize the overall performance of the network. It is affected by many factors that are hard to control and especially the complexity due to the several types of equipment and the variety of manufacturers [65]. The configuration and adaptation of network components require a deep knowledge and understanding of the operation and interactions between several processes and physical components [66]. Many steps are needed when an Information Technology (IT) administrator needs to monitor or modify the behavior of the different devices in a traditional network. First, he will have to manually configure multiple devices on a device by device basis. The next step is using device level management tools to update numerous configuration settings. This configuration approach makes it very complex for an administrator to deploy a consistent set of policies related to management problems.

Other fundamental problems with the current PON maintenance process is that operators own a variety of equipment from different vendors. Each vendor uses different programming languages and models with different semantics, making this a very difficult problem that impedes the interoperability of end-to-end solutions [44]. For this, to successfully complete a process, an administrator will therefore need an extensive knowledge of all present device types. This lack of adaptability to the network state makes the current network management unable to cope with the requirement of emerging services and time critical applications, characterized by highly demanding traffic, varying traffic patterns, and QoS requirements.

Considering that human beings administration is viewed as a total dependency and a full time need for intervention in deployment and maintenance actions and that manual control of network management processes may generate risks in terms of possible human errors and conflicts between different taken decisions, developing approaches intended to practical and cost effective survivability of maintenance mechanisms with a minimum of human intervention is becoming a key challenge to the continued development of viable management solutions [88].

As a reaction to this, reference can be made to the autonomic networking concept [60] as a suitable solution to be involved into the current and emerging maintenance systems. Autonomic networking focuses on the application of self-management principles to make networks and applications more intelligent by enabling them to make decisions without having to consult a human administrator. The motivation behind autonomic networking is to identify those functions that can be done with the minimum help of human being and to reduce the dependence on skilled resources. This not only lowers operational costs, but also enables valuable resources to be used for more complex tasks. It also enables the network to respond faster and more accurately to the changing needs of users and businesses.

The migration towards an autonomic maintenance of the whole PON from an operator perspective is not a simple task. Each application of an autonomic property including Self-Configuration, Self-Optimization, Self-Healing and Self-Protection, requires a deep knowledge and should be well studied to avoid any conflict possibility.

In this thesis, we will delve into the self-optimization aspects of PONs. Self-optimization can be a solution for network operators to improve the Quality of Experience (QoE) and solve the problems that can trouble their subscribers by improving the resources usage and reducing the costs in the network. A particular attention is given to the self-optimization of PON upstream resources. This is justified by the limits we identified in the reviewed work regarding the upstream resources allocation mechanism, especially the lack of the operator vision and the difficulty of changing the whole mechanism for each update due to the dependency on a specific vendor.

Current PONs rely on what is called a Dynamic Bandwidth Allocation (DBA) mechanism to efficiently assign, for each Optical Network Unit (ONU) connected directly to the end user, the corresponding upstream bandwidth according to its demand and taking into account the traffic status of other ONUs. The DBA was the subject of many discussions in the research world [15] with the aim to converge towards an optimal design in terms of bandwidth utilization, link capacity, packet loss, and upstream delay. By cons, despite the efforts made in this regard, an important criterion has not been taken into account regarding the implementation nature of the DBA. Indeed, in PON, the DBA which resides in the Optical Line Termination (OLT) is classified as a closed control protocol that cannot be directly modified by the network operator without recourse to the equipment supplier. It requires vendor specific developments and might not be easily integrated in operator's equipment each time there is an optimization. Thus, any new proposal by researchers in this respect regardless of the operator vision cannot be selected to be directly integrated.

To address the aforementioned issue, we propose in this work a new architecture for optimizing the DBA operation based only on the dynamic adjustment of some Service Level Agreement (SLA) parameters considered as the DBA external manageable inputs, while the DBA algorithm itself remains intact. The adjustment is performed based on learning users' daily life bandwidth

usage. Using some machine learning techniques, three classes of users are identified: heavy, light, and flexible. Then, based on the traffic profile of each user, heavy users will benefit from a bandwidth surplus which is determined based on an estimate of the bandwidth unused by light users at a specific day period. Finally, this surplus will be added to the heavy users' Peak Information Rates (PIRs) as the SLA parameters that represent the maximum bandwidth that can be provided.

The task of the dynamic reconfiguration of SLA parameters is accomplished by the network itself in such a way that they are automatically handled by self-decision elements without any direct human intervention based on the autonomic and cognitive networking fundamentals. This work represents a new leap forward in the field of the optimization of PONs upstream resources with a minimum level of human being involvement.

1.2 Contributions

The contributions of the thesis are summarized as follows:

- **Management architecture for optimizing PON upstream resources:** having studied the different proposals regarding the improvement of PON upstream resources dynamic allocation, we identify an opportunity to derive a new optimization approach based on the dynamic adjustment of SLA parameters by referring to customers' traffic behavior during the day. This traffic behavior is determined following a users' classification phase based on machine learning techniques which is performed to assign each user to a particular class (heavy, light, and flexible) depending on his historical upstream bandwidth usage. With regard to what has been presented before by researchers in this domain, the proposed architecture comes with two major contributions. First, it can be directly and easily integrated in the PON management system without the need to modify the DBA control algorithm at the OLT level. Second, as it focuses only on manageable parameters, the proposed approach gives us the opportunity to apply in a more advanced stage the autonomic & cognitive paradigm in order to enrich the network with self-decision capabilities that leave the task of the dynamic reconfiguration of the SLA parameters to the network itself with the minimum of direct human intervention. Although proposed for improving the reallocation of PON upstream resources, the proposed architecture is adaptable in a way that it can be applied also in the downstream direction. The idea of classifying users based on their traffic behavior using machine learning techniques and performing an SLA parameters' adjustment can be also used in wireless mobile networks.
- **A forecasting approach for studying PON users traffic tendency:** based on the issue related to users' classification identified in the proposed architecture which consists

on an unbalanced users' distribution where the majority of users are designated to be flexible, it is preferable to have a significant number of heavy and light users in order to maximize the efficiency of the bandwidth usage in the network. To resolve this issue and to cope with the unexpected traffic behavior of some PON users by learning more about the traffic tendency of flexible ones, we propose to enhance the proposed approach by adding a forecasting module. This module can effectively extract the trend information from ambiguity and can be considered as a part of a second user classification step that represents an improvement of the first classification method.

- **A GANA oriented architecture for self-optimized PON:** as the adjustment process of the SLA parameters is supposed to be performed at different day periods and due to the nature of the current management of these parameters which is under the responsibility of human operation teams, this may lead to slow, complex and inefficient reaction. To address this issue, we propose to self-handle the task of the dynamic reconfiguration of SLA parameters which will be accomplished by the network itself without direct human intervention. Autonomic & cognitive networking fundamentals are used as reference in this context to allow the injection of self-decision elements and the automation of the different machine learning techniques.
- **Evaluation of the self-optimized PON architecture:** to demonstrate the applicability of the proposed architecture, an evaluation study is conducted based on two use cases: (i) a demonstration use case based on a real network scenario inspired by a real PON infrastructure using collected traffic traces as historical data. (ii) A traffic usage variation use case where different bandwidth usage scenarios are considered (e.g. low bandwidth usage, average bandwidth usage, high bandwidth usage). The efficiency of the architecture is validated through a set of individual evaluation criteria.

1.3 Thesis Organization

The remainder of this thesis is organized as follows:

Chapter 2 “Context” presents some background on FTTH access network and its main architectures and standards. It presents how this kind of network is managed, how it can be maintained, what are the limits of the traditional maintenance and what are the existing approaches that can overcome these limitations. A critical study is presented to show which approach is better than the others and should be applied in our use case.

Chapter 3 “Management Architecture for Optimizing PON Upstream Resources Allocation” introduces a novel mechanism for optimizing the allocation of upstream PON resources based

on customers' estimated traffic patterns.

Chapter 4 "Forecasting PON Users Traffic Behavior" presents an extension of the proposed architecture based on a forecasting module.

Chapter 5 "A GANA Oriented Architecture for Self-Optimized PON" proposes a novel architecture for the PON traffic management model based on the autonomic/cognitive networking fundamentals.

Chapter 6 "Use Cases - Evaluation of the Self Optimized PON Architecture" provides an evaluative study to demonstrate the applicability and the efficiency of the proposed architecture.

Chapter 7 "Conclusions & Perspectives" gives a summary of the thesis and recalls the main contributions proposed. It concludes this dissertation with a vision on the perspectives of our work and some interesting directions that may improve the thesis.

Contents

2.1	Introduction	37
2.2	FTTH Access Network Overview	38
2.2.1	FTTH Network Architectures	38
2.2.2	PON Overview	40
2.3	Traditional PON Maintenance	44
2.4	Towards Autonomic Proactive Maintenance	47
2.5	Conclusion	51

2.1 Introduction

In the last years, the development of telecommunications networks and information technologies has become increasingly fast. A change of paradigm in fixed access networks has been experienced due to the growth and the heterogeneity of communication terminals and the introduction of new services such as VoD, IPTV, video conferencing, 3D video services, IoT, and UHD services. This evolution has been associated with an exponential growth of data traffic and a large demand for higher data speeds. In this context, ensuring network efficiency, preparing and structuring fixed access networks to converge voice, data, and video traffic with a focus on higher bandwidth applications is strongly recommended.

FTTH networks represent currently the most reliable fixed access network infrastructure and they are considered as the most suitable solution for building smart cities and smart homes. Nevertheless, service providers and telecommunications operators should always adapt in a dynamic fashion the offered services to their customers' requirements [31]. The transition towards

new standards increasingly sophisticated that offer higher bitrates is not an easy task. It will be associated to an increasing of the Operational Expenditure (OPEX), a complexity of network management which is already not easy to handle and it requires a lot of time in the deployment phase. Thus, until this time, it is much more preferable to ensure the maintenance of the current FTTH network architecture in order to optimize it and make it as efficient as possible.

This chapter presents some background on FTTH access network and its main architectures and standards. It presents how this kind of network is managed, how it can be maintained, what are the limits of the traditional maintenance and what are the existing approaches that can overcome these limitations. A critical study is presented to show which approach is better than the others and should be applied in our use case.

2.2 FTTH Access Network Overview

Conventional fixed access networks based on twisted-pair copper cable have very limited bandwidth distance products [63]. According to the standard [38] related to the Very-high-bit-rate Digital Subscriber Line (VDSL) technology (the most advanced copper-cable based technology), the maximum bitrate that can be reached in each direction is 100 Mbit/s over a distance of 500 meters with the requirement of using highly sophisticated transmission technologies. Copper-wires are also constrained by some electrical properties including the attenuation of signals during propagation along the cable length. Thus, it is essential to use another transmission medium that offers much more guaranteed bandwidth regardless of distances. Optical fibers look the most suitable solution for transmission supports in advanced fixed access networks. They outclass copper-wires in terms of signal range, resistance, low attenuation when routing the light signal over a long distance and insensitivity to external electrical disturbances. Optical fibers provide also an unlimited transmission bandwidth and the offered bitrates depend only on the equipment placed at the end. Fiber networks can be classed as green networks since the energy needed to convey a pulse of light is very small compared to that needed to transmit an electrical signal. To fully leverage these benefits, operators and service providers tend to provide an optical fiber from the Central Office (CO) to each customer premise or home. This type of network is commonly referred to as Fiber to the Home/Premise (FTTH/P) network [63]. The manner in which the optical fiber reaches the subscriber may vary according to the deployed FTTH network architecture. An overview of these architectures is outlined in the following.

2.2.1 FTTH Network Architectures

2.2.1.1 Point-to-Point (PtP) Architecture

In a Point-to-Point (PtP) architecture, all subscribers are connected to the CO via dedicated fibers (direct link). A high throughput is guaranteed by this technology for the subscriber since

there is no link sharing as well as a low attenuation generated by the fiber. Nevertheless, the PtP architecture is not considered as the best solution due to many practical constraints. The implementation of several dedicated fiber connections requires a very high number of fibers in the whole access network, which causes high costs for fiber rollout and handling. In addition, dealing with such large number of fiber connectors and optical splices is currently much more difficult than with copper wires, especially when the distance between customers and CO is more than 10 km [14][97]. In light of these constraints, PtP architecture seems inadequate especially in the case of a large-scale access network and it is preferred to look for point to multi-points architectures.

2.2.1.2 Point to Multi-Points (PtMP) Architecture

Unlike the PtP architecture, the Point to Multi-Points (PtMP) architecture offers one or more additional aggregation layers between the subscriber location and the CO in the access network [14]. Several subscribers share the same transmission support in order to reduce the high number of used fibers. Generally, two PtMP architectures can be distinguished: Active Optical Network (AON) and Passive Optical Network (PON).

- AON: it's characterized by an active aggregation element (e.g. Ethernet switch) able to route the signal and allow the operator to direct the different flows in order to ensure a better traffic management. AON allows a reduction of the fiber count in the access network compared to a PtP solution. Nevertheless, it increases the number of required Optical-Electrical-Optical (OEO) interfaces and simultaneously the power consumption in the network [14].
- PON: it does not include any active elements between the CO and the customer premises. Only optical fibers and optical passive elements which do not require electrical power or active management are present in the architecture. Traffic aggregation is based on passive components called optical power splitters. The PON architecture enables fiber reduction as well as the optimization of energy and the power consumption compared to a PtP architecture [63][97]. The PON fiber infrastructure is also future-proof, as optical fibers and splitters in the field do not need to be replaced when new PON technologies are installed at the CO and customer premises.

When designing their infrastructure, network operators look for the technology which fits their needs and those of its customers. Given the advantages of the PON architecture in terms of economies made on the amount of used fibers, infrastructure mutualization and the number of used interfaces, it was selected by Orange for the implementation of its FTTH access network. Thus, our work will be focused solely on this kind of architecture. We provide in the following more details about the PON architecture.

2.2.2 PON Overview

2.2.2.1 PON Architecture

The main building blocks of the PON architecture are outlined in Fig.2.1.

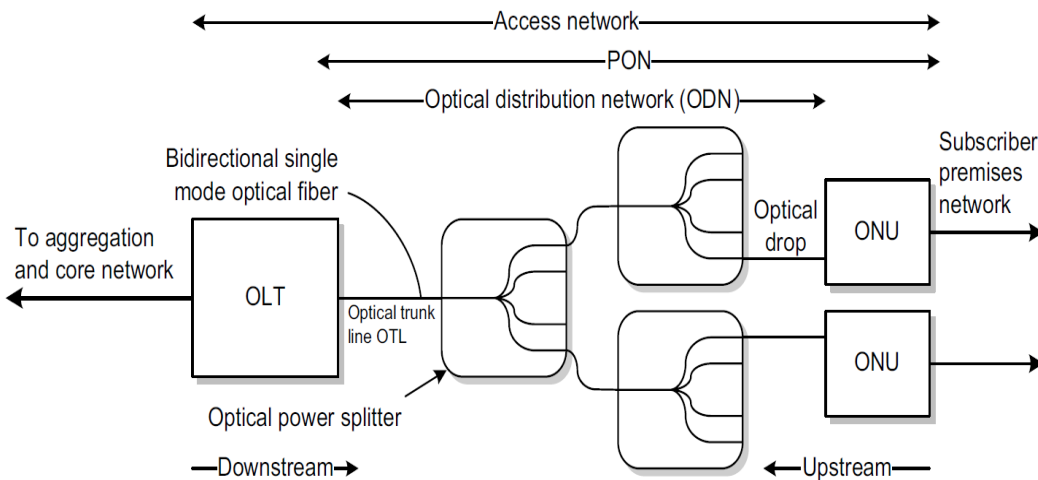


Figure 2.1: Passive Optical Network architecture [47]

The PON architecture relies on three main elements:

- **OLT:** an active equipment which connects the service provider to the optical distribution network. Its major functions consist in ensuring the conversion of the standard signals used by the service provider to the framing used by the PON system, the management and the registration of the different ONUs and the allocation of the bandwidth.
- **ONU:** an active equipment used to terminate the fiber line from the subscriber side. Its major functions are the aggregation of the different types of data coming from the customer and their transmission to the OLT through the Optical Distribution Network (ODN).
- **ODN:** it represents all optical fibers and passive splitters used to connect the OLT to the different ONUs.

Two major standard groups develop standards for PONs along with the help of other industry organizations, respectively the Institute of Electrical and Electronics Engineers (IEEE) and the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T). The implemented standards in Orange follow the ITU-T group. The next section provides an overview of the current state and future generations of these standards.

2.2.2.2 PON Standards

The current-generation of ITU-T PON standards relies on the Gigabit-capable Passive Optical Network (GPON) norm. GPON [35] was standardized for the first time in 2003 and several versions were proposed after with some improvements. It provides 2.5 Gb/s downstream and 1.25 Gb/s upstream capacities. These rates are shared between 64 users maximally, depending on the splitter options used. ONUs transmit their data to the OLT on the same wavelength. A Time-Division Multiple Access (TDMA) method is used to ensure the sharing of the speaking time between several subscribers in order to avoid the collision between the different signals in the case of simultaneous transmission [95]. The OLT assigns each ONU an interval of few microseconds in turn. During this time interval, the corresponding ONU is the only one authorized to transmit data. In case the ONU wants to transmit more data, an additional time interval may be allocated [35] if available. A DBA mechanism is used to enhance effective users bandwidths by allocating unused time slots to busier ONUs [47].

With the expected bandwidth growth in the next years, operators are looking at gigabit broadband speeds to their customers and there was an agreement that GPON needs eventually to be upgraded. The first method for upgrading GPON system consists in keeping the PON architecture untouched and increasing the transmission data rate shared by all users [63]. This approach has been standardized by the ITU-T in 2010 under the name of 10 Gigabit-Passive Optical Network (XG-PON) and called also Next-Generation Passive Optical Network 1 (NG-PON1) [36]. XG-PON is a mid-term solution compatible with legacy GPON ODNs. It can be seen as an enhanced TDMA-GPON release with two options: asymmetric XG-PON with rates of 10Gb/s in the downstream and 2.5Gb/s in the upstream. Symmetric XG-PON (XGS-PON) [34] with rates of 10Gb/s in both, downstream and upstream directions. XG-PON inherits the PtMP architecture, the framing and the management from GPON. A wavelength coupler and a 10 Gb interface board will be added at the OLT level to ensure the coexistence with the GPON system. In parallel, all ONUs shall have an integrated filter in order to eliminate the interferences between GPON and XG-PON wavelengths [36].

The second method for upgrading GPON systems consists in adding wavelength channels in both, upstream and downstream directions [63]. This method was taken into account by the ITU-T and was standardized for the first time in 2013 under the name of Next-Generation Passive Optical Network 2 (NG-PON2) [37]. NG-PON2 is a long-term solution which can be also deployed over existing ODNs. It outclasses XG-PON in terms of bandwidth capacity and downstream/upstream bitrates. Time and Wavelength Division Multiplexing (TWDM) technology has been selected as the approach to be employed in NG-PON2. By using four bidirectional wavelength channels, each of 10 Gb/s downstream and 2.5 Gb/s or 10Gb/s upstream, NG-PON2 will increase the PON capacity to 40 Gb/s downstream and 10 Gb/s or 40Gb/s upstream. In downstream, tunable filters will be used by ONUs in order to get only the desired wavelength.

In upstream, tunable lasers will be placed at each ONUs and will be dynamically assigned to a wavelength. Given that some customers may not be interested by very high bitrates and the deployment phase may take a lot of time, the coexistence between GPON, XG-PON and NG-PON2 must be ensured by operators to meet the needs of all customers. For this, the architecture of NG-PON2 includes a coexistence element which is a passive equipment placed in the CO allowing to combine the different wavelengths associated with each technology generation [37].

2.2.2.3 PON Management System

In order to be efficient and reliable during their life cycle, PONs require a level of monitoring, supervision, and management. Referring to [52], the complexity of network management can be handled by splitting the management functionality into five hierarchical layers as shown in Fig.2.2, each of which reflects particular aspects of management arranged by different levels of abstraction.

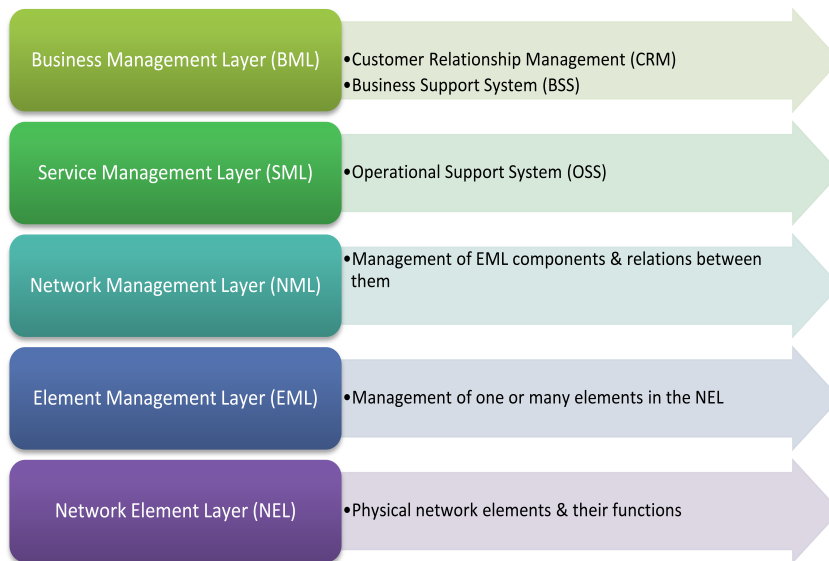


Figure 2.2: Model for layering management functions

The PON management system follows the same pattern described above. We highlight its modelling in Fig.2.3.

The lowest layer of the management hierarchy consists of two main devices previously described: the OLT and the ONU. The management of the ONUs is ensured by the OLT while the management of OLTs is provided by a Graphical User Interface (GUI)-based network management package called Element Management System (EMS). The EMS resides on a server and

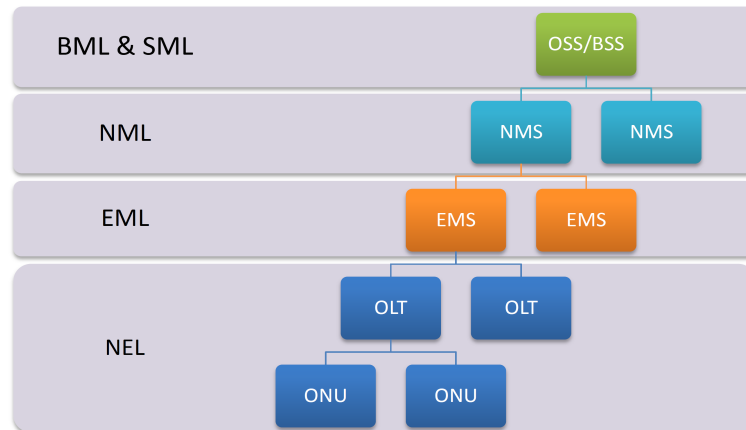


Figure 2.3: PON management system

supports a multiclient architecture. The intuitive interface in the EMS gives human provider an efficient control and management of the Network Element Layer (NEL). EMS capabilities include configuration management for OLTs and ONUs, fault detection, isolation and correction of errors, data analysis on network element behavior, real-time monitoring, and the ability to perform system backups and software upgrades. Administrators may connect to the EMS server through an EMS client or a separate interface [80]. The EMS could be embedded within the Network Element (NE) as it could be specific to a NE with respect to the manager-agent paradigm. The EMS could be also common to multiple NEs coming from the same vendor and supporting multiple technologies [106]. It can back up all provisioned and configured parameters and different information from OLTs, store them in its database, and ensure their transmission to the Network Management System (NMS) server.

The NMS is the above part of the EMS and it is used to manage the whole network. It could be common to multiple EMSs from different domains, or a specific one is needed for each EMS per domain manager [106]. The NMS uses the data obtained from different EMSs to provide an overview of the entire network. Based on the collected data, it performs several key management functions such as event creation based on faults, modification of running configurations, dynamic network discovery, and performance reports. The proper integration of EMS data in NMS solutions is ensured by the fact that the information exchanged is standard-based (Simple Network Management Protocol (SNMP), Simple Object Access Protocol (SOAP), Network Configuration Protocol (NETCONF), etc.). An adaptation layer should perform the conversion to a common format that is supported by the NMS. The NMS consists also of a northbound interface component that provides interfaces to higher layer support systems such as Operational Support System (OSS) or Business Support System (BSS).

OSS and BSS include systems used to support the daily operations of network operator. OSS

systems provide the service management of the network, while BSS solutions are more customer-centric and include functionalities such as billing, Customer Relationship Management (CRM), sales, etc. For both OSS and BSS, a model is used to represent operational and business requirements, namely the extended Telecommunications Operations Map (eTOM). It is a very extensive model, which defines three main functionalities for network operations: fulfilment, assurance and billing. Fulfilment is defined as the set of operations used to fulfil the customer's order or needs. Within OSS, this includes service provisioning and inventory management, while within BSS, this includes order management and sales. Assurance is the set of operations required to assure that services are offered according to the expected performance. The billing functionality involves ensuring the reception of payment for the offered services [80].

It should be noted that there is another method allowing human administrator to manage OLTs without passing by the EMS which consists on a direct or remote connection to the OLT system control unit through Command Line Interface (CLI) or a management protocol such as SNMP, Telnet, etc. This enables to use thousands of commands to control, monitor, and configure the PON system.

For reasons of innovation and control of technological developments, operators should keep their own PON infrastructure and management system effective for a long time. Indeed, a network failure or a low throughput means unsustainable consequences for the network operator including a loss of revenue and dissatisfied customers. Thus, a maintenance system becomes an increasingly important need and should be as reliable as possible to avoid any negative circumstances.

2.3 Traditional PON Maintenance

By deploying a passive optical network and building infrastructures capable of delivering the expected levels of service, operators have two challenges to face: on the one hand, providing simple and functional management interfaces, and on the other hand, ensuring the reliability, robustness, and performance of the network to satisfy the end user.

Indeed, in an environment where the opinion and experience of the user have a direct impact on the reputation of the operator (e.g. social media), a requested service by the customer should be guaranteed and irreproachable. Thus, the most important thing after the deployment phase is how to maintain the PON and make it work in the most efficient way.

Maintenance can be defined as a process that aims to maintain, improve or restore a specified state of the network in order to ensure the proper network operation. It's performed at every layer from Network Element Layer (NEL) to Service Management Layer (SML). This includes all actions related to monitoring, optimization, diagnosis, troubleshooting, testing, security, alarm detection, fault repair, etc.

In relation with PONs, OLTs and ONUs are expected to monitor and collect data required for making tests and to report any malfunction or any violations of optical parameters [47]. PON maintenance key actions include fault diagnosis and optimization of network resources. Fault diagnosis is a central aspect in managing a network since a fault that is qualified as a root cause can generate several anomalies in the PON. Thus, an efficient diagnostic system can find as quickly as possible the root causes of the observed anomalies that interrupt or degrade the QoS provided to subscribers [76]. Optimizing the resources allocation mechanism is also a key process in improving PON functioning. This action has direct impacts on the QoE since it allows adjusting the different bitrates offered to customers whenever it is possible.

Generally, network operators are always looking for practical and cost effective maintenance mechanisms to the continued development of viable PON solutions [88]. The current PON maintenance system requires a continuous intervention to provide the operational environment required for services to run smoothly, recover from faults, and maximize the overall performance of the network. It is affected by many factors that are hard to control and especially the complexity due to the several types of equipment and the variety of manufacturers [65]. The configuration and adaptation of network components require a deep knowledge and understanding of the operation and interactions between several processes and physical components. Many steps are needed when an IT administrator needs to monitor or modify the behavior of the different devices in a traditional network. First, he will have to manually configure multiple devices on a device by device basis. The next step is using device level management tools to update numerous configuration settings. This configuration approach makes it much more complex for an administrator to deploy a consistent set of policies related to management problems.

Manual network maintenance, whereby a human operator takes care of the continuous observation, operation, testing and optimization, is therefore very costly. The human factor is the dominant one in the network maintenance process given that human administrators were still responsible for manually interpreting and analyzing the different collected data and then performing various service and component configuration through a set of interfaces [92]. For instance, an optimization approach of PON resources that relies on customers' traffic behavior needs a human expert to analyze the collected information and take the right decision according to the observations.

This lack of automation leads to slow and inefficient reaction. The difficulty to manage this complex environment seriously affects the process of introducing and managing new products and services for the network operators customers [106]. Configuration management in such cases is also prone to configuration inconsistencies that may result in failures or performance degradation. In general, traditional configuration practices cannot effectively provide configuration management services to large-scale networks spreading across several management domains [12]. Current management approaches involve distribution of configuration data across hetero-

geneous databases present in several management layers (NEL, Element Management Layer (EML), Network Management Layer (NML), SML, Business Management Layer (BML)). They perform the configuration of network elements on an individual basis, which is a major drawback in the configuration and management of large scale networks and emerging network services.

Another limitation is that even simple operations such as the introduction of a new network variable or an entry in different tables with its relation to other variables, require a manual introducing of the planned changes, which makes the performance of the managed networks constrained by the expertise of human operators [92].

Other fundamental problems with the current PON maintenance process is that operators own a variety of equipment of different vendors, each vendor uses different programming languages and models with different semantics, making this a very difficult problem that impedes the interoperability of end-to-end solutions. For this, to successfully complete a process, an administrator will therefore need extensive knowledge of all present device types. This lack of adaptability to the network state makes the current network management unable to cope with the requirement of emerging services and time critical applications, characterized by highly demanding traffic, varying traffic patterns and quality of service requirements.

In the near future, the situation is expected to become worse since the new standards (e.g, NG-PON2) will increase the complexity in the network by adding more wavelengths and more equipment which simply cannot be handled by traditional networking, management and maintenance schemes. The arrival of these emerging technologies, despite their advantages in rising throughputs, satisfying the needs of customers in the near future and achieving a high level of security, performance and mobility, will not solve the complex problem of managing the network infrastructures and by cons, they may have a negative influence on the network management efficiency.

Actually, whether current or emerging technologies, both are always sensitive and brittle regarding network issues caused by hardware failures and human errors. Significant economic losses may be caused either by the failures themselves or by the time required to fix them without forgetting that the pressure to quickly restore operations may result in human configuration errors which may lead to a future network failure. In this case, instead of optimizing their networks, operators are decreasing the efficiency and the QoS in an indirect way. For this, human beings administration is viewed as a total dependency and a full time need for intervention in deployment, maintenance and optimization actions.

If we consider that manual control of network maintenance processes is costly, time consuming and generates a risk in terms of possible human errors and conflicts between different taken decisions, developing approaches for the maintenance of network systems with a minimum human intervention is becoming a key challenge to the continued development of viable management solutions. In this case, operators are carried to prepare a well-defined strategy in order to improve

the effectiveness of their network maintenance mechanisms.

Thus, there is a critical need for automating maintenance functionalities which should adapt to environmental changes both reactively and proactively and should be able to react rapidly to network breakdowns [65][94]. The automation of configuration management aims to address the aforementioned issues by providing effective and consistent configuration of large-scale networks and reducing as much as possible the degree of human involvement in the maintenance of PONs. This represents a priority for any operator seeking to improve the performance of its network and reduce its OPEX [33]. The automation must be associated with an adaptation to the context and the changes in the network environment in order to have a self-managed and context-aware network at the same time.

2.4 Towards Autonomic Proactive Maintenance

Based on the limits of the traditional PON maintenance discussed previously, it is necessary to rely on advanced approaches and efficient tools to automate the maintenance tasks, decrease the complexity in operators' networks and enable the PON network to be self-aware.

The introduction of the agent technology may help in this area. Software agents intended for the network management system bring along new operational capabilities allowing achieving asynchronous control and management decentralization. Software agents can be defined as persistent software entities dedicated to a specific purpose and capable of controlling their own decision making with the ability to perform the majority of their problem solving tasks without direct human intervention [57]. The use of agents provides the way to migrate the management code from the manager to the managed network elements. This in turn, provides a customized distribution of configuration and management programs [84].

Along with software agents, policy management systems have been introduced to formalize the description of different operations of various network elements in response to changes in the environment in order to simplify the administration of large scale networks [113]. Policies prescribe a set of rules defined by network administrators that guide the behavior of network components. The great benefit of policy-based systems is their ability to easily change the system behavior by modifying the applied policies at runtime without stopping the system operation. A network operator can influence the behavior of the system through an adaptation of the policies or by assigning a different role into a group of users [9].

Another concept related to the automation of maintenance tasks that can be quoted is the automatic networking. Automatic networking relies on a repetitious and self-regulating process that carries out a specified program or sequence of commands in order to perform some specific actions as a consequence of some occurring events or known problems. The automated process will always exhibit the same behavior for the same input. It tells the managed element when

to turn on and off, and exactly what it is expected to do [2]. It is to highlight that automatic networks are often created for one operating environment and will not function in another one. They will do nothing outside of their defined boundaries and do not have any knowledge outside their predefined area of functioning [93].

Concepts such as context-awareness may remedy this situation by making networks able to adapt their operations to the current context without explicit user intervention, i.e. without conscious or active involvement of the end-user [1]. Context-aware networks have the advantage to dynamically change or adapt their behavior based on the context and to have a good knowledge and judgement about environment changes over time. They allow systems to specialize their external behaviors depending on the perceived needs of their users which are captured in turn from context sources, such as sensors in the user environment. This can be particularly useful in the case of mobile networks given that end-users move all the time with their personal handheld devices and for this, network must react specifically to their current location, time and other environment attributes and adapt their behavior according to the changing circumstances [8][96].

The above-mentioned technologies are divided between those whose goal consists in automating the deployment of different management strategies for performing the necessary management operations, e.g. software agents, policy-based systems, automatic networking, and those which aim to adapt the network to environmental changes and context in a reactive manner, e.g. context awareness. The approaches designed for automating some tasks in the network may be optimal for particular settings, but operators are under the obligation to precisely describe and program their desired behavior (i.e. continuously modify their behavior in response to changes in the environment). These automated-management approaches are also limited by the inability to evolve and by the lack of situation and context adaptability [92]. For policy management systems, even if their injection into the network can be considered as an adaptation to the context and the network environment changes, they do not include self-management functions and they are limited to the control part handled by the operator. Regarding context awareness, despite the fact that it solves the problem of retrieving the context and making the network capable of reacting to changes in its environment, it does not offer advanced self-management properties.

Given the drawbacks of the aforementioned approaches, solutions that allow the network to be self-managed on the one hand and able to learn and adapt to changes in its environment and context on the other hand are strongly recommended. In that case, reference can be made to the “autonomic network management” paradigm as a promising approach that makes networks able to monitor their behavior, learn about their status, and then execute decisions that are in line with operators’ business goals and policies.

The concept of autonomic networks consists in a network empowerment through self-x capabilities, which help operators to handle the increasing amount of devices, data, management

and control operations and at the same time, having a certain level of supervision by keeping the overall control of the network [106]. Autonomic networking [60] has been proposed as a promising approach to reduce the cost of complexity of managing network infrastructure and to increase the human ability to manage the systems properly. Inspired by the autonomic nervous system in the human body, it enables devices to be automatically handled while having the ability to adapt to their environment and attempts to lead the human administration out of the control loop by leaving the management tasks to be performed by the network itself, respecting some rules and policies provided by the administrator [77][96]. The interaction of human administrators with the network is limited to the control and the provisioning of high level operational goals [9]. The motivation behind autonomic networking is to identify those functions that can be done with the minimum of human intervention and to reduce the dependence on skilled resources. This not only lowers operational costs, but also enables these valuable resources to be used for more complex tasks. It also enables the network to respond faster and more accurately to the changing needs of users and businesses. The ability to manage network operation and to take the appropriate set of actions based on the network environment is provided by a control loop (Fig.2.4) called also feedback loop which represents the heart of any autonomic system [40].

