ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών
Δίκτυα Επικοινωνιών Νέας Γενιάς και Κατανεμημένα Περιβάλλοντα Εφαρμογών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σύγχρονες τοπολογίες δικτύων κέντρων δεδομένων βασισμένες σε οπτικά δίκτυα

Αικατερίνη Δ. Τερζάκη
Α.Μ. 21009

Εισηγητής: Δρ Αντώνιος Μπόγρης, Καθηγητής

(Κενό φύλλο)

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σύγχρονες τοπολογίες δικτύων κέντρων δεδομένων βασισμένες σε οπτικά δίκτυα

Αικατερίνη Δ. Τερζάκη
Α.Μ. 21009

Εισηγητής: Antonios Bogris Digitally signed by Antonios Bogris Date: 2023.12.08 08:26:56 +02'00'

Δρ Αντώνιος Μπόγρης, Καθηγητής

Εξεταστική Επιτροπή:

Παναγιώτης Καρκαζής, Αναπληρωτής Καθηγητής
Νικόλαος Ψαρράς , Λέκτορας

Ημερομηνία εξέτασης 07/12/2023

Panagiotis Karkazis Digitally signed by Panagiotis Karkazis Date: 2023.12.11 14:03:12 +02'00'

Digitally signed by Nikolaos Psarras Date: 2023.12.12 00:16:22 +02'00'

(Κενό φύλλο)

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Τερζάκη Αικατερίνη του Δημητρίου, με αριθμό μητρώου 21009 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών Δίκτυα Επικοινωνιών Νέας Γενιάς και Κατανεμημένα Περιβάλλοντα Εφαρμογών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:
«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».
Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι …… και έπειτα από αίτηση μου στη Βιβλιοθήκη και έγκριση του επιβλέποντα καθηγητή.

Η Δηλούσα

Τερζάκη Αικατερίνη

(Κενό φύλλο)

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Αντώνιο Μπόγρη για την πολύτιμη βοήθεια και καθοδήγησή του σε όλη τη διάρκεια υλοποίησης της διπλωματικής μου εργασίας.

Ακόμα ένα μεγάλο ευχαριστώ στην οικογένειά μου που με στήριξε όλο αυτό το διάστημα και συνεχίζει να με στηρίζει σε κάθε μου προσπάθεια.

(Κενό φύλλο)

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία ασχολείται με τις ανάγκες που υπάρχουν στα σύγχρονα κέντρα δεδομένων και τους τρόπους αντιμετώπισης αυτών των αναγκών. Η αύξηση εφαρμογών και υπηρεσιών όπως υπηρεσίες υπολογιστικής νέφους, εφαρμογές μηχανικής μάθησης, υπηρεσίες κοινωνικής δικτύωσης και εφαρμογές ροής πολυμέσων υψηλής ευκρίνειας καθώς και ο τεράστιος όγκος δεδομένων έχει αλλάξει τις απαιτήσεις των κέντρων δεδομένων.

Οι τάσεις αυτές δημιουργούν νέες προκλήσεις. Ορισμένες από αυτές τις προκλήσεις είναι η κλιμάκωση των κέντρων δεδομένων, η παροχή υψηλού εύρους ζώνης, η χαμηλή καθυστέρηση και η αξιοποίηση με τον καλύτερο δυνατό τρόπο των διαθέσιμων πόρων. Ακόμα δημιουργείται η ανάγκη μείωσης της ενέργειας που καταναλώνεται, του κόστους και της πολυπλοκότητας καθώς επίσης η ανάγκη διαχείρισης και κάλυψης των απαιτήσεων πολλαπλών πελατών μέσα στο ίδιο κέντρο δεδομένων.

Τα κλασσικά ιεραρχικά δίκτυα που αποτελούνται από ηλεκτρικούς διακόπτες δεν μπορούν να ανταποκριθούν στις υψηλές απαιτήσεις των σύγχρονων κέντρων δεδομένων. Μια καλή πρακτική για την αντιμετώπιση αυτών των αναγκών είναι η χρήση οπτικών δικτύων. Οι οπτικές διασυνδέσεις και οι οπτικοί διακόπτες προσφέρουν υψηλότερο εύρος ζώνης και χαμηλή καθυστέρηση. Ακόμα η κλιμάκωση των δικτύων δημιουργεί την ανάγκη για βελτίωση της αξιοποίησης των διαθέσιμων πόρων πράγμα το οποίο οδηγεί στη χρήση διαχωρισμένων κέντρων δεδομένων. Οι αυστηρές απαιτήσεις που εισάγουν τα διαχωρισμένα κέντρα δεδομένων κάνει ακόμα πιο έντονη την ανάγκη για χρήση της οπτικής τεχνολογίας.

Στα πλαίσια της διπλωματικής εργασίας έχει γίνει διερεύνηση σύγχρονων τοπολογιών οπτικών δικτύων για την εξυπηρέτηση κέντρων δεδομένων μεγάλης κλίμακας.

ABSTRACT

The present thesis concerns the needs that exist in modern data centers and the ways to deal with these needs. The growth of applications and services such as cloud computing services, machine learning applications, social networking services and high-definition media streaming applications as well as the huge amount of data has changed the requirements of data centers.

These trends create new challenges. Some of these challenges are scaling data centers, providing high bandwidth, low latency, and making the best use of available resources. There is also the need to reduce energy consumption, cost and complexity as well as the need to manage and meet the demands of multiple clients within the same data center.

Classic hierarchical networks consisting of electrical switches cannot meet the high demands of modern data centers. A good practice to address these needs is to use optical networks. Optical interfaces and optical switches offer higher bandwidth and low latency. Also, the scaling of networks creates the need to improve the utilization of available resources, which leads to the use of disaggregated data centers. The strict requirements that disaggregated data centers introduce make the need for optical technology even more intense.

In the context of the present thesis, modern topologies of optical networks have been investigated to serve large-scale data centers.

ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ: Δίκτυα Επικοινωνιών
ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: προκλήσεις σύγχρονων δικτύων κέντρων δεδομένων, οπτικά δίκτυα, διαχωρισμένα κέντρα δεδομένων, SDN, NFV.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

AWS Amazon Web Services

IoT Internet of Things

ToR switch Top of Rack switch

TCAM Ternary Content-Addressable Memory

RDMA Remote Direct Memory Access

IEEE Institute of Electrical and Electronics Engineers

IETF Internet Engineering Task Force

DCE Data Center Ethernet

CEE Converged Enhanced Ethernet

QSFP-DD Quad Small Form-factor Pluggable Double Density

OSFP Octal Small Form-factor Pluggable

SDN Software Defined Networking

API Application Programming Interface

ONF Open Networking Foundation

TLS Transport Layer Security

SSL Secure Sockets Layer

SNMP Simple Network Management Protocol

CLI Command Line Interface

NFV Network Functions Virtualization

IANA Internet Assigned Numbers Authority

OSPF Open Shortest Path First

IS-IS Intermediate System – Intermediate System

VTEPs Virtual Tunnel End Points

ASIC Application Specific Integrated Circuit

ToC Top of Cluster

OCS Optical Circuit Switching

OPS Optical Packet Switching

AWG Arrayed Waveguide Grating

WSS Wavelength Selective Switch

SSS Spectrum Selective Switch

FPGA Field Programmable Gate Array

IM/DD Intensity Modulation / Direct Detection

NRZ-OOK Non Return to Zero On Off Keying

EDB Electrical Duo-Binary

PAM4 Four-level Pulse Amplitude Modulation

SDM Spatial-Division Multiplexing

WDM Wavelength-Division Multiplexing

SMF Single-Mode Fiber

MMF Multi-Mode Fiber

VCSEL Vertical-Cavity Surface-Emitting Laser

SiP Silicon Photonic

MEMS Micro-Electromechanical Systems

ΚΕΦΑΛΑΙΟ 1 : Κέντρα δεδομένων

1.1 Εισαγωγή στα κέντρα δεδομένων

Το κέντρο δεδομένων είναι ένας φυσικός χώρος που μπορεί να είναι ένα δωμάτιο, ένα κτίριο ή ολόκληρη εγκατάσταση η οποία στεγάζει υποδομή πληροφορικής με σκοπό την παροχή, εκτέλεση ή δημιουργία εφαρμογών και υπηρεσιών καθώς επίσης και την αποθήκευση και διαχείριση δεδομένων.

Τα σύγχρονα κέντρα δεδομένων διαφέρουν πολύ σε σχέση με τα παλαιότερα. Η παραδοσιακή υποδομή πληροφορικής με φυσικούς διακομιστές σε ιδιωτικές εγκαταστάσεις με αυστηρούς ελέγχους για χρήση αποκλειστικά από μία εταιρεία έχει αντικατασταθεί από κέντρα δεδομένων σε απομακρυσμένες εγκαταστάσεις ή κέντρα δεδομένων που ανήκουν σε παρόχους υπηρεσιών νέφους που προσφέρουν εικονική υποδομή πληροφορικής για κοινή χρήση από πολλούς πελάτες.

1.1.1 Τύποι κέντρων δεδομένων

Υπάρχουν διάφοροι τύποι κέντρων δεδομένων και οι εταιρείες μπορούν να χρησιμοποιήσουν περισσότερους από έναν τύπους ανάλογα με τις ανάγκες τους. Στη συνέχεια θα περιγράψουμε τέσσερις κύριους τύπους κέντρων δεδομένων [1].

Κέντρα δεδομένων επιχειρήσεων. Σε αυτό τον τύπο κέντρου δεδομένων όλη η υποδομή και τα δεδομένα βρίσκονται μέσα σε εγκαταστάσεις της εταιρείας. Η εταιρεία είναι υπεύθυνη για τη δημιουργία, τη συντήρηση, την παρακολούθηση και τη διαχείριση του κέντρου δεδομένων. Οι εταιρείες συνήθως επιλέγουν αυτό τον τύπο όταν θέλουν να έχουν μεγαλύτερο έλεγχο για την ασφάλεια των δεδομένων τους.

Ένας άλλος τύπος είναι τα κέντρα δεδομένων διαχειριζόμενων υπηρεσιών. Σε αυτή την περίπτωση η εταιρεία νοικιάζει την υποδομή και τον εξοπλισμό από κάποιον πάροχο κέντρου δεδομένων. Η διαχείριση του κέντρου δεδομένων γίνεται από τον πάροχο ή από κάποιο τρίτο για λογαριασμό της εταιρείας.

Ακόμα υπάρχουν κέντρα δεδομένων συντοπισμού. Στα κέντρα δεδομένων συντοπισμού η εταιρεία μισθώνει ένα χώρο εντός ενός κέντρου δεδομένων

που ανήκει σε κάποιον άλλο. Το κέντρο δεδομένων συντοπισμού παρέχει την υποδομή όπως το φυσικό χώρο, τα συστήματα ψύξης, ασφάλειας, εύρος ζώνης, δικτυακή υποδομή και η εταιρεία παρέχει το υλικό όπως τους διακομιστές, τα μέσα αποθήκευσης, το τείχος προστασίας (firewall) και τη διαχείριση.

Οι δύο αυτοί τύποι κέντρων δεδομένων συνήθως επιλέγονται όταν οι εταιρείες δεν διαθέτουν τον χώρο, την τεχνογνωσία ή το προσωπικό που χρειάζεται για τη δημιουργία και διαχείριση ενός κέντρου δεδομένων.

Τέλος θα αναφερθούμε στα κέντρα δεδομένων νέφους. Σε αυτή την περίπτωση οι πάροχοι κέντρου δεδομένων φιλοξενούν εφαρμογές και δεδομένα και παρέχουν πόρους υποδομής για κοινή χρήση μεταξύ μεγάλου αριθμού πελατών (δεκάδες έως εκατομμύρια) μέσω του διαδικτύου. Παραδείγματα μεγάλων παρόχων νέφους είναι η Amazon Web Services (AWS), η Microsoft (Azure), η Google Cloud Platform ή το IBM Cloud [2].

1.1.2 Εξέλιξη υπολογιστικής υποδομής

Τα τελευταία 65 χρόνια η υπολογιστική υποδομή έχει περάσει από τρία βασικά στάδια εξέλιξης.

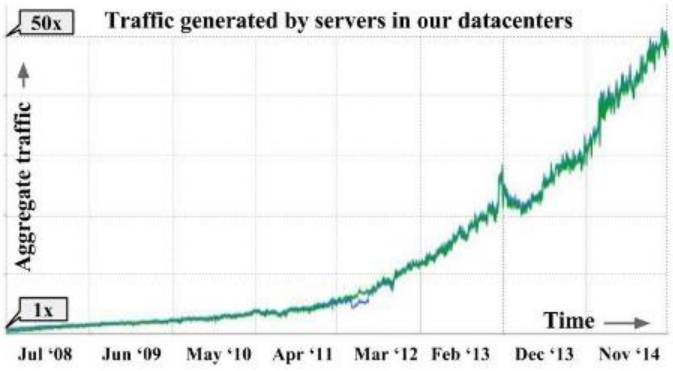
Το πρώτο στάδιο περιλαμβάνει τη μετάβαση από τη χρήση μεγάλων ιδιόκτητων υπολογιστών γνωστών με το όρο mainframes σε διακομιστές x86 οι οποίοι τοποθετούνται σε εσωτερικές εγκαταστάσεις και η διαχείριση τους γίνεται από εσωτερικά τμήματα πληροφορικής των ίδιων των εταιρειών.

Το δεύτερο στάδιο περιλαμβάνει την εισαγωγή της εικονικοποίησης της υποδομής που υποστηρίζει τις εφαρμογές, επιτρέποντας με αυτό τον τρόπο την καλύτερη και αποδοτικότερη αξιοποίηση των πόρων και την μεταφορά του φορτίου εργασίας σε ομάδες φυσικής υποδομής.

Και τέλος το τρίτο στάδιο αφορά τη μετάβαση στο νέφος το οποίο κυριαρχεί στις μέρες μας και αφορά εφαρμογές που δημιουργήθηκαν στο νέφος.

1.1.3 Βασικά στοιχεία ενός κέντρου δεδομένων

Ο σχεδιασμός ενός κέντρου δεδομένων αποτελείται από εκατοντάδες χιλιάδες διακομιστές, δρομολογητές, διακόπτες, συστήματα αποθήκευσης και ζεύξεις υψηλής ταχύτητας, ελεγκτές παράδοσης εφαρμογών και τείχη προστασίας (firewalls). Επιπλέον για την υποστήριξη του είναι απαραίτητη η διασφάλιση ύπαρξης συστημάτων ψύξης, εξαερισμού, καταστολής πυρκαγιάς καθώς επίσης και εφεδρικών γεννητριών, υποσυστημάτων τροφοδοσίας και τροφοδοτικών για την αδιάλειπτη παροχή ενέργειας. Κατά τον σχεδιασμό ενός κέντρου δεδομένων θα πρέπει να λαμβάνεται υπόψιν η τοποθεσία δημιουργίας του κέντρου δεδομένων, το κόστος της ενέργειας, η αγορά ακίνητης περιουσίας και οι απαιτήσεις υλικού και λογισμικού. Στόχος είναι η μεγιστοποίηση της απόδοσης του κέντρου δεδομένων διατηρώντας το κόστος χαμηλό.



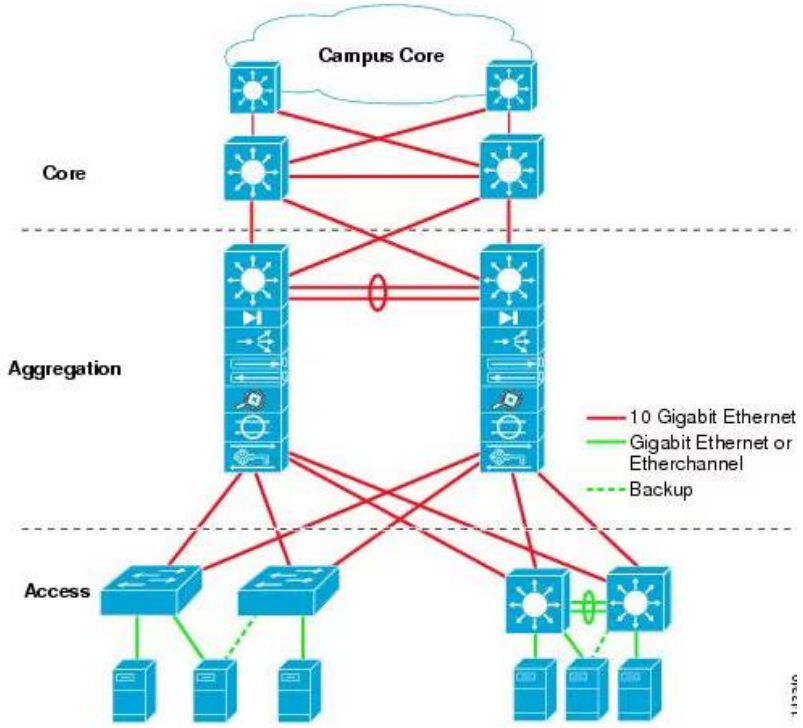
Εικόνα 1.1: Συγκεντρωτική κίνηση διακομιστών στα κέντρα δεδομένων της Google ("Juniper Rising: A Decade of ClosTopologies and Centralized Control in Google's Datacenter Network", Singh et.al., 2015)

Οι χωρητικότητες των κέντρων δεδομένων αυξάνονται συνεχώς με πολύ γρήγορους ρυθμούς. Οι απαιτήσεις εύρους ζώνης διπλασιάζονται κάθε 12 με 15 μήνες [3]. Η αύξηση αυτή είναι αποτέλεσμα ορισμένων τάσεων που υπάρχουν. Οι εφαρμογές και οι υπηρεσίες ιστού αυξάνονται και προσφέρουν αποτελέσματα υψηλότερης ποιότητας. Ακόμα το σύνολο των δεδομένων αυξάνεται συνεχώς. Υπάρχουν περισσότερα αρχεία καταγραφής και αρχεία με βίντεο και εικόνες. Επίσης λόγω του Διαδικτύου των Πραγμάτων (IoT)

αυξάνεται συνεχώς ο αριθμός των αισθητήρων που συνδέονται στο Διαδίκτυο. Αυτό έχει ως αποτέλεσμα να αυξάνεται το σύνολο των δεδομένων που συλλέγονται και επεξεργάζονται. Ακόμα η αξιοποίηση ομάδων φυσικής υποδομής για την κάλυψη αναγκών των εφαρμογών και υπηρεσιών δημιουργεί έντονη χρήση μηνυμάτων επικοινωνίας και κίνηση δεδομένων [4].