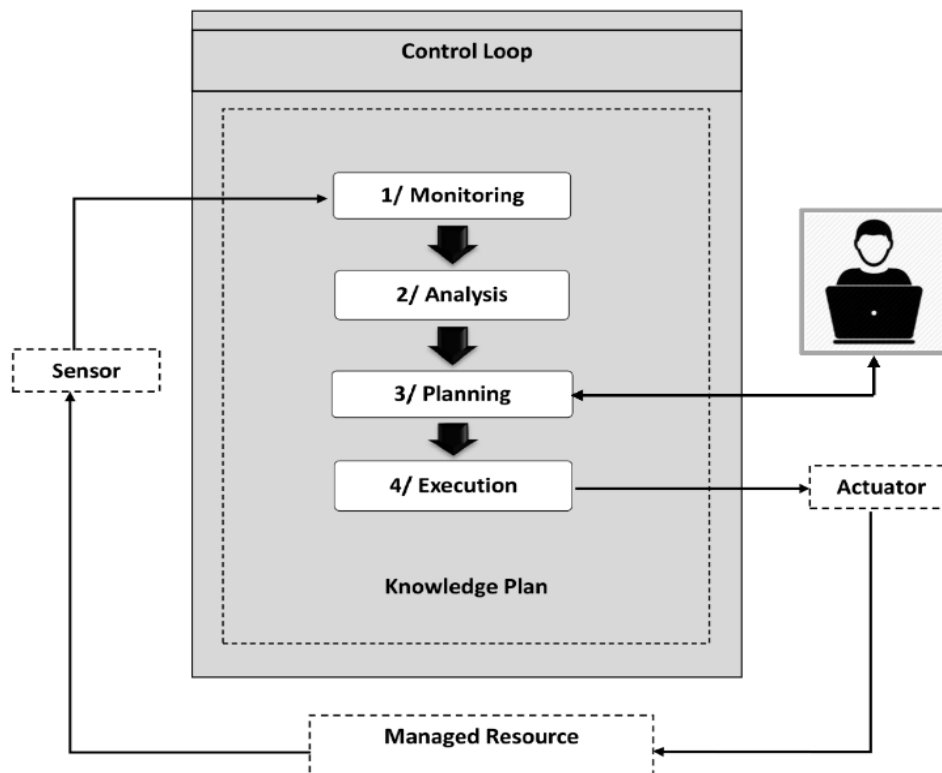


Figure 2.4: Autonomic control loop

The mechanism of this loop is as follows [9]:

1. Monitoring: retrieving and aggregating data through sensors
2. Analysis: observing and analyzing data to determine if any correction to the monitored managed resource is needed
3. Planning: planning the corresponding corrections if actions should be taken
4. Execution: executing corrections using actuators which translate commands to the managed resource

The knowledge plan is used by network elements to define the rules and policies provided by operators, coordinate the different taken decisions, and ensure the cooperation of all autonomic elements. The governance interface is used to define the control loop objectives and visualize its behavior (evaluation, problem identification, etc.).

By linking the autonomic concept to different maintenance actions, autonomic maintenance can be ensured so through a number of self-management properties that allow to improve PON network performances. The best known are the following [40][60]:

- Self-Configuring: capacity of the network to configure and reconfigure itself according to operators' needs and high-level policies.
- Self-Healing: self-diagnosis and self-reaction to system dysfunctions, knowing the expected system behavior, observing system failures, and applying suitable corrections to continue the functioning smoothly
- Self-Optimizing: capacity of networks to continually seeking opportunities to improve their own performance and efficiency and ensuring a better use of resources
- Self-Protecting: detection of attack situations, protection without disturbing users, and maintaining the overall system security and integrity

Thus, autonomic maintenance focuses on the application of these self-management properties to make PON maintenance mechanisms and applications more intelligent by enabling them to make decisions without having to consult a human administrator or user. From an operator perspective, the migration towards an autonomic maintenance of the whole network is not a simple task. Each application of an autonomic property requires a deep knowledge and should be well studied to avoid any conflict possibility.

In our thesis, the focus will be on the self-optimization property. Self-optimization can be a solution for network operators to improve the QoE and solve the problems that can

trouble their subscribers by improving the resource usage and reducing the costs in the network. It is recommended to let the PON operate in its best level of efficiency and utilize the advantages of optical fibers to better leverage existing network resources and to achieve a balance between coverage, capacity, and QoS. **We give a particular attention to the self-optimization of PON upstream resources.** This is justified by the limits we identified in the reviewed work regarding the upstream resources allocation mechanism, especially the lack of the operator vision and the difficulty of changing the whole mechanism for each update due to the dependency on a specific vendor.

However, is it desirable to reinforce the traditional maintenance system by a proactive self-optimization mechanism or a reactive one? Indeed, as in many industrial and transport sectors, the question is no longer to be able to react effectively in case of a network optimization issue, but to make sure that this issue does not simply happen. Therefore, the primary mission of the maintenance is no longer to optimize even very quickly, but to ensure the proper operation of network components by anticipating all kinds of problems in order to protect the network from any risks that may occur. Thus, an action of optimizing network resources must be considered in a perspective of logic of anticipation and regular supervision of the regular bandwidth usage. This means assessing the probability of using resources by customers, then estimating whether an optimization is needed in order to determine the optimal action plan based on predefined criteria and objectives.

“Autonomic maintenance should be more proactive, and not merely reactive”

With the trend of using machine learning and data mining techniques, network maintenance has reached a new milestone. The ability to handle and analyze large amounts of data of all kinds allows going much further in understanding and anticipating network failures and inconsistencies. Thus, any new contribution in the field of optimizing PON upstream resources should take into account the opinion and the experience of the end user to take another major step towards the goal of self-optimizing PON network resources.

2.5 Conclusion

This chapter provides a holistic view of research in FTTH access network with a more focus on the PON architecture. Traditional PON maintenance was discussed in order to motivate the migration towards an autonomic proactive maintenance mechanism.

The work carried out as part of this thesis provides contributions to the problematic of self-optimizing upstream resources in PON. Proposing a new approach in this area taking into account what has been mentioned before requires an in-depth analysis of customer behavior. The

proposed model must be intelligent enough to allow the network to understand the traffic pattern of each PON user and take the necessary actions (e.g. load balancing, throughputs adjustment, etc.). It should be also prepared to deal with some specific cases such as the lack of some data points and the inaccuracy or the ambiguity of some data generated by the equipment. It should be also kept in mind that at the level of the decision-making part, the proposed approach must act only on network parameters which are manageable and configurable by the network operator in order to be able to integrate the autonomic concept in the PON management system.

We propose in this thesis a new mechanism for self-optimizing PON upstream resources based on users traffic patterns. We try to exploit the extra bandwidth theoretically untapped by the current resources allocation mechanism in order to share it among a specific class of users. Machine learning techniques are used to estimate the traffic pattern of each customer. The approach is characterized by its ability to be implemented directly in the network without any recourse to a material modification taking into account the operator vision.

CHAPTER 3

MANAGEMENT ARCHITECTURE FOR OPTIMIZING PON UPSTREAM RESOURCES ALLOCATION

Contents

3.1	Introduction	53
3.2	Related Work	54
3.2.1	Upstream Bandwidth Allocation	54
3.2.2	Dynamic Bandwidth Allocation Optimization Approaches	58
3.2.3	Limits & Motivations	60
3.3	Management Architecture for Optimizing PON Upstream Resources	61
3.3.1	Overall Architecture	61
3.3.2	Detailed Architecture	63
3.4	Users Classification	68
3.4.1	Clustering Tendency	68
3.4.2	Clustering Algorithms	69
3.4.3	Experimental Evaluation	70
3.5	Conclusion	78

3.1 Introduction

Due to the widespread integration of IoT, the expected arrival of 5G networks at the horizon 2020, and the usage of high-speed services such as VoD, IPTV, video conferencing, and UHD video services, the control of the bandwidth allocation mechanism becomes a key task in PONs to ensure that available resources are continuously used in the most efficient manner.

A particular attention may be given to the control of the upstream direction. Indeed, the majority of work contributing to the improvement of the upstream bandwidth resources usage in PONs doesn't take into account the operator vision. The update of the upstream resources allocation mechanism which is classified as a closed control protocol between the OLT and ONUs in the PON [24][115] should be performed by the equipment supplier each time there is a new theoretical proposal. This makes it more complicated and less practical for operators which tend to be independent of specific vendors, especially with the evolution towards more and more softwarized networks in the near future.

To resolve this issue and avoid being bound to the constraints of equipment supplier, this chapter introduces a novel mechanism for optimizing the allocation of upstream PON resources based on customers' estimated traffic patterns. According to the proposed mechanism, the focus is only on the management of the input parameters of the upstream bandwidth allocation algorithm, known as the SLA parameters, which are allowed to be modified by the network operator. The idea is so to try to dynamically adjust these SLA parameters in order to exploit and allocate in a more efficient manner the bandwidth available in the network, without modifying the implementation of the upstream bandwidth allocation mechanism itself as it was proposed in the majority of the reviewed work.

This chapter proceeds as follows. Section 3.2 presents work related to the upstream bandwidth allocation mechanism in PON, the associated optimization approaches, and motivates the usefulness of the proposed management mechanism. Section 3.3 introduces the proposed model for optimizing PON upstream resources allocation. An overview is given followed by a detailed explanation of each module. A particular attention is given to the users' classification module in section 3.4 where an evaluation study is conducted to select which classification approach should be selected. Finally, section 3.5 concludes the chapter.

3.2 Related Work

3.2.1 Upstream Bandwidth Allocation

In PONs, and specifically in GPONs, upstream bandwidth allocation is a control mechanism ensured by the OLT which defines for each ONU, the length and the time of transmission of each upstream burst. This mechanism can be also performed by ONUs, so-called distributed bandwidth allocation [53]. However, it is rarely known and not very common compared to the centralized bandwidth allocation. TDMA is used in PON upstream transmission to exploit the shared ODN. Thus, only a single ONU has permission to transmit data at a given time [99]. The size of the allocated time slot depends on the bandwidth allocation scheme used. Basically, PON upstream transmission can be achieved following two schemes: static or dynamic. Fig.3.1 shows the concepts of both, Static Bandwidth Allocation (SBA) and Dynamic Bandwidth Allocation

(DBA).

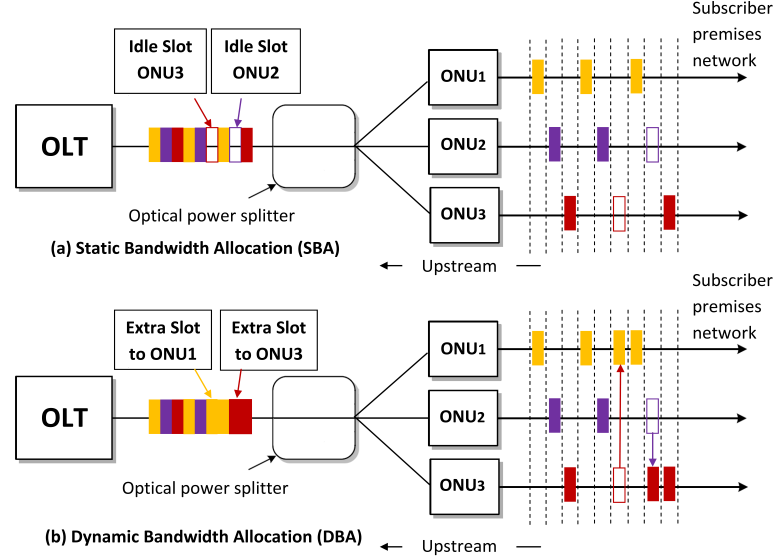


Figure 3.1: Static bandwidth allocation and dynamic bandwidth allocation mechanisms

3.2.1.1 Static Bandwidth Allocation (SBA)

SBA allows the different ONUs to have the same transmission slots in each cycle (whether used or not) while not taking into account the possible traffic fluctuations during the day [45]. Whereas this approach has the advantage of being simple to implement at the OLT level, the risk of congestion or bottleneck in the network is extremely high. In fact, customers' traffic patterns are not the same during an allocation cycle, leading to idle time slots that occupy the network and so, many ONUs buffers reach their saturation and can no longer store packets which will be automatically dropped. This can result in a degradation of the overall network performance and a customer dissatisfaction towards his service provider [44]. SBA can be classified as a QoS-unaware allocation mechanism and can be optimal only in the case where all network services require a constant bitrate which is simply not probable. As service providers need to adapt their network performances based on the dynamicity of users' traffic patterns, a DBA mechanism is so needed in order to make the best use of the unallocated bandwidth.

3.2.1.2 Dynamic Bandwidth Allocation (DBA)

DBA allows the bandwidth which is not consumed by some ONUs to be assigned to the other ones, resulting in a more flexible, efficient, and fair bandwidth allocation mechanism while respecting the SLA constraints of each subscriber. ONUs can send their pending data only within

the limits of the allocated grants which are specified in the Upstream Bandwidth Map (US-BWmap) field [53]. These grants are transmitted by the OLT to the different ONUs in the downstream traffic frames [31]. They are updated regularly in each DBA cycle which represents the duration that covers all transmissions from all ONUs and may vary depending on the state of the ONUs packet queues [44].

GPON (the PON standard studied in this thesis) [35] supports a mixture of services (e.g. Voice over Internet Protocol (VoIP), internet browsing, business services...) which do not have the same bandwidth requirements in the upstream (e.g. VoIP requires a constant bitrate). Thus, the DBA needs to take this into account in order to satisfy all customer needs. In this context, ONUs may support multiple Class of Services (CoSs). Each CoS has its own buffer with a specific QoS profile [110]. Therefore, the bandwidth allocation is no longer assured per ONU, but rather per ONU per service. To guarantee the different QoS levels, GPON uses the Transmission Container (T-CONT) mechanism which allows different ONUs GPON Encapsulation Method (GEM) flows to be gathered in the same CoS [53][81][82]. Table 3.1 shows the five containers that can be assigned to a single ONU depending on which services the latter had requested [11][32]. The activity status of each container can be obtained either through a status or a non-status reporting mechanism.

Table 3.1: T-CONT types

T-CONT Type	Associated Services
Type 1	Fixed bandwidth services with a high priority Constant bit rate applications sensitive to jitter and delay
Type 2	Assured bandwidth services with a well-defined bitrate Variable bit rate applications
Type 3	Assured bandwidth services similar to T-CONT2 Possibility to take advantage of non-assured bandwidth
Type 4	Best effort bandwidth services No specific requirement on QoS (no delay sensitivity) Served only when there is an available bandwidth
Type 5	Mixed type or a combination of two or more T-CONTs

The Status Reporting (SR)-DBA mode relies on the data queue occupancy reported by each ONU to perform the calculation process of the US-BWmap. Each ONU sends its current data queue status to the OLT in a Dynamic Bandwidth Report upstream (DBRu) which is a field of the upstream frame and in return, the OLT reckons the number of grants for each T-CONT taking into account the SLA constraints, builds and sends the US-BWmap to the different ONUs [11][31]. Fig.3.2 provides a simplified schema of the SR-DBA mode.

In the Non-Status Reporting (NSR)-DBA mode, ONU data queue occupancy information is not provided to the OLT. The latter should hinge on the actual transmission in the previous

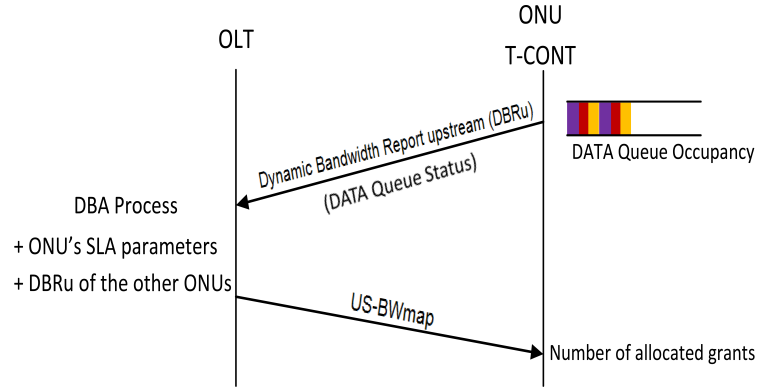


Figure 3.2: Status reporting DBA mode

cycle and the information given by the GEM headers to estimate the ONUs queue status. In fact, when the ONU does not have data to send, idle frames are generated and sent during the allocation interval. These frames are observed by the OLT which understands that there is no transmission for the corresponding ONU and automatically decreases its bandwidth allocation in the following cycle [11][31][44].

The SR-DBA based algorithms are widely implemented in OLTs compared to NSR-DBA ones. Thanks to their ability to reduce the latency (traffic can be granted for transmission once reported) and to avoid the overestimation or the underestimation of data queues, we will only consider the SR-DBA mechanism in the following. SR-DBA algorithms are based on the combination of the queue occupancy information transmitted from each T-CONT depending on its activity status and the provisioned SLA of each T-CONT [11].

SLA parameters are defined in a service contract which represents a formal commitment between the provider and the subscriber. They define the specific objectives expected and the QoS profile that a client asks to obtain from the service provider. The main parameters specified in the customer SLA are respectively the Committed Information Rate (CIR), the Excess Information Rate (EIR), and the PIR [31]. The CIR defines the guaranteed bandwidth provided by the network when delivering the data frames. According to the GPON G.984.3 standard [53], the CIR represents the sum of two sub-parameters known as Fixed Information Rate (FIR) and Assured Information Rate (AIR). The FIR represents the fixed bandwidth which is allocated to the T-CONT type1 supported by services that require strict throughput and delay. Ordinarily, the DBA should not touch this parameter and therefore, the corresponding bandwidth cannot be increased, decreased or allocated to another T-CONT [32]. The AIR represents the assured bandwidth that is allocated to the T-CONT type 2 and type 3 [32][67]. The difference between

the FIR and the AIR lies in the fact that the FIR is reserved cyclically regardless of demand whereas the AIR may not be given without demand. The service provider must guarantee that the network will always support the CIR in order that at any given time, the available bandwidth for a specific subscriber should not fall below the CIR. Regarding the EIR, it specifies the additional bandwidth above that of the CIR that a subscriber may use if available. It is generally allocated to the T-CONT type 3 and type 4 supported by services experiencing the lowest priority (best-effort) [32]. Finally, the PIR represents the maximum bandwidth that can be provided. Table 3.2 gives more details about the SLA parameters discussed above [53].

Table 3.2: SLA parameters

SLA Parameters		Significance
Peak Information Rate (PIR)	Committed Information Rate (CIR)	Fixed Information Rate (FIR) *Fixed bandwidth allocated to T-CONT type 1 *Untouchable parameter *Reserved cyclically regardless of demand
		Assured Information Rate (AIR) *Assured bandwidth allocated to T-CONTs type 2 and 3 *Not given without demand
PIR = CIR + EIR	CIR = FIR + AIR	
	Excess Information Rate (EIR)	*Allocated to T-CONTs type 3 and 4

3.2.2 Dynamic Bandwidth Allocation Optimization Approaches

As the DBA directly affects the performance of the network in the upstream, many works [7, 20, 73, 79, 98] have been proposed in the literature with the objective to enhance its operation. These works can be classified according to:

- The decision-making time. For online approaches, the bandwidth allocation is decided upon the reception and the processing of the queue status report from one ONU. Regarding offline algorithms, the OLT shall wait until all ONUs send their reports. Then, it will take the appropriate bandwidth allocation scheme [26].
- The entity designated to perform the bandwidth allocation per CoS. In distributed approaches, the OLT grants the bandwidth per ONU and the latter, using an intra-ONU bandwidth scheduling algorithm, distributes the bandwidth among each CoS. In centralized approaches (the most common), the OLT is the sole responsible for allocating the upstream bandwidth per ONU CoS.
- The ability to take or not into consideration the constraints agreed in the SLA contract, i.e. SLA awareness.

In our thesis work, and with respect to the operator vision, only the most relevant enhanced DBA proposals which respect the minimum and the maximum of bandwidth that must be provided to subscribers according to their SLA parameters are taken into account.

In [79], an SLA aware dynamic bandwidth allocation algorithm (SLA-DBA) was presented. In this algorithm, the OLT allocates in a first stage a number of bytes to an ONU queue depending on its bandwidth report. Then taking into account the SLA constraints, an adjustment will be performed in such a way that the number of assigned bytes never exceeds the maximum rate provided by the SLA (i.e. a basic DBA algorithm that respects SLA constraints).

In [20], the proposal introduces an approach called cyclic-polling-based dynamic bandwidth allocation with service level agreements (CPBA-SLA) based on a two-layer allocation scheme: SLA-layer and ONU-layer. In the SLA-layer allocation, ONUs are distributed on two or more groups based on their SLA parameters. Then, the OLT assigns the bandwidth for each CoS in each group in a way that high-priority traffic can be served earlier. In the ONU-layer allocation, bandwidth is assigned to ONUs which belong to the same group where a group can be defined as a set of ONUs that require the same average packet delay of a specific CoS (e.g. high-priority, delay-sensitive traffic, etc.) for a specific cycle time.

Authors in [73] propose a fair bandwidth allocation algorithm based on network utility maximization. In addition to the requested bandwidth by each ONU in its report message and the guaranteed bandwidth coming from the SLA contract, the algorithm takes into account an additional factor that corresponds to the importance of an ONU regarding the conditions specified in its QoS requirements. Then, it distributes the extra bandwidth available in the network in a fair manner according to the definition of fairness in the network utility maximization model.

In [91], a Hierarchical Sharing Dynamic Bandwidth Allocation (SHDBA) algorithm is illustrated. The algorithm can be classified as a distributed DBA algorithm where two mechanisms for the bandwidth allocation are integrated inside the OLT and the ONU. The algorithm that resides in the OLT allocates the bandwidth per ONU depending on its requests. Then, taking into account the SLA constraints, the ONU shares the allocated bandwidth among data and video using an SLA ratio.

As the DBA grants the bandwidth by traffic class in a separate cycle, idle periods can be distinguished between each cycle resulting in a throughput limitation. In this case, authors in [27] propose the Delay-Aware Window Sizing (DAWS) approach to provide a constant delay variation to high priority traffic with high throughput, and the Delay-Aware Grant Sizing (DAGS)

approach for medium priority traffic. To take into account the SLA awareness criterion, authors propose in [26] an enhancement of the DAWS approach, namely the Modified DAWS (MDAWS) in which, users who have a strict delay-bound requirement are polled more frequently whereas before, users with different delay bound requirements are treated identically.

3.2.3 Limits & Motivations

Although the majority of reviewed work contributes to the optimization of the DBA mechanism, they do not take into account the fact that the DBA is a closed control protocol between the OLT and ONUs in the GPON network [24][115]. Thus, the update of the DBA should be performed by the equipment supplier each time there is a new theoretical proposal. This makes it more complicated and less practical for operators which tend to be independent of specific vendors, especially with the evolution towards more and more softwarized networks in the near future.

To resolve this issue and allow operators to optimize their networks without being bound to the constraints of specific vendors, it's recommended to focus only on the input parameters of the DBA algorithm which can be managed by the operator, i.e. SLA parameters. The idea is so to try to optimize the management of these parameters in order to exploit in a more efficient manner the bandwidth available in the network, without modifying the implementation of the DBA mechanism itself as it was proposed in the majority of the reviewed work. Thus, we propose in our thesis a novel management procedure for optimizing the allocation of GPON resources based on the dynamic adjustment of the SLA parameters according to estimated customer' traffic patterns without interference of the DBA algorithm. We propose to identify these patterns through the use of clustering techniques based on real-time and historical data. The idea came from the fact that during the day, customers are not connected to the network in the same way. There are times when some of them need more than their PIRs, but they cannot afford an extra bandwidth by the DBA due to the SLA parameters constraints. Thus, adjusting dynamically these parameters during the day can improve the QoS and maximize the overall satisfaction of users. To summarize, we formulate the problem as follows:

- **Given:** The initial SLA parameters and the activity status of each ONU based on real-time and historical transmission data
- **Objective:** Optimize the upstream resource allocation by adjusting the SLA parameters depending on the daily life customer bandwidth usage
- **Output:** New SLA parameters for each ONU at a specific period of time

3.3 Management Architecture for Optimizing PON Upstream Resources

To address the aforementioned issues and achieve the main objective introduced in the previous section, we present a new management architecture for optimizing the upstream bandwidth allocation in PON. This is the first time, an optimization approach based on the dynamic adjustment of SLA parameters and the customer’ traffic behavior during the day is applied to PON networks. With regard to what has been presented before by researchers in this domain, the proposed architecture comes with two major contributions. First, it can be directly and easily integrated in the PON management system without the need to update the DBA control algorithm at the OLT level. Second, as it focuses only on manageable parameters, the proposed approach gives us the opportunity to apply in a more advanced stage the autonomic & cognitive paradigm in order to enrich the network with self-decision capabilities that leave the task of the dynamic reconfiguration of the SLA parameters to the network itself with a minimum of direct human intervention.

3.3.1 Overall Architecture

The design of the proposed architecture is illustrated in Fig.3.3.

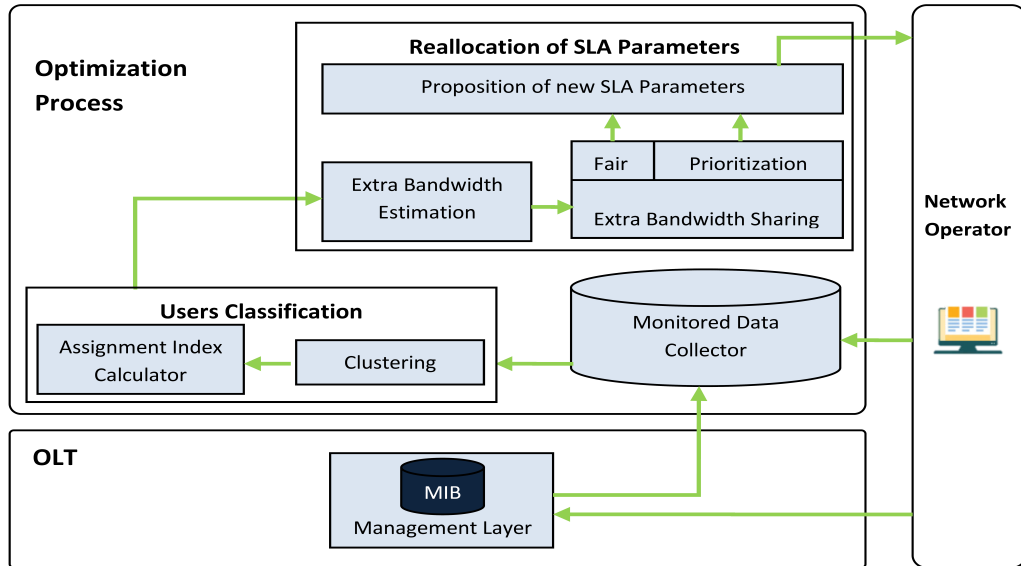


Figure 3.3: Management architecture for optimizing PON upstream resources

In daily life, traffic patterns of different users may change several times per day. In this case, it is highly possible to have some customers who consume more bandwidth than others for a specific day period. The current DBA allows allocation of PON resources, depending on the

instantaneous demand of each ONU. However, each user is limited by predefined PIR, which represents one of the SLA parameters which defines the maximum bitrate that a user might benefit from. In this case, when the extra available bandwidth resulting from the presence of ONUs with low resources usage is greater than the maximum bandwidth authorized to be allocated to high bandwidth usage ONUs, a part of the extra bandwidth will be lost and not exploited. For this reason, the idea was to try to exploit this extra bandwidth theoretically untapped by the DBA in order to share it among high bandwidth usage ONUs without being limited by the constraints of the SLA parameters. As the goal is to propose an implementable approach by the operator in which the DBA algorithm should not be touched, the challenge is then to be able to model the functioning of the DBA while using the historical data provided by the operator and acting only on the DBA externally manageable parameters, i.e. SLA parameters. In this context, four main components were proposed namely, Monitored Data Collector, Clustering Module, Assignment Index Calculator, and a mechanism for Reallocating SLA Parameters.

The Monitored Data Collector gathers the traffic data for each ONU by requesting the Management Information Base (MIB) at regular intervals. This module connects also to the network operator in order to store the historical transmission data. As the DBA refers to the packet queue status of each ONU to know its needs, two complementary modules responsible for the classification of different users depending on their historical transmission data are proposed. The clustering module classifies users into 3 classes: heavy, light, and flexible.

- Heavy users: users characterized by a high bandwidth usage. They need an extra bandwidth if possible and an adjustment of their PIRs is desirable.
- Light users: users characterized by a low bandwidth usage and generally satisfied by their PIRs. Their unused bandwidth can be reallocated.
- Flexible users: users characterized by a moderate bandwidth usage while they can be sometimes heavy and sometimes light.

The clustering module is supposed to be applied per time interval (e.g. 5 minutes). So, to be able to classify all users based on the entire time series, a second module called Assignment Index Calculator is proposed. This module aims to provide the probability for each user to be either heavy, light or flexible. For each day and for each time interval, it analyzes the clustering results. If the user belongs to the heavy class in a given time interval, his probability to be a heavy will increase and likewise for the light class. Then, a calculation process of the average of all probabilities associated to a standard deviation calculation (for validation) is assured to finally assign each user to a specific class. The final users' classification will be used then by the Reallocation of SLA Parameters module to define for each heavy user new temporary upstream bitrate allocation in a specific day period. The modification of the SLA parameters at the MIB

level is performed by the network operator which has the control of these parameters and may stop the optimization process whenever a problem is detected. For a better explanation, the operation of the different modules is detailed in the following.

3.3.2 Detailed Architecture

3.3.2.1 Monitored Data Collector

The Monitored Data Collector represents the first entity in the proposed model design. As its name suggests, this module is responsible for gathering all data required by the other modules in a local repository. These data come in three forms: the initial SLA parameters of each subscriber, the historical data needed for the first classification of different users, and the real data that can be seen as a regular update of the historical data. The initial SLA parameters are retrieved only once for each ONU from the network operator, and then stored in a specific table. When a new user is added, the network operator is responsible for notifying the Monitored Data Collector to update its SLA table. The historical data is provided as a result of traffic collections made in advance by the network operator using probes and other collection tools. The metric used in this work is the mean upstream bitrates allocated to each subscriber. The monitoring interval is configurable by the network operator. In general, the used tools can provide the instantaneous allocated upstream bitrate (i.e. per second). So, the mean values can be calculated according to the desired time interval. Once the optimization mechanism is running, historical data can be updated by the Monitored Data Collector using the real traffic data. The latter is collected from the MIB using SNMP GET requests at regular intervals parameterized by the network operator in order to get information about daily users' upstream transmission data. To make the optimization approach more efficient, every time the classification process is complete, the historical data can be deleted respecting the limitations of the local repository in terms of storage and the network capacity in terms of memory and processing.

3.3.2.2 Users Classification

3.3.2.2.1 Clustering Module

As the traffic utilization of PON users may change several times per day since traffic patterns are linked to daily life, defining different time periods of the day when users have a similar traffic utilization can be very useful in the optimization of PON resources allocation. The easiest way to classify subscribers according to their online traffic behavior over a time period is to look at the differences between the traffic data of each user in order to create common groups based on the traffic similarity. If we specify for subscribers the minimum or maximum threshold that each one shall reach to be assigned to a specific class (i.e. static assignment), then this approach can be the most simple to be adopted. However, in PON networks, the usage segments can be very

varied and dynamic since all users do not have the same behavior in terms of upstream data bitrate during the day. Thus, there is no exact definition for a specific class of users.

To overcome these limitations, we will refer to the clustering analysis approach which identifies the attachment of a user to a specific group according to his online behavior during the day without any prior assumption. The main objective is to identify heavy users and try to improve their satisfaction in terms of QoS by giving them a surplus of bitrate on their PIRs. Consequently, for all day periods, the clustering algorithm will identify clusters of heavy users with a high upstream transmission bitrate and clusters which contain light users with a low upstream bitrate. Clustering analysis is a very common approach in the field of data analysis and related disciplines and can be very useful in the case of grouping customers into different classes regarding their traffic data. Given a specific data set, the objective of clustering is to find segments where objects are similar based on a measure of similarity (e.g. size, density, shape, etc.) [55]. The input of the proposed clustering module is basically the historical transmission data collected over some time horizon and provided by the Monitored Data Collector (updated regularly). The collection is achieved using a time series format with the average bitrates for each ONU over a parametrizable monitoring interval (e.g. 5min). The clustering process is planned to be executed for the whole OLT for each time interval in order to have as many as possible of data points. Table 3.3 presents our proposal of decomposing each day into four time slots. In each time slot, three classes of users are distinguished: heavy, light, and flexible.

Table 3.3: Clustering explanation

Day Period	Time Slot	User Traffic Classes	Cluster Index
Morning	6 a.m. to 12 p.m.	Heavy Users	C_{M_H}
		Light Users	C_{M_L}
		Flexible Users	C_{M_F}
Afternoon	12 p.m. to 6 p.m.	Heavy Users	C_{A_H}
		Light Users	C_{A_L}
		Flexible Users	C_{A_F}
Evening	6 p.m. to 11 p.m.	Heavy Users	C_{E_H}
		Light Users	C_{E_L}
		Flexible Users	C_{E_F}
Night	11 p.m. to 6 a.m.	Heavy Users	C_{N_H}
		Light Users	C_{N_L}
		Flexible Users	C_{N_F}

3.3.2.2.2 Assignment Index Calculator

As the results of the clustering module are provided according to the chosen monitored time interval, we introduce an Assignment Index (AI) to classify the user’s traffic by day period. This index ranges from 0 to 1 and represents the degree of attachment of one ONU to a certain

traffic class in a specific day period. For instance, if during a day period, a user has been classified as heavy 5 times over 10 time intervals, then its AI to the heavy users class will be 0.5. The mathematical formulation for this calculation process is detailed in algorithm 1 while the notations are provided in table 3.4.

Table 3.4: Table of notations

Notation	Meaning
U_i	User index, $i \in [1, n]$, with n:total number of users
t	Interval of time in a day period, e.g. with a 5 minutes step
$NumDays$	Number of supervised days
$Days$	The set of supervised days
DP	The set of day periods (i.e. morning, afternoon, evening, and night)
C_{t_H}	The cluster which contains heavy users for a specific interval t
C_{t_L}	The cluster which contains light users for a specific interval t
$AI_{C_{j_H}}^d(U_i)$	AI of U_i to the heavy users cluster in the day period j of the day d
$AI_{C_{j_L}}^d(U_i)$	AI of U_i to the light users cluster in the day period j of the day d
$PIR(U_i)$	Peak Information Rate of U_i
$PIR'_j(U_i)$	Optimized Peak Information Rate of U_i for a specific day period
$BR_t^{avg}(U_i)$	Mean upstream bitrate of U_i over a specific interval t
$ExBw_{j_d}^{avg}$	The average of the extra bandwidth by day period for a specific day
$ExBw_j^{avg}$	The average of the extra bandwidth by day period for all supervised days
η_j	The multiplicative factor of heavy users PIRs by day period for all supervised days
α_j	The percentage increase in the PIR of each heavy user by day period for all supervised days

Algorithm 1 Assignment index calculation

```

1: for  $d$  in  $Days$  do
2:   for  $j$  in  $DP$  do
3:     for each  $U_i$  do
4:        $nbItr \leftarrow 0$  #  $nbItr$ : number of time intervals in a day period  $j$ 
5:        $x \leftarrow 0$  #  $x$ : number of times the  $U_i$  was considered as heavy
6:        $y \leftarrow 0$  #  $y$ : number of times the  $U_i$  was considered as light
7:       for  $t$  in  $j$  do
8:         if  $U_i \in C_{t\_H}$  then
9:            $x \leftarrow x + 1$ 
10:        end if
11:       if  $U_i \in C_{t\_L}$  then
12:          $y \leftarrow y + 1$ 
13:       end if
14:        $nbItr \leftarrow nbItr + 1$ 
15:     end for
16:      $AI_{C_{j\_H}}^d(U_i) \leftarrow \frac{x}{nbItr}$ 
17:      $AI_{C_{j\_L}}^d(U_i) \leftarrow \frac{y}{nbItr}$ 
18:   end for
19: end for
20: end for

```

In a first phase, the AI is computed upon the results of the clustering analysis. The latter is performed on the historical data provided by the service provider over a specified time horizon. To get a more accurate index, it is preferable to have a longer monitoring period as well as time intervals with smaller steps in a day period. As the index is calculated per day and per day period, we calculate the average of all daily AIs per day period for each user (the average may be calculated upon a set of particular days, e.g. a set of Saturdays, as it can be calculated upon all days):

$$\begin{cases} AI_{C_{j-H}}^{avg}(U_i) = \frac{\sum_{d \in Days} AI_{C_{j-H}}^d(U_i)}{NumDays} \forall j \in DP, \forall i \in [1, n] \\ AI_{C_{j-L}}^{avg}(U_i) = \frac{\sum_{d \in Days} AI_{C_{j-L}}^d(U_i)}{NumDays} \forall j \in DP, \forall i \in [1, n] \end{cases}$$

Since the $AI_{C_{j-H}}^{avg}(U_i)$ and $AI_{C_{j-L}}^{avg}(U_i)$ may not be too meaningful in the case where the different indexes are very dispersed, an additional step consisting in calculating the standard deviation will be strongly recommended to decide whether or not the mean value can be taken into account:

$$\begin{cases} AI_{C_{j-H}}^{sd}(U_i) = \sqrt{\frac{1}{NumDays} \sum_{d \in Days} (AI_{C_{j-H}}^d(U_i) - AI_{C_{j-H}}^{avg}(U_i))^2} \\ AI_{C_{j-L}}^{sd}(U_i) = \sqrt{\frac{1}{NumDays} \sum_{d \in Days} (AI_{C_{j-L}}^d(U_i) - AI_{C_{j-L}}^{avg}(U_i))^2} \\ \forall j \in DP, \forall i \in [1, n] \end{cases}$$

We propose that once the $AI_{C_{j-H}}^{sd}(U_i)$ and $AI_{C_{j-L}}^{sd}(U_i)$ reach a certain threshold AI_{TH}^{sd} , the $AI_{C_{j-H}}^{avg}(U_i)$ and $AI_{C_{j-L}}^{avg}(U_i)$ will not be taken into account and hence the U_i will be assigned to flexible user clusters (i.e. C_{j-F}). The AI_{TH}^{sd} value should be as close as possible to 0 to assert that the calculated averages are significant. Hence, the assignment index of a user during a long period of time can be a good criterion to classify it as heavy or light for a specific day period. Finally, to decide to which cluster the user belongs based on its mean AI, we suppose that:

$$\begin{cases} U_i \in C_{j-H} & \text{if } AI_{C_{j-H}}^{avg}(U_i) \geq 0.5 \forall j \in DP, \forall i \in [1, n] \\ U_i \in C_{j-L} & \text{if } AI_{C_{j-L}}^{avg}(U_i) \geq 0.5 \forall j \in DP, \forall i \in [1, n] \\ U_i \in C_{j-F} & \text{Otherwise } \forall j \in DP, \forall i \in [1, n] \end{cases}$$

To improve the accuracy of the AI process, we continuously consider the results of the clustering module. This is achieved based on the new data collected by the Monitored Data Collector module which is regularly updated. As a first step, we consider a weekly update. The AI is used as a reference in the reallocation of the SLA parameters in the network and the determination of the additional bandwidth that can be provided to heavy users.

3.3.2.3 Reallocation of SLA Parameters

The main objective of the reallocation of SLA parameters in the network is to try to maximize as well as possible the satisfaction of all users in a specific day period. Ordinarily, the DBA assigns to each ONU the amount of data to be transferred regarding its packet queue status and the whole state of the network. The problem is when there is a majority of light or offline users in the PON, the excess available bandwidth to be assigned by the DBA to heavy users may exceed their PIRs. Thus, heavy users do not take advantage of a part of the additional bandwidth due to their SLA parameters constraints. The idea is therefore to extend the PIRs for this class of users in such a way they can benefit from a bandwidth supplement whenever possible. We propose a calculation and a sharing process of this extra bandwidth in algorithm 2. This algorithm is performed by PON port, i.e. the considered users (whether light or heavy) are all part of the same PON port. We consider that offline users are part of light users cluster. The notations are provided in table 3.4.

Algorithm 2 Extra bandwidth calculation and sharing

```

1: for  $d$  in  $Days$  do
2:   for  $j$  in  $DP$  do
3:      $ExBw = 0$  #  $ExBw$ : extra bandwidth calculated over all time intervals in  $j$ 
4:      $nbItr = 0$  #  $nbItr$ : number of time intervals in a day period  $j$ 
5:     for  $t$  in  $j$  do
6:       for  $U_i$  in  $C_{j\_L}$  do
7:          $ExBw = ExBw + (PIR(U_i) - BR_t^{avg}(U_i))$ 
8:       end for
9:        $nbItr = nbItr + 1$ 
10:    end for
11:     $ExBw_{j\_d}^{avg} = \frac{ExBw}{nbItr}$ 
12:  end for
13: end for
14: for  $j$  in  $DP$  do
15:    $ExBw_j^{avg} = \frac{\sum_{d \in Days} (ExBw_{j\_d}^{avg})}{NumDays}$ 
16:    $ExBw_j^{avg} = \sum_{U_i \in C_{j\_H}} PIR'_j(U_i) - PIR(U_i)$ 
17:   for  $U_i$  in  $C_{j\_H}$  do
18:      $PIR'_j(U_i) = \eta_j \times PIR(U_i)$ 
19:   end for
20:    $ExBw_j^{avg} = \sum_{U_i \in C_{j\_H}} \eta_j \times PIR(U_i) - PIR(U_i)$ 
21:    $ExBw_j^{avg} = (\eta_j - 1) \times \sum_{U_i \in C_{j\_H}} PIR(U_i)$ 
22:    $\eta_j = 1 + \frac{ExBw_j^{avg}}{\sum_{U_i \in C_{j\_H}} PIR(U_i)}$ 
23:    $\alpha_j = (\eta_j - 1) \times 100$ 
24: end for

```

In algorithm 2, the extra bandwidth $ExBw_j^{avg}$ is calculated by day period and represents the average of all extra bandwidths over a specific set of days (days can be a set of a particular day, e.g. a set of Saturdays, as it can represent all weekdays). It will be shared between heavy users following a fair sharing where the additional bandwidth is added to all ONUs in the same way (a prioritization of some heavy users can be also performed while the extra bandwidth will be shared according to network operator objectives). The resulting parameters η_j and α_j are used then to define the new PIRs of heavy users by day period. The new configuration of the SLA parameters will be then planned to be executed at a specific time. This reconfiguration is performed as follows: new DBA profiles are created firstly at the OLT level using CLI. Then, for each profile, the new values of PIRs are specified depending on the new SLA parameters configuration. Once DBA profiles are created, each heavy user will be assigned to his new optimized profile. To avoid the creation of many DBA profiles each time a new configuration is proposed, the Object Identifier (OID) of each profile is localized at the MIB level and using SNMP SET requests, the PIR values can be directly updated. The Monitored Data Collector is informed by the new SLA parameters and stores them in the same table as the initial parameters. In the event that a problem has occurred, a rapid return to the initial configuration can be performed.

In the remainder of this chapter, a particular attention is given to the users' classification modules (i.e. clustering & AI calculation) as they represent the key modules in the proposed architecture. The objective is to study and evaluate the different clustering mechanisms that exist in the literature in order to select which one can be the most suitable for our use case. Then, based on the selected algorithm, we test the AI calculation module using some traffic traces that we have already collected in order to generate users' distribution, and see if it can be used for SLA parameters reallocation or if an upgrade of the model is required.

3.4 Users Classification

3.4.1 Clustering Tendency

Before applying any clustering technique to the monitored data, it is recommended first of all to know if there are actually clusters in the data or not, i.e. clustering tendency. To determine whether the monitored data contains meaningful clusters and not a uniform structure, we will refer to the Hopkins Statistic method [48]; a spatial statistic that tests the spatial randomness of a variable as distributed in a space [42]; in order to assess the clustering tendency of the monitored data. Referring to [10], the Hopkins Statistic is the most easy, intuitive and explicit technique to be used compared to other approaches such as Holgate [46] or Eberhardt [28] approaches. Hopkins Statistic starts from a simple hypothesis that objects in the data set are uniformly distributed in multidimensional space and try to know if the analyzed distribution differs from this assumption. It works as follows: from a real dataset D , n points are sampled: P_1, \dots, P_n .

For each point $P_i \in D$, let P_j be its nearest neighbor and $d(p_i)$ the distance between P_i and P_j . In the space of the dataset D (limited by the maximum and the minimum of the coordinates of D points), n artificial points are generated following a random uniform distribution: q_1, \dots, q_n . For each point $q_i \in D$, let q_j be its nearest real neighbor and $d(q_i)$ the distance between q_i and q_j . The artificial and real points are randomly selected over the whole space [10][42][121]. The Hopkins statistic H is calculated according to the following formula:

$$H = \frac{\sum_{i=1}^n d(q_i)}{\sum_{i=1}^n d(q_i) + \sum_{i=1}^n d(p_i)}$$

This statistic compares the nearest-neighbor distribution of randomly selected real points to that for the randomly selected artificial points in the same dimensional space [10]. The Hopkins statistic will be computed for several times with new random points and the average of the calculated values will be taken into account in the decision phase. If the average of the H values is close to 0.5, it means that $d(q_i)$ and $d(p_i)$ are close to each other and thus, points are uniformly distributed and the initial assumption will be confirmed (no meaningful clusters). If the average is greater than 0.75, then there is a high probability to have a well-defined clustered data. For the values which are much less than 0.5, the analyzed data can be supposed to be regularly spaced [10][85].

3.4.2 Clustering Algorithms

After ensuring the existence of significant clusters in the monitored data, the main issue is on which clustering algorithm should we rely to classify the different users. Indeed, many clustering algorithms exist in the literature and several categorization criteria are taken into account, making it difficult to consider a specific approach. For this reason, we will focus in this work on two major fundamental clustering approaches that can be applied in our use case which are respectively, partitioning methods and density-based ones [43, 54, 109, 117].

3.4.2.1 Partitioning Methods

Partitioning methods represent the simplest and fundamental way to group points in clustering analysis. They take a set of n data points and try to construct k clusters ($k \leq n$) while each data point belongs to exactly one cluster. The metric used to differentiate clusters in most of partitioning methods is the distance between the objects. Therefore, objects that belong to the same cluster are probably close [43]. The most well-known partitioning algorithm in the literature is the K-means. Even if it was proposed a long time ago, K-means [71] remains one of the most used partitioning algorithms by researchers thanks to its simplicity and ease of implementation. The K-means algorithm performs as follows. Given a data set D containing n objects and k the number of clusters, the algorithm selects arbitrarily k points from D called cluster centroids.

Then, each remaining point in D will be assigned to one of the k clusters in such a way that the Euclidean Distance between the points and the centroids is minimized. This process will be repeated for several times and in each time, the centroids values will be updated based on the objects that belong to the corresponding cluster. Then, all objects will be re-assigned taken into account the new cluster centers. Once the centroids values and the objects assignments are immutable, the algorithm process is stopped and clusters are formed [55].

3.4.2.2 Density-Based Methods

Density based methods represent clustering approaches that divide a set of data points into dense regions separated by scattered ones. The capacity of a given cluster continues to increase as long as the number of objects, i.e. density, in the “neighborhood” reaches some threshold [43]. One of the most common and cited density-based clustering algorithms is Density Based Spatial Clustering of Applications with Noise (DBSCAN). It was proposed for the first time in 1996 [29]. The idea of DBSCAN is to find core data points which have dense neighborhoods in order to connect them and form the corresponding clusters. Unlike partitioning algorithms, DBSCAN does not require any specification regarding the number of clusters. Only two parameters need to be taken into account: the maximum radius of each cluster “ ϵ ” and the smallest number of cluster points “ $MinPts$ ” which represents the density threshold of a dense region [68]. The process of the DBSCAN algorithm is as follows. Given a data set D , ϵ and $MinPts$, all points are marked as unvisited. The algorithm selects a randomly unvisited point P and marks it as visited. Then it calculates the number of points within the ϵ – neighborhood of P (ϵ – neighborhood = $\{Q \in D / dist(P, Q) \leq \epsilon\}$). If the number of points is larger than $MinPts$, a new cluster is created containing all identified points. Else, P is considered as noise point and does not belong to any cluster. The same process will be repeated until all points are assigned to a cluster or considered as noise [5][29].

3.4.3 Experimental Evaluation

3.4.3.1 Environment Settings

In order to analyze and evaluate the clustering algorithms presented above and choose which one is more suitable for our use case, we rely on real traffic traces collected based on a tool developed within Orange France called *OTARIE*. This probe is based on a software written in C that works on various PC-based hardware architectures equipped with a Data Acquisition and Generation (DAG) traffic capture card which has an Application Programming Interface (API) that allows to read the packets as they arrive on the network interface. The traffic traces were collected for 3447 ONUs belonging to the same OLT equipment over a period of one month between the 2nd of November and the 3rd of December 2016. However, they do not cover the

whole day. For this reason, we have chosen to work on a specific day period between 9p.m and 12a.m with a 5 minutes interval since the clustering approach becomes very advantageous in the case of busy hours when a significant number of subscribers are online. The accomplishment of the experiment relies on the use of Python scripts and the Sklearn library [83] provided for Python developers to evaluate the different algorithms. Matplotlib and Seaborn modules have been used for plotting the different points and clusters in the most convenient way.

3.4.3.2 Results & Analysis

To have a rough vision on the general behavior of customers in the upstream, we show in Fig.3.4 an histogram of users' rate by common slice of mean upstream bitrate for all time intervals of a given day period. For display reasons and better representation, we present the results in this section for the day of November 14, 2016.

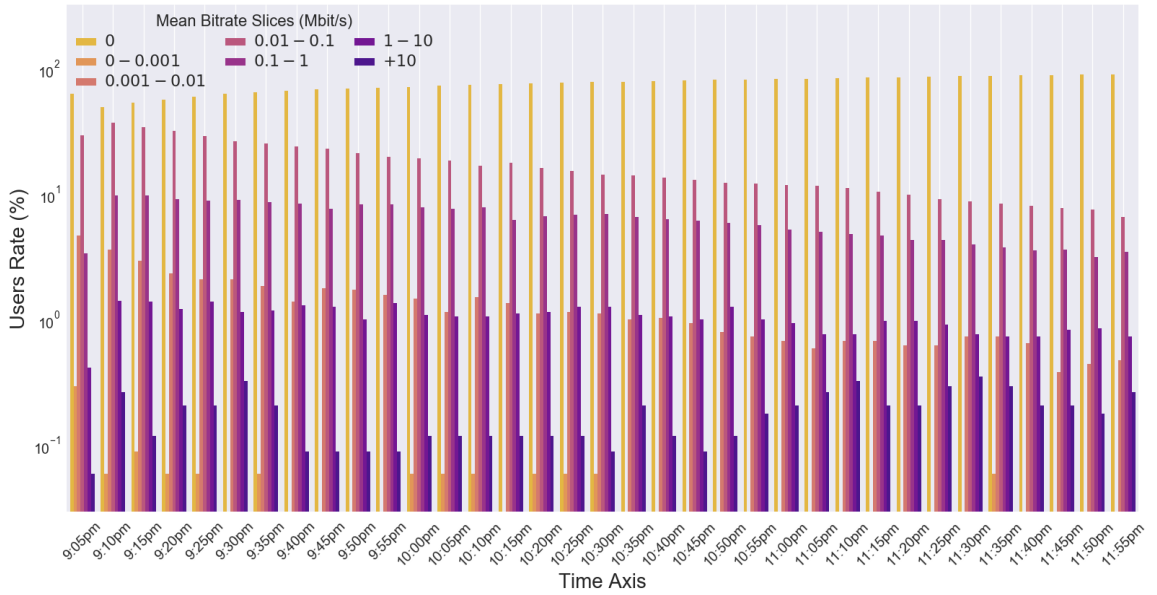


Figure 3.4: Distribution of users per mean upstream bitrate slices in a given day period

Before evaluating the clustering approaches, we should determine first of all whether the monitored data contains meaningful clusters or not (as it was explained in the clustering tendency section). Thus, we proceed to the calculation of the Hopkins Statistic index as illustrated in Fig.3.5.

The obtained values for the different time intervals are greater than 0.75. Thus, respecting the Hopkins theory, there is a high probability to have a well-defined clustered data. In the following, we present the results of the application of the K-means and DBSCAN algorithms to our dataset.

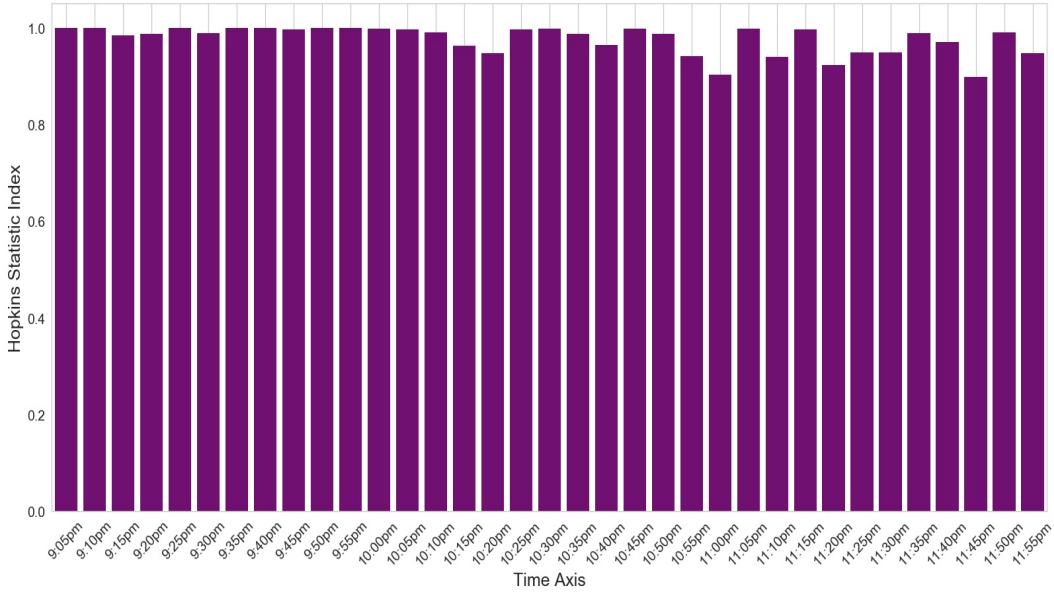


Figure 3.5: Hopkins statistic index

The K-means algorithm takes as input the number of possible clusters k . In our dataset, k represents the number of user classes per a parametrized interval of time t (see Table 3.4). We would expect that for each time interval t , a maximum of 3 user classes may exist: heavy, light and flexible users. Therefore, we evaluated the K-means algorithm with $k = 3$ for each time interval. Two metrics were used to calculate the Euclidean Distance used by K-means to differentiate user classes: the first consists in directly calculating the difference between two ONUs mean upstream bitrates. The second applies the decimal logarithm on the ONUs mean upstream bitrates. The results are presented respectively in Fig.3.6 and Fig.3.7. In all intervals, the three requested clusters are visualized. In Fig.3.6, the K-means algorithm has assigned the majority of users with a very low mean upstream bitrate to the same cluster. In Fig.3.7, the use of the decimal logarithm metric allows to have a more balanced distribution between the different clusters.

Unlike K-means, DBSCAN does not have any requirements regarding the specification of the number of clusters before analyzing the dataset. However, the most difficult task lies in how to choose the ϵ and $MinPts$ parameters. We assume that a user class can contain at least one user (heavy or light), that is, even if only one user has been assigned to a group, the DBSCAN algorithm must take it into account and not consider it as noise. It is possible to have a minority of users who are online during a day period. If we add that the final assignment of a user to a specific cluster depends on the calculation of his AI (i.e. the $MinPts$ value does not reflect the final percentage of heavy or light users), we can propose as a first hypothesis that $MinPts=1$. Regarding ϵ , we tried to vary it in order to choose which value can produce the best clustering

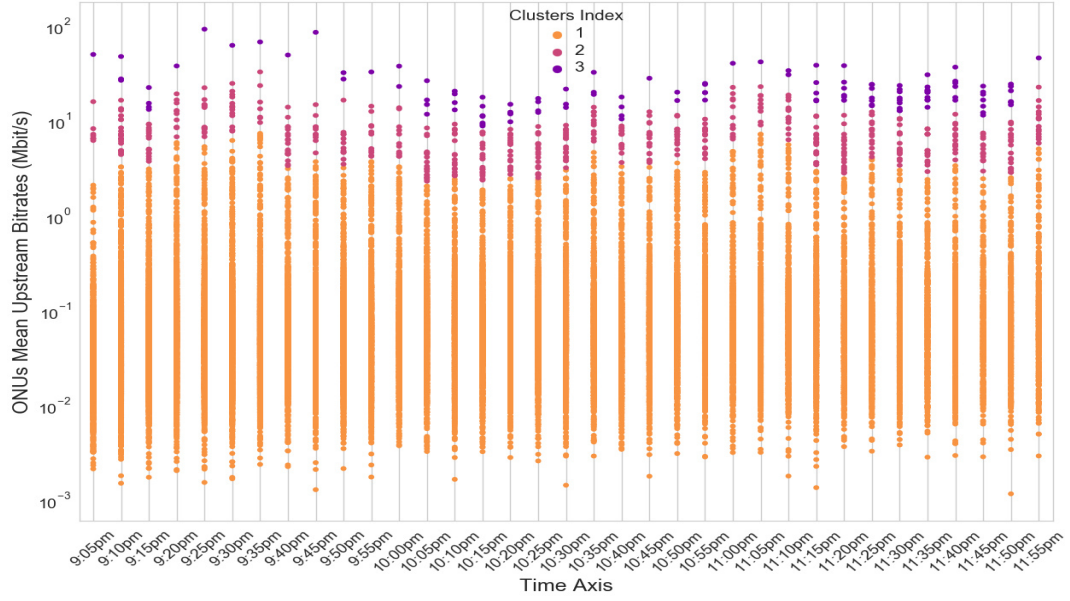


Figure 3.6: K-means results (k=3, metric: ONUs mean upstream bitrates)

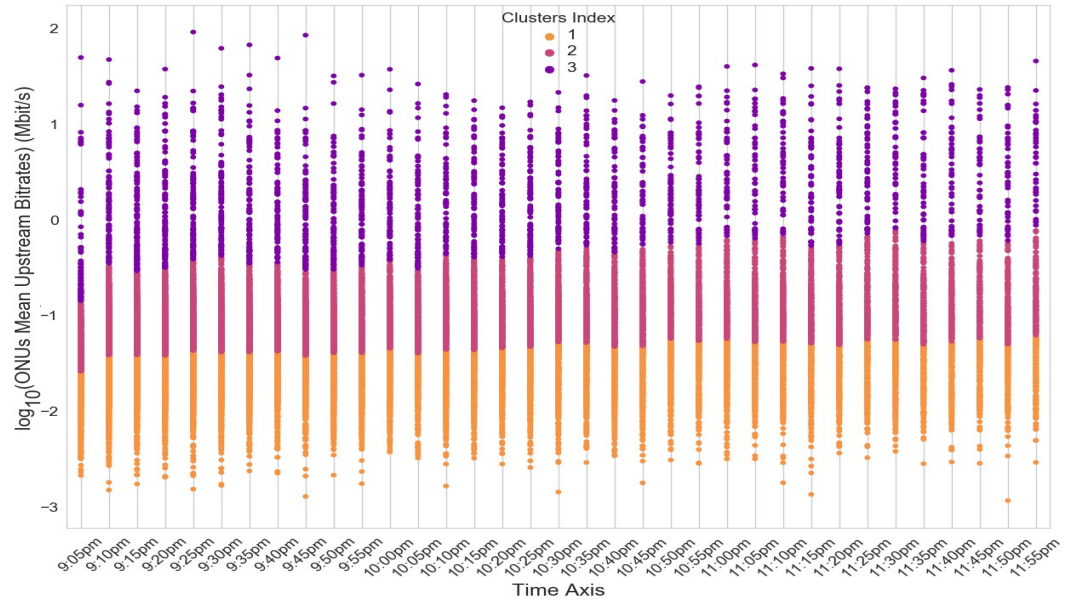


Figure 3.7: K-means results (k=3, metric: decimal logarithm of ONUs mean upstream bitrates)

results. In our case, ϵ can be understood as the gap between two user classes. Thus, the wider the gap, the easier the interpretation of the clusters. The tested values range from 0.1 Mbit/s to 20 Mbit/s. When the ϵ value decreases, more clusters are formed by the DBSCAN algorithm. The number of resulted clusters is a very important criterion to take into consideration in traffic patterns classification. Indeed, the higher it is, the more complicated the interpretation will be.

Between the different evaluated ϵ values, we have chosen to show the results for $\epsilon=10$ Mbit/s (see Fig.3.8) since the number of the identified clusters is the same as the number of clusters initialized to K-means (i.e. $k = 3$).

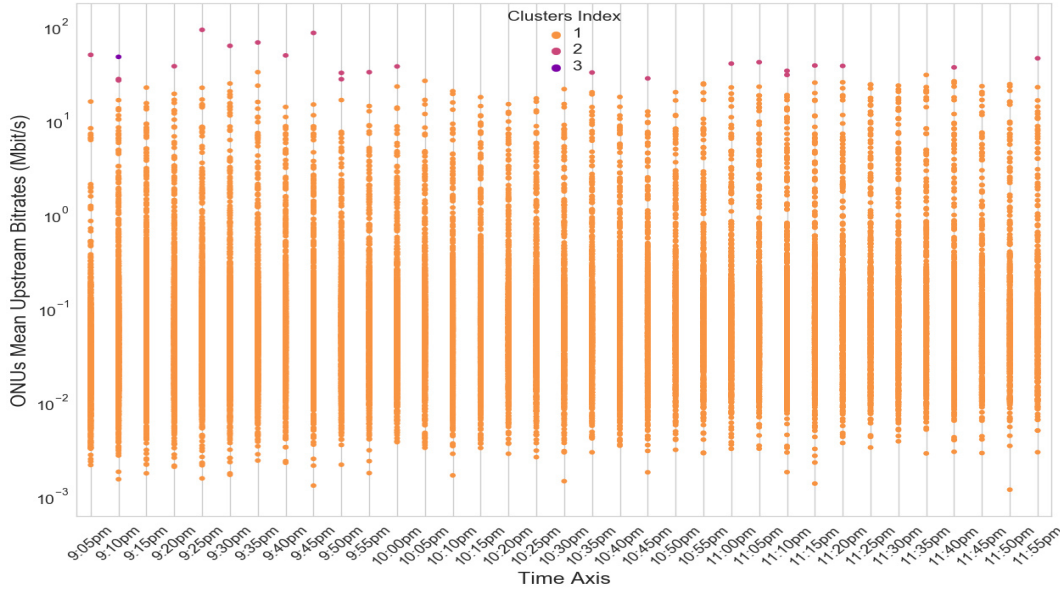


Figure 3.8: DBSCAN results ($\epsilon=10$ Mbit/s, $MinPts = 1$)

The evaluation of the K-means and DBSCAN algorithms allows us to visualize and differentiate users according to their traffic patterns. DBSCAN, thanks to its concept based on the differentiation of points according to their density, makes it possible to determine heavy and light user classes. However, it has a strong tendency to identify only a very small number of heavy users compared to the number of light ones using a 5 minutes time interval. This may seem as a major drawback for this algorithm in a policy of maximizing satisfaction for all network users. The other problem with DBSCAN is that we can fall in a case where one of the identified clusters contains heavy and light users in the same time. This case can occur when there are no differences between the densities of points in one time interval in such a way that the epsilon parameter has no effect (e.g. the interval of 10:05 pm in Fig.3.8). Concerning K-means, this algorithm gives a much more balanced distribution in terms of users' number between heavy and light. The use of the decimal logarithm metric allowed having a better classification results compared to the other metric. Nonetheless, K-means has the disadvantage of depending on the initial positions of different centroids and the need to specify for it the possible number of clusters before each process.

We consider in the rest of this work that K-means using a decimal logarithm metric is the most suitable algorithm for our model. The combination of the clustering analysis and the assignment index calculation allows having an overall vision of the traffic profile of each user and

makes it possible to estimate the possible behavior of a specific user at a specific time. As the customer traffic pattern is linked to daily life where the online behavior in the weekends may not be the same as the other weekdays, we propose that the users' classification phase should be performed per weekday and per weekend. In the following, we proceed to an evaluation of the users' classification modules where we illustrate the resulting users' distribution for a list of Wednesdays as a weekday and a list of Saturdays as day of the weekend.

Fig.3.9 and Fig.3.10 show the users' rate for each class (heavy, light, and flexible) for Wednesdays and Saturdays of the supervised period. These rates are calculated for each day based on the assignment index of each user to the heavy or light classes, calculated once the clustering process based on the K-means algorithm is finished. This index represents the probability of belonging to a specific class of users. The selected threshold 0.5 is the minimum value that must be chosen to remove any ambiguity concerning the classification step. Indeed, the sum of the assignment indexes to the heavy and light classes is always lower than 1. If we choose a threshold lower than 0.5, we may have cases where both indexes are above the selected threshold and therefore, users will be assigned to both classes at the same time. To remind, this index is calculated once the clustering process is finished. Indeed, for each 5 minutes time interval, the clustering module based on the K-means algorithm generates a classification of users depending on their mean upstream bitrates. Depending on this classification, assignment indexes of each user are calculated.

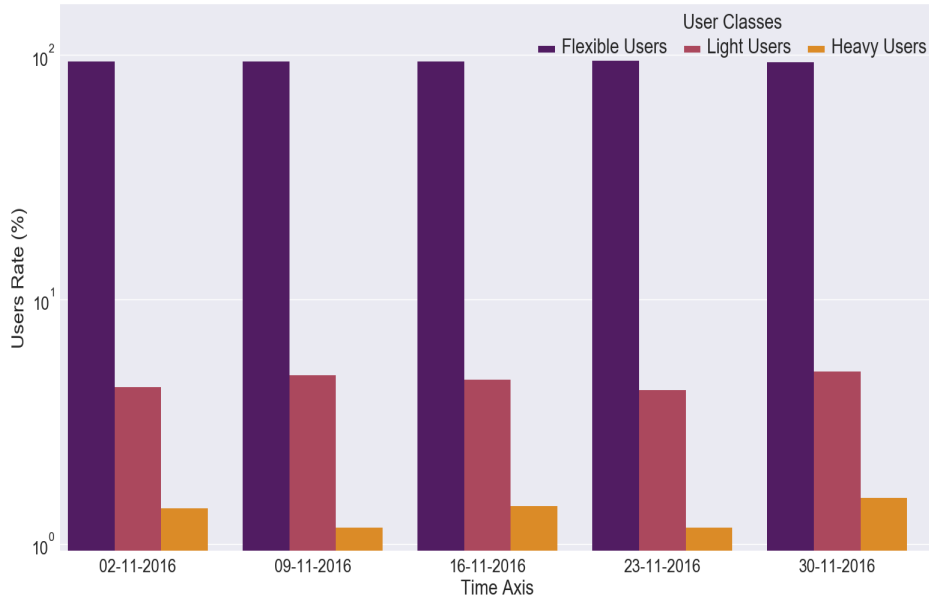


Figure 3.9: OLT user distribution for Wednesdays

The analysis of the resulting user rates shows an absolute majority of flexible users compared to light and heavy ones. This is confirmed for Wednesdays as well as Saturdays. Since our

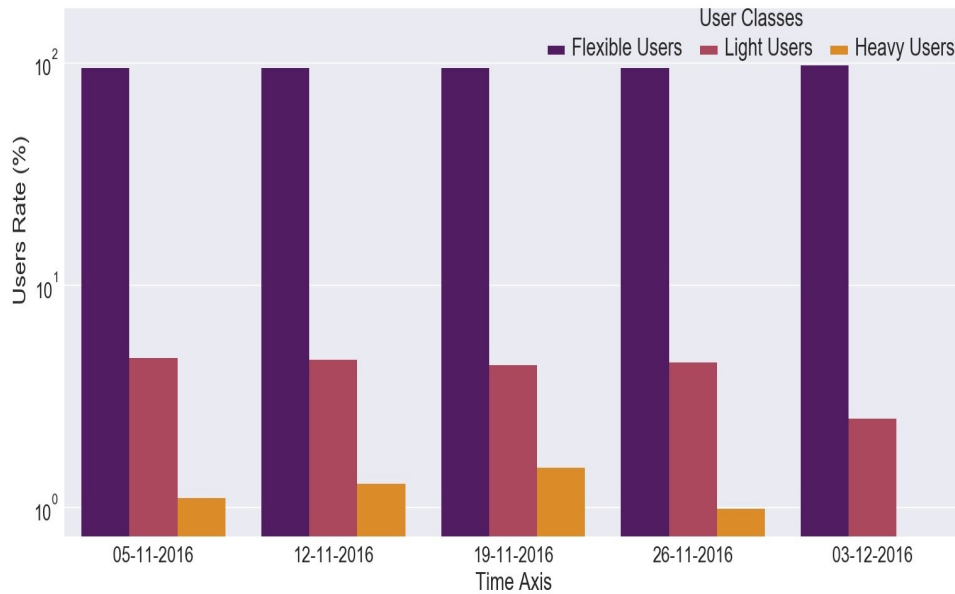


Figure 3.10: OLT user distribution for Saturdays

optimization approach is supposed to be applied per PON port, we choose to work on a specific OLT PON port instead of working on the whole OLT. Fig.3.11 and Fig.3.12 show the users' distribution for a PON port that connects 32 subscribers. Fig.3.13 presents the final distribution of the PON port users based on the average of their assignment indexes over all supervised days.

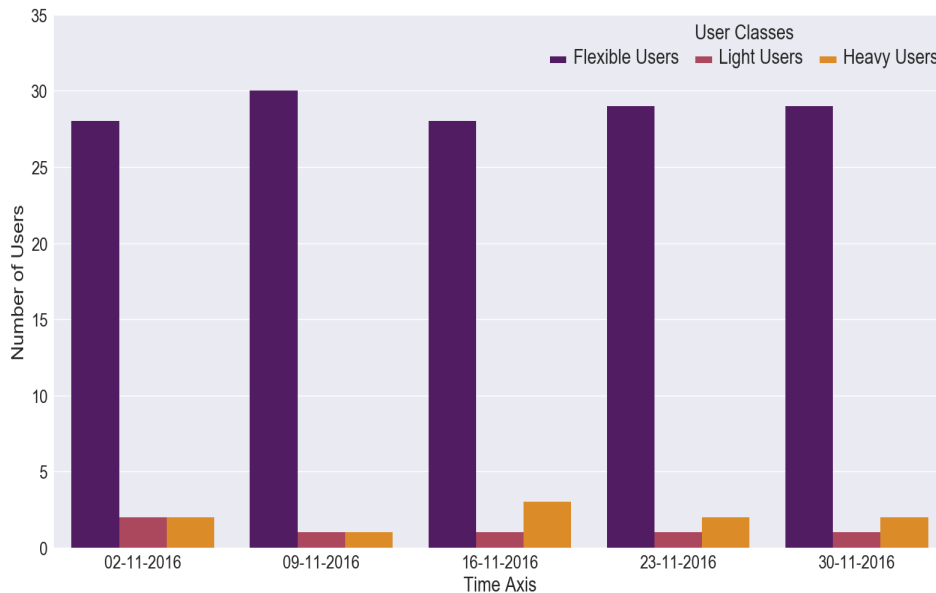


Figure 3.11: A PON port user distribution for Wednesdays

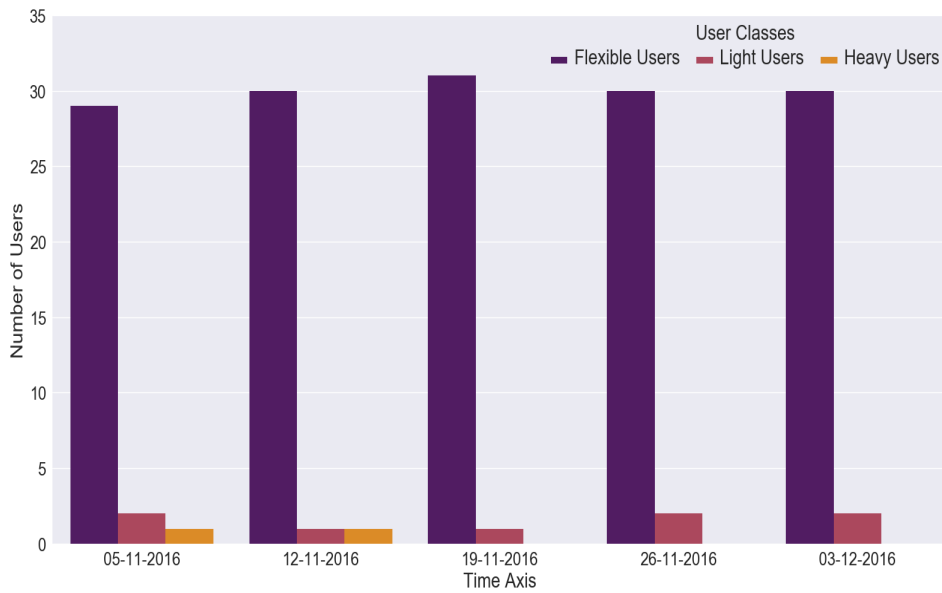


Figure 3.12: A PON port user distribution for Saturdays

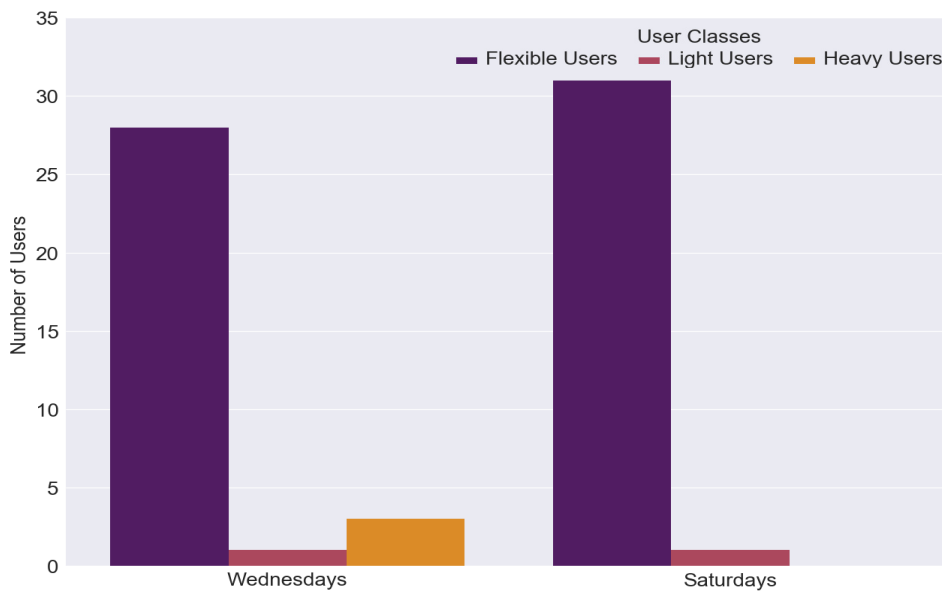


Figure 3.13: A PON port user distribution for Wednesdays and Saturdays

3.4.3.3 Discussion

The purpose of using clustering analysis associated with an assignment index calculation process is to classify PON users depending on their traffic patterns. The results that we have achieved in relation with users' distribution are characterized by a high accuracy in the assignment of a PON user to a certain traffic class. For that, the majority of users were assigned to the flexible class

with a percentage of 87,5% for Wednesdays and 96,87 % for Saturdays, and the minority were assigned to the heavy and light classes. We can proceed to the calculation of the extra bandwidth that can be allocated to heavy users based on the algorithm proposed in the Reallocation of SLA Parameters Module. However, we think that in the purpose of maximizing as much as possible the satisfaction for the majority of customers, the resulted unbalanced user distribution looks like an obstacle in our optimization approach. Indeed, it is preferable to have a significant number of heavy and light users in order to maximize the efficiency of the bandwidth usage in the network. Thus, the additional bandwidth that can be estimated will be greater if the number of light users is increased and in parallel, the number of users who will benefit from the bandwidth supplement will increase if the number of heavy users increases in turn.