1.2 Κλασσικός σχεδιασμός κέντρου δεδομένων

Ένας κλασσικός σχεδιασμός δικτύου κέντρου δεδομένων βασίζεται σε μία ιεραρχική δομή που μοιάζει με δέντρο και αποτελείται από δύο ή τρία επίπεδα [5]. Η αρχιτεκτονική τριών επιπέδων χρησιμοποιείται συνήθως για να υποστηρίξει μεγαλύτερα δίκτυα και περιλαμβάνει το επίπεδο πυρήνα, το επίπεδο συνάθροισης και το επίπεδο πρόσβασης.



Εικόνα 1.2: Αρχιτεκτονική κέντρου δεδομένων. Βασικός σχεδιασμός σε επίπεδα (www.cisco.com, 2023)

Στο επίπεδο πρόσβασης γίνεται η σύνδεση των διακομιστών στο δίκτυο. Συνήθως υπάρχουν 20 έως 40 διακομιστές τοποθετημένοι σε κάθε ικρίωμα οι οποίοι συνδέονται στο δίκτυο μέσω των διακοπτών πρόσβασης που βρίσκονται στην κορυφή κάθε ικριώματος (Top of Rack - ToR switch) [6],[7]. Η σύνδεση αυτή γίνεται μέσω θυρών GigE. Κάθε διακόπτης πρόσβασης συνδέεται με δύο διακόπτες συνάθροισης στο επίπεδο δύο, οι οποίοι με τη σειρά τους συνδέονται μέσω πολλαπλών διαδρομών με το επίπεδο πυρήνα έτσι ώστε να μην υπάρχει μοναδικό σημείο αστοχίας. Οι διακόπτες συνάθροισης αποτελούνται από θύρες υψηλών ταχυτήτων 10 GigE και παρέχουν σημαντικές λειτουργίες και υπηρεσίες όπως ορισμός τομέα επιπέδου 2, εκτέλεση πρωτοκόλλου Spanning Tree, καθορισμός εναλλακτικής προεπιλεγμένης πύλης, εξισορρόπηση φορτίου, ανάλυση δικτύου και ανίχνευση εισβολών, λειτουργίες τείχους προστασίας και άλλα. Στη κορυφή της ιεραρχίας το επίπεδο πυρήνα αποτελείται και αυτό από ζεύξεις υψηλής ταχύτητας 10 GigE καθώς είναι αυτό που μεταφέρει τις ροές μέσα και έξω από το κέντρο δεδομένων. Οι συσκευές αυτού του επιπέδου χρησιμοποιούν γνωστά πρωτόκολλα όπως Open Shortest Path First και Enhanced Interior Gateway Protocol για την εσωτερική δρομολόγηση και μπορούν να κάνουν εξισορρόπηση φορτίου μεταξύ των δύο ανώτερων επιπέδων χρησιμοποιώντας τον αλγόριθμο Cisco Express Forwarding.

1.2.1 Μειονεκτήματα κλασσικής αρχιτεκτονικής

Η κλασσική αυτή αρχιτεκτονική των κέντρων δεδομένων παρουσιάζει όπως θα δούμε στη συνέχεια αρκετά μειονεκτήματα [7] τα οποία οδήγησαν στη αναζήτηση και δημιουργία νέων αρχιτεκτονικών. Ένα από τα σημαντικότερα μειονεκτήματα είναι η δημιουργία συμφόρησης λόγω της αναλογίας υπερσυνδρομής. Η αναλογία υπερσυνδρομής υπολογίζεται από το συνολικό εύρος ζώνης σύνδεσης των διακομιστών προς το συνολικό εύρος ζώνης ανερχόμενης ζεύξεις. Η αναλογία αυτή στη κλασσική αρχιτεκτονική αυξάνεται γρήγορα καθώς προχωράμε στα ανώτερα επίπεδα και δημιουργεί συμφόρηση στους διακόπτες συνάθροισης και πυρήνα. Η αναλογία υπερσυνδρομής των διακομιστών θα έπρεπε να είναι όσο το δυνατόν πιο κοντά στο 1:1 έτσι ώστε η επικοινωνία να γίνεται κάνοντας χρήση του συνολικού εύρους ζώνης. Ακόμα λόγω της συμφόρηση που δημιουργείται από την υπερσυνδρομή μπορεί να

γεμίσει η μνήμη των διακοπτών και αυτό θα έχει ως αποτέλεσμα να αρχίσουν να πετάνε πακέτα. Το ίδιο συμβαίνει και στην περίπτωση που ένας διακόπτης λαμβάνει ταυτόχρονα πακέτα από πολλούς αποστολείς. Δεν υπάρχει κάποιος μηχανισμός για να προλαμβάνει την πτώση των πακέτων.

Ένα ακόμη μειονέκτημα της κλασσικής αρχιτεκτονικής είναι η χαμηλή χρησιμοποίηση των πόρων. Αυτό οφείλεται στη χρήση του πρωτοκόλλου Spanning Tree σύμφωνα με το οποίο αν και υπάρχουν πολλαπλές διαδρομές αξιοποιείται μόνο μία για αποφυγή βρόγχων. Επίσης ένα επιπλέον θέμα είναι η εξισορρόπηση φορτίου. Οι ζεύξεις στα ανώτερα επίπεδα επιβαρύνονται πολύ περισσότερο ενώ θα έπρεπε το φορτίο να κατανέμεται ομοιόμορφα σε όλο το δίκτυο. Σημαντικό ρόλο παίζει η ανοχή σε σφάλματα η οποία μπορεί να αφορά κάποιο διακομιστή, κάποιο διακόπτη ή τμήμα ζεύξης. Στα ανώτερα επίπεδα μια αστοχία για παράδειγμα ενός διακόπτη θα οδηγήσει σε σημαντική πτώση της απόδοσης του δικτύου λόγω της χαμηλής φυσικής συνδεσιμότητας που υπάρχει.

Τα σύγχρονα κέντρα δεδομένων θα πρέπει να μπορούν να επεκταθούν ώστε να μπορούν να υποστηρίζουν ολοένα και περισσότερους διακομιστές. Σε μία συμβατική τοπολογία κέντρου δεδομένων για να αυξήσουμε τη χωρητικότητα χρειάζεται να αναβαθμίσουμε ή να αντικαταστήσουμε τον εξοπλισμό με άλλο υψηλότερων δυνατοτήτων κάτι το οποίο κοστίζει και σε χρήμα και σε χρόνο. Οι διακόπτες του επιπέδου συνάθροισης και πυρήνα λόγω των σημαντικών λειτουργιών που παρέχουν όπως αναφέραμε πιο πάνω και των θυρών που έχουν είναι πολύ ακριβοί και καταναλώνουν πολύ ενέργεια με αποτέλεσμα να αυξάνεται τόσο το κεφαλαιουχικό όσο και λειτουργικό κόστος του κέντρου δεδομένων [8].