3.5 Conclusion

In this chapter, a new mechanism for reallocating SLA parameters based on users' traffic patterns in a PON network has been proposed. The idea was to try to exploit the extra bandwidth theoretically untapped by the DBA in order to share it among heavy users without being limited by the constraints of the SLA parameters. The approach is characterized by its ability to be implemented directly in the network without the need to update the DBA at the OLT level. A particular attention has been granted to the clustering module which allows to classify subscribers according to their daily online behavior during the day in order to have clusters that contain heavy users who have a high upstream transmission bitrate and clusters which contain light users with a low transmission upstream bitrate. Two clustering approaches have been evaluated, namely K-means and DBSCAN algorithms. The analysis that we have conducted is based on the ability of each algorithm to differentiate or not the PON users according to their mean upstream bitrates by time interval in a day period. The results showed that the two algorithms can meet this requirement. However, K-means based on the decimal logarithm metric shows a better classification results compared to DBSCAN in terms of balanced distribution between heavy and light users. The association between the clustering analysis and the assignment index calculation proposed in this work has shown that a high accuracy selection may lead to an unbalanced users' distribution where the majority were affected to the flexible class.

In the next chapter, we will try to fix this issue by studying more the flexible users' behavior. The forecasting paradigm will be used as a reference in this context to have a vision about the possible traffic trend of flexible users (whether heavy or light) and then, have a more balanced users' distribution per day period.

CHAPTER 4

FORECASTING PON USERS TRAFFIC BEHAVIOR

Contents

4.1	Introduction	79
4.2	Related Work	80
4.2.1	Auto-Regressive Integrated Moving Average (ARIMA)	81
4.2.2	Artificial Neural Networks (ANNs)	83
4.2.3	Grey Theory	85
4.3	Forecasting PON Users Traffic Behavior	87
4.4	Evaluation	89
4.4.1	Environment Settings	89
4.4.2	Results & Analysis	90
4.4.3	Discussion	100
4.5	Conclusion	100

4.1 Introduction

In the previous chapter, we have proposed a new model for optimizing PON upstream resources that stems from a very simple idea: analyzing the past customers' behavior based on their historical data to estimate their upstream bitrates in the future, and efficiently exploit the bandwidth available in the network. Clustering analysis was used to identify heavy and light users based on their mean upstream bitrates for specific time intervals (e.g., 5 minutes). Then, an assignment index calculator module was proposed to assign each user to a particular class (heavy, light, or flexible) for all the time series collected by the network operator. The combination of the clustering analysis and the assignment index calculation allows having an overall vision of

the traffic profile of each user and makes it possible to estimate the possible behavior of a specific user at a specific time. In this case, the reallocation of the SLA parameters can be very useful and advantageous in the context of optimizing PON upstream resources for a specific day period. Although this approach is characterized by its high accuracy in the assignment of a PON user to a certain traffic class, it can lead to an unbalanced users' distribution where the majority of ONUs will be assigned to the flexible users class. This may be an impediment to the overall objective which consists in maximizing as much as possible the satisfaction for the majority of customers. Thus, it is preferable to have a significant number of heavy and light users in order to maximize the efficiency of the bandwidth usage in the network. To resolve this issue and to cope with the unexpected traffic behavior of some PON users by learning more about the traffic tendency of flexible ones, we assume that the future online behavior of a PON user depends on his former behavior. Thus, forecasting approaches that can deal with the analysis and the prediction of network users' traffic attitude are strongly recommended. In this context, we extend the architecture that we have proposed in the previous chapter by adding a new module that relies on a forecasting theory. This module is considered as a part of a second users' classification step that we suggest to be an improvement of the first classification method and can effectively extract the trend information from ambiguity.

The remaining of this chapter is structured as follows. Section 4.2 discusses some work related to forecasting techniques. Section 4.3 presents the enhanced version of the initial model using a forecasting module while emphasizing the method used in this regard. In section 4.4, we evaluate the forecasting module to show its impact on the users' classification phase. Finally, we summarize our work in section 4.5.

4.2 Related Work

When we design an adaptive model that relies on a proactive management scheme to plan in advance new SLA parameters for PON users, forecasting approaches become good candidates to improve the decision-making process. A simple and practical way to define forecasting is to consider it as a construction of stochastic models to predict time series values based on the historical ones [3]. Traditionally, this can be performed following two techniques, respectively qualitative and quantitative approaches. Qualitative technique relies on the knowledge and the experience of the forecaster who will take the final decision about the expected trend of data. The quantitative one aims to identify the data patterns from the historical dataset in order to predict the future values [107][116]. Quantitative approaches may be also classified into causal relationship methods and time series ones. While the first category tries to make a relationship between many factors in order to generate the forecasted values, the second is limited to the statistical data that was observed and collected over regular time intervals such as hourly,

daily, weekly, monthly, etc [58]. Since the historical data required by the network operator to forecast users' traffic behavior can be easily obtained and processed with the aim to classify the different customers, the focus in this work will be on time series methods. Time series forecasting methods represent one of the most important forecasting approaches in the literature. Thanks to the collection and the analysis of past observations, the time series methods try to extract the relationship between the different points in order to propose an extrapolation in the future of the time series under study [119]. Several time series forecasting approaches exist in the literature. Traditional ones and the most simple in terms of understanding, evaluation, and implementation remain the linear models (e.g. autoregressive models, moving average, Auto-Regressive Integrated Moving Average (ARIMA), etc.) which have been mainly used in the past decades. The idea behind these models is the use of linear functions to forecast future values based on past ones. When dealing with time series that have a non linear pattern, a non linear model is preferred. In that case, Artificial Neural Networks (ANNs) are the best candidates to be selected thanks to their flexibility in modelling and pattern identification. As it's usual that linear models as well as ANNs require a huge amount of data to perform a good forecasting accuracy while in most cases researchers have a limited datasets, another forecasting theory was proposed, namely "Grey theory". The latter was designed to deal with systems characterized by incomplete information or poor datasets. We present in the following an explanation of ARIMA model, ANNs, and Grey theory. The three models are supposed to be the reference for our forecasting module in learning the tendency of PON users in the future.

4.2.1 Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA model represents one of the most efficient linear time series forecasting models. ARIMA is built on the assumption that the time series is stationary while stationarity can be defined as the maintaining of the time series statistical properties (i.e. statistical properties such as the mean or the variance remain constant) over time. Thus, if a time series follows a particular behavior during a given time period, it's highly probable that it continues to follow the same behavior in the next time period [13][119]. As the name suggests, the ARIMA model is supposed to have three main components, while each one has its corresponding parameter [13]:

- The auto regressive component AR(p): this component supposes that the model uses the past values as reference to forecast the future value. Accordingly, p represents the number of autoregressive terms or the lags used in the model. The mathematical representation of the AR(P) component can be modelled as follows:

$$Y_t = c + \sum_{i=1}^p \alpha_i Y_{t-i} + \epsilon_t \quad (4.1)$$

where Y_t is the value to be predicted based on the Y_{t-i} past values, c is a constant, α_i is the autoregressive coefficient, and ϵ_t is a white noise.

- The integrated component I(d): this component is responsible for differencing the time series. That is, correcting the non-stationarity of the time series in order to stabilize it. Thus, the d parameter refers to the number/degree of differences required for ensuring stationarity
- The moving average component (MA(q)): this component supposes that the time series presents some fluctuations around the mean value. Thus, the best method to estimate the future value is to consider it as the average of the past observed values plus q errors assumed to impact the made observations. The mathematical representation of the MA(q) component can be modelled as follows:

$$Y_t = \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (4.2)$$

where ϵ_{t-i} is a white noise and θ_i the moving average coefficient.

Thus, an ARIMA(p,d,q) model can be modelled as follows:

$$Y_t = c + \sum_{i=1}^p \alpha_i Y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (4.3)$$

By looking to the ARIMA (p,d,q) mathematical formulation, there is a need to identify the model optimal order as well as the autoregressive and moving average parameters after validating the stationarity properties. By referring to the Box–Jenkins model [13], these steps are called model identification, parameter estimation, and diagnostic checking. Model identification involves checking the stationarity/non-stationarity of the time series. Parameter estimation has to determine the three parameters p, d, and q as well as the autoregressive and moving average parameters. For estimating p and q, Akaike’s Information Criterion (AIC) can be very useful insofar as the lower the AIC, the higher quality of the model [50]. The mathematical formulation of the AIC is as follows:

$$AIC = -2 \log(L) + 2 \times (p + q + k + 1) \quad (4.4)$$

where L represents the likelihood of the data ; $k = 0$ if $c = 0$ and $k = 1$ if $c \neq 0$.

For determining d, the best way is to refer to the Auto Correlation Function (ACF) plot and the Partial Auto Correlation Function (PACF) plot in order to decide whether data need to be differentiated or not. Regarding the autoregressive and moving average coefficients (i.e. α_i and θ_i), Maximum Likelihood Estimation (MLE) is the best method to be used thanks to its capability to maximize the probability of having the observed data [50]. In the final step, the diagnostic checking is ensured by the Box–Jenkins model which validates that the error ϵ_t is small enough to be considered [59]. As it was mentioned before, ARIMA is one of the most known linear forecasting models in the research world. Thus, it is very advantageous when dealing with

datasets that have linear patterns. But it is not suited to non linear behaviors of some time series. In this context, a well-known non-linear forecasting paradigm, namely ANN, is discussed in the following section.

4.2.2 Artificial Neural Networks (ANNs)

ANNs represent one of the most popular forecasting paradigms [75] that deal with nonlinear problems. Classified as a machine learning approach, ANNs have the ability to learn from complicated data and deduce its pattern and tendency. They can be very appropriate in the context of a knowledge-based learning mechanism that is difficult to specify [120]. Generally, ANNs are used to identify data patterns which are not easily perceptible by classical forecasting methods. However, due to the fact that ANNs principle relies on finding the periodicity within the data, there is a need for long time series and huge amount of data to ensure a good forecasting accuracy [75][120]. By analogy to the human brain, ANNs ensure the information process through the interaction of artificial neurons (or nodes) and can interpret the future behavior of a dataset despite the existence of noisy information [116]. Each neuron obtains the information as an input signal from other neurons and through a transfer function, a local process is performed to generate an output signal that can be used by other neurons or external outputs.

As many types of ANNs are proposed in the literature to deal with a variety of problems, the focus in the rest of this section will be on the Multilayer Perceptron Model (MLP) as a well-known class of ANNs. The MLP paradigm is built on the basis of a layering system with no less than 3 layers of neurons. The first layer of the MLP (the lowest), namely the “input layer”, is responsible for collecting the external information. The last layer (the highest), namely the “output layer”, is in charge of providing the problem solution. In a time series forecasting problem, MLP inputs can be seen as the data points already collected such as historical data or past observations while the outputs represent the desired future data points. Between these two layers, one or more intermediate layers called “hidden layers” may be present. The different neurons of the adjacent layers are connected acyclically from the lower to the higher layer as shown in Fig.4.1 [4].

Thus, an important step before forecasting with MLP is to determine these connections weights (usually known as arc weights). This step, so called “training process”, is essential insofar as it prepares the MLP to achieve the non-linear mapping from input layers to output layers and then, specify the way in which the network should learn. Weights are chosen randomly at the beginning of the process then, a supervised learning mechanism is performed in a manner that desired outputs and their associated inputs are provided. To explain more, let’s suppose that we have N past data points y_1, y_2, \dots, y_N in the training set and k input nodes; $k < N$; in the MLP. Then, we have $N - K$ training patterns. The first will contain y_1, \dots, y_k as inputs and y_{k+1} as the desired output while the last will be composed of y_{N-k}, \dots, y_{N-1} as input nodes and

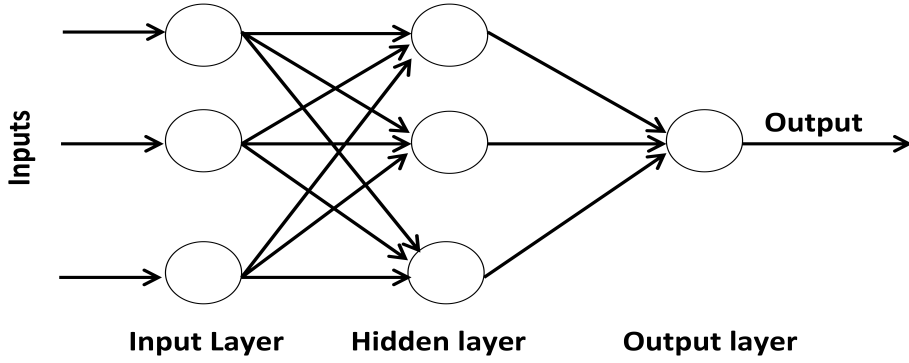


Figure 4.1: A typical MLP ANN structure with one hidden layer and one output

y_N as the desired output [120]. For each training pattern, the feed-forward scheme presented in Fig.4.2 is adopted.

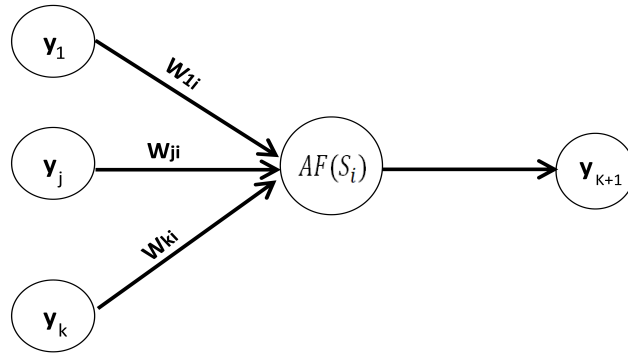


Figure 4.2: A feed-forward scheme for training process in MLP

where w_{ji} is the weight associated to the input node j and the hidden layer node i , AF represents the activation function selected to perform the transformation towards another node, and $S_i = \sum_{h=1}^k w_{hi}y_h$. This function can be a linear, hyperbolic, sine, cosine or logistic. The logistic function (called also sigmoid) remains the most used by researchers [61] and can be expressed as follows:

$$f(x) = (1 + \exp(-x))^{-1} \quad (4.5)$$

If there are many hidden layers, the activation value will pass from a node to another undergoing the same transformation (summation of received values and modification using the activation function). The predicted value using the feed-forward scheme above is then compared to the actual output provided to the network. The error function (e.g. Mean Squared Error (MSE) or Sum of Squares Error (SSE)) will be then used following a backward error propaga-

tion (i.e. Gradient Descent algorithm) to adjust the different weights. The training process is repeated several times until weights for different connections are optimized by minimizing the error function. Once, this process is finalized, the MLP ANN can be considered as ready to forecast. As it was mentioned before that ANNs require a large sample of data to perform well, it is interesting to look for alternative approaches that can deal with small sample sizes. In the following, a good candidate namely “Grey Model” that can overcome ANNs limit in terms of dataset size is discussed.

4.2.3 Grey Theory

The majority of the forecasting approaches (including ANNs explained before) need in general a huge quantum of history data to perform a good forecasting accuracy. Thus, these approaches may not be suitable for predicting future time series data points when we fall in case of a limited amount of past data [105]. To address this issue, Grey forecasting theory was proposed by Deng in 1982 [25]. Thanks to its ability to estimate the possible data behavior based only on a few information samples [64], Grey theory becomes one of the most popular prediction approaches used in the research world [112]. Grey model is very advantageous when dealing with systems characterized by incomplete (partially known) and/or poor information (minimum of 4 data points) and does not impose any assumptions related to the data statistical distribution [49]. The core and the most commonly used model of Grey is known as Grey Model First Order One Variable (GM(1,1)) [58][75]. The concept of this model relies in identifying the mathematical relationship between different points to learn more about the behavior of the dataset and to make the right decision about the future trend [116]. We illustrate the mathematical formulation of the Grey forecasting Model GM(1,1) in the following. We assume the initial data series with n ($n \geq 4$) non-negative values as follows:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (4.6)$$

To reduce the randomness of data, the Accumulated Generating Operation (AGO) is then applied since the initial data series may change randomly while there is a need to know its regular pattern [25]:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \quad (4.7)$$

where $x^{(1)}(k) = \sum_{m=1}^k x^{(0)}(m)$, $k \in [1, n]$.

The original form of the GM(1,1) model is:

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (4.8)$$

where a is the developing coefficient and b is the grey input according to the Grey theory. $x^{(1)}(k)$ can be replaced then by the average of two consecutive neighbours $x^{(1)}(k)$ and $x^{(1)}(k-1)$:

$$x^{(0)}(k) + az^{(1)}(k) = b, k \in [2, n] \quad (4.9)$$

where $z^{(1)}(k) = 0,5 \times (x^{(1)}(k) + x^{(1)}(k - 1))$.

According to the least square method, a and b can be identified as follows:

$$A = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \quad (4.10)$$

where:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \cdot & \\ \cdot & \\ \cdot & \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdot \\ \cdot \\ \cdot \\ x^{(0)}(n) \end{bmatrix}$$

By regarding the following differential equation as a shadow of Eq. (4.9):

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b \quad (4.11)$$

The GM(1,1) can be therefore established:

$$\hat{x}^{(1)}(k + 1) = (x^{(1)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (4.12)$$

As we have applied the AGO in the equation 4.7, we apply the inverse (IAGO) to obtain the forecasted value of the original data $x^{(0)}$:

$$\hat{x}^{(0)}(k + 1) = (1 - e^a)(x^{(1)}(1) - \frac{b}{a})e^{-ak} \quad (4.13)$$

Each of the three forecasting approaches discussed above has its own advantages and drawbacks and all of them can be applied in a context of upstream traffic tendency forecasting problem. ARIMA works very well for stable and stationary time series while there is a need to specify the (p,d,q) parameters to fit the model. ANNs are appropriate for non linear patterns whereas there is a need for long time series and huge amount of data. GM(1,1) is very suitable for prediction of short time series, nevertheless the selection of the initial $x^{(0)}(1)$ may impact directly the model accuracy. At this level of work, we do not make any choice between GM(1,1), ARIMA, and ANNs since none of them can be considered as a universal approach or the most efficient method in a general way. Thus, we leave the task of which approach to select to the forecasting module that we describe in the following section.

4.3 Forecasting PON Users Traffic Behavior

In this section, we introduce the design of the enhanced model inspired by the first model proposed in the previous chapter while the main novelty lies in the integration of a forecasting module. Fig.4.3 shows the architecture of the new proposed model.

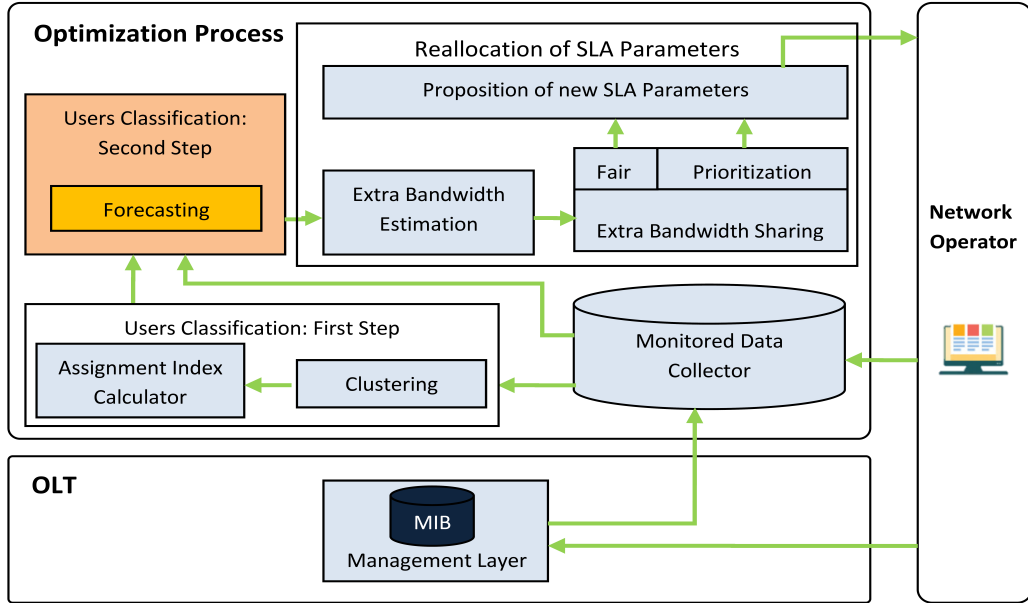


Figure 4.3: Enhanced model using a forecasting module

There are no changes performed on the modules inherited from the initial model (we keep the same concepts). The only modification is supposed to be at the level of the users' classification. In fact, a forecasting module is added. This module will be considered as a second step of the users' classification phase as it depends on the results of the clustering analysis and assignment index calculator modules. It will be applied particularly to the assignment indexes of flexible users in order to forecast new values for next weeks, while heavy and light ones will not be involved. Indeed, in our approach, a user can be flexible only if the averages of his assignment indexes to heavy and light users over the supervised days are smaller than 0.5 (the selected threshold). That is, the user does not tend to be neither heavy nor light. Since the index is calculated on the basis of historical data, cases like missing data or the lack of a vision on the traffic trend of users in the future will be often encountered. This can impact the calculation of the assignment index and subsequently the classification of users. In this context, the enhancement of the initial model by using a forecasting approach can be beneficial to have a more reliable and useful approach where a part of flexible users can be assigned to one of the other two classes. Accordingly, the extra bandwidth estimated and the number of beneficiary heavy users will increase automatically.

We illustrate the proposed method for forecasting flexible users assignment indexes in Fig.4.4.

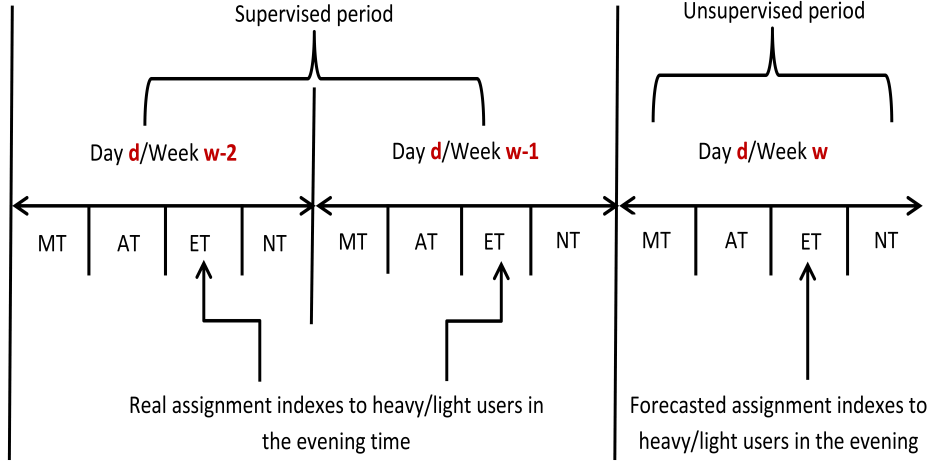


Figure 4.4: Proposed method for forecasting flexible users assignment indexes

where MT:Morning Time, AT: Afternoon Time, ET: Evening Time, NT:Night Time

The proposed method works as follows:

- **Step 1:** The assignment indexes of flexible users generated by the assignment index calculator module on the basis of the Clustering module results are recovered.
- **Step 2:** For each flexible user, the dataset presented in Table.4.1 is instantiated.

Table 4.1: Dataset instantiated for each flexible user

Day: d, Week: w-n	...	Day: d, Week: w-2	Day: d, Week: w-1
$AI_{C_{MT_H}}^{d,w-n}$.	$AI_{C_{MT_H}}^{d,w-2}$	$AI_{C_{MT_H}}^{d,w-1}$
$AI_{C_{AT_H}}^{d,w-n}$.	$AI_{C_{AT_H}}^{d,w-2}$	$AI_{C_{AT_H}}^{d,w-1}$
$AI_{C_{ET_H}}^{d,w-n}$.	$AI_{C_{ET_H}}^{d,w-2}$	$AI_{C_{ET_H}}^{d,w-1}$
$AI_{C_{NT_H}}^{d,w-n}$.	$AI_{C_{NT_H}}^{d,w-2}$	$AI_{C_{NT_H}}^{d,w-1}$
$AI_{C_{MT_L}}^{d,w-n}$.	$AI_{C_{MT_L}}^{d,w-2}$	$AI_{C_{MT_L}}^{d,w-1}$
$AI_{C_{AT_L}}^{d,w-n}$.	$AI_{C_{AT_L}}^{d,w-2}$	$AI_{C_{AT_L}}^{d,w-1}$
$AI_{C_{ET_L}}^{d,w-n}$.	$AI_{C_{ET_L}}^{d,w-2}$	$AI_{C_{ET_L}}^{d,w-1}$
$AI_{C_{NT_L}}^{d,w-n}$.	$AI_{C_{NT_L}}^{d,w-2}$	$AI_{C_{NT_L}}^{d,w-1}$

n: Supervised Weeks AI: Assignment Index C__H: Heavy Class C__L: Light Class

- **Step 3:** Using the dataset generated, forecasting is performed for $w + k; k \geq 0$. For instance, the forecasting model uses the assignment indexes already calculated for Wednesday evenings in $w_i; i \in [1, n]$ to forecast the assignment indexes for next Wednesday evenings.
- **Step 4:** If the forecasted assignment index to the heavy or light class value exceeds 0.5 (the selected threshold), then the flexible user will be classified temporarily as heavy or light for the next week. Otherwise, he will remain as flexible. It should be noted that a flexible user may be finally and not temporarily assigned to the heavy or light class. This can be done by re-calculating the new averages of the assignment indexes using the new monitored data collected during the week and comparing the resulted values to the selected threshold.

As previously mentioned, we are in a situation where we need to forecast on the basis of a time series that contains data points collected at constant time intervals, i.e. Time Series Forecasting. In section 4.2, we broached this topic by discussing three well-known time series forecasting approaches respectively, ARIMA Model, ANNs, and Grey Model GM(1,1). However, we did not make a choice between these three approaches. This is due to the fact that theoretically, there is no universal approach or an approach that outperforms others in terms of performance in a general way. The best approach would still depend on the dataset which is under study. In that case, we will proceed to a comparative study from which we will select the most suitable forecasting model.

4.4 Evaluation

In this section, we proceed to an evaluation of the user classification module enhanced in this chapter by adding a forecasting mechanism. We start first of all by a comparative study in order to opt for the most suitable model in terms of forecasting accuracy. Then we investigate the performance in terms of a more balanced users' distribution and top customers satisfaction that can be achieved through the use of the selected forecasting model.

4.4.1 Environment Settings

The reference dataset used relies on real traffic traces collected within the Orange France network. The data collection was ensured by the use of a probe called *OTARIE* and equipped with a DAG traffic capture card which has an API that allows reading the packets as they arrive on the network interface. 3447 ONUs belonging to the same OLT were supervised over a period of

one month between the 2nd of November and the 3rd of December 2016. Given that the traffic traces do not cover the whole day, the day period theoretically qualified as the busiest which is between 9p.m and 12a.m was selected in order to analyze the behavior of the majority of subscribers.

As the customer traffic pattern is linked to daily life where the online behavior in the weekends may not be the same as the other weekdays and similar to what we did in the previous chapter, we classify customers per weekday and per weekend. For display reasons, we decide to show the results of the application of forecasting models for a list of Wednesdays. The accomplishment of the experiment relies on the use of Python scripts and libraries to evaluate the different algorithms and *Matplotlib* and *Seaborn* libraries to plot the different results in the most convenient way.

4.4.2 Results & Analysis

As mentioned in section 4.3, the forecasting module takes into account the different users' distributions resulting from the combination of clustering analysis and assignment index calculation. While the heavy and light users are clearly characterized by the mean assignment index calculation, the flexible ones which represent the majority may tend to one of the other classes if we extend the dataset and generate more indexes. This can influence the final users' distribution and consequently the extra bandwidth that will be estimated to be shared among heavy customers.

As it was explained before, in theory, there is no universal model for time series forecasting. Thus, we cannot affirm that for instance, ANN is better than GM(1,1) or ARIMA is better than both. For this, it is recommended to go through a test step in order to evaluate the different approaches using the dataset under possession. Then, using some forecasting accuracy metrics, the approach that outclasses the others will be selected as reference in the proposed forecasting module. However, a problem that is often encountered in relation to the evaluation of forecasting models is the number of data points of the studied dataset. In general, the larger the dataset, the more effective the evaluation. In relation with our work, ANN, GM(1,1), and ARIMA allow to achieve our main goal concerning the forecasting of customers' traffic behavior. However, we expect that only GM(1,1) and ARIMA remain for the moment the most appropriate for the evaluation phase. This is due to the fact that the dataset we have is limited, which does not represent a problem for ARIMA model or Grey systems which can even work with incomplete data (although it is always better to have the maximum of data points). However, ANN require a very large amount of data (not less than 50 data points as mentioned in [75]) to ensure that the forecasted values are statistically accurate [16][116]. This on the one hand makes the learning speed slower and in the other hand, excludes the ANN track from the evaluative part. In the following, we evaluate the GM(1,1) and ARIMA model using the metrics presented below:

- **Residual (R)**: this metric measures the difference between the real value of the assignment index and the forecasted one.

$$R = Real\ Value - Forecasted\ Value$$

- **Forecast Error (FE)**: this metric measures the absolute value of the difference between the real value of the assignment index and the forecasted one. It can be stated as a simple value or a percentage. We use it in this work as a percentage as indicated in the following formula:

$$FE = (|Real\ Value - Forecasted\ Value| / Real\ Value) \times 100$$

- **Forecast Accuracy (FA)**: this metric is a decreasing function of the forecast error in terms of percentage. In case of large forecast error values (above 100%), forecasting accuracy is assigned to 0 to avoid a negative value, resulting in an inaccurate forecast.

$$FA = \max(0, 100 - FE)$$

- **Mean Residuals (MR)**: this metric measures the average of the calculated residuals over the time series periods. In general, the MR is used to identify the tendency of the forecasting model. While a positive value denotes that the model tends to under-forecast, a negative one indicates an over-forecasting trend. The ideal value is supposed to be 0.

$$MR = \sum_{i=1}^n (Real\ Value_i - Forecasted\ Value_i) / n$$

- **Mean Absolute Deviation (MAD)**: this metric measures the average of the forecast errors (as absolute values) over the entirety of the forecast time periods. The MAD should be as small as possible to ensure the correct accuracy of the forecasting model.

$$MAD = \sum_{i=1}^n |Real\ Value_i - Forecasted\ Value_i| / n$$

- **Tracking Signal (TS)**: this metric measures the ratio of the cumulative sum of residuals to the mean absolute deviation. The tracking signal is used generally to check if the forecasting model needs or not to be reviewed. Values within $+/- 4$ assume that the model does not need adjustment.

$$TS = \sum_{i=1}^n (Real\ Value_i - Forecasted\ Value_i) / MAD$$

To remind, the forecasting model is applied only to the assignment indexes of flexible users in order to learn more about their traffic trends. That is, for each flexible user, we predict his possible assignment index to the heavy and the light class for the next week (even for the next weeks) and depending on the calculated value, a temporary reassignment to the heavy or light classes is performed. For more homogeneity, we begin our evaluative study from where we stopped in the previous chapter. We recall the final users' classification that we generated for a PON port in Fig.4.5.

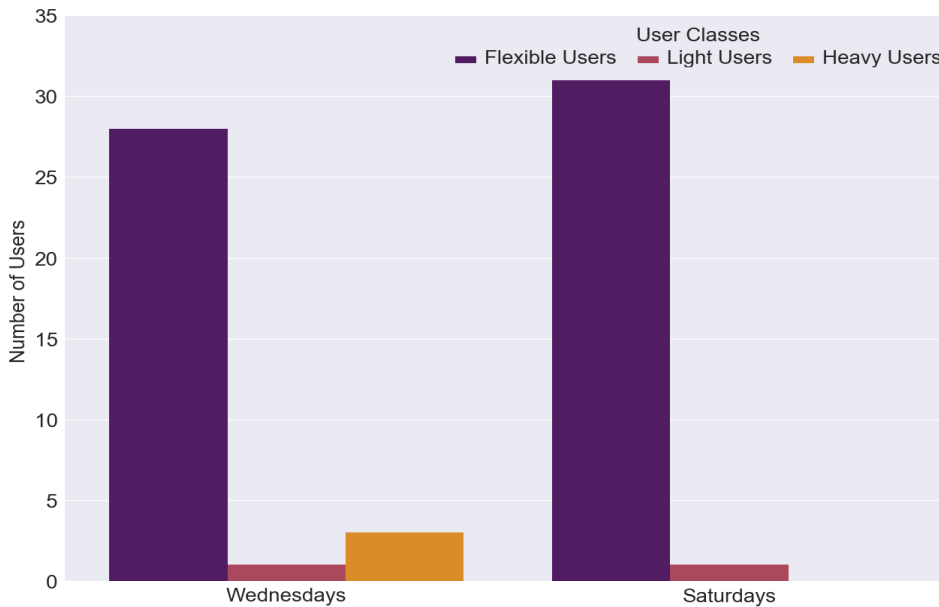


Figure 4.5: A PON port user distribution for Wednesdays and Saturdays

We conduct the comparative study between GM(1,1) and ARIMA model on the basis of a flexible user profile for a list of Wednesdays as a first step, then for all weekdays (from Monday to Friday) covered by the collected traffic traces as a second step.

In the following, we highlight for a flexible user, the real values of the assignment index to the heavy class for all Wednesdays in the supervised period, and the forecasted values while extending the dataset by 4 weeks to have a vision on the traffic tendency.

For ARIMA model, we visualize firstly our data and then, we tested several configurations using the AIC, ACF, and PACF explained before in section 4.2.1. The best results were obtained for ARIMA(1,0,0) and ARIMA(2,0,1). So, the dataset in the two cases does not require a differencing phase before the prediction. Fig.4.6, Fig.4.7, and Fig.4.8 show the different results obtained by using respectively, ARIMA(1,0,0), ARIMA(2,0,1), and GM(1,1). Tables 4.2, 4.3, and 4.4 give more details about the forecasting results as well as the values of the different metrics explained before.

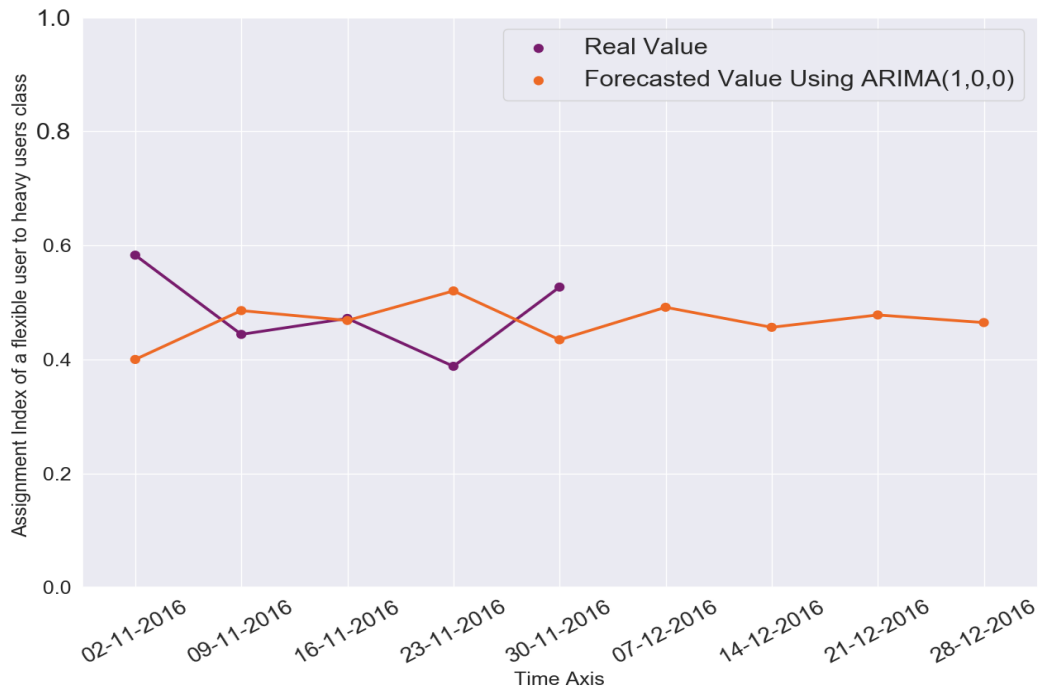


Figure 4.6: Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using ARIMA(1,0,0)

Table 4.2: Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using ARIMA(1,0,0)

Day	AI to Heavy Users Real Value	AI to Heavy Users Forecasted Value	Residual	Forecast Error (%)	Forecast Accuracy (%)
02-11-2016	0,583	0.4	0.183	31.39	68.61
09-11-2016	0,444	0.486	-0.042	9.46	90.54
16-11-2016	0,472	0.468	0.004	0.847	99.153
23-11-2016	0,388	0.520	-0.132	34.02	65.98
30-11-2016	0,527	0.435	0.092	17.46	82.54
07-12-2016	-	0.492	-	-	-
14-12-2016	-	0.456	-	-	-
21-12-2016	-	0.478	-	-	-
28-12-2016	-	0.465	-	-	-
Mean Residual (MR) :			0.021		
Mean Absolute Deviation (MAD) :			0.0906		
Tracking Signal (TS) :			1.159		

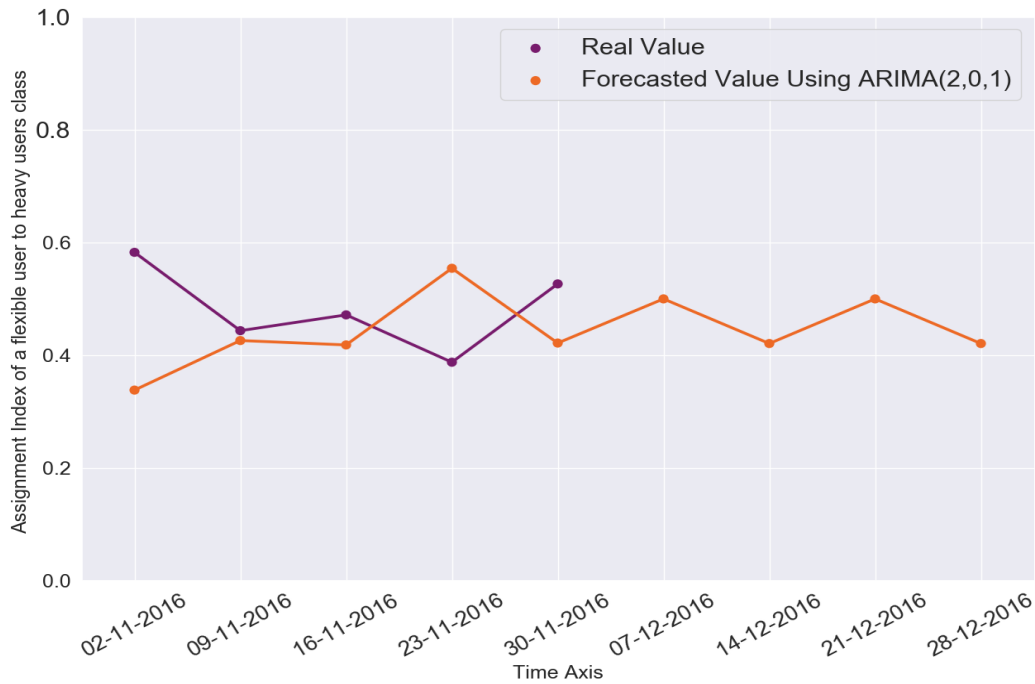


Figure 4.7: Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using ARIMA(2,0,1)

Table 4.3: Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using ARIMA(2,0,1)

Day	AI to Heavy Users Real Value	AI to Heavy Users Forecasted Value	Residual	Forecast Error (%)	Forecast Accuracy (%)
02-11-2016	0,583	0.339	0.244	41.85	58.15
09-11-2016	0,444	0.426	0.018	4.05	95.95
16-11-2016	0,472	0.419	0.053	11.23	88.77
23-11-2016	0,388	0.555	-0.167	43.04	56.96
30-11-2016	0,527	0.422	0.105	19.92	80.08
07-12-2016	-	0.5	-	-	-
14-12-2016	-	0.421	-	-	-
21-12-2016	-	0.5	-	-	-
28-12-2016	-	0.421	-	-	-
Mean Residual (MR) :			0.0506		
Mean Absolute Deviation (MAD) :			0.117		
Tracking Signal (TS) :			2.16		

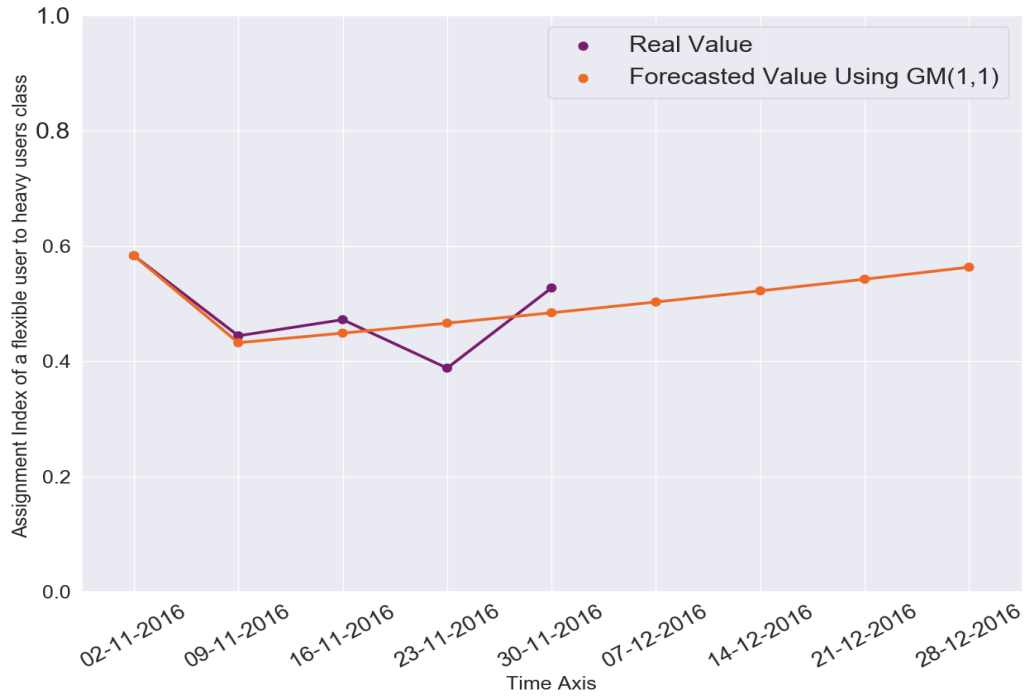


Figure 4.8: Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using GM(1,1)

Table 4.4: Assignment index (AI) of a flexible user to the heavy class: real and forecasted values using GM(1,1)

Day	AI to Heavy Users	AI to Heavy Users	Residual	Forecast	Forecast
	Real Value	Forecasted Value		Error (%)	Accuracy (%)
02-11-2016	0,583	0,583	0	0	100
09-11-2016	0,444	0,432	0,012	2,702	97,298
16-11-2016	0,472	0,448	0,024	5,084	94,916
23-11-2016	0,388	0,465	-0,077	19,845	80,155
30-11-2016	0,527	0,483	0,044	8,349	91,651
07-12-2016	-	0,502	-	-	-
14-12-2016	-	0,522	-	-	-
21-12-2016	-	0,542	-	-	-
28-12-2016	-	0,563	-	-	-
Mean Residual (MR) :			0,0006		
Mean Absolute Deviation (MAD) :			0,0314		
Tracking Signal (TS) :			0,095		

To simplify the comparison between both models, we recapitulate in table 4.5, the main forecasting metrics for each approach.

Table 4.5: Comparative table of forecasting models

	Mean FA (%)	Max FA (%)	MR	MAD	TS
ARIMA (1,0,0)	81,364	99.153	0.021	0.0906	1.159
ARIMA (2,0,1)	75,982	95.95	0.0506	0.117	2.16
GM(1,1)	92.804	97.298	0.0006	0.0314	0.095

By looking to the comparative table and the different obtained results, the different metrics used to evaluate the forecasting models show high performances of GM(1,1) compared to ARIMA model with its two versions. With a good forecast accuracy (the average is 92.804 %) and a very low MAD, GM(1,1) presents itself as a good candidate for our forecasting module when dealing with a limited number of supervised days. The tracking signal is generally used to decide if the forecasting model need or not to be reviewed. The smallest value for this parameter was obtained using GM(1,1) which brings confirmation of the good quality of this model. For ARIMA(1,0,0) and ARIMA(2,0,1), the values of MR, MAD, and TS despite being above those of GM(1,1), remain in the standards and can be considered as acceptable. However, the average obtained for the forecast accuracy (81.364% for ARIMA (1,0,0) and 75.982% for ARIMA(2,0,1)) is less than the value obtained by GM(1,1), which leaves doubt about the stability of this model using a small dataset.

Based on this analysis, GM(1,1) will be used to generate the new users' distribution. From table 4.4, the calculation of the average of the assignment indexes to the heavy class for the flexible user demonstrates that it increases from 0.4828 (using real values for the supervised days) to 0.505 (using real values for the supervised days and forecasted ones for the next 4 weeks). This qualifies the user as a heavy user instead of a flexible one. We performed the same approach for all flexible users belonging to the same PON port for a list of Wednesdays and Saturdays. The new users' classification is shown in Fig.4.9. Whereas before using the forecasting model, the rate of flexible users was 87,5% for Wednesdays and 96,87 % for Saturdays, the application of the GM(1,1) shows a more balanced distribution where the flexible users rate decreases to 68,75% for Wednesdays and 75% for Saturdays. By looking to the new users' distribution resulting from the application of the Grey model, it's clear that the additional bandwidth that can be estimated will be greater since the number of light users has increased both for Wednesdays and Saturdays. Additionally, the number of users who will benefit from the bandwidth supplement, i.e. heavy users, has also increased, which asserts that the use of the GM(1,1) looks essential and advantageous in a context of satisfying the majority of customers in our approach.

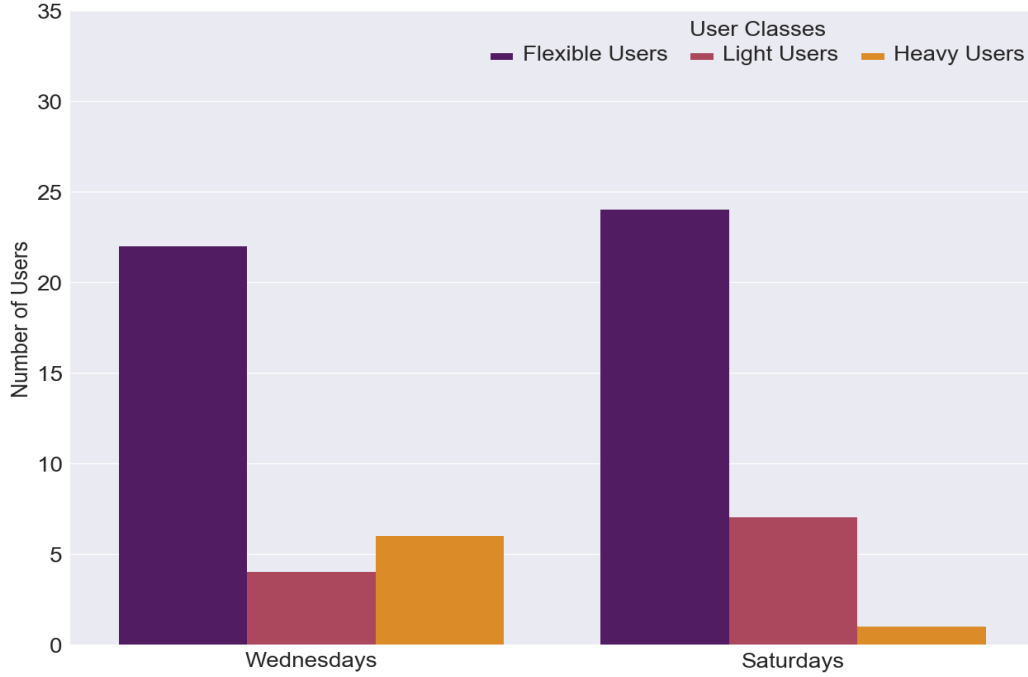


Figure 4.9: A PON port user distribution for Wednesdays And Saturdays Using GM(1,1)

Nevertheless, the evaluative study conducted was based only on five data points representing the historical assignment indexes for a list of five supervised Wednesdays. Thus, a question that can be asked is if the number of data points is greater (which is expected because our optimization mechanism collects data continuously), it may be that ARIMA gives better forecasting results. For this reason, we take advantage of our real dataset and we improve the comparative study between GM(1,1) and ARIMA model using all weekdays (from Monday to Friday) covered by the collected traffic traces instead of a list of specific day. Thus, in place of having only 5 data points, we can reach 23 data points (all weekdays between the 2nd of November 2016 and the 3rd of December 2016).

In the following, we highlight for the same flexible user studied before, the real values of the assignment index to the heavy class for all weekdays in the supervised period, and the forecasted values while extending the dataset by 5 days (to have a vision about user's behavior tendency). As it was performed before, ARIMA configurations which give the best results after the testing phase (section 4.2.1) are selected for the comparative study. Fig.4.10 and Fig.4.11 represent real and forecasted assignment indexes using respectively ARIMA(1,0,0) and ARIMA(2,0,1) (the best forecasting results were performed by these two configurations compared to other ARIMA configurations), while Fig.4.12 shows the results corresponding to the GM(1,1) model. We summarize in table 4.6 the evaluation results of forecasting metrics.

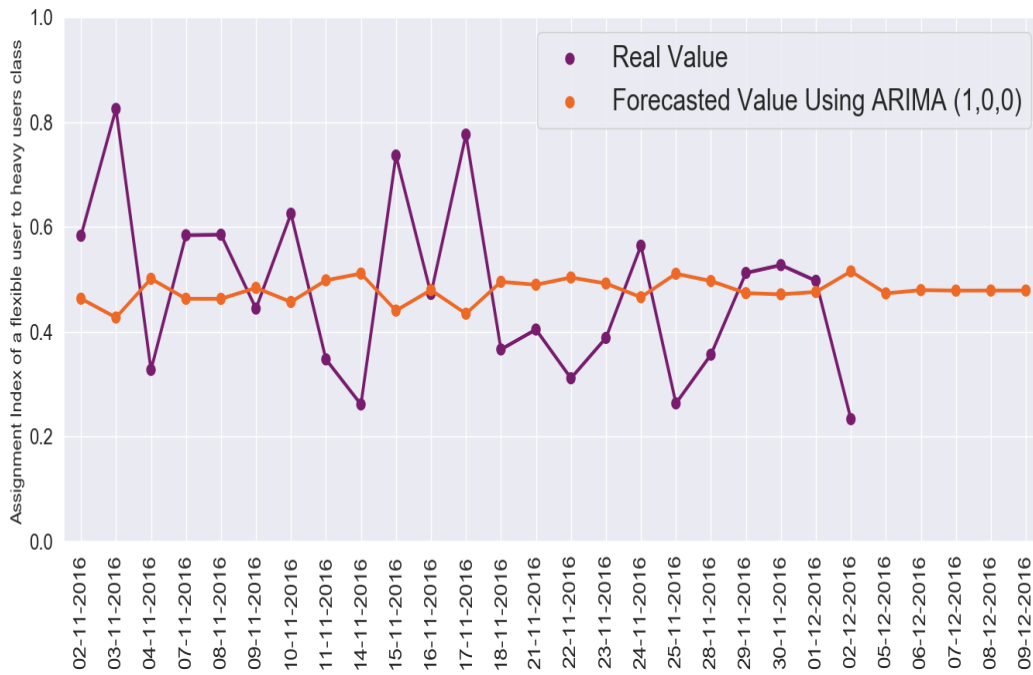


Figure 4.10: Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using ARIMA(1,0,0)

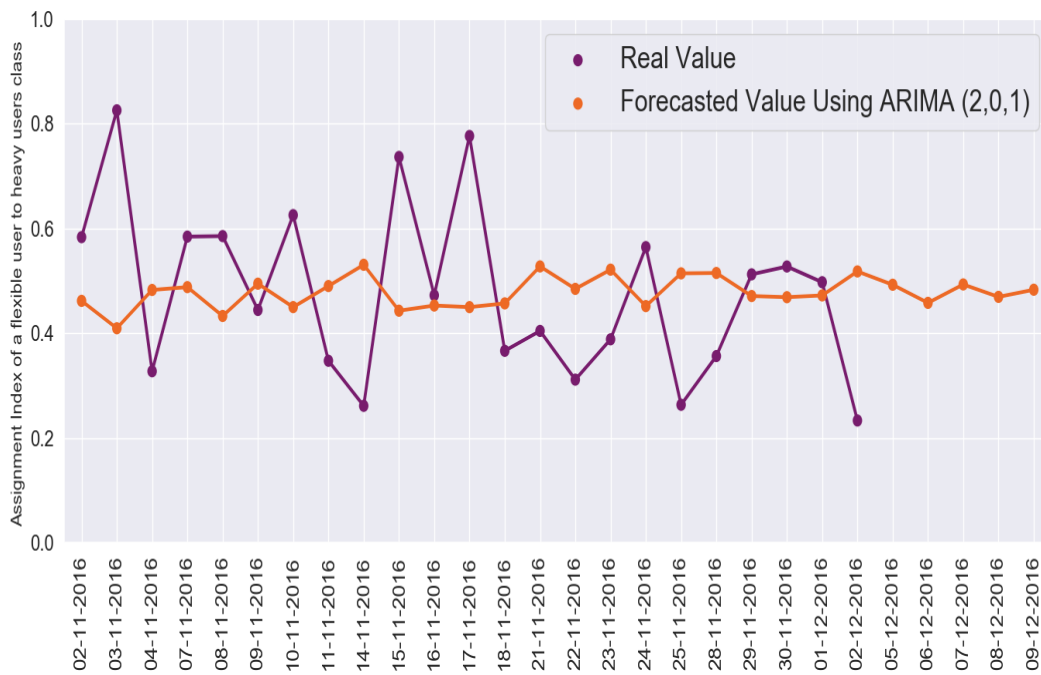


Figure 4.11: Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using ARIMA(2,0,1)

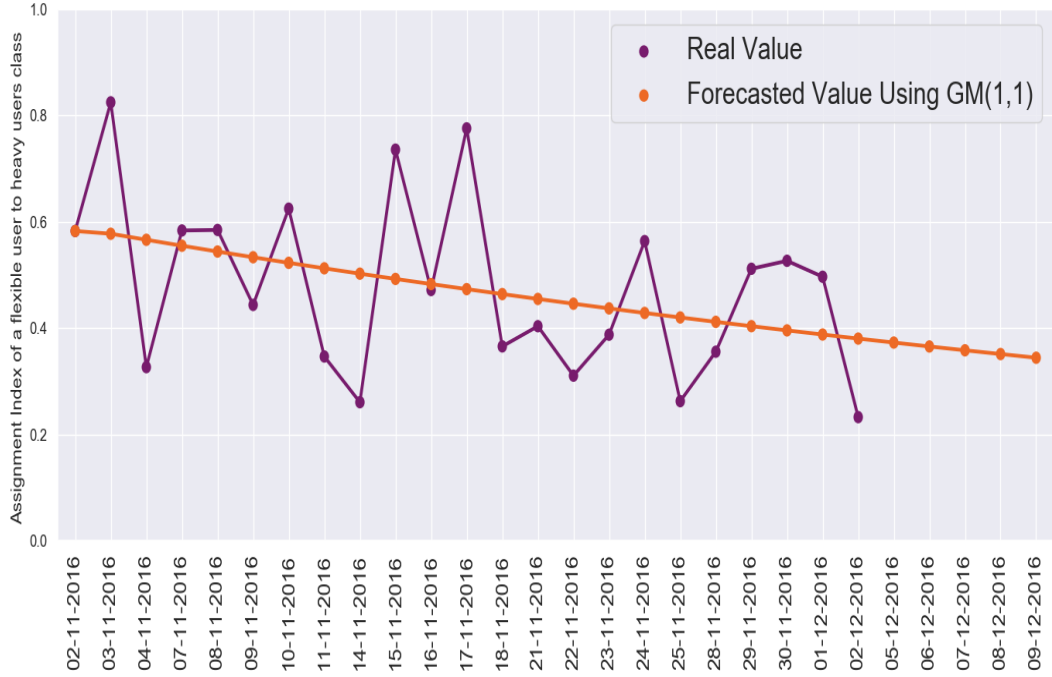


Figure 4.12: Forecasted and real assignment indexes of a flexible user to the heavy users class for all weekdays using GM(1,1)

Table 4.6: Comparative table of forecasting models for all weekdays

	Mean FA (%)	Max FA (%)	MR	MAD	TS
ARIMA (1,0,0)	63.34	98.43	-0.0009	0.155	-0.134
ARIMA (2,0,1)	62.72	95.86	0.0003	0.159	0.049
GM(1,1)	69.87	97.628	0.0001	0.125	0.0213

By looking to the comparative table, the different metrics used to evaluate the forecasting models show that GM(1,1) still outclasses ARIMA model with its two versions using a large number of assignment indexes. GM(1,1) presents a better mean forecast accuracy with a value of 69.87% and a lower MAD compared to ARIMA model. The mean residuals as well as the tracking signal used to decide if the forecasting model need or not to be reviewed present small values for both GM(1,1) and ARIMA with its two configurations. Generally speaking, GM(1,1) looks more efficient than ARIMA model using more real assignment indexes. The GM(1,1) aims by the end to give the tendency of a specific user (whether light or heavy) which met our initial objective and helps to assign a flexible user temporarily to one of the other two classes.

4.4.3 Discussion

The evaluative study conducted in this chapter was based in a first step only on five data points representing the historical assignment indexes for a list of five supervised Wednesdays. Then, it was extended using all weekdays (from Monday to Friday) covered by the collected traffic traces to test the studied forecasting models for larger dataset. In the two cases, GM(1,1) outperformed the tested ARIMA configurations in terms of forecasting metrics. It was shown that from the performed results, GM(1,1) gives a clearer idea about the upstream behavior trend of a flexible user (whether heavy or light). As in our approach, the objective of using a forecasting module is to anticipate the behavior of a flexible user in order to maximize the efficiency of the bandwidth usage in the network, we think that we can rely on GM(1,1) as the most suitable model for our forecasting module. Nonetheless, this shall not prevent from testing other models (e.g. ANNs) in the future; subject to having collected more traffic traces; and see if they can perform better results than GM(1,1).

4.5 Conclusion

In this chapter, a new approach for enhancing PON users' classification based on their traffic patterns has been proposed.

In the previous chapter, we have proposed a mechanism for reallocating SLA parameters in a PON network based on users' traffic behavior. This mechanism has as objective to optimize the upstream bandwidth allocation process without modifying the DBA algorithm itself. The idea was to classify PON users into 3 classes, heavy, light, and flexible and then, try to add an extra bandwidth to heavy users for a specific day period. The classification mechanism was designed based on clustering analysis and an assignment index calculation method. This mechanism is limited by the fact that the majority of users were assigned to the flexible class, which looks like an obstacle in our optimization approach. In this chapter, we enhanced the PON users' classification module by adding a forecasting module to learn more about flexible users traffic trends. Three forecasting approaches were studied while only two were evaluated due to dataset limitations.

The obtained results show that we can refer to the Grey forecasting theory (GM(1,1)) as the best candidate to predict the possible traffic behavior of flexible users in the future with the aim to have a more balanced distribution. We have showed the advantage of using this predictive approach to improve the final users' distribution which impacts directly the extra bandwidth estimation and the number of beneficiary customers.

This enhanced model proposed in this chapter cannot be yet considered as a final version before moving on to the whole assessment in terms of QoS. Indeed, it has to dynamically reconfigure the SLA parameters and simultaneously notify the network operator of the new

modifications. As the current management of the SLA parameters is under the responsibility of operation teams, the adjustment process supposed to be performed at different day periods may lead to slower and less efficient reactions as the management system becomes more and more complex. In general, in a context of network optimization, manual management should be avoided as much as possible in order to provision effective and efficient network services. In our case, it looks essential that the task of the dynamic reconfiguration of the SLA parameters table should be accomplished by the network itself without direct human intervention.

In the next chapter, we will treat this issue by proposing a novel architecture for our PON traffic management model based on the autonomic and cognitive networking fundamentals. SLA parameters will be automatically handled by self-decision elements and the automation of the different machine learning techniques allowing estimating the different users' traffic behaviors will be achieved through a set of cognitive process decision elements.

CHAPTER 5

A GANA ORIENTED ARCHITECTURE FOR SELF-OPTIMIZED PASSIVE OPTICAL NETWORK

Contents

5.1	Introduction	103
5.2	Related Work	104
5.2.1	Autonomic Network Approaches	106
5.2.2	Autonomic Network Models	107
5.2.3	Cognitive Networks: An Alternative Vision of Autonomic Networks	115
5.3	A Comparative of Autonomic models	116
5.4	GANA-Oriented Architecture for Self-Optimizing Passive Optical Network Upstream Resources Allocation	119
5.4.1	Proposed Model	119
5.4.2	Runtime Scenario	122
5.5	Conclusions and Future Work	124

5.1 Introduction

Ensuring the efficiency of the bandwidth allocation mechanism in PONs is a key challenge in the future for network operators. We have proposed in chapter III a new architecture for optimizing the upstream resources allocation mechanism (i.e. DBA) operation based only on the dynamic adjustment of some SLA parameters. This architecture represented a new leap forward in the field of the optimization of network mechanisms which are not under the control of network operators. The idea was very simple: if we cannot change the algorithm itself, why not change

the manner in which we manage its external parameters that are under the network operator's control (i.e. SLA parameters) in order to change the DBA overall operation?

To develop this idea, the overriding question was on which basis are we going to adjust the SLA parameters. In this context, we have proposed to estimate users' traffic patterns based on learning their daily life bandwidth usage. Using clustering analysis and an assignment index calculation module, three classes of users are then identified: heavy, light, and flexible. Then, based on the traffic profile of each user, heavy users will benefit from a bandwidth surplus which is determined based on an estimate of the bandwidth unused by light users at a specific day period. Finally, this surplus will be added to the heavy users' PIRs as the SLA parameters that represent the maximum bandwidth that can be provided. To enhance the learning mechanism, we have proposed in chapter IV an extended version of the model by adding a forecasting module to learn more about flexible users' behaviors and their possible future trend. This last version (as well as the first one) has to dynamically reconfigure the SLA parameters and simultaneously notify the network operator of the new modifications.

As the current management of SLA parameters is under the responsibility of operation teams, the adjustment process performed at different day periods may lead to slow and inefficient reactions as the management system becomes more and more complex. In general, in a context of network optimization, manual management should be avoided as much as possible in order to provision effective and efficient network services. In our case, it looks essential that the task of the dynamic reconfiguration of SLA parameters should be accomplished by the network itself without direct human intervention.

In this chapter, we propose a novel architecture for our PON traffic management model based on the autonomic/cognitive networking fundamentals. SLA parameters will be automatically handled by self-decision elements and the automation of the different machine learning techniques allowing estimating the different users' traffic behaviors will be achieved through a set of specific autonomic decision elements.

This chapter is structured as follows: Section 5.2 presents some work related to the field of autonomic and cognitive networks. A comparative study between the different studied models is provided in section 5.3. Section 5.4 presents the new version of our model inspired by the self-management properties. Finally, we conclude the chapter in section 5.5.

5.2 Related Work

The automation of the configuration management has become a priority for network operators. Manual network management whereby operation teams take care of the continuous observation, configuration and optimization may result in fault management with performance degradation and inefficient reactions.

The model that we have proposed in chapter III (presented in Fig.5.1) and enhanced in the previous chapter (presented in Fig.5.2) acts ultimately on some SLA parameters that are currently manually managed (i.e. under the responsibility of operation teams). It has to dynamically reconfigure these SLA parameters and simultaneously notify the network operator of the new modifications. Since the adjustment process is planned to be performed at different day periods, this may lead to slow and inefficient reactions as the management system becomes more and more complex.

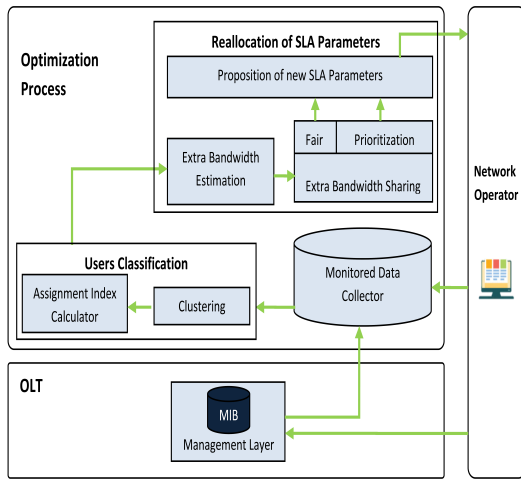


Figure 5.1: The proposed model for optimizing PON upstream resources: initial version

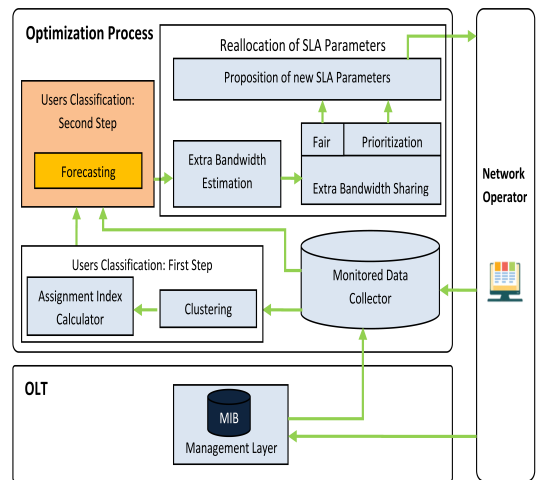


Figure 5.2: The proposed model for optimizing PON upstream resources: enhanced version

Thus, the proposed model should rely on an automated management functionality to improve its performance and achieve the goal towards a self-optimized PON. In this context, SLA parameters should be automatically handled by self-decision elements in a way that their dynamic reconfigurations should be accomplished by the network itself without direct human intervention. In parallel, the deployment and the coordination of the different machine learning techniques allowing estimating the different users' traffic behaviors should be also achieved through a set of specific decision elements.

Many advanced approaches and efficient tools exist in the literature with the aim to address the evoked limitations. A review of these approaches was provided in chapter II and showed that we can refer to the autonomic network management as the most advanced and efficient concept in the automation of network processes and the best solution that may be integrated in our model.

5.2.1 Autonomic Network Approaches

Inspired originally by the autonomic nervous system in the human body, autonomic networking [60] is a promising approach that enables devices to be automatically handled with the ability to adapt to their environment and leave the management tasks to the network itself respecting some rules and policies imposed by network operator. The autonomic concept can be integrated in PONs management systems through centralized, distributed, or hybrid solutions. Centralized solutions are adapted for slow control loops (slow reaction), distributed solutions using control loops that are embedded in the equipment are intended for fast reactions, whereas hybrid solutions use slow control loops to govern the fast ones through defined rules. With any of the three approaches, the challenge is that all network elements must converge towards a consistent system configuration with a minimum help from human administrator (human involvement can be limited to the control and the provision of operator's objectives including rules and policies). We give in the following, an overview of each approach.

5.2.1.1 Centralized Autonomic Approach

In a centralized architecture, the Decision making Element (DE) is located in a centralized server in the Operation & Management system while network elements (e.g, OLT, ONU) have only to execute the received commands. The main advantages of such approach are the capability and the flexibility to support and modify sophisticated and powerful strategies based on the correlation of different parameters, information and performance indicators collected from the equipment. Likewise, centralized approach is characterized by the ease of deployment since autonomic functions and decisional algorithms are located in a single place which facilitate the handling of a multi-vendor network [90]. Nevertheless, the centralized approach has also some drawbacks. First, a central controller makes the network less robust when connection to it is lost [18] and limits scalability, because it represents a bottleneck for processing and communication functions, especially in environments that require frequent configuration changes. Then, it requires a large bandwidth and a large data exchange in order to have available inputs for self-management functions. Finally, its processing is slow compared to other approaches since the control loop functions, i.e. Monitor-Analyze-Plan-Execute over a shared Knowledge (MAPE-K), are insured for all network elements which may entail delays in data recovery and processing [90][94].

5.2.1.2 Decentralized/Distributed Autonomic Approach

In a decentralized architecture, the DE is located at each network element which is able to operate fully autonomously, in a stand-alone fashion. Autonomic functions and algorithms are decoupled with the ability to make local decisions based on the knowledge of their network context (internal

status, information coming from EMS...) in agreement with the overall network purpose and goal and the quality constraints imposed by users [33][94]. This approach has the advantage of fast reaction since control loops are embedded within network elements, which reduces the time to retrieve information, process data, and take the appropriate decision. Another advantage is that a large bandwidth is not required given that data and information are collected locally [90]. However, the distributed approach may introduce a level of complexity in the network since the implementation of the different algorithms in network elements requires vendor specific mechanisms, which makes it difficult to modify strategies and functions at this level when it is needed [87].

5.2.1.3 Hybrid Autonomic Approach

The hybrid approach is a mixture of centralized and decentralized approaches. It pushes some functionality from a central system into the network element. Although that this approach gathers the advantages of the two other approaches like the fast control loops implemented in each equipment and the facility to integrate new strategies and new functions in the network, it has also its disadvantages since it should rely on vendor specific mechanisms in link with embedded DEs and an efficient system to coordinate autonomic functions that reside in each network element [90][94].

We summarize the main advantages and drawbacks of autonomic network approaches in table 5.1.

	Centralized Approach	Decentralized Approach	Hybrid Approach
Control loops' nature	Slow	Fast	Slow & Fast
Modification of strategies and functions	Flexible	Inflexible	Flexible
Ease of deployment	Yes	No	Average
Configuration Time	Long	Short	Short
Rate of exchanged data & requested bandwidth	Large	Small	Medium
Common DE algorithms for multi vendor solutions	Yes	No	Yes & No

Table 5.1: Comparison of autonomic approaches

It is clear that each autonomic approach has its advantages and drawbacks according to several criteria. The aim of this comparison is to understand the different autonomic models architectures that are presented in the following.

5.2.2 Autonomic Network Models

Due to the highly distributed, complex, and heterogeneous nature of current networking systems, no single common representation can be specified for autonomic models. Roughly, two classes

for autonomic networks can be distinguished, respectively clean slate models and evolutionary ones [70][101][104]. Clean slate models ignore the existing research findings in terms of previous autonomic modelling efforts and aim to redesign everything from scratch while starting with new set of ideas and eliminating existing commitments (e.g. [41]). Regarding evolutionary models, they enable to take into consideration the different research attempts with the aim to address their limits and converge towards a standardized solution. Thus, they enable new ideas to evolve while simultaneously emphasizing backwards compatibility with the existing approaches. We limit our research on the evolutionary models as this autonomic class is more advanced, more realistic at the implementation level from an operator vision compared to clean slate models, and could be applied within legacy systems. Broadly speaking, the idea behind evolutionary models is the IBM MAPE-K control loop [60], which relies on fundamental concepts and building blocks linked to monitoring, analysis, planning, execution, and knowledge phases. IBM MAPE-K model can be seen as a reference model, basis and fundamental architecture for all evolutionary models. It introduces an abstract architecture for autonomic networking that relies on fundamental concepts and building blocks with the aim to inject self-management properties into network management systems [60]. The IBM MAPE-K model is built based on two main building blocks: the autonomic manager and the managed resource. The managed resource represents any manageable hardware or software component in the network. It is managed by the mean of an autonomic manager which is responsible for collecting and analyzing data, formalizing and executing the remedy plans in order to correct the unwanted behavior according to the control loop objectives. The functions performed by the autonomic manager are implemented through the MAPE-K control loop. To define policies and operator needs and adjust objectives, IBM proposes a knowledge plan which is common and shared by MAPE-K functions [51][96]. We highlight the IBM model architecture in Fig.5.3.