Σημαντική κατανάλωση ενέργειας γίνεται επίσης από τους διακομιστές καθώς και από τα συστήματα ψύξης που απαιτούνται. Σύμφωνα με την έρευνα [9] ένας διακομιστής καταναλώνει περίπου το μισό της πλήρους ισχύς του ακόμη και όταν είναι σε αδράνεια. Ακόμη έχει αποδειχθεί ότι η μέση χρήση διακομιστών στα κέντρα δεδομένων είναι συνήθως γύρο στο 30%. Μια καλή πρακτική θα ήταν η δυναμική ανακατανομή των πόρων μεταξύ των διακομιστών ώστε να συγκεντρωθούν οι λειτουργίες στο 30% των

διακομιστών και να κλείσουν οι υπόλοιποι. Αυτό θα είχε ως αποτέλεσμα την εξοικονόμηση ενέργειας. Ωστόσο λόγω της ιεραρχικής δομής και της μεγάλης αναλογίας υπερσυνδρομής στο επίπεδο συνάθροισης και πυρήνα οι πόροι είναι κατακερματισμένοι και απομονωμένοι και δεν μπορεί να γίνει δυναμική ανακατανομή των διακομιστών μεταξύ των εφαρμογών του κέντρου δεδομένων.

1.3 Κατηγορίες αρχιτεκτονικών

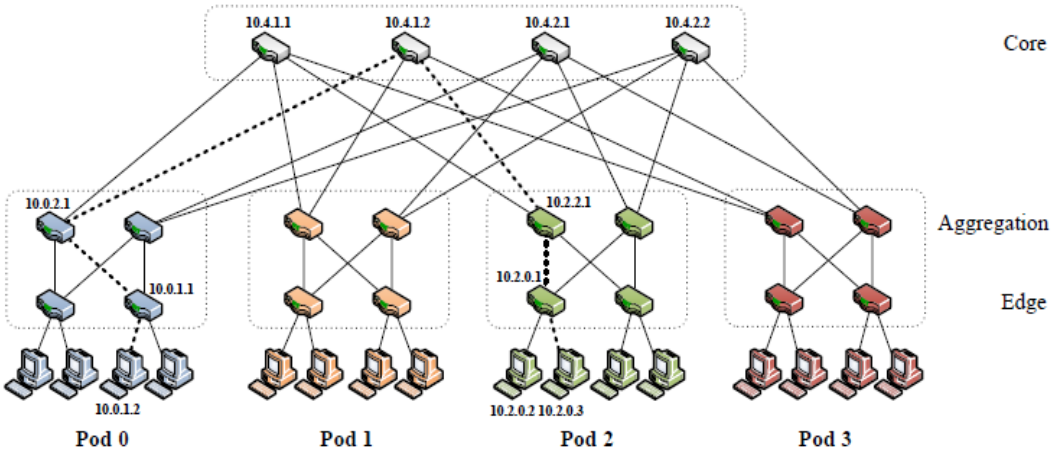
Τα μειονεκτήματα της κλασσικής αρχιτεκτονικής οδήγησαν στη δημιουργία νέων αρχιτεκτονικών. Οι αρχιτεκτονικές των κέντρων δεδομένων μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες [6]. Η πρώτη κατηγορία περιέχει τις αρχιτεκτονικές οι οποίες χρησιμοποιούν τους διακόπτες ως βασικά στοιχεία για την διασύνδεση και δρομολόγηση. Σε αυτή την κατηγορία ανήκει η κλασσική αρχιτεκτονική. Άλλες τέτοιου τύπου αρχιτεκτονικές είναι οι Fat-tree [10] την οποία θα εξηγήσουμε παρακάτω, VL2, Portland και άλλες. Στη δεύτερη κατηγορία ανήκουν αρχιτεκτονικές οι οποίες χρησιμοποιούν διακομιστές με πολλαπλές κάρτες δικτύου για τη διασύνδεση και δρομολόγηση πακέτων. Παραδείγματα τέτοιων αρχιτεκτονικών είναι η Bcube, Dcell, Ficon και άλλες. Επιπλέον με βάση την τεχνολογία υποδομής που χρησιμοποιείται ταξινομούμε τα κέντρα δεδομένων σε ηλεκτρικά και οπτικά. Υπάρχουν επίσης και υβριδικές καταστάσεις με ηλεκτρο – οπτικά κέντρα δεδομένων.

1.3.1 Αρχιτεκτονική Fat-tree

Στη συνέχεια θα περιγράψουμε την τοπολογία fat-tree [10]. Η τοπολογία fat-tree είναι μία ειδική περίπτωση μίας Clos τοπολογίας. Αποτελείται από k ομάδες, ο αριθμός των οποίων καθορίζεται από τον αριθμό των θυρών των διακοπτών. Οι διακόπτες που χρησιμοποιούνται είναι ίδιοι σε όλα τα επίπεδα δίνοντας έτσι το πλεονέκτημα χρήσης απλών, χαμηλού κόστους και χαμηλής κατανάλωσης διακοπτών.

Κάθε ομάδα σχηματίζεται από k/2 διακόπτες συνάθροισης και k/2 διακόπτες άκρου. Κάθε διακόπτης άκρου συνδέεται με k/2 διακομιστές και k/2 διακόπτες συνάθροισης. Επίσης κάθε διακόπτης συνάθροισης συνδέεται με k/2

διακόπτες πυρήνα και k/2 διακόπτες άκρου. Υπάρχουν $(k/2)^2$ διακόπτες πυρήνα. Η i θύρα κάθε διακόπτη πυρήνα συνδέεται στη i ομάδα. Κάθε ομάδα συνδέεται με όλους τους διακόπτες πυρήνα. Ο συνολικός αριθμών διακομιστών που μπορούν να υποστηριχθούν είναι $k^3 / 4$.