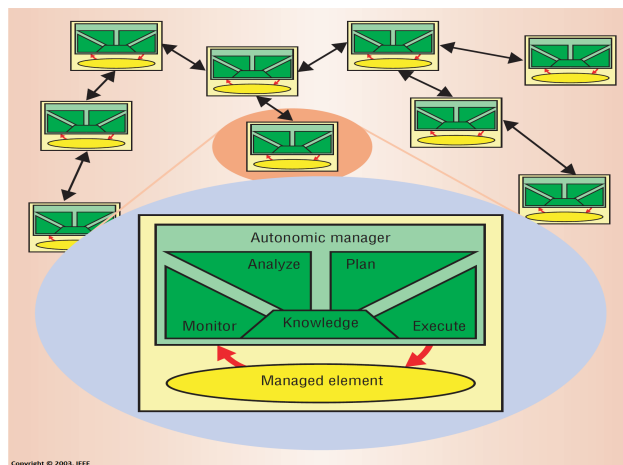


Figure 5.3: IBM MAPE-K reference model architecture [60]

Thanks to their advantages, evolutionary models were subject of many research projects in the recent years.

In [56], an autonomic network management architecture called Foundation, Observation, Comparison, Action, Learn, rEason (FOCALE) is proposed. FOCALE integrates distributed self-management algorithms and allows each managed entity to transform itself into an autonomic component by embedding an autonomic manager in all network elements [101][102]. The FOCALE main architecture is presented in Fig.5.4.

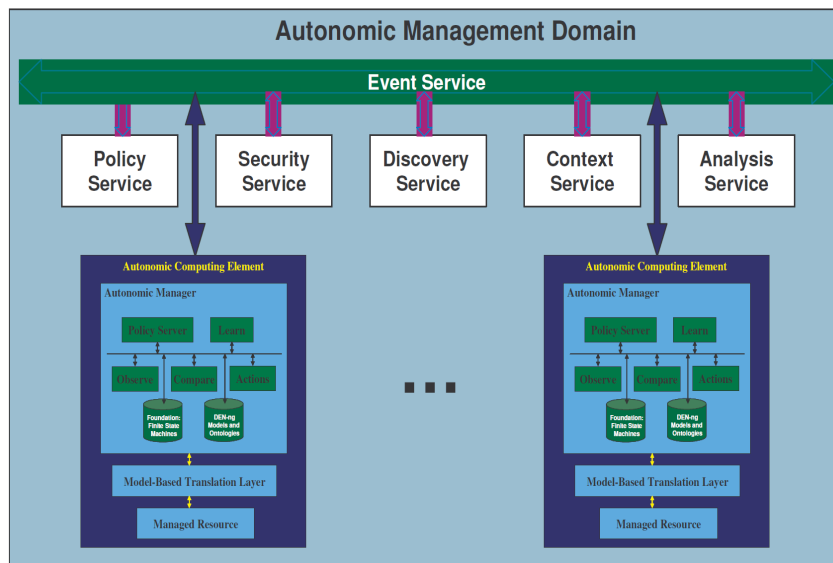


Figure 5.4: FOCALE architecture [102]

In addition to the main autonomic properties, FOCALE comes with three new concepts that differentiate it from the IBM MAPE-K reference model:

- Directory Enabled Network-next generation model (DEN-ng): it gathers information and ontologies that allow to build the foundation component in the FOCALE architecture. It relies on a set of Finite State Machines (FSMs) used to retrieve the current state of a managed entity and identify the transition state desired for adjusting its behavior.
- Model Based Translation Layer (MBTL): it represents the mediation layer between the autonomic manager and the managed resource. Its main role is to translate vendor-specific sensed data into a form that the autonomic manager can understand and in the same way, it translates the actions taken by the autonomic manager into vendor-specific commands.
- Policy server: it allows the autonomic element to interface to the outside world (e.g. humans and machines) and offers operators and service providers the opportunity to control

the action of each element and govern the network with provided objectives in order to converge the system to the best performing state.

In [23], an approach to autonomic and cognitive network management called Generic Architecture of Self-Organized Networks (GARSON) is proposed in the context of COMMUNE Celtic project. The GARSON architecture inherits the Autonomic Monitoring (AMON), the Autonomic Actuating (ACT), and the Knowledge Plane (KNOW) features from the reference model as shown in Fig.5.5.

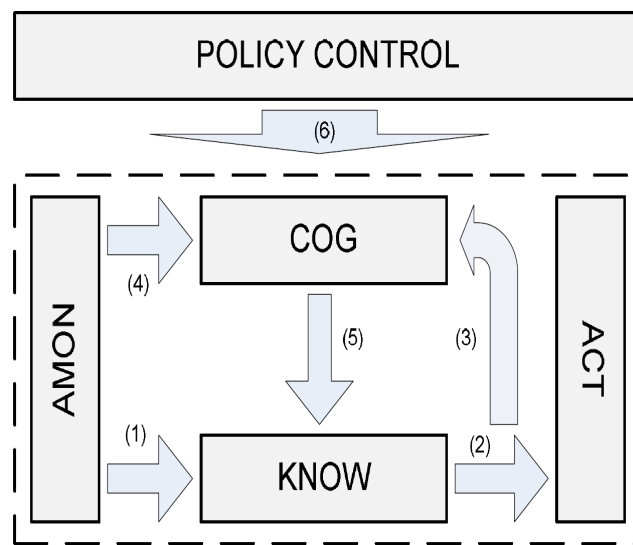


Figure 5.5: GARSON architecture [23]

The main novelty of GARSON consists in introducing a Cognitive Plane (COG) and a Policy Control Plane (POL):

- Cognitive Plane: its main role is to control the behavior of the Knowledge Plane and observe the taken decisions and the results obtained via the monitoring component. On this basis, an adaptation mechanism of the Knowledge Plane algorithms is provided [62].
- Policy Control Plane: this plane is proposed by GARSON to allow the operator to have an indirect impact on the network behavior in the event that a policy requirement is fulfilled and a specific action needs to be executed. Policy Control Plane may be also used to control the behavior of all architecture elements [62].

According to [23][62], the GARSON architecture is flexible in such a way that it adapts to any approach based on the behavior desired by the operator.

Authors in [103][104] propose a promising solution towards self-managed networks called Unified Management Framework (UMF). UMF was proposed as part of the ambitious and pragmatic UniverSelf research project funded by the 7th EU Framework Program. UMF is an extensible management framework with the goal to adopt and deploy autonomic solutions in the management of operators' networks. It allows moving research into practice by developing migration strategies that will consolidate existing initiatives and allows the convergence toward self-managed networks in the immediate and the mid-term [103]. The architecture of UMF illustrated in Fig.5.6 relies on a set of Functional Blocks (FBs) that can be seen as an extension to the MAPE-K actions with new aspects and more explanations. An emphasis was given to the embedding of the control and management intelligence in current management systems to avoid possible conflicts between different control loops which target the same parameters in the managed elements [21][39].

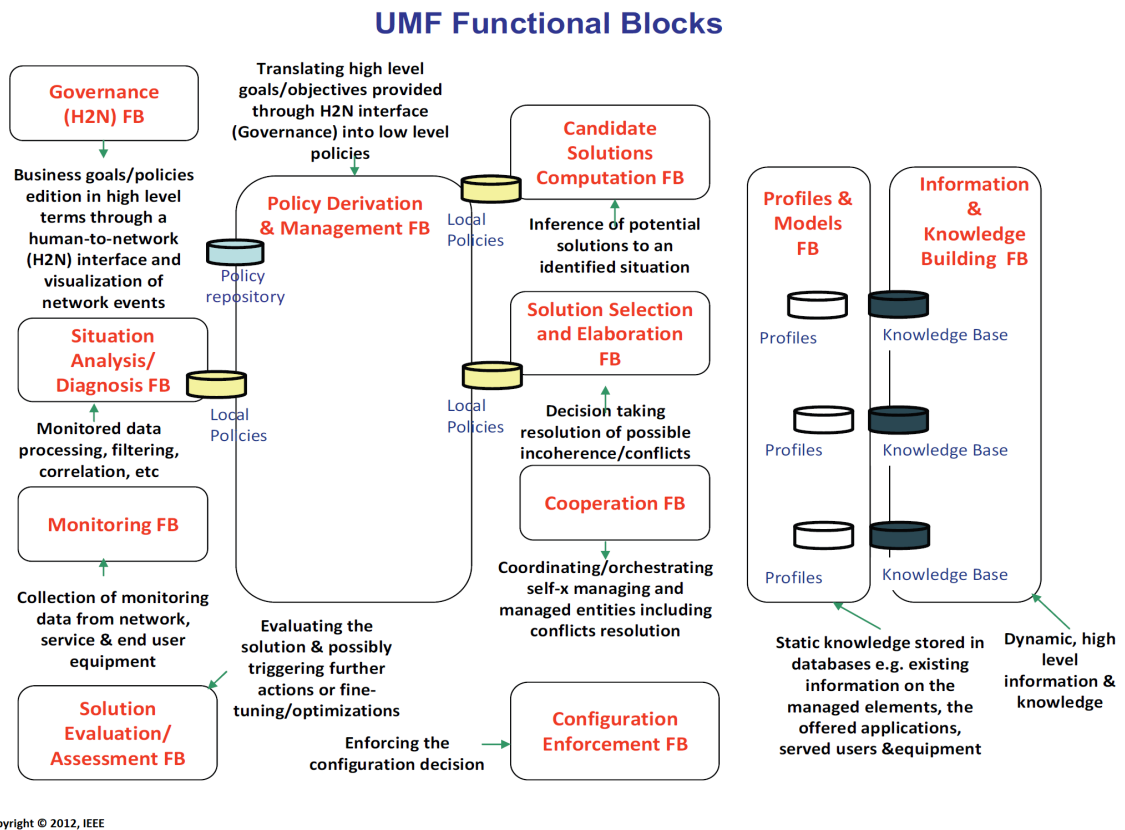


Figure 5.6: UMF functional blocks [103]

The main functionalities of each FB defined by the UMF are as follows [103]:

- Knowledge Management
 - Monitoring FB (M-FB): monitoring & collection of measurements regarding network services and customer related data
 - Situation Analysis/Diagnosis FB (SAD-FB): analyzing events in order to take appropriate decisions
 - Information and Knowledge Building FB (IKB-FB): making reference to all functions including storage, update, modification, and exploitation related to knowledge management
 - Profiles and Models FB (PM-FB): representing static knowledge accumulated in databases (i.e. FSMs)
- Intelligence
 - Candidate Solutions Computation FB (CSC-FB): discovering and identifying best solutions to be implemented
 - Solution Selection and Elaboration FB (SSE-FB): making decisions and resolving possible conflicts between different involved entities in the same domain
 - Cooperation FB (C-FB): coordinating self-managed entities and resolving possible conflicts between them
 - Solution Evaluation/Assessment FB (SEA-FB): evaluating solutions with the possibility to take appropriate actions for future decisions
- Enforcement
 - Configuration Enforcement FB (CE-FB): applying configuration decisions
- Governance
 - Governance FB (G-FB): defining the objectives of the network with high level policies, collecting, and visualizing the results through a human-to-network interface
 - Policy Derivation and Management FB (PDM-FB): translating of high level objectives into low level ones through human-to-network interface

The major advantage of the UMF architecture is that it takes into account the current management framework by defining the way in which the control and management intelligence can be embedded into the existing management chain (OSS/BSS-NMS-EMS-NE) [106]. According to [39], UMF has the ability to avoid possible conflicts between control loops, especially in the case of optimization algorithms which target the same parameters in the managed elements. To ensure the interworking among FBs of the UMF and the interfacing with network/service

operator side as well as with managed elements/resources, a mapping of the UMF FBs into the different entities of the system layout is required. An example of this mapping is proposed in [22].

In [19, 30, 108, 114], a Generic Autonomic Networking Architecture (GANA) is proposed. GANA is an evolvable architecture that prescribes the design and the operational principles of introducing autonomicity within devices and network architectures [6][18][114]. It enables autonomicity in the network by instrumenting managed entities with reusable generic specialized DEs. These DEs collaboratively work together to achieve the self-management functionalities such as self-healing, self-optimization, and self-discovery of network objectives, policies and resource capabilities specified by the operator [18]. As shown in Fig.5.7, GANA proposes four main levels of control loops respectively, protocol level, function level, and node level for fast control loops, and network level for slow control loops. The different DEs are orchestrated into these four levels in such a way that each DE manages one or more lower-level DEs associated with it. Thus, the lower-level DEs are considered as Managed Entities (MEs) of upper DEs. To define what the network is supposed to do and to offer services and rules to different network equipment, GANA introduces a knowledge plane at the network level which can be also seen as the higher cognitive level of GANA.

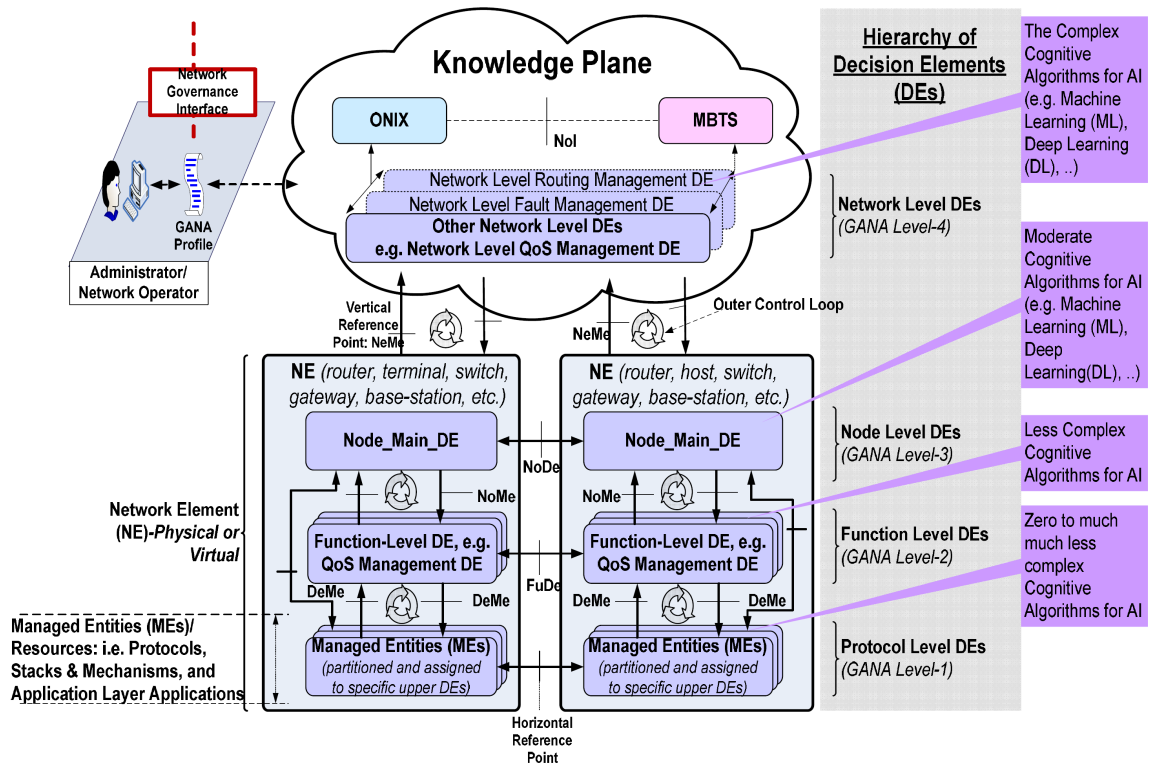


Figure 5.7: GANA architecture [108]

Referring to [6, 18, 69, 86], we can summarize the main functions of each level as follows:

- Protocol Level DE (lowest level/level 1): this DE represents protocols, services and any other mechanisms related to the MEs. In GANA specification, it is recommended to focus more on the three higher levels (from level 2 to level 4) when introducing the autonomic concept in network architectures.
- Function Level DE (level 2): this DE manages the DEs of the underlay resource layer by sending commands, objectives and policies to the MEs and receiving feedbacks in the form of monitoring information or other type of knowledge. It can be determined according to the functions required by the network element. The following DEs are among those specified by GANA for the function level: routing management-DE; forwarding management-DE; quality of service management-DE; mobility management-DE; monitoring management-DE, and service & application management-DE.
- Node Level DE (level 3): this DE ensures the coordination of the operation of network functions and the sharing of the knowledge with the other nodes. It gathers the aspects that cover and restrict the behavior of the whole network element. GANA specifies four DEs for this level: security management DE, fault management DE, auto configuration & discovery DE, and resilience & survivability DE.
- Network Level DE (highest level/level 4): this DE is one of the FBs of the Knowledge Plane next to the Overlay Network Information Exchange (ONIX) and the Model Based Translation Service (MBTS) modules.
 - Knowledge Plane: a system within the network responsible for defining what the network is supposed to do with the aim of offering services and advices to all network elements.
 - ONIX: a distributed scalable system of information servers with the aim to ensure an exchange of information and to enable network elements to query, find, and store knowledge.
 - MBTS: an intermediation layer between the knowledge plane and the network elements aiming at translating the information and the commands & responses.

Given that they are part of the knowledge plane, Network Level DEs allow to collect network polices imposed by service providers and translate and disseminate them to lower DEs, which make operators able to govern, control, and orchestrate the operation of its network. For this reason, Network Level DEs are designed to be centralized and so, the control-loops at this level operate within a slow time-scale (in contrast to the other levels DEs which are designated as fast control-loops).

GANA allows also the use of cognitive algorithms (e.g. machine learning, deep learning) in the DE logic [108]. These algorithms can be designed following a particular degree of cognition complexity as shown in Fig.5.7. The hierarchical paradigm of DEs (i.e. control-loops) facilitates the efficient design of large-scale complex autonomic networks due to the logical separation of complex decision making processes, the ability to resolve conflicts, and the inherent stability and scalability attributes [6]. The management of the operation of the whole network in GANA is ensured by the network administrator through a network governance interface responsible for validation, encapsulation and submission of high-Level network objectives, policies and configuration-data in the form of a GANA profile towards the Knowledge Plane [30].

5.2.3 Cognitive Networks: An Alternative Vision of Autonomic Networks

Along with autonomic networks, another concept that can be considered as very close with the same overall objectives is investigated in this work, namely “Cognitive Networks”. Referring to [72][74][118], a cognitive network can be defined as a combination of 3 main elements. The Monitoring Element which makes the network aware by providing the current conditions, the Software Adaptable Elements which make the network able to plan, decide, and modify its configuration and the Cognitive Process which learns from the previous decisions and the historical data in order to enhance future decision making. Thus, the structure of cognitive networks makes them able to be self-optimized while anticipating the probable changes in network behavior (i.e., proactive network). Thanks to their flexibility and their ability to be applied in various fields, cognitive networks have drawn the attention of many researchers in several areas. We limit our research on proposals related to optical networks.

Authors in [118] present a new cognitive architecture intended to optical networks called “COGNITION” with the aim to enhance their modelling. All network layers (from application to physical), devices, systems, protocols, mechanisms, etc., are supposed to take into consideration the different environmental conditions in order to learn, self-adapt, and perform more efficiently.

Neto *et al.* have proposed in [78][89] a cognitive architecture for broadband optical-wireless network. Regarding the increase in the number of connected devices and the desire for convergent access networks, the proposed approach takes advantage of Cognitive Radio (CR) and Radio over Fiber (RoF) concepts to deal with Dynamic Resource Allocation and Spectral Sensing in order to improve the network performance in terms of QoS metrics.

In [17], a cognitive technique for estimating traffic changes in Optical Wireless Network is highlighted. A probabilistic inference method is used to estimate node parameters such as link capacity and offered traffic load based on observations. The approach allows providing an optimal configuration over time and a dynamic topology reconfiguration faster than before.

Despite the fact that the concept of cognitive networks is very close to that of autonomic

ones and that both share some ingredients in common, some researchers do not consider them the same. Referring to [77], authors explain that the difference between autonomic and cognitive networks lies in how networks perform changes taking into consideration the monitoring knowledge i.e., adaptation approach, and do they take or not into consideration the learning mechanisms to converge towards an optimal solution, i.e. learning ability. The comparative study conducted in [77] shows that the majority of autonomic networks rely on policy-based approaches to perform the self-adaptation thanks to the possibility to modify at runtime the applied policies (using a human to network interface) while the network continues to operate ordinarily (without stopping). Regarding cognitive networks, learning mechanisms are the most used based on historical data and network experience acquired from previous situations in order to decide about the possible changes in network behavior. Thus, autonomic and cognitive networks start from a common base with similar goals. While cognitive networks focus more on learning mechanisms and shift more emphasis to the cognitive process, autonomic networks aim to give a particular attention to the decision-making processes hierarchy and their coordination to facilitate the design of large-scale complex networks.

To our way of thinking, both autonomic and cognitive networks as defined, seem interesting since the proposed approach with its two versions needs to learn regularly and should make decisions automatically in the most efficient manner with the minimum of external intervention. However, we do not think that the way in which autonomic and cognitive networks were defined may present an issue for selecting which concept to be implemented. We still think that cognitive networks are part of autonomic ones as it was proposed in [108] and differentiating them is a debate that should be avoided as it does not give any particular contribution to the domain. Another remark regarding this problem is that contributions related to cognitive networks are more dedicated to well-defined use cases, whereas those linked to autonomic networks are more generalizable and tend to converge towards a universal solution. This brings confirmation to what has been discussed before and let us focus more on autonomic models.

5.3 A Comparative of Autonomic models

We have presented in section 5.2.2 a summary of the best-known autonomic models in the research world. We present in the following the main advantages and drawbacks of each proposal in order to select which approach should be taken.

The FOCAL architecture has the advantage of focusing on the information model DEN-ng as well as the MBTL module which represents an important evolution compared to the reference model. Nevertheless, it does not define how to inject control loops in the autonomic network and how these control loops can interact.

The GARSON model has the advantage of proposing several scenarios for the integration of autonomic concept into network management system (FTTH network is one of these scenarios) and introducing a cognitive plane into the autonomic network management system. Despite these benefits, this model does not show how to convert data models and specific commands to each device (e.g. MBTL in FOCALÉ) and does not provide any information about how to resolve possible conflicts between different decisions taken. Moreover, GARSON shows that there is a coordination between the different devices and the different autonomic functions in the network [23], however it does not provide detailed specifications about this mechanism and how it works.

The UMF approach has the advantage of offering the operator flexibility and a freedom in the implementation of the autonomic functions. It aims to provide a management framework applicable to existing and future autonomic management architectures. Coordination and conflicts resolution are provided by this model, which gives the operator more trust and confidence. Translating high level policies and goal/objectives is provided through the use of human to network interfaces. A migration path from legacy to autonomic managed network is also proposed by UMF and there are attempts that show how to integrate autonomic functions into the management system (OSS-BSS-NMS-EMS-NE).

Regarding GANA, the major advantage of this model compared to the others is that it is standardized by the ETSI, which makes it easier for operators to implement the model. Its hierarchical control loops' architectural designing facilitates the task of implementing decision-oriented operational elements in different network equipment. GANA framework offers the possibility for operators to govern their network through Network-Level-DEs which enable network governance via providing appropriate interfaces with a human operator. These DEs allow to define policies, encapsulate them into GANA profiles and push these profiles for further distribution into the network elements. Compared to other alternatives, GANA can be considered as the most advanced model in terms of hierarchical control loops design. Its framework dissociates the different decision-oriented elements and facilitates the implementation task in network equipment while offering a governance interface to network operator for policy management and control reasons. The level of cognition complexity introduced in GANA represents a major advantage for this model as the complex cognitive algorithms are performed in high levels while low levels may include simple or less complicated algorithms. The weak point of GANA lies in the nature of its knowledge plane which is supposed to be centralized for all network equipment. This may conduct to slow control loops and a huge processing time with a requirement to have a strong coordination mechanism.

We summarize in table 5.2 the properties of the different models according to some criteria.

From table 5.2, we can conclude that GANA & UMF outperform the others models according to the selected criteria and the choice may be limited to these two models. Thus, we must still confine our research in order to select a single model to be applied. By making a link with our

	IBM MAPE-K	FOCALE	GARSON	UMF	GANA
Category of Architecture	Hybrid	Decentralized	Hybrid	Hybrid	Hybrid
Governance & Operator's Control	Yes	Yes	Yes	Yes	Yes
Network Convergence Mechanism	No	No	No	Yes	Yes
Translating mechanism	No	Yes	No	Yes	Yes
Coordination	No	No	No	Yes	Yes
Resolving conflicts caused by different decision elements	No	No	No	Yes	Yes

Table 5.2: Evaluation of autonomic models

work, our objective is to propose a self-optimized architecture (third version of our model) in order to improve the upstream resources allocation mechanism while SLA parameters will be self-adjusted and the classification mechanisms will be self-performed. So, we are dealing with the self-optimization property of autonomic networks. Self-optimization can be performed by both UMF and GANA. However, there is a criterion that may favor one model over the other in this context, which is simply the manner in which each model behaves regarding conflict issues. GANA and UMF, both propose methods to resolve possible conflicts in the case of self-optimization issues.

Conflict avoidance approaches proposed by UMF aim at resolving conflicts either on a time-scale basis or by empowering the optimization processes to solve the conflicts through direct interaction. In GANA, the hierarchical architecture based on four levels allows to resolve the possible conflicts by giving authority to each higher DE (e.g. the Node Level DE can resolve conflicts of the Function Level DE) to take the appropriate decision (e.g. assignment of priority to one optimization process where all optimization processes in the lower level do not make any decision before returning to the higher DE). The problem with the UMF approach is that it requires that all optimization processes are aware of all potentially conflicting processes that are active in the shared area. It also requires a high level of simultaneity for the execution of optimization processes, which puts additional constraints on the processes [9]. For GANA, the higher DE is the only one that must be aware for possible conflicts in the lower level, which makes it easier than UMF in the implementation phase. Thus, we can affirm that GANA outperforms UMF in terms of conflicts avoidance and ease of implementation in a case of self-optimization issue and can be the most suitable solution to be selected for our model.

In that case, we propose in the following section, a GANA oriented architecture for self-optimized PON inspired by the main concepts of GANA autonomic model. We try to take advantage of the advancement in research efforts in the autonomic area in order to converge towards smarter and proactive passive optical network. Thus, we inherit the autonomic decision elements hierarchy from the GANA model based on its advantages discussed before.

5.4 GANA-Oriented Architecture for Self-Optimizing Passive Optical Network Upstream Resources Allocation

5.4.1 Proposed Model

To achieve the main objective discussed previously, we present in Fig.5.8 a GANA-oriented architecture for self-optimizing passive optical network upstream resources allocation.

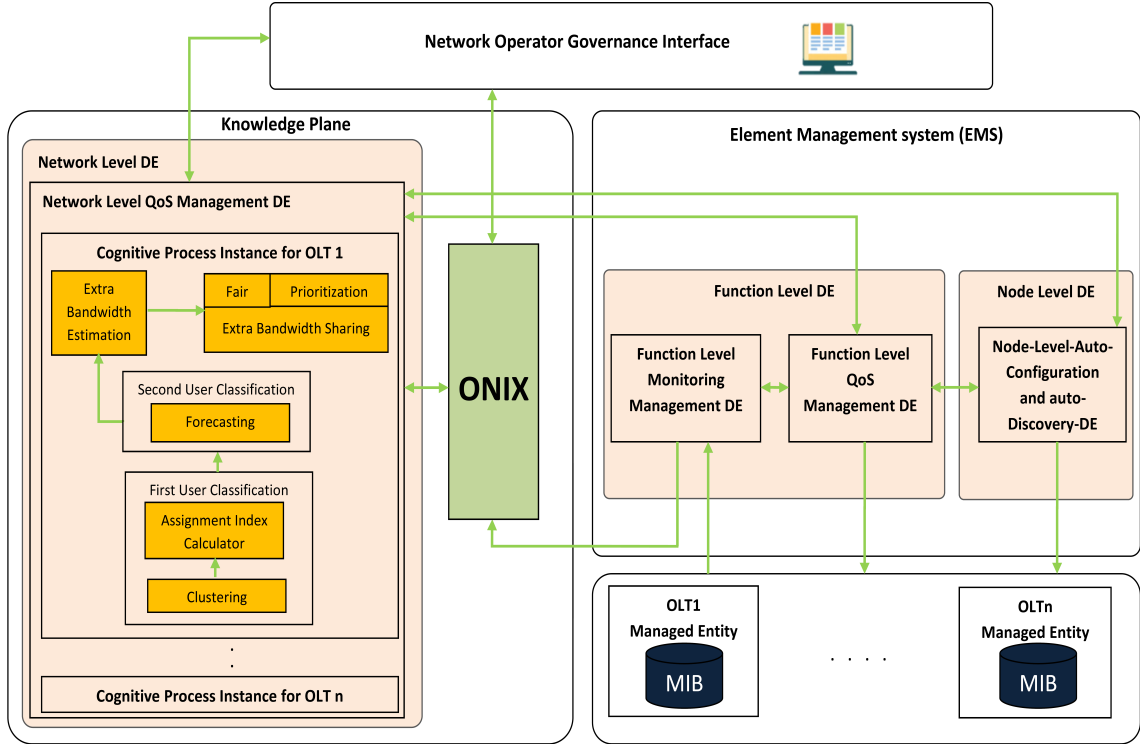


Figure 5.8: GANA-oriented architecture for self-optimizing passive optical network upstream resources allocation

The design of the architecture is inspired by the GANA autonomic architecture. Several DEs ensure the autonomic management of the optimization process. The whole architecture is controlled by a Network Operator Governance Interface that allows injecting policies, modifying parameters, controlling and stopping, if needed, the self-optimization process.

The core element of the proposed architecture is the cognitive process which takes into account users' historical transmission data (updated regularly) to optimize the upstream bandwidth allocation by determining the possible expected online users' behavior for a specific day period. The cognitive process can be defined as a set of modules relying on machine learning techniques as well as extra bandwidth estimation and sharing. The hierarchy of the different cognitive process modules is defined in such a way that the outputs of each module are used as

inputs for another module. Thus, the interaction between the different modules allows converging by the end to an optimal solution that the network should take into account.

The model presented in Fig.5.8 shows where the cognitive process is supposed to be executed, i.e. the Network Level QoS Management DE. This choice is justified by the fact that the cognitive process is performed only once per week upon the collection of new traffic information, thus, it can be classed as a slow control loop that needs much more time for processing compared to other DEs. The selection of the Network Level QoS Management DE above all other Network Level DEs is supported by the description of this DE proposed in [108], which shows that all logics and algorithms related to QoS guarantee and improvement should be provided by the Network Level QoS Management DE.

In this context, the cognitive process should be performed at the Network Level QoS Management DE in order to provision new configurations to fast control loop DEs (i.e. levels 3,2, and 1) which in turn, apply the corresponding decisions when the scheduled time arrives. In general, the implementation of the cognitive process is possible through a centralized, distributed or hybrid approach. While the cognitive process is expected to run as a single instance for all OLTs that are under the control of the EMS (remotely connected to all OLTs and responsible for their management) in a centralized approach, several instances for each OLT are supposed to be created in distributed and hybrid approaches. The difference between the two latter lies in the modelling of the data repository which will be centralized for all cognitive processes in the hybrid solution whereas it's the opposite for the distributed one (each cognitive process will have its own data repository). In our approach, we try to find a compromise between the processing and configuration time, the rate of exchanged data, and the ease of deployment. Thus, we opt for the hybrid approach where each OLT will be associated with its Cognitive Process Instance (CPI) and the data repository will be centralized for all CPIs. In GANA terminology, the data repository refers to the so called ONIX.

The ONIX is used in our architecture to store and provide knowledge and information related to PON users profiles and traffic behavior (i.e. dissemination). Thus, implemented DEs can retrieve the needed data during their operations from the ONIX system whenever they want. The results of the DEs processes are published via reports into the ONIX. This allows the Network Operator Governance Interface to access, view, and analyze the generated reports in order to govern the behavior of the overall network and take the appropriate decision (e.g. stopping the self-optimization process at the Network Level DE) if needed. The Network Operator Governance Interface serves also to notify the ONIX system whenever a PON user' profile is created or deleted. So, the Network Level QoS Management DE should take this into consideration in order to adjust users' classification. For the first implementation, the historical data is provided to the ONIX through the OSS as a result of traffic data collections made in advance by the network operator using probes and other collection tools. Once the optimization mechanism is

running, historical data can be updated using a Monitoring Management DE, representing one of the GANA Function Level DEs, that connects to the OLTs MIBs in order to get information about users upstream transmission data using SNMP GET requests. Thus, a connection between the ONIX and the Function Level Monitoring Management DE is assured to allow CPIs taking advantage of the new collected data. The information to be gathered and exchanged within the ONIX system is linked to users' bandwidth usage, e.g. number of bytes sent and number of packets sent, allowing to have an idea about users bitrates.

For accuracy reasons, the frequency at which the information changes over time is supposed to be fast. Intervals around 1 second should be avoided for efficiency purposes. In the designed architecture, we propose a 10 seconds time interval for updating information in the ONIX (i.e. the Function Level Monitoring Management DE sends the results of SNMP GET requests to the ONIX each 10 seconds while CPIs perform the calculation of the average upstream bitrates using a parameterizable time interval (e.g. 5min as used in in this work)).

For the sake of more control and flexibility, the Network Level QoS Management DE is supposed to handle the different CPIs and ensure their interactions with the other modules. Once CPIs deliver their results, the Network Level QoS Management DE will serve to notify the Network Operator Governance Interface that new decisions were made (it's just a notification, the access to the parameters is possible only via the ONIX). These decisions (i.e. new SLA parameters) are stored in the ONIX and then, sent to the Function Level QoS Management DE as they impact directly the QoS and the QoE in the network. The Function Level QoS Management DE is responsible for the implementation of the decisions (i.e. modification of SLA parameters) taken by the CPIs at the Network Level QoS Management DE using SNMP SET requests acting directly on OLTs MIBs.

As these decisions are performed based on an estimation and a prediction of users' traffic behavior, we can fall in some cases where users change their behavior compared to what has been predicted (reported by the Function Level Monitoring Management DE). Thus, to avoid any kind of QoS degradation in the network, an Auto-Configuration and Auto-Discovery DE that was proposed in GANA as a Node Level DE (i.e. acts directly on the node) is used in our model. This DE aims to ensure a coordination with the Function Level QoS Management DE in order to intervene whenever a problem regarding overall OLTs performances is identified. This intervention is limited to a quick return to the original settings (original SLA parameters) and a sending of a notification to the Network Level QoS Management DE.

To conclude, the association between the concept of machine learning techniques applied to the analysis of users behavior and GANA model adapted to our model offered both, an efficient design of the autonomic hierarchy due to the logical separation of complex decision making processes and a continuous learning mechanism that allows the convergence towards an optimal

solution with the evolution of time.

5.4.2 Runtime Scenario

To demonstrate how the model works, the runtime scenario presented in the following (Fig.5.9) explains how to ensure the different model tasks.

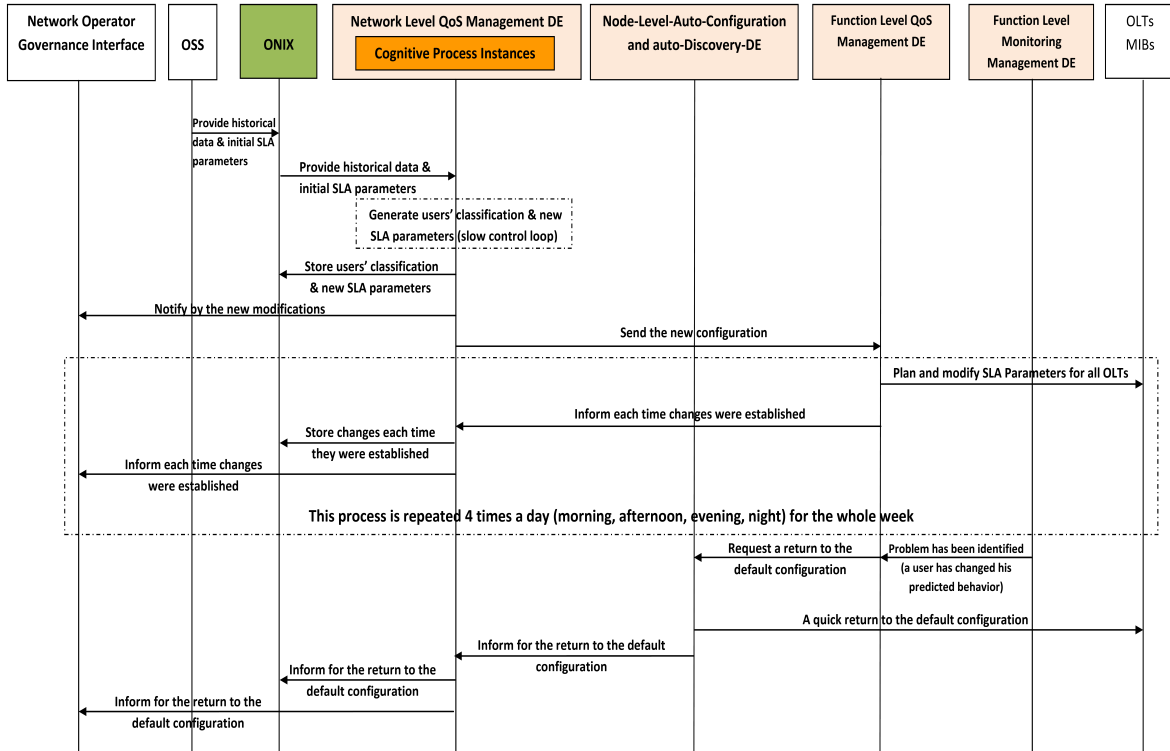


Figure 5.9: Runtime scenario

To allow the optimization process to operate for the first time, The first task is then to store historical data as well as the initial SLA parameters in the ONIX system by providing a connection to the OSS. The historical data as well as the initial SLA parameters will serve the CPIs to propose optimized SLA parameters (more specifically PIRs) for each day period for all PON users for the duration of one week. The Network Level QoS Management DE is responsible for checking the availability of data at the ONIX level and sharing it among the different CPIs. Once CPIs deliver their results, they will inform the Network Level QoS Management DE which in turn informs the network operator by the new modifications and stores the new SLA parameters as well as users classification and their assignment indexes in a specific table at the ONIX level.

The hierarchy of the different DEs inspired by GANA and designed in the architecture shows that the results of the optimization process should be sent from the Network Level QoS

Management DE towards the Function Level DE and more specifically, the QoS Management DE. The latter is expected to store all possible actions into specific files while each file includes the modifications that need to be established in a day period time-scale (i.e. 4 files a day, 28 files a week). These files are classed by time priority and used by the Function Level QoS Management DE to plan and execute the different modifications during the week. Then, after each day period, the Function Level QoS Management DE sends confirmation to the Network Level QoS Management DE that changes were taken into account and new SLA parameters were customized at the OLTs MIBs level.

In parallel to all these tasks, the update of information related to users traffic behavior is ensured continuously during the week. Indeed, the Monitoring Management DE is expected to request the MIB of each OLT for data transmission linked to each user every 10 seconds. Then, these data are transferred to the ONIX to be stored for a further use by CPIs as shown in Fig.5.10.

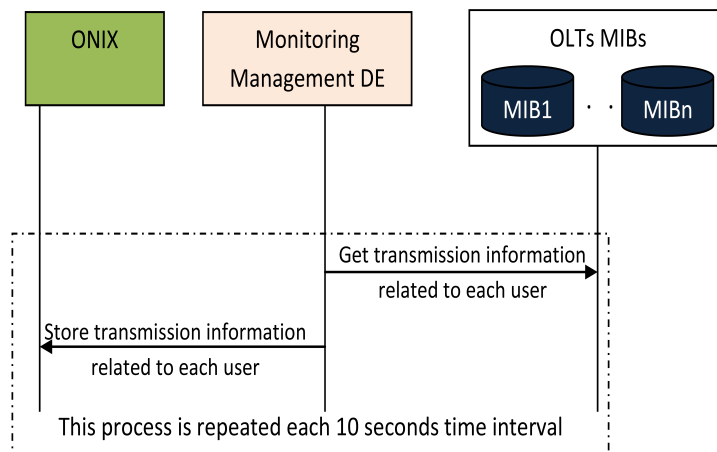


Figure 5.10: Continuous monitoring phase

By the end of the week (precisely, Sunday at midnight), the optimization process is re-launched using the new data collected. An update of assignment indexes as well as users classification is then performed, and new SLA parameters are generated for the next week.

It should be noted that network administrators may decide at any time to switch off the optimization process if any issue was observed. Also, if the Function Level Monitoring Management DE observes that there is a change in users behaviors compared to what has been predicted, a report is sent to the Function Level QoS Management DE. In the two cases, a quick return to the default configuration via the Node Level Auto-Configuration and Auto-Discovery DE is forced.

5.5 Conclusions and Future Work

A GANA oriented architecture for the self-optimization of passive optical network is presented in this chapter. The concept is to improve the modelling of the optimization mechanism proposed in the previous chapter which allows to enhance the upstream bandwidth usage efficiency based on the dynamic adjustment of some SLA parameters without modifying the DBA mechanism. Autonomic and cognitive networking fundamentals were used as a reference to automatically handle the parameters configuration.

At this stage, we can affirm that the model version proposed in this chapter meets the thesis objectives towards a more intelligent passive optical network and can be considered as adequate to be smoothly implemented within existing/legacy network.

In the next chapter, we proceed to an evaluation phase where we test our optimization mechanism for some PON scenarios.

CHAPTER 6

USE CASES - EVALUATION OF THE SELF OPTIMIZED PASSIVE OPTICAL NETWORK ARCHITECTURE

Contents

6.1	Introduction	125
6.2	Evaluation Environment	126
6.3	Demonstration Use Case	126
6.3.1	Use Case Objective	126
6.3.2	Experimental Scenario	126
6.3.3	QoS Assessment	129
6.4	Traffic Usage Variation Use Case	132
6.4.1	Use Case Objective	132
6.4.2	Experimental Scenario & QoS Assessment	132
6.4.3	Discussion	138
6.5	Conclusion	139

6.1 Introduction

To illustrate the functioning of our model and motivate the usefulness of our optimization mechanism in improving the upstream bandwidth usage efficiency, we proceed in this chapter to an evaluation phase in order to assess the QoS performances. Two use cases are considered to demonstrate the applicability of the proposed methodology: (i) a demonstration use case based on a real network scenario inspired by a real PON infrastructure using collected traffic traces as historical data. (ii) A traffic usage variation use case where different bandwidth usage scenarios

are considered (e.g. low bandwidth usage, average bandwidth usage, high bandwidth usage). The efficiency of the architecture is validated through a set of individual evaluation criteria.

This chapter is organized as follows: section 6.2 describes the evaluation environment used to assess the two experimental scenarios. Section 6.3 presents the first evaluation scenario which can be seen as a demonstration scenario showing the applicability and the efficiency of our optimization mechanism. Section 6.4 shows the second evaluation scenario with different configurations enabling to test the optimization mechanism in many cases that may occur in the PON. Section 6.5 concludes the chapter.

6.2 Evaluation Environment

To assess the two experimental scenarios presented in this chapter before and after applying the optimization approach, we conducted our tests within a real GPON access network used by Orange Labs France for test purposes. In accordance with the GPON standard specifications, we consider a 1.25 Gbit/s upstream and 2.5 Gbit/s downstream bandwidth capacities. For traffic generation, we injected traffic loads in the PON using the traffic generation tools provided by the Spirent Test Center testbed platform [100][111]. In order to have a realistic reasoning, these traffic loads were based on the traffic profile of each ONU. According to the learning study that was conducted, light and flexible users do not ask for upstream data transmission beyond their SLA constraints. To show the impact of our optimization approach, we expect that heavy users require more data transmission compared to what they are allowed to send. Their traffic loads are supposed to exceed their PIRs. Traffic loads are varied every 5 minutes for all ONUs. In order to properly represent the results, we have limited the evaluation time to one hour.

6.3 Demonstration Use Case

6.3.1 Use Case Objective

The demonstration use case is considered as a validation case study of our optimization approach. It aims to show that learning the daily bandwidth usage of different users and performing a self-adjustment of some SLA parameters based on their traffic patterns can be very advantageous in terms of improving upstream bitrates and bandwidth usage efficiency.

6.3.2 Experimental Scenario

To be more realistic and to avoid starting from scratch, we decided to conduct the experiment on the basis of traffic traces that we have already collected before. To remind, these traffic traces captured on an operational FTTH GPON network of Orange France, were collected for 3447 ONUs over a one month period between the 2nd of November and the 3rd of December 2016.

They will serve us as a database for the classification of our customers. However, as we cannot perform tests for all ONUs since we evaluate our approach in a laboratory environment, we select a representative PON port on which we will test our approach and assess the QoS performances. This port connects 12 subscribers with the same features. The classification of the 12 ONUs is suggested to be performed per day period and per weekday due to the fact that customers' traffic patterns are linked to their daily life and it is difficult to find similarities between each day period traffic profile. To explain more, based on the traffic traces at our disposal, we select for example the day period when the majority of customers are online, theoretically between 9p.m and 12a.m. Then, we select a weekday such as Wednesday for example and we perform our classification study based on the list of Wednesdays covered by the traffic traces. So in final, the result will be the traffic trend of each subscriber based on his mean upstream bitrates for supervised Wednesdays between 9p.m and 12a.m (i.e. heavy, light or flexible). Taking into account these results, our optimization mechanism adjusts for next Wednesdays the PIRs of heavy users between 9p.m and 12a.m. So, each one will benefit from a bandwidth surplus calculated on the basis of the estimation of the bandwidth unused by light users. Simultaneously, light and flexible users will not feel any degradation in terms of QoS since their SLA parameters remain untouched. So to conclude, what will happen in the test phase is supposed to be performed for Wednesdays from 9p.m to 12a.m in a real traffic scenario using our optimization mechanism.

We start our experimental study by highlighting firstly the upstream SLA parameters of each ONU in table 6.1. The CIR refers to guaranteed bandwidth provided by the network while the PIR represents the maximum bandwidth that can be allocated.

Table 6.1: ONUs upstream SLA parameters

ONU	Traffic Class	CIR (Kbit/s)	PIR (Mbit/s)
ONU1	Heavy	512	100
ONU2	Light	512	100
ONU3	Heavy	512	100
ONU4	Heavy	512	100
ONU5	Light	512	100
ONU6	Flexible	512	100
ONU7	Flexible	512	100
ONU8	Flexible	512	100
ONU9	Flexible	512	100
ONU10	Flexible	512	100
ONU11	Flexible	512	100
ONU12	Flexible	512	100

Fig.6.1 illustrates the classification of the 12 ONUs following the learning mechanisms explained in chapters III and IV.

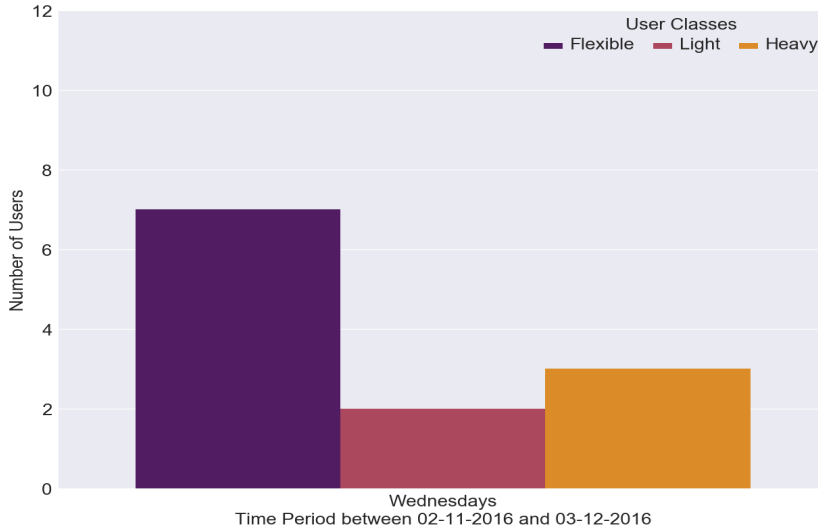


Figure 6.1: Users classification

According to this classification, table 6.2 shows the results of the calculation of the estimated bandwidth used by light users on average (more details about the estimation algorithm can be found in chapter III, section 3.3.2).

Table 6.2: Estimation of the bandwidth used by light users

ONU	Average Bandwidth Used (Mbit/s)
ONU2	1.512
ONU5	0.8

To share the unused bandwidth among heavy users and to guarantee that flexible and light users will not feel any degradation in terms of QoS, the following criteria are taken into consideration:

- Flexible and light users’ PIRs remain untouched.
- In the case where a light user changes his behavior (to heavy or flexible), if there is an available bandwidth, his request will be established. If not (i.e. saturated network), the usual operation of the DBA algorithm allows for a load balancing where all ONUs will have a percentage of their PIRs in a manner that all allocations cannot exceed the total upstream capacity.
- The PIR represents the maximum bandwidth that can be allocated to each user. It can be defined as a best effort traffic rate. Thus, cases where the sum of all users’ PIRs exceeds

the total upstream capacity (e.g. more than 12 users with a PIR=100 Mbit/s while the PON upstream bandwidth capacity is 1.25 Gbit/s) cannot create any problems insofar as the DBA does not allow exceeding the total upstream capacity (i.e. PIR shaping).

The calculation of the extra bandwidth available to be reallocated to heavy users as well as η (the multiplicative factor of heavy users PIRs), and α (the percentage increase in PIR) are performed following the algorithm 2 presented in chapter III. The results are shown in table 6.3.

Extra bandwidth	197.688 Mbit/s
η	1.65896
α	65.896%

Table 6.3: Extra bandwidth, η , and α values in a demonstration use case scenario

Therefore, each heavy user will have a 65.896% increase in his PIR. We illustrate the new SLA parameters of the different users in table 6.4.

Table 6.4: ONUs new upstream SLA parameters

ONU	Traffic Class	CIR (Kbit/s)	Optimized PIR (Mbit/s)
ONU1	Heavy	512	165.896
ONU2	Light	512	100
ONU3	Heavy	512	165.896
ONU4	Heavy	512	165.896
ONU5	Light	512	100
ONU6	Flexible	512	100
ONU7	Flexible	512	100
ONU8	Flexible	512	100
ONU9	Flexible	512	100
ONU10	Flexible	512	100
ONU11	Flexible	512	100
ONU12	Flexible	512	100

6.3.3 QoS Assessment

The main objective of our optimization approach is to improve the allocation of the available upstream resources by estimating the bandwidth unused by light users and share it among heavy ones. Thus, the first investigated parameter in this work is the upstream bitrate allocated by the DBA to each ONU. In general, this parameter doesn't raise any problem for light or flexible

users. However, it may not satisfy heavy users' needs. Fig.6.2 and Fig.6.3 illustrate the upstream bitrates (calculated on the basis of the average of the instantaneous upstream bitrates by one minute interval at the OLT level) that were allocated to the 12 ONUs depending on their traffic demands before and after applying our optimization mechanism between 9pm and 10pm (we limit the evaluation time to an hour for better representation)

The main observation that can be drawn from these results is that using the same traffic scenario, increasing the PIRs of heavy users has allowed this class of users to transmit more data than authorized and therefore to have an improved upstream bitrates. Heavy users were limited by their SLA constraints in the first case (without optimization) and although they requested more than 100 Mbit/s, they were not permitted to exceed their PIRs (the maximum upstream bitrate reached is around 96 Mbit/s). In the second case where our optimization mechanism was involved, heavy users' upstream bitrates were increased up to 160 Mbit/s following their new PIRs. Thus, on the one hand, heavy users benefited from a surplus of bandwidth and a much better quality than before without paying extra costs and on the other hand, the bandwidth capacity has been efficiently handled by reducing the unused portion.

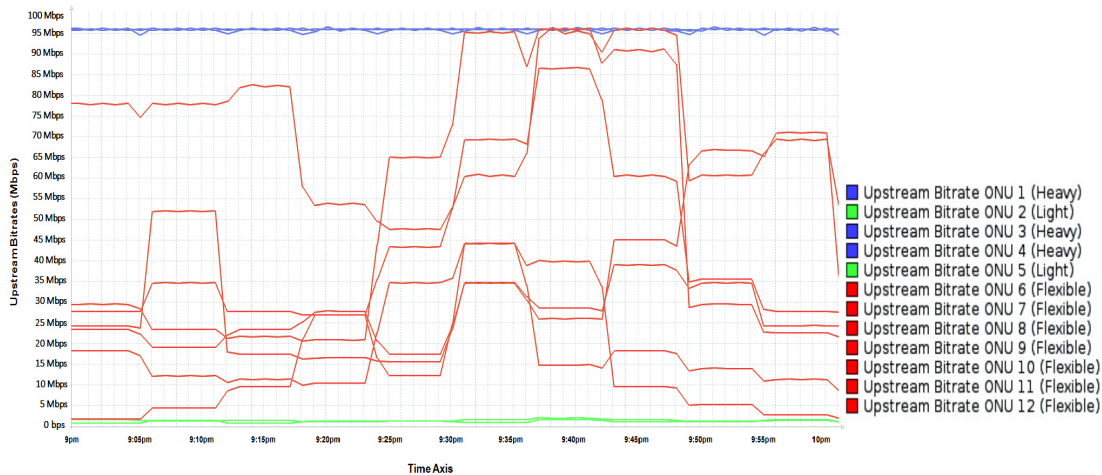


Figure 6.2: Allocated upstream bitrates before applying the optimization mechanism

As it is known in general that techniques enabling performances improvement by offering more bitrates may negatively impact other QoS metrics, we decided to study the impact of our optimization approach on the latency measurements. Table 6.5 shows the average latency measures for each ONU (performed by the Spirent Test Center) before and after applying the adjustment of the SLA parameters.

We notice that the average latency values are almost the same for all ONUs with a slight increase estimated around 0.1 us. This low value can be considered as negligible without a

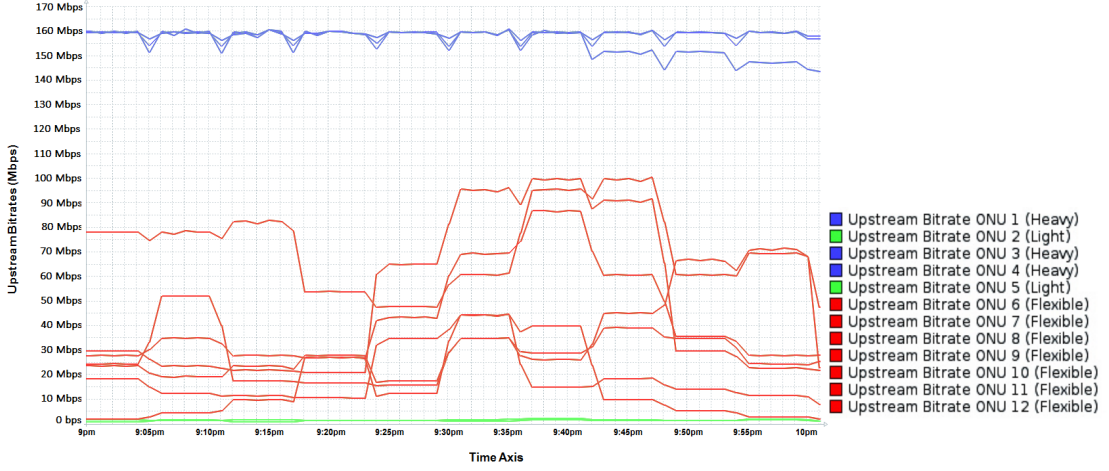


Figure 6.3: Allocated upstream bitrates after applying the optimization mechanism

Table 6.5: Average latency measures

ONU	Taffic Class	Average Latency Without Optimization (us)	Average Latency With Optimization (us)
ONU1	Heavy	17.777	17.937
ONU2	Light	15.842	15.93
ONU3	Heavy	17.579	17.76
ONU4	Heavy	17.921	18.107
ONU5	Light	16.436	16.667
ONU6	Flexible	16.51	16.669
ONU7	Flexible	16.638	16.667
ONU8	Flexible	16.566	16.653
ONU9	Flexible	15.014	15.133
ONU10	Flexible	21.719	21.945
ONU11	Flexible	15.02	15.139
ONU12	Flexible	15.017	15.15

negative impact on the QoS performances. We also studied the number of lost packets in the upstream direction in order to evaluate the impact of our approach on this metric. The results show that no packets were lost during the evaluation. This proves that the adopted optimization approach has no negative influence on the upstream resources allocation performance.

Thus to conclude, the test phase that was conducted with the different results presented allows us to infer that learning the daily bandwidth usage of different users and performing a self-adjustment of some SLA parameters based on their traffic patterns is very advantageous in terms of improving upstream bitrates and bandwidth usage efficiency.

6.4 Traffic Usage Variation Use Case

6.4.1 Use Case Objective

After showing the applicability and the efficiency of the proposed approach within a real PON scenario, we decided to go further in the evaluation study while providing some traffic scenarios that may occur in the network. The objective of this use case is to evaluate the impact of the injection of different traffic loads configurations into the network on the efficiency of the upstream bandwidth allocation before and after applying the optimization approach. We propose three evaluation scenarios respectively, low bandwidth usage, average bandwidth usage, and high bandwidth usage. From one scenario to another, we vary the number of users for each class whether light, flexible, or heavy. This aims to show if the optimization mechanism can be always useful regardless of the bandwidth usage behavior.

6.4.2 Experimental Scenario & QoS Assessment

As we conducted the evaluation study of the demonstration use case within a real PON connecting 12 subscribers, we continue with the same PON configuration in this scenario (i.e. 12 ONUs) due to laboratory environment limits. However, this time, the classification step is not performed upon the historical data already collected (i.e. traffic traces). We propose a classification phase of the 12 ONUs on the basis of three scenarios that we outline below:

- Low bandwidth usage: this scenario consists in a low number of heavy users, a high number of light users, and some flexible users.
- Average bandwidth usage: this scenario consists in a fair distribution between different users.
- High bandwidth usage: this scenario consists in a majority of heavy users with some flexible and light users.

These three scenarios may occur in the PON at any time. Thus, evaluating our optimization approach for each scenario allows by the end to demonstrate its usefulness and contribution to enhance the upstream resources allocation for different users' distributions. We propose in table 6.6 the users' classification for each scenario.

The estimation of the amount of bandwidth unused by light users is ordinarily performed upon the historical data already collected. As we don't use these data as reference for users classification in this use case, and inspired by the analyzes already performed during learning phases upon users traffic behavior, we assume that light users do not exceed in general 5%

Table 6.6: Users classification for traffic usage variation use case scenarios

Scenario	Number of light users	Number of flexible users	Number of heavy users
Low bandwidth usage	7	3	2
Average bandwidth usage	4	4	4
High bandwidth usage	2	3	7

of their PIRs (e.g. a maximum upstream bitrate of 5Mbit/s for a PIR=100Mbit/s). Thus, we propose that for each 5 minutes time interval, the average bandwidth used by a light user is a random value that ranges from 0% to 5% of his PIR. The ranges of different injected loads for each traffic class are shown in table 6.7. These loads are injected randomly respecting the minimum and the maximum of each range associated to a traffic class.

Table 6.7: Ranges of injected loads for different traffic classes

Traffic class	Range of injected loads (Mbit/s)
Light users	0 – 5
Flexible users	0 – 100
Heavy users	70 – 400

The QoS assessment is ensured regarding the upstream bandwidth utilization ratio metric. The latter can be defined as the percentage of the used bandwidth compared to the total bandwidth capacity in the PON. According to each traffic variation scenario and its associated users' classification, we present in the following, the different configurations as well as the resulting bandwidth utilization ratio before and after applying the PIRs adjustment.

6.4.2.1 Low Bandwidth Usage

In a low bandwidth usage scenario, we suppose that approximately 30% of the total PON upstream bandwidth is used. Thus, a high number of light users is considered compared to heavy and flexible ones as mentioned in table 6.6. The standard SLA parameters are the same for all ONUs with a CIR of 512 Kbit/s and a PIR of 100 Mbit/s. Table 6.8 summarizes the classification of different ONUs.

Table 6.8: ONUs classification in a low bandwidth usage scenario

ONU	Traffic class
ONU1, ONU2	Heavy
ONU3, ONU4, ONU5	Flexible
ONU6,ONU7,ONU8,ONU9 ONU10,ONU11,ONU12	Light

The estimation of the bandwidth used by light users on average as well as the calculation of the extra bandwidth available to be reallocated to heavy users, η (the multiplicative factor of heavy users PIRs), and α (the percentage increase in PIR) are performed following the algorithm 2 presented in chapter III. The results are shown respectively in table 6.10 and table 6.11.

Table 6.9: Results of bandwidth estimation and sharing in a low bandwidth usage scenario

Table 6.10: Estimation of the bandwidth used by light users in a low bandwidth usage scenario

ONU	Average Bandwidth Used (Mbit/s)
ONU6	2.89
ONU7	2.535
ONU8	1.22
ONU9	1.76
ONU10	2.685
ONU11	1.88
ONU12	3.04

Table 6.11: Extra bandwidth, η , and α values in a low bandwidth usage scenario

Extra bandwidth	683.99 Mbit/s
η	4.42
α	342%

Because of the significant number of light users in this scenario, each heavy user will have a 342% increase in his PIR. Thus, the new configuration of heavy users' SLA parameters let them enjoy a new PIR value of 442 Mbit/s, while flexible and light users parameters remain untouched. We present in Fig.6.4 the results of the bandwidth usage ratio before and after applying the optimization mechanism.

We can see from Fig.6.4 that for the same injected traffic loads, the maximum upstream bandwidth utilization ratio has reached 0.43 (i.e. 43% of the total capacity) after applying the PIR adjustment while it was limited to 0.27 before the optimization process. The users' distribution including a majority of light ONUs allows to have a significant amount of additional bandwidth and consequently, a better QoS is assured by reducing the wasted bandwidth untapped before applying the optimization mechanism.

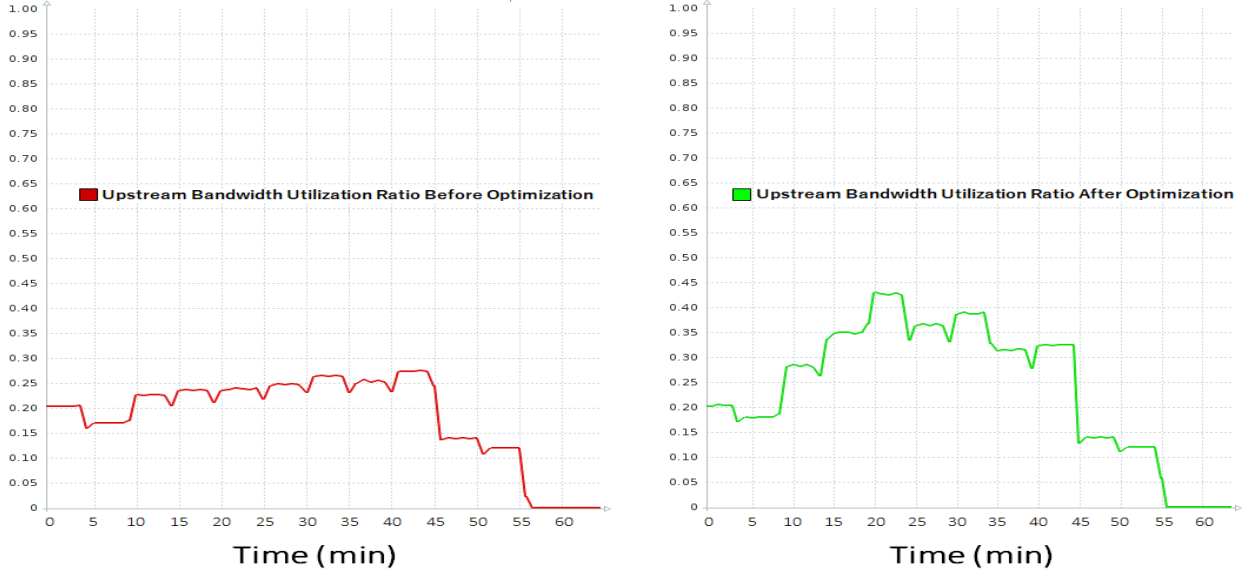


Figure 6.4: Upstream bandwidth usage ratio before and after applying the optimization mechanism in a low bandwidth usage scenario

6.4.2.2 Average Bandwidth Usage

The average bandwidth usage scenario evaluates the impact of loading about half of the PON capacity on the bandwidth usage ratio. An equitable distribution between PON users is considered with an assignment of four ONUs to each traffic class as shown in table 6.12. The standard SLA parameters are the same for all ONUs with a CIR of 512 Kbit/s and a PIR of 100 Mbit/s.

Table 6.12: ONUs classification in an average bandwidth usage scenario

ONU	Traffic class
ONU1, ONU2, ONU3, ONU4	Heavy
ONU5, ONU6, ONU7, ONU8	Flexible
ONU9, ONU10, ONU11, ONU12	Light

The results of the estimation of the bandwidth used by light users on average as well as the calculation of the extra bandwidth available to be reallocated to heavy users, η , and α are shown respectively in table 6.14 and table 6.15.

The resulting extra bandwidth calculated upon the equitable users' distribution allows each heavy user to benefit from a 97.5% of his PIR as a bandwidth supplement. To show the impact

Table 6.13: Results of bandwidth estimation and sharing in an average bandwidth usage scenario

Table 6.14: Estimation of the bandwidth used by light users in an average bandwidth usage scenario

ONU	Average Bandwidth Used (Mbit/s)
ONU9	2.84
ONU10	3.04
ONU11	1.83
ONU12	2.2

Table 6.15: Extra bandwidth, η , and α values in an average bandwidth usage scenario

Extra bandwidth	390.09 Mbit/s
η	1.975
α	97.5%

of this new configuration, Fig.6.5 outlines the results of the upstream bandwidth usage ratio in the PON before and after applying the optimization mechanism.

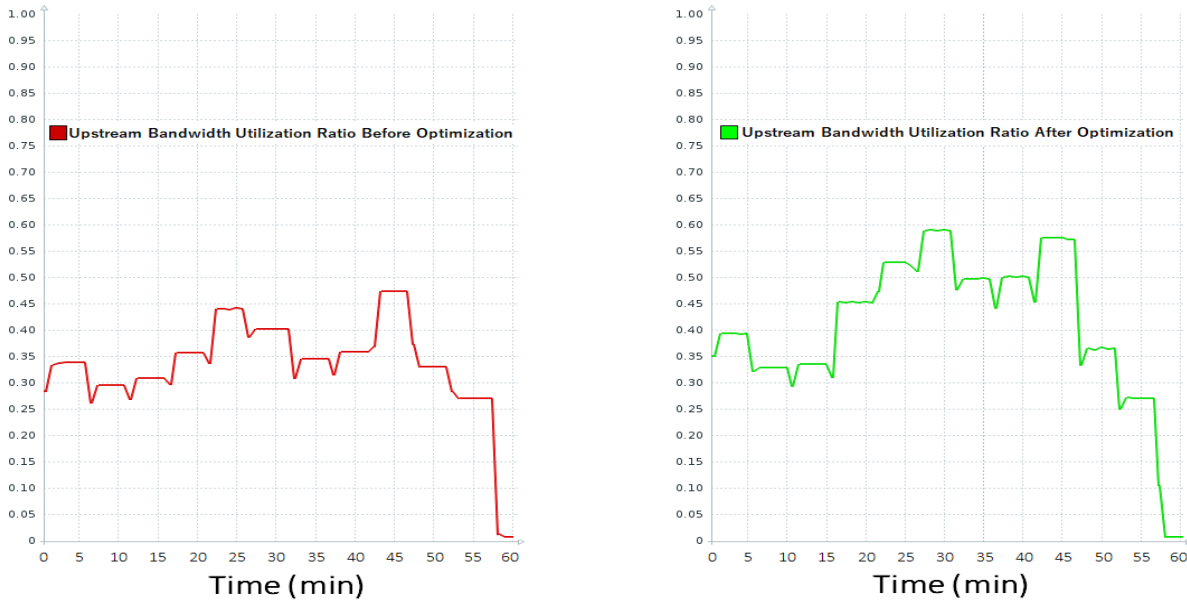


Figure 6.5: Upstream bandwidth usage ratio before and after applying the optimization mechanism in an average bandwidth usage scenario

The results of the bandwidth usage ratio for the equitable users distribution show that the optimization mechanism ensures high values of bandwidth usage with a maximum of 59% observed at 28 min of the evaluation time (a gain of 19% compared to the same time before the optimization). Over the entire evaluation time, an improvement of the bandwidth usage and a reduction of the wasted bandwidth can be distinguished in the majority of time intervals. Thus, with the same number of users in each traffic class, the performed PIRs adjustment stills efficient and advantageous in terms of enhancing PON upstream resources allocation.

6.4.2.3 High Bandwidth Usage

Having evaluated our optimization mechanism in low and average bandwidth usage scenarios, we conclude the test phase by a high bandwidth usage scenario where about 90% of the PON capacity is supposed to be loaded. 90% is the maximum capacity that could be used in the PON test environment. As configured, beyond 90% of the total upstream capacity, a traffic shaping is ensured by the OLT to avoid congestion and packets discarding. Thus, allocated upstream bitrates will be automatically adjusted respecting the total PON upstream capacity.

To ensure the implementation of this scenario, we consider a users' distribution with a majority of heavy users compared to light and flexible ones as shown in table 6.16. Unlike the two previous scenarios, the standard SLA parameters are kept the same only for flexible and light ONUs with a CIR of 512, Kbit/s and a PIR of 100 Mbit/s. Heavy users' standard PIRs are set to 200 Mbit/s while their CIRs are the same as for the other users. This aims to assure the usage of the requested bandwidth capacity.

Table 6.16: ONUs classification in a high bandwidth usage scenario

ONU	Traffic class
ONU1,ONU2,ONU3,ONU4 ONU5,ONU6,ONU7	Heavy
ONU8, ONU9, ONU10	Flexible
ONU11, ONU12	Light

The results of the estimation of the bandwidth used by light users on average as well as the calculation of the extra bandwidth available to be reallocated to heavy users, η and α are shown respectively in table 6.17 and table 6.18.

Table 6.17: Estimation of the bandwidth used by light users in a high bandwidth usage scenario

ONU	Average Bandwidth Used (Mbit/s)
ONU11	2.345
ONU12	1.71

Table 6.18: Extra bandwidth, η , and α values in a high bandwidth usage scenario

Extra bandwidth	195,945 Mbit/s
η	1.14
α	14%

The resulting extra bandwidth leads to an increase of 14% in each heavy user PIR. Fig.6.6 highlights the results of the upstream bandwidth usage ratio in the PON before and after applying the optimization mechanism.

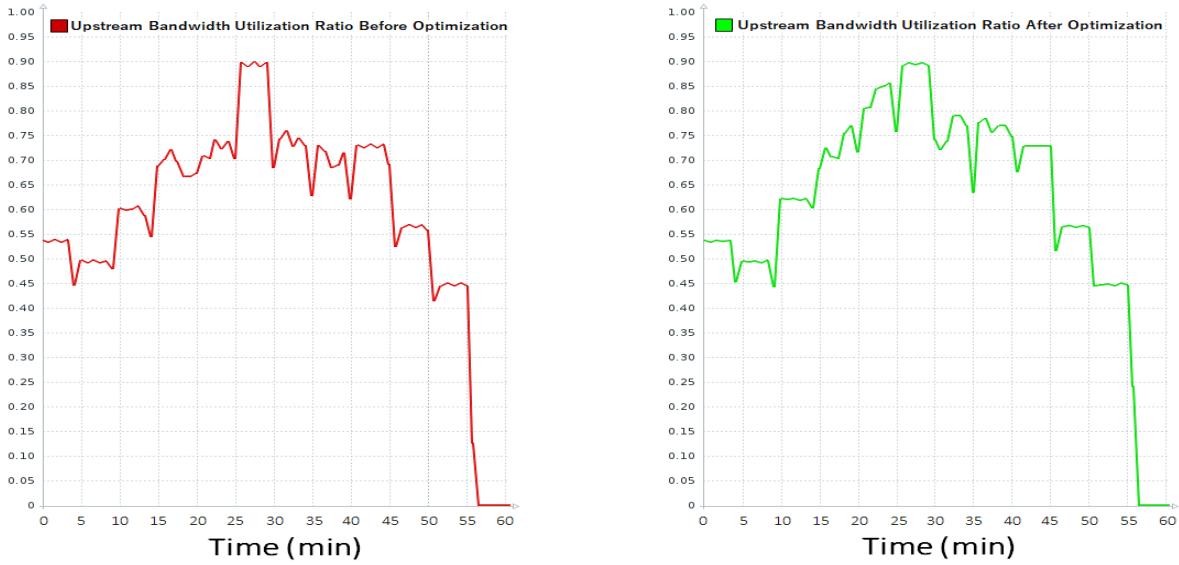


Figure 6.6: Upstream bandwidth usage ratio before and after applying the optimization mechanism in a high bandwidth usage scenario

In this last use case, the resulting upstream bandwidth usage ratio shows that no significant changes can be observed (except the time interval between 20 min and 25 min where the optimization mechanism has improved the bandwidth usage by 11%). For a distribution that includes a majority of heavy users and a minority of light ones, the additional bandwidth that heavy users may benefit from is less important than the other distributions. Thus, the supplement of bitrate which is added to each heavy user PIR is not enough to show a significant contribution of the optimization process.

6.4.3 Discussion

After evaluating the optimization mechanism for different bandwidth usage configurations, we proceed in the following to a comparative study which aims to show when the optimization mechanism is more reliable and useful and has a significant impact on the PON upstream bandwidth allocation efficiency.

The bandwidth utilization ratio graphics presented in Fig.6.7 and Fig.6.8 show that the maximum and the average of upstream bandwidth usage vary from one scenario to another. The best results are obtained for the low bandwidth usage scenario where we achieved overall gains of roughly 16% regarding the maximum of bandwidth usage and 5.7% regarding the average bandwidth usage when compared to the non-optimized PON. For the average bandwidth usage scenario, the optimization mechanism allows to achieve a higher bandwidth utilization estimated at 0.59 in a scale of 1 and ensure a gain of 11.5% in comparison to the non-optimized PON.