Εικόνα 1.3: Τοπολογία fat-tree (“A Scalable, Commodity Data Center Network Architecture”, Al-Fares et.al., 2008)

Για να μπορέσουμε να αξιοποιήσουμε το μέγιστο εύρος ζώνης πρέπει να εκμεταλλευτούμε τη δομή της τοπολογίας η οποία μας παρέχει πολλαπλά μονοπάτια ίσου κόστους με σκοπό το διαμοιρασμό της κίνησης. Προκειμένου να επιτευχθεί αυτό γίνεται χρήση ενός πίνακα δρομολόγησης δύο επιπέδων. Οι IP διευθύνσεις που εκχωρούνται ακολουθούν μια συγκεκριμένη μορφή. Στους διακόπτες των ομάδων δίνονται IP διευθύνσεις της μορφής 10.pod.switch.1 όπου το pod δείχνει την ομάδα στη οποία ανήκει ο διακόπτης και το switch υποδεικνύει τη θέση του διακόπτη στη ομάδα. Οι IP διευθύνσεις των διακοπτών πυρήνα είναι της μορφής 10.k.j.i όπου τα j και i υποδηλώνουν τις συντεταγμένες του διακόπτη στο πλέγμα διακοπτών πυρήνα. Τέλος στους διακομιστές εκχωρούνται διευθύνσεις της μορφής 10.pod.switch.ID όπου το ID υποδηλώνει τη θέση του διακομιστή στο υποδίκτυο. Αυτός ο τρόπος διευθυνσιοδότησης απλοποιεί τη δημιουργία των πινάκων δρομολόγησης.

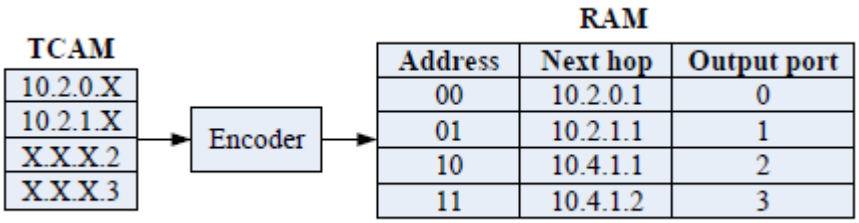
Οι πίνακες δρομολόγησης έχουν τροποποιηθεί, ώστε να γίνεται αναζήτηση προθέματος δύο επιπέδων. Κάθε εγγραφή σε έναν αρχικό πίνακα δρομολόγησης ενδέχεται να έχει έναν δείκτη σε ένα δεύτερο μικρότερο πίνακα εγγραφών. Η αναζήτηση ενός προθέματος στον αρχικό πίνακα μπορεί να δείχνει απευθείας σε μία θύρα εξόδου και να τερματίζει. Αν η αναζήτηση δώσει ένα μη τερματικό αποτέλεσμα τότε χρησιμοποιείτε το επίθημα με την μεγαλύτερη αντιστοίχιση που βρίσκεται στον δεύτερο πίνακα. Στην παρακάτω εικόνα φαίνεται ένα παράδειγμα όπου ένα εισερχόμενο πακέτο με IP διεύθυνση προορισμού 10.2.1.2 προωθείται στη θύρα 1, ενώ ένα πακέτο με IP διεύθυνση προορισμού 10.3.0.3 προωθείται στη θύρα 3.

Prefix	Output port
10.2.0.0/24	0
10.2.1.0/24	1
0.0.0.0/0	

Suffix	Output port
0.0.0.2/8	2
0.0.0.3/8	3

Εικόνα 1.4: Παράδειγμα πίνακα δύο επιπέδων. (“A Scalable, Commodity Data Center Network Architecture”, Al-Fares et.al., 2008)

Η αναζήτηση δύο επιπέδων εισάγει μια μικρή καθυστέρηση. Για το λόγο αυτό η υλοποίηση της γίνεται κάνοντας χρήση κατάλληλου υλικού. Η τριαδική μνήμη διεύθυνσης περιεχομένου (TCAM) μπορεί να αποθηκεύει διευθύνσεις και να εκτελεί παράλληλες αναζητήσεις μεταξύ των εγγραφών. Οι μνήμες αυτές έχουν χαμηλή πυκνότητα αποθήκευσης, είναι ακριβές και καταναλώνουν πολύ ενέργεια ωστόσο μπορούν να χρησιμοποιηθούν σε αυτή την αρχιτεκτονική γιατί οι πίνακες δρομολόγησης είναι σχετικά μέτριου μεγέθους. Στην παρακάτω εικόνα απεικονίζεται η υλοποίηση της αναζήτησης δύο επιπέδων με TCAM.



Εικόνα 1.5: Υλοποίηση πίνακα δρομολόγησης δύο επιπέδων με TCAM (“A Scalable, Commodity Data Center Network Architecture”, Al-Fares et.al., 2008)

Η TCAM μνήμη αποθηκεύει διευθύνσεις προθεμάτων και επιθεμάτων οι οποίες έχουν έναν δείκτη σε μία μνήμη RAM που αποθηκεύει την IP διεύθυνση για το επόμενο άλμα και την θύρα εξόδου για αυτή την διεύθυνση. Η έξοδος της μνήμης TCAM κωδικοποιείται ώστε το αποτέλεσμα να είναι η εγγραφή με το μικρότερο ταίριασμα διεύθυνσης και να ικανοποιείται η ανάγκη για αναζήτηση δύο επιπέδων.

Η αρχιτεκτονική λοιπόν του fat-tree η οποία παρουσιάζεται από τους Al-Fares, Loukissa και Vahdat (2008) δίνει λύση στα προβλήματα υπερσυνδρομής, επεκτασιμότητας, υψηλού κόστους των διακοπτών συνάθροισης και πυρήνα και ανοχής σε σφάλματα.

1.4 Προσεγγίσεις κλιμάκωσης

Όταν οι απαιτήσεις χωρητικότητας αυξάνονται τότε υπάρχουν δύο δυνατοί τρόποι αντιμετώπισης.

Η μία λύση είναι η αναβάθμιση ή η επιλογή ενός μηχανήματος με μεγαλύτερες δυνατότητες. Αυτή την προσέγγιση ακολουθούσαν όπως αναφέραμε πιο πάνω στα παραδοσιακά κέντρα δεδομένων. Είναι γνωστή ως Scale Up. Η λύση αυτή κάνει πιο απλή και εύκολη τη διαχείριση ωστόσο έχει κάποια μειονεκτήματα. Τα μηχανήματα μεγάλων δυνατοτήτων είναι πολύ πιο ακριβά σε σχέση με τα μικρά μηχανήματα. Επίσης ακόμα και τα μεγάλα μηχανήματα επαρκούν μέχρι ένα όριο το οποίο μπορεί να μην καλύπτει τις ανάγκες μας για επέκταση. Ακόμα είναι ακριβό, πολύπλοκο και σε κάποιες περιπτώσεις μπορεί να μην είναι δυνατό να αλλάξουμε το μέγεθος ενός μηχανήματος.

Τέλος η χρήση ενός μεγάλου μηχανήματος δεν μας προσφέρει ανοχή σε σφάλματα. Στην περίπτωση μίας αστοχίας συστήματος η υπηρεσία θα διακοπεί και έτσι δεν καλύπτεται η ανάγκη για διαθεσιμότητα [4].

Η δεύτερη λύση είναι η ομαδοποίηση πολλαπλών μικρών μηχανημάτων χρησιμοποιώντας κατάλληλο λογισμικό το οποίο σχηματίζει μια ενοποιημένη προβολή συστήματος. Η λύση αυτή είναι γνωστή ως οριζόντια κλιμάκωση - Scale Out . Χρησιμοποιείται σε περιπτώσεις που η ζήτηση για νέες υπηρεσίες και αποθήκευση ξεπερνάει τη χωρητικότητα ενός μηχανήματος. Το σημαντικό πλεονέκτημα είναι ότι με αυτό τον τρόπο μπορούμε να έχουμε τη ζητούμενη χωρητικότητα με πολύ μικρότερο κόστος. Χρησιμοποιείται στα σύγχρονα κέντρα δεδομένων σε πολλές περιπτώσεις για να καλύψει ανάγκες διακομιστών ιστού, βάσεων δεδομένων, εφαρμογών, αρχείων και μπλοκ αποθήκευσης παρέχοντας με αυτό τον τρόπο βελτιωμένη απόδοση και χωρητικότητα.

Αντίστοιχα για την δικτύωση των κέντρων δεδομένων οι παραδοσιακή αρχιτεκτονική ακολουθεί μία προσέγγιση scale up. Χρησιμοποιεί μεγάλους, ακριβούς διακόπτες συνάθροισης και πυρήνα, οι οποίοι καταναλώνουν μεγάλα ποσά ενέργειας και δημιουργούν όπως έχουμε αναφέρει θέματα συμφόρησης. Για την αύξηση της χωρητικότητας απαιτείται αναβάθμιση ή αντικατάσταση του εξοπλισμού. Οι σύγχρονες αρχιτεκτονικές εφαρμόζουν μία οριζόντια κλιμάκωση όπως για παράδειγμα είδαμε στην αρχιτεκτονική fat-tree, αντικαθιστώντας τους μεγάλους και ακριβούς διακόπτες με πολλούς μικρούς διακόπτες προσφέροντας σημαντικά οφέλη τόσο στην απόδοση όσο και στο κόστος.

1.5 Ανάγκες σύγχρονων κέντρων δεδομένων

Οι ανάγκες των σύγχρονων κέντρων δεδομένων διαφέρουν πολύ σε σχέση με παλαιότερα. Στα παραδοσιακά κέντρα δεδομένων το κύριο φορτίο εργασίας ήταν μεταξύ πελάτη και εξυπηρετητή με αποτέλεσμα το μεγαλύτερο ποσοστό της κίνησης να είναι από έξω προς το εσωτερικό του κέντρου δεδομένων και αντίστροφα. Είναι αυτό που ονομάζουμε κίνηση μεταξύ Βορρά-Νότου. Στα σύγχρονα κέντρα δεδομένων το κύριο φορτίο εργασίας είναι μεταξύ των διακομιστών [11]. Υπάρχει πολύ έντονη επικοινωνία μεταξύ δεκάδων,

εκατοντάδων έως και χιλιάδων διακομιστών κατά μήκος όλου του δικτύου του κέντρου δεδομένων. Είναι αυτό που ονομάζουμε κίνηση μεταξύ Ανατολής-Δύσης. Η κίνηση αυτή δημιουργεί τεράστια ανάγκη για χαμηλή καθυστέρηση στο δίκτυο, υψηλή απόδοση και αποφυγή απώλειας πακέτων ώστε να επιτευχθεί η ζητούμενη απόδοση των εφαρμογών [12][13].

1.6 Αρχιτεκτονική Leaf- Spine

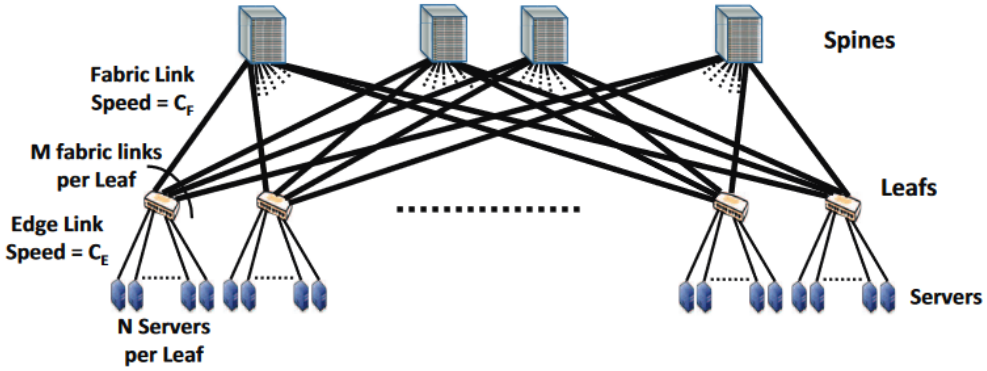
Η αρχιτεκτονική leaf-spine χρησιμοποιείτε στα σύγχρονα κέντρα δεδομένων. Έχει σχεδιαστεί για να παρέχει με ένα προβλέψιμο τρόπο χαμηλή καθυστέρηση στο δίκτυο και κλιμακούμενη απόδοση. Προσεγγίζει το ιδανικό δίκτυο όπου όλοι οι διακομιστές είναι απευθείας συνδεδεμένοι σε ένα μεγάλο διακόπτη χωρίς αποκλεισμό [14]. Καλύπτει ανάγκες διασύνδεσης υψηλής ταχύτητας εντός του κέντρου δεδομένων και αντιμετωπίζει προβλήματα λόγω της γρήγορης αύξησης της κίνησης και της συνεχούς κλιμάκωσης του κέντρου δεδομένων.