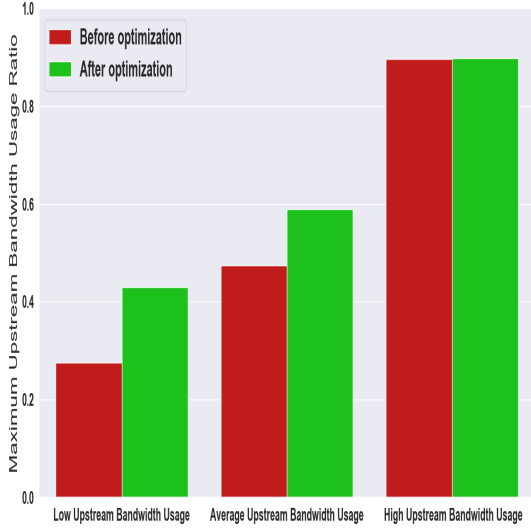


Figure 6.7: Maximum upstream bandwidth usage ratio for each traffic variation scenario before and after applying the optimization mechanism

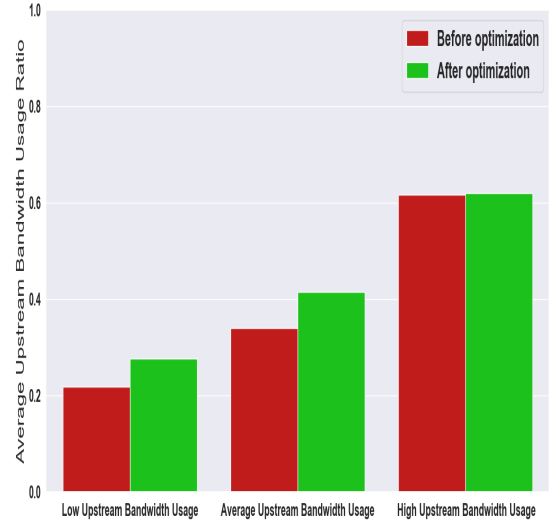


Figure 6.8: Average upstream bandwidth usage ratio for each traffic variation scenario before and after applying the optimization mechanism

Given the high bandwidth usage scenario, no significant changes are identified. This is logical since when the number of heavy users increases, the share of each one in terms of additional bandwidth becomes less important. Thus, the PIRs adjustment will not have a major impact on the bandwidth usage efficiency.

From these results, it's clear that the proposed optimization mechanism achieves good QoS performances when the users' classification tends to have a majority of light users or an equitable distribution. With a significant number of heavy users and when we are getting closer to the total upstream capacity of the PON, the optimization mechanism becomes less efficient as the estimated supplement of bandwidth decreases. To conclude, the optimization approach is very beneficial when there is an important amount of upstream bandwidth which is theoretically untapped by the DBA and wasted due to the SLA constraints. Thus, the higher the unused bandwidth by light users, the more efficient the optimization mechanism.

6.5 Conclusion

This chapter demonstrates the applicability of the proposed methodology, the functioning of our model and motivates the usefulness of our optimization mechanism in improving the upstream bandwidth usage efficiency. An experimental evaluation study was conducted based on two use cases: (i) a demonstration use case based on a real network scenario inspired by a real PON infrastructure using collected traffic traces as historical data. (ii) A traffic usage variation use

case where different bandwidth usage scenarios are considered (e.g. low bandwidth usage, average bandwidth usage, high bandwidth usage). The adjustment of heavy users' SLA parameters has shown its benefits in terms of improving the QoS by enhancing the allocated upstream bitrates without any degradation of quality for light and flexible users.

Contents

7.1 Summary	141
7.2 Perspectives	143

7.1 Summary

PONs represent one the most reliable and deployed FTTH access network infrastructures. They are characterized by their ability to offer high-speed services and satisfy the needs of the majority of customers in terms of QoS. To handle the dynamicity of customers' traffic patterns and maximize the efficiency of the bandwidth usage particularly in the upstream, PONs rely on a DBA mechanism to assign for each ONU, the corresponding amount of upstream bandwidth according to its demand. Thus, it is recommended to let the DBA operates in its best level of efficiency to better leverage existing network resources and achieve a balance between capacity and offered QoS.

In the literature, a significant amount of research and many discussions have been proposed with the aim to converge towards an optimal DBA design. The majority of these studies does not take into account that each new DBA proposal requires vendor specific involvement and might not be easily integrated in operator's equipment. Thus, the operator vision was not taken into account when integrating the DBA mechanism in the OLT equipment.

The objective of the study carried out in this thesis was to propose a new architecture for optimizing the upstream resources allocation while the DBA mechanism itself remains untouchable. In this context, four contributions were presented in this work:

1. **Management architecture for optimizing PON upstream resources:** after surveying the different proposals regarding the improvement of PON dynamic upstream resources allocation, we have noticed the lack of the operator vision in these studies. Based on this issue, we have identified a new opportunity to improve the DBA functioning by adjusting dynamically some of SLA parameters with reference to customers' traffic behavior during the day. Machine learning techniques were used to determine the traffic pattern of each user whether heavy, light or flexible using his historical upstream bandwidth usage. The proposed approach was designed in such a way that it can be directly and easily integrated in the PON management system without the need to update the DBA at the OLT level.
2. **A forecasting approach for studying PON users traffic tendency:** after evaluating the proposed architecture in terms of users' classification, an issue related to users' distribution was identified. The results have shown that the majority of users were designated to be flexible, leading to an unbalanced users' distribution. Since it is preferable to have a significant number of heavy and light users in order to maximize the efficiency of the bandwidth usage in the network, we enhanced the proposed approach by adding a forecasting module as a part of a second user classification step. This module allowed coping with the unexpected traffic behavior of some PON users by learning more about the traffic tendency of flexible ones and the results have demonstrated its usefulness.
3. **A GANA oriented architecture for self-optimized PON:** to enrich the network with self-decision capabilities that leave the task of the dynamic reconfiguration of the SLA parameters to the network itself with the minimum of direct human intervention, we proposed an enhancement of the proposed approach based on the autonomic & cognitive principles. Particularly, we have opted for GANA autonomic model to perform the adjustment process of SLA parameters at different day periods.
4. **Evaluation of the self-optimized PON architecture:** to demonstrate the applicability and the usefulness of the proposed architecture, we conducted an evaluation study based on two use cases: (i) a demonstration use case based on a real network scenario inspired by a real PON infrastructure using collected traffic traces as historical data. (ii) A traffic usage variation use case where different bandwidth usage scenarios were considered (e.g. low bandwidth usage, average bandwidth usage, high bandwidth usage). The efficiency of the architecture was validated through a set of individual evaluation criteria such as the measured upstream bitrate, the latency, and the bandwidth usage ratio.

7.2 Perspectives

The work presented in this thesis was intended for optimizing resources allocation in PONs. A particular attention was given to the upstream traffic allocation due to the issue identified within the reviewed DBA proposals. The adopted methodology to readjust some of SLA parameters based on machine learning techniques is a novelty in the domain. Thus, it can be used in many areas closely linked to the topic of this thesis. In the following, we present some of the potential research perspectives and future directions of this work.

- **Reinforcement learning module:** in order to classify users based on their traffic patterns, many parameters such as time intervals, clustering parameters, assignment index thresholds, etc. were fixed in the learning process. However, we think that the selection phase of these parameters should be more investigated to converge towards the most efficient users' classification. For this, the addition of a reinforcement learning module to the optimization mechanism could help to improve the parameters' selection through iterated experiments, trials, and feedbacks from previous actions and experiences. Reinforcement learning is a machine learning technique that uses computer agents which interact with the environment, learn, and take the appropriate decision based on the past actions.
- **Context aware module :** in this thesis, the proposed optimization approach learns users' traffic patterns from their historical data. Then, it proposes an adjustment of SLA parameters based on the learning process decisions. A suggested improvement is to add a context aware module which in addition to historical collected data, it could feed the learning process by contextual data such as weather conditions, weekly event, beginning and end of holidays, etc. This kind of data could improve the decision making process and enhance the users' classification.
- **Downstream bandwidth allocation :** although the optimization process proposed in this thesis was designed to improve the upstream resources allocation, future work could investigate its applicability in the downstream direction. In PONs, downstream traffic is broadcasted from the OLT to all ONUs. Then, each ONU selects its corresponding traffic using a specific encryption mechanism and discards the rest. SLA parameters are used also in the downstream direction to ensure the compliance with users requirements. Thus, the learning process proposed in this thesis based on machine learning and forecasting techniques could be used as reference to improve downstream resources allocation.

- [1] Gregory D Abowd et al., “Towards a better understanding of context and context-awareness”, *in: International Symposium on Handheld and Ubiquitous Computing*, Springer, 1999, pp. 304–307.
- [2] Nazim Agoulmine, “Autonomic network management principles: from concepts to applications”, *in: Academic Press*, 2010, p. 5.
- [3] V. Alarcon-Aquino and J. A. Barria, “Multiresolution FIR neural-network-based learning algorithm applied to network traffic prediction”, *in: IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36.2 (2006), pp. 208–220.
- [4] A. Sermet Anagun, “Development of Committee Neural Network for Computer Access Security System”, *in: Computational Science and Its Applications - ICCSA 2006*, ed. by Marina Gavrilova et al., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 11–20, ISBN: 978-3-540-34076-8.
- [5] I. V. Anikin and R. M. Gazimov, “Privacy preserving DBSCAN clustering algorithm for vertically partitioned data in distributed systems”, *in: 2017 International Siberian Conference on Control and Communications (SIBCON)*, 2017, pp. 1–4, DOI: 10.1109/SIBCON.2017.7998473.
- [6] G. Aristomenopoulos et al., “Autonomic mobility and resource management over an integrated wireless environment - A GANA oriented architecture”, *in: 2010 IEEE Globecom Workshops*, 2010, pp. 545–550, DOI: 10.1109/GLOCOMW.2010.5700379.
- [7] C. M. Assi et al., “Dynamic bandwidth allocation for quality-of-service over Ethernet PONs”, *in: IEEE Journal on Selected Areas in Communications* 21.9 (2003), pp. 1467–1477.

-
- [8] Matthias Baldauf, Schahram Dustdar, and Florian Rosenberg, “A Survey on Context& Aware Systems”, *in: Int. J. Ad Hoc Ubiquitous Comput. 2.4* (June 2007), pp. 263–277, ISSN: 1743-8225, DOI: 10.1504/IJAHUC.2007.014070, URL: <http://dx.doi.org/10.1504/IJAHUC.2007.014070>.
- [9] Tobias Bandh, “Coordination of autonomic function execution in Self-Organizing Networks”, *in: Techn. Univ. München, Lehrstuhl Netzarchitekturen und Netzdienste*, 2013, pp. 34–40.
- [10] A. Banerjee and R. N. Dave, “Validating clusters using the Hopkins statistic”, *in: 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, vol. 1, 2004, 149–153 vol.1, DOI: 10.1109/FUZZY.2004.1375706.
- [11] Hakjeon Bang et al., “Dynamic bandwidth allocation method for high link utilization to support NSR ONUs in GPON”, *in: 2010 The 12th International Conference on Advanced Communication Technology (ICACT)*, vol. 1, 2010, pp. 884–889.
- [12] R. Boutaba, S. Omari, and A. P. S. Virk, “SELFCON: An architecture for self-configuration of networks”, *in: Journal of Communications and Networks 3.4* (2001), pp. 317–323, ISSN: 1229-2370, DOI: 10.1109/JCN.2001.6596962.
- [13] George EP Box et al., *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.
- [14] D. Breuer et al., “Opportunities for next-generation optical access”, *in: IEEE Communications Magazine 49.2* (2011), s16–s24, ISSN: 0163-6804, DOI: 10.1109/MCOM.2011.5706309.
- [15] Rizwan Aslam Butt et al., “A Survey of Dynamic Bandwidth Assignment Schemes for TDM-Based Passive Optical Network”, *in: Journal of Optical Communications* ().
- [16] James Cannady, “Artificial neural networks for misuse detection”, *in: National information systems security conference*, 1998, pp. 368–81.
- [17] V. W. S. Chan, “Cognitive optical wireless network”, *in: 2016 18th International Conference on Transparent Optical Networks (ICTON)*, 2016, pp. 1–4, DOI: 10.1109/ICTON.2016.7550316.
- [18] R. Chaparadza et al., “Implementation Guide for the ETSI AFI GANA model: A Standardized Reference Model for Autonomic Networking, Cognitive Networking and Self-Management”, *in: 2013 IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 935–940, DOI: 10.1109/GLOCOMW.2013.6825110.

- [19] R. Chaparadza et al., “Implementation Guide for the ETSI AFI GANA model: A Standardized Reference Model for Autonomic Networking, Cognitive Networking and Self-Management”, in: *2013 IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 935–940, DOI: 10.1109/GLOCOMW.2013.6825110.
- [20] S. i. Choi and J. Park, “SLA-Aware Dynamic Bandwidth Allocation for QoS in EPONs”, in: *IEEE/OSA Journal of Optical Communications and Networking 2.9* (2010), pp. 773–781, ISSN: 1943-0620, DOI: 10.1364/JOCN.2.000773.
- [21] Laurent Ciavaglia et al., “Coordination of Self-Organizing Network Mechanisms: Framework and Enablers”, in: *International Conference on Mobile Networks and Management*, Springer, 2011, pp. 174–184.
- [22] Laurent Ciavaglia et al., “Coordination of Self-Organizing Network Mechanisms: Framework and Enablers”, in: *International Conference on Mobile Networks and Management*, Springer, 2011, pp. 174–184.
- [23] Celtic Project: Deliverable D.3.1, “Specification of cognitive network management framework and functions”, in: *Celtic Project* (2012), pp. 1–109.
- [24] E. Dai and W. Dai, “Towards SDN for optical access networks”, in: *Spring Technical Forum Proceedings*, 2016, pp. 1–6.
- [25] J. L. Deng, “Introduction to Grey System Theory”, in: *J. Grey Syst. 1.1* (Nov. 1989), pp. 1–24, ISSN: 0957-3720.
- [26] A. Dixit et al., “Dynamic bandwidth allocation with SLA awareness for QoS in ethernet passive optical networks”, in: *IEEE/OSA Journal of Optical Communications and Networking 5.3* (2013), pp. 240–253, ISSN: 1943-0620, DOI: 10.1364/JOCN.5.000240.
- [27] Abhishek Dixit et al., “Jitter performance for QoS in Ethernet passive optical networks”, in: *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on, IEEE*, 2011, pp. 1–3.
- [28] LL Eberhardt, “Some developments in ‘distance sampling’”, in: *Biometrics* (1967), pp. 207–216.
- [29] Martin Ester et al., “A density-based algorithm for discovering clusters in large spatial databases with noise”, in: *Kdd*, vol. 96, 34, 1996, pp. 226–231.
- [30] ETSI GS AFI 002 V1.1.1, “Autonomic network engineering for the self-managing Future Internet (AFI); Generic Autonomic Network Architecture (An Architectural Reference Model for Autonomic Networking, Cognitive Networking and Self-Management)”, in: *ETSI Technical Specification* (2013), pp. 11–13.

-
- [31] M. Feknous et al., “Status reporting versus non status reporting dynamic bandwidth allocation”, *in: 2015 6th International Conference on the Network of the Future (NOF)*, 2015, pp. 1–7, DOI: 10.1109/NOF.2015.7333289.
- [32] Moufida Feknous et al., “Multi-Criteria Comparison Between Legacy and Next Generation Point of Presence Broadband Network Architectures”, *in: Advances in Computer Science: an International Journal 4.3* (2015), pp. 126–140.
- [33] L. Florio and E. D. Nitto, “Gru: An Approach to Introduce Decentralized Autonomic Behavior in Microservices Architectures”, *in: 2016 IEEE International Conference on Autonomic Computing (ICAC)*, 2016, pp. 357–362, DOI: 10.1109/ICAC.2016.25.
- [34] ITU-T G.9807.1, “G.9807.1 : 10-Gigabit-capable symmetric passive optical network (XGSPON)”, *in: ITU-T standards* (June 2016).
- [35] ITU-T G.984.1, “Gigabit-capable Passive Optical Networks (G-PON): General characteristics”, *in: ITU-T standards* (March 2008).
- [36] ITU-T G.987, “10-Gigabit-capable passive optical network (XG-PON)”, *in: ITU-T standards* (January 2010).
- [37] ITU-T G.989.1, “40-Gigabit-capable passive optical networks (NG-PON2)”, *in: ITU-T standards* (March 2013).
- [38] ITU-T G.993.2, “Very high speed digital subscriber line transceivers 2 (VDSL2)”, *in: ()*.
- [39] A. Galani et al., “Core functional and network empower mechanisms of an operator-driven, framework for unifying autonomic network and service management”, *in: 2012 IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2012, pp. 191–195, DOI: 10.1109/CAMAD.2012.6335330.
- [40] A. G. Ganek and T. A. Corbi, “The Dawning of the Autonomic Computing Era”, *in: IBM Syst. J. 42.1* (Jan. 2003), pp. 5–18, ISSN: 0018-8670, URL: <http://dx.doi.org/10.1147/sj.421.0005>.
- [41] Albert Greenberg et al., “A Clean Slate 4D Approach to Network Control and Management”, *in: SIGCOMM Comput. Commun. Rev. 35.5* (Oct. 2005), pp. 41–54, ISSN: 0146-4833, DOI: 10.1145/1096536.1096541, URL: <http://doi.acm.org/10.1145/1096536.1096541>.
- [42] Jiawei Han, Jian Pei, and Micheline Kamber, *Data mining: concepts and techniques*, Elsevier, 2011, pp. 484–486.
- [43] Jiawei Han, Jian Pei, and Micheline Kamber, *Data mining: concepts and techniques*, Elsevier, 2011, pp. 443–476.

- [44] Onn Haran and Amir Sheffer, “The importance of dynamic bandwidth allocation in GPON networks”, *in: PMC-Sierra inc., White paper 1* (2008).
- [45] Paola Garfias Hernandez, “Resource Management Research in Ethernet Passive Optical Networks”, PhD thesis, Universitat Politecnica De Catalunya, 2013.
- [46] P Holgate, “Some new tests of randomness”, *in: The Journal of Ecology* (1965), pp. 261–266.
- [47] Dave Hood and Elmar Trojer, *Gigabit-capable passive optical networks*, John Wiley & Sons, 2012.
- [48] Brian Hopkins and John Gordon Skellam, “A new method for determining the type of distribution of plant individuals”, *in: Annals of Botany* 18.2 (1954), pp. 213–227.
- [49] Che-Chiang Hsu and Chia-Yon Chen, “Applications of improved grey prediction model for power demand forecasting”, *in: Energy Conversion and management* 44.14 (2003), pp. 2241–2249.
- [50] Rob J Hyndman and George Athanasopoulos, *Forecasting: Principles and Practice*, Monash University, Australia, 2018.
- [51] IBM, “An architectural blueprint for autonomic computing”, *in: IBM White Paper* 31 (2006).
- [52] ITU-T, “M3010: Principles for a telecommunications management network”, *in: ITU-T standards* (2000).
- [53] ITU-T G.984.3, “Gigabit-capable Passive Optical Networks (G-PON): Transmission convergence layer specification”, *in: ITU-T standards* (January 2014).
- [54] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering: A Review”, *in: ACM Comput. Surv.* 31.3 (Sept. 1999), pp. 264–323, ISSN: 0360-0300, DOI: 10.1145/331499.331504.
- [55] Anil K Jain, “Data clustering: 50 years beyond K-means”, *in: Pattern recognition letters* 31.8 (2010), pp. 651–666.
- [56] B. Jennings et al., “Towards autonomic management of communications networks”, *in: IEEE Communications Magazine* 45.10 (2007), pp. 112–121, ISSN: 0163-6804, DOI: 10.1109/MCOM.2007.4342833.
- [57] Nicholas R Jennings and Michael J Wooldridge, “Software agents”, *in: IEE review* (1996), pp. 17–20.
- [58] R. Jouini et al., “Employing Grey Model forecasting GM(1,1) to historical medical sensor data towards system preventive in smart home e-health for elderly person”, *in: 2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2016, pp. 1086–1091.

- [59] A. Kavousi-Fard, “A Hybrid Accurate Model for Tidal Current Prediction”, *in: IEEE Transactions on Geoscience and Remote Sensing* 55.1 (2017), pp. 112–118, ISSN: 0196-2892, DOI: 10.1109/TGRS.2016.2596320.
- [60] J. O. Kephart and D. M. Chess, “The vision of autonomic computing”, *in: Computer* 36.1 (2003), pp. 41–50, ISSN: 0018-9162, DOI: 10.1109/MC.2003.1160055.
- [61] C.C Klimasauskas, “Applying neural networks. Part 3: Training a neural network”, *in: PC-AI* (1991), pp. 20–24.
- [62] S. Kukliński et al., “GARSON: Management performance aware approach to autonomic and cognitive networks”, *in: 2012 IEEE Globecom Workshops*, 2012, pp. 914–918, DOI: 10.1109/GLOCOMW.2012.6477698.
- [63] C. H. Lee, W. V. Sorin, and B. Y. Kim, “Fiber to the Home Using a PON Infrastructure”, *in: Journal of Lightwave Technology* 24.12 (2006), pp. 4568–4583, ISSN: 0733-8724, DOI: 10.1109/JLT.2006.885779.
- [64] J. Li, W. Dai, and H. Pan, “GM(1,1) Model - Ameliorated Based on Genetic Algorithm and Its Application”, *in: Third International Conference on Natural Computation (ICNC 2007)*, vol. 4, 2007, pp. 23–27, DOI: 10.1109/ICNC.2007.407.
- [65] Marja Liinasuo et al., “Human operator perspective to autonomic network management”, *in: Proceedings of the 5th International Conference on Advances in Computer-Human Interaction (ACHI 2012)*, IEEE Computing Society, Citeseer, 2012, pp. 128–134.
- [66] S. Lin and H. Yin, “A Novel Optical Network Model for Resource Balance and Congestion Reduction”, *in: 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, 2011, pp. 1–4, DOI: 10.1109/wicom.2011.6040243.
- [67] Ioannis Loumiotis et al., “Optimal Backhaul Resource Management in Wireless-Optical Converged Networks”, *in: Internet of Things. IoT Infrastructures: First International Summit, IoT360 2014, Rome, Italy, October 27-28, 2014, Revised Selected Papers, Part II*, ed. by Raffaele Giaffreda et al., Cham: Springer International Publishing, 2015, pp. 254–261, ISBN: 978-3-319-19743-2.
- [68] C. Lu et al., “Data Mining Applied to Oil Well Using K-Means and DBSCAN”, *in: 2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 2016, pp. 37–40, DOI: 10.1109/CCBD.2016.018.
- [69] M. Lu and W. Wang, “An autonomic wireless video transmission system based on GANA”, *in: 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, vol. 02, 2012, pp. 826–830, DOI: 10.1109/CCIS.2012.6664291.

- [70] M. Luoto et al., “Distributed decision engine - An information management architecture for autonomic wireless networking”, *in: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, pp. 713–719, DOI: 10.1109/INM.2015.7140360.
- [71] James MacQueen et al., “Some methods for classification and analysis of multivariate observations”, *in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, 14, Oakland, CA, USA., 1967, pp. 281–297.
- [72] Qusay Mahmoud, “Cognitive networks: towards self-aware networks”, *in: John Wiley & Sons*, 2007, pp. 23–51.
- [73] N. Merayo et al., “Fair bandwidth allocation algorithm for PONS based on network utility maximization”, *in: IEEE/OSA Journal of Optical Communications and Networking 9.1* (2017), pp. 75–86, ISSN: 1943-0620, DOI: 10.1364/JOCN.9.000075.
- [74] I. de Miguel et al., “Cognitive dynamic optical networks [invited]”, *in: IEEE/OSA Journal of Optical Communications and Networking 5.10* (2013), A107–A118, ISSN: 1943-0620, DOI: 10.1364/JOCN.5.00A107.
- [75] J. Milojković and V. Litovski, “ANN versus Grey theory based forecasting methods implemented on short time series”, *in: 10th Symposium on Neural Network Applications in Electrical Engineering*, 2010, pp. 117–122, DOI: 10.1109/NEUREL.2010.5644094.
- [76] Serge Romaric Tembo Mouafo, “Applications de l’intelligence artificielle à la détection et l’isolation de pannes multiples dans un réseau de télécommunications”, PhD thesis, Ecole nationale supérieure Mines-Télécom Atlantique, 2017.
- [77] Z. Movahedi et al., “A Survey of Autonomic Network Architectures and Evaluation Criteria”, *in: IEEE Communications Surveys Tutorials 14.2* (2012), pp. 464–490, ISSN: 1553-877X, DOI: 10.1109/SURV.2011.042711.00078.
- [78] E Raimundo Neto et al., “Dynamic spectral sensing and resource allocation in optical-wireless network”, *in: SDR-WInnComm, Schaumburg-Illinois, United States* (2014).
- [79] D. Nowak, P. Perry, and J. Murphy, “Bandwidth allocation for service level agreement aware Ethernet passive optical networks”, *in: Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 3, 2004, pp. 1953–1957, DOI: 10.1109/GLOCOM.2004.1378334.
- [80] Asgeir Thor Oskarsson, “Broadband services virtual operator for bitstream open-access networks: business case and infrastructure”, PhD thesis, Universitat Politècnica De Catalunya, 2011, pp. 18–22.

- [81] Joanna Ozimkiewicz et al., “Evaluation of Dynamic Bandwidth Allocation Algorithms in GPON Networks”, *in: WSEAS Trans. Cir. and Sys.* 9.2 (Feb. 2010), pp. 111–120, ISSN: 1109-2734.
- [82] J Ozimkiewicz et al., “Dynamic bandwidth allocation in GPON networks”, *in: 4th WSEAS international conference on Circuits, systems, signal and telecommunications*, 2010, pp. 182–187.
- [83] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python”, *in: Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [84] Vu Anh Pham and Ahmed Karmouch, “Mobile software agents: an overview”, *in: IEEE Communications magazine* 36.7 (1998), pp. 26–37.
- [85] JA Fernández Pierna and DL Massart, “Improved algorithm for clustering tendency”, *in: Analytica Chimica Acta* 408.1 (2000), pp. 13–20.
- [86] Arun Prakash, Alexej Starschenko, and Ranganai Chaparadza, “Auto-discovery and Auto-configuration of Routers in an Autonomic Network”, *in: Access Networks: 5th International ICST Conference on Access Networks, AccessNets 2010 and First ICST International Workshop on Autonomic Networking and Self-Management in Access Networks, SELFMAGICNETS 2010, Budapest, Hungary, November 3-5, 2010, Revised Selected Papers*, ed. by Róbert Szabó et al., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 311–324, ISBN: 978-3-642-20931-4, DOI: 10.1007/978-3-642-20931-4_24, URL: http://dx.doi.org/10.1007/978-3-642-20931-4_24.
- [87] C. Prehofer and C. Bettstetter, “Self-organization in communication networks: principles and design paradigms”, *in: IEEE Communications Magazine* 43.7 (2005), pp. 78–85, ISSN: 0163-6804, DOI: 10.1109/MCOM.2005.1470824.
- [88] M. M. Rad et al., “Passive optical network monitoring: challenges and requirements”, *in: IEEE Communications Magazine* 49.2 (2011), s45–S52, ISSN: 0163-6804, DOI: 10.1109/MCOM.2011.5706313.
- [89] E. Raimundo-Neto et al., “Cognitive broadband optical-wireless network”, *in: 2015 International Workshop on Telecommunications (IWT)*, 2015, pp. 1–5, DOI: 10.1109/IWT.2015.7224555.
- [90] Juan Ramiro and Khalid Hamied, “Self-organizing networks (SON): self-planning, self-optimization and self-healing for GSM, UMTS and LTE”, *in: John Wiley & Sons*, 2011, pp. 50–53.

- [91] S. Kh. Sadon et al., “Enhanced Hierarchical Dynamic Bandwidth Allocation Algorithm with Excess Bandwidth and SLA Techniques for Upstream Ethernet PON”, *in: 2009 Second International Conference on Computer and Electrical Engineering*, vol. 2, 2009, pp. 147–151, DOI: 10.1109/ICCEE.2009.175.
- [92] N. Samaan and A. Karmouch, “Towards Autonomic Network Management: an Analysis of Current and Future Research Directions”, *in: IEEE Communications Surveys Tutorials* 11.3 (2009), pp. 22–36, ISSN: 1553-877X, DOI: 10.1109/SURV.2009.090303.
- [93] Stefan Schmid, Manolis Sifalakis, and David Hutchison, “Towards autonomic networks”, *in: Autonomic Networking*, Springer, 2006, pp. 1–11.
- [94] S. Schuetz et al., “Autonomic and decentralized management of wireless access networks”, *in: IEEE Transactions on Network and Service Management* 4.2 (2007), pp. 96–106, ISSN: 1932-4537.
- [95] F. Selmanovic and E. Skaljo, “GPON in Telecommunication Network”, *in: International Congress on Ultra Modern Telecommunications and Control Systems*, 2010, pp. 1012–1016, DOI: 10.1109/ICUMT.2010.5676500.
- [96] Boris Shishkov, Martijn Warnier, and Marten Van Sinderen, “On the Application of Autonomic and Context-aware Computing to Support Home Energy Management.”, *in: ICEIS (3)*, 2010, pp. 307–313.
- [97] P. W. Shumate, “Fiber-to-the-Home: 1977-2007”, *in: Journal of Lightwave Technology* 26.9 (2008), pp. 1093–1103, ISSN: 0733-8724, DOI: 10.1109/JLT.2008.923601.
- [98] B. Skubic et al., “A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON”, *in: IEEE Communications Magazine* 47.3 (2009), S40–S48, ISSN: 0163-6804, DOI: 10.1109/MCOM.2009.4804388.
- [99] Björn Skubic et al., “Dynamic bandwidth allocation in EPON and GPON”, *in: Convergence of Mobile and Stationary Next-Generation Networks* (2010), pp. 227–252.
- [100] “SPIRENT TESTCENTER PACKET GENERATOR AND ANALYZER BASE PACKAGE”, *in: Spirent Communications* (2010).
- [101] J. Strassner, J. W. K. Hong, and S. van der Meer, “The design of an Autonomic Element for Managing Emerging Networks and Services”, *in: 2009 International Conference on Ultra Modern Telecommunications Workshops*, 2009, pp. 1–8, DOI: 10.1109/ICUMT.2009.5345533.
- [102] John Strassner, Nazim Agoulmine, and Elyes Lehtihet, “Focale: A novel autonomic networking architecture”, *in: Latin American Autonomic Computing Symposium (LAACS)*, 2006, pp. 1–13.

-
- [103] K. Tsagkaris et al., “Identifying standardization opportunities of an operator-driven, framework for unifying autonomic network and service management”, in: *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 6921–6925, DOI: 10.1109/ICC.2012.6364850.
- [104] K. Tsagkaris et al., “Operator-Driven Framework for Establishing and Unifying Autonomic Network and Service Management Solutions”, in: *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, 2011, pp. 684–689, DOI: 10.1109/GLOCOMW.2011.6162540.
- [105] S. Uma, D. E. Geetha, and T. V. S. Kumar, “Prediction of Web Service response time using Grey system model”, in: *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016, pp. 139–143, DOI: 10.1109/ICATCCT.2016.7911980.
- [106] “Univerself project deliverable D2.1: UMF Specifications Release 1”, in: *Univerself deliverables*, 2011.
- [107] Muzaffer Uysal and John L Crompton, “An overview of approaches used to forecast tourism demand”, in: *Journal of Travel Research* 23.4 (1985), pp. 7–15.
- [108] ETSI TS 103 195-2 V1.1.1, “Autonomic network engineering for the self-managing Future Internet (AFI); Generic Autonomic Network Architecture; Part 2: An Architectural Reference Model for Autonomic Networking, Cognitive Networking and Self-Management”, in: *ETSI Technical Specification* (2018).
- [109] I. A. Venkatkumar and S. J. K. Shardaben, “Comparative study of data mining clustering algorithms”, in: *2016 International Conference on Data Science and Engineering (ICDSE)*, 2016, pp. 1–7, DOI: 10.1109/ICDSE.2016.7823946.
- [110] A. Walid and A. Chen, “Self-adaptive dynamic bandwidth allocation for GPON”, in: *Bell Labs Technical Journal* 15.3 (2010), pp. 131–139, ISSN: 1089-7089, DOI: 10.1002/bltj.20461.
- [111] Jie Wang et al., “DDoS attacks traffic and Flash Crowds traffic simulation with a hardware test center platform”, in: *Internet Security (WorldCIS), 2011 World Congress on*, IEEE, 2011, pp. 15–20.
- [112] Xingqi Wang et al., “Grey System Theory based prediction for topic trend on Internet”, in: *Engineering Applications of Artificial Intelligence* 29 (2014), pp. 191–200.
- [113] René Wies, “Policies in network and systems management—Formal definition and architecture”, in: *Journal of Network and Systems Management* 2.1 (1994), pp. 63–83, ISSN: 1573-7705, DOI: 10.1007/BF02141605, URL: <http://dx.doi.org/10.1007/BF02141605>.

- [114] M. Wodczak et al., “Standardizing a reference model and autonomic network architectures for the self-managing future internet”, *in: IEEE Network* 25.6 (2011), pp. 50–56, ISSN: 0890-8044, DOI: 10.1109/MNET.2011.6085642.
- [115] H. Woesner and D. Fritzsche, “SDN and OpenFlow for converged access/aggregation networks”, *in: 2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2013, pp. 1–3, DOI: 10.1364/OFC.2013.OTu3E.4.
- [116] L. Yi, X. Ke, and S. Junde, “Research on Forecasting and Early-Warning Methods”, *in: 2013 IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks*, 2013, pp. 570–576.
- [117] Mohamed Zait and Hammou Messatfa, “A comparative study of clustering methods”, *in: Future Generation Computer Systems* 13.2-3 (1997), pp. 149–159.
- [118] G. S. Zervas and D. Simeonidou, “Cognitive optical networks: Need, requirements and architecture”, *in: 2010 12th International Conference on Transparent Optical Networks*, 2010, pp. 1–4, DOI: 10.1109/ICTON.2010.5549176.
- [119] G Peter Zhang, “Time series forecasting using a hybrid ARIMA and neural network model”, *in: Neurocomputing* 50 (2003), pp. 159–175.
- [120] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu, “Forecasting with artificial neural networks: The state of the art”, *in: International journal of forecasting* 14.1 (1998), pp. 35–62.
- [121] K. Zhang, L. Huang, and Y. Chai, “An algorithm to adaptive determination of density threshold for density-based clustering”, *in: 2016 35th Chinese Control Conference (CCC)*, 2016, pp. 3929–3935, DOI: 10.1109/ChiCC.2016.7553965.

Titre: Maintenance autonome du réseau programmable d'accès optique de très haut débit

Mot clés : FTTH, PON, Apprentissage Automatique, Prédiction, Réseaux Autonomiques

Resumé : Les réseaux optiques passifs (PONs, Passive Optical Networks) représentant l'une des solutions les plus performantes du réseau d'accès FTTH ont été largement déployés par les opérateurs grâce à leur capacité d'offrir des services de très haut débit. Cependant, en raison de la dynamique du trafic des différents clients, ces réseaux doivent s'appuyer sur un mécanisme efficace pour l'allocation de ressources, plus particulièrement dans le sens montant. Ce mécanisme est actuellement limité par la nature statique des paramètres SLA (Service Level Agreement). Ceci peut avoir une influence négative sur la qualité de service ressentie par les utilisateurs. L'objectif de cette thèse est de proposer une nouvelle architecture pour optimiser l'allocation de ressources dans les réseaux PON tout en agissant uniquement sur les paramètres SLA, désignés comme des paramètres gérables par l'opérateur. Des techniques de classification basées sur l'apprentissage automatique et la prédiction sont utilisées pour analyser le comportement des différents utilisateurs et déterminer leurs tendances de trafic. Un ajustement dynamique sur la base du concept autonome de certains paramètres SLA est ensuite effectué afin de maximiser la satisfaction globale des clients vis-à-vis du réseau.

Title: Autonomic maintenance of high programmable optical access network

Keywords : FTTH, PON, Machine Learning, Forecasting, Autonomic Networking

Abstract : Passive Optical Network (PON) representing one of the most attractive FTTH access network solutions, have been widely deployed for several years thanks to their ability to offer high-speed services. However, due to the dynamicity of users traffic patterns, PONs need to rely on an efficient upstream bandwidth allocation mechanism. This mechanism is currently limited by the static nature of Service Level Agreement (SLA) parameters which can lead to an unoptimized bandwidth allocation in the network. The objective of this thesis is to propose a new management architecture for optimizing the upstream bandwidth allocation in PON while acting only on manageable parameters to allow the involvement of self-decision elements into the network. To achieve this, classification techniques based on machine learning approaches are used to analyze the behavior of PON users and to specify their upstream data transmission tendency. A dynamic adjustment of some SLA parameters is then performed to maximize the overall customers' satisfaction with the network.