Στην αρχιτεκτονική αυτή γνωρίζουμε πάντα πόσα άλματα θα κάνει κάθε πακέτο, το φορτίου του δικτύου και την απόσταση που θα διανύσει με αποτέλεσμα να μειώνεται η καθυστέρηση στο δίκτυο [15]. Όπως και στην fat-tree αρχιτεκτονική έτσι και εδώ το πρωτόκολλο Spanning Tree έχει διαγραφεί και χρησιμοποιούνται τα πολλαπλά μονοπάτια για κλιμάκωση του εύρους ζώνης [14][15].

Η αποτελεσματικότητα δικτύου του κέντρου δεδομένων βελτιώνεται σημαντικά εφαρμόζοντας την leaf-spine αρχιτεκτονική, καθιστώντας την για αυτό το λόγο κατάλληλη για κέντρα δεδομένων υψηλής απόδοσης ή υψηλής πυκνότητας εύρους ζώνης. Όταν η κίνηση στο δίκτυο αυξάνεται σημαντικά μπορούν να προστεθούν επιπλέον διακόπτες για να μπορεί να γίνει εξισορρόπηση φορτίου και να αποφεύγονται καταστάσεις συμφόρησης. Παρέχει δυνατότητες υψηλής κλιμάκωσης και μπορεί να προσαρμοστεί σε μικρά, μεσαία και μεγάλα κέντρα δεδομένων.

Χρησιμοποιεί μία σχεδίαση δύο επιπέδων. Στο υψηλότερο επίπεδο είναι οι διακόπτες σπονδυλικής στήλης (spine switches) οι οποίοι μπορούν να θεωρηθούν ως διακόπτες πυρήνα. Στο χαμηλότερο επίπεδο είναι οι διακόπτες

φύλλου (leaf switches) οι οποίοι μπορούν να θεωρηθούν ως διακόπτες πρόσβασης μέσω των οποίων συνδέονται οι διακομιστές στο δίκτυο. Η δομή αυτή βοηθάει να μειωθεί ο αριθμός των αλμάτων και μειώνει την καθυστέρηση στο δίκτυο.



Εικόνα 1.6: Leaf-Spine αρχιτεκτονική κέντρου δεδομένων. (“On the Data Path Performance of Leaf-Spine Datacenter Fabrics”, Alizadeh et.al., 2013)

Οι διακόπτες σπονδυλικής στήλης αποτελούνται από μεγάλο αριθμό θυρών για λόγους επεκτασιμότητας. Κάθε θύρα τους χρησιμοποιείται για σύνδεση με ένα διακόπτη φύλλου καθορίζοντας με αυτό τον τρόπο το μέγιστο αριθμό διακοπτών φύλλου που μπορεί να υπάρξουν, οι οποίοι με τη σειρά τους καθορίζουν το μέγιστο αριθμό διακομιστών που μπορούν να συνδεθούν. Το πλήθος των διακοπτών σπονδυλικής στήλης καθορίζεται από ένα συνδυασμό του αριθμου των θυρών τους, της επιθυμητής απόδοσης μεταξύ των διακοπτών φύλλου και του ζητούμενου πλεονασμού [16]. Οι διακόπτες σπονδυλικής στήλης θα πρέπει να έχουν υψηλή χωρητικότητα, μεγάλη κρυφή μνήμη και εικονική απόδοση. Συνήθως αποτελούνται από θύρες 10G, 25G, 40G και 100G, πλήρη πρωτόκολλα και κατάλληλες λειτουργίες εφαρμογών για να μπορεί να επιτευχθεί γρήγορη ανάπτυξη του δικτύου.

Οι διακόπτες φύλλου είναι κοινά χρησιμοποιούμενες συσκευές οι οποίες συνδέονται με κάθε διακόπτη σπονδυλικής στήλης και παρέχουν όπως έχουμε αναφέρει πρόσβαση των διακομιστών στο δίκτυο. Συνήθως χρησιμοποιούνται θύρες 40G ή 100G για ανοδικές ζεύξεις και θύρες 10G,

25G, 40G, 50G, 100G για καθοδικές ζεύξεις. Καθώς το πλήθους των διακομιστών αυξάνεται συνεχώς είναι καλό να επιλέγονται διακόπτες με μεγαλύτερους ρυθμούς και μεγαλύτερο πλήθος θυρών. Επιπλέον για να αποφεύγονται καταστάσεις συμφόρησης η αναλογία υπερσυνδρομής δεν πρέπει να ξεπερνά το 3:1 [15][16].

1.7 Πρότυπα επικοινωνιών δικτύωσης

Τα σύγχρονα κέντρα δεδομένων γίνονται όλο και πιο πυκνά, πιο εικονικοποιημένα, δημιουργώντας ανάγκες για μεγαλύτερη απόδοση, υψηλότερο εύρος ζώνης, αξιοποίηση πολλαπλών μονοπατιών, προβλεπόμενη συμπεριφορά, μικρότερες καθυστερήσεις, μικρότερη κατανάλωση ενέργειας και απομόνωση κατηγορία κίνησης. Το παραδοσιακό Ethernet δεν μπορεί να καλύψει αυτές τις ανάγκες. Δεν μπορεί να υποστηρίξει καλά τις απαιτήσεις κλιμάκωσης. Όσο το κέντρο δεδομένων κλιμακώνεται , το κόστος μεγαλώνει εκθετικά, η απόδοση μειώνετε και αυξάνεται η πολυπλοκότητα διαχείρισης.

Ωστόσο υπάρχουν άλλο πρότυπα τα οποία έχουν σχεδιαστεί από την αρχή για να υποστηρίζουν την κλιμάκωση των κέντρων δεδομένων. Το InfiniBand είναι ένα τέτοιο πρότυπο καθώς επίσης και το Fiber Channel σε κάποιο βαθμό. Χρησιμοποιούν τοπολογίες τύπου mesh και υποστηρίζουν τη χρησιμοποίηση πολλαπλών μονοπατιών μέσω ενός διαχειριστή υποδομής αξιοποιώντας έτσι το διαθέσιμο εύρος ζώνης και διασφαλίζοντας παράλληλα την έλλειψη βρόγχων. Επιπλέον διαθέτουν ικανότητες διαχωρισμού κατηγορίας κίνησης [4].

1.7.1 InfiniBand

Το πρότυπο InfiniBand αναπτύχτηκε από την InfiniBand Trade Association η οποία ιδρύθηκε το 1999 με σκοπό την ανάπτυξη του προτύπου. Είναι μία ομάδα που αποτελείται από περισσότερες από 180 εταιρείες και διευθύνεται από μία διακεκριμένη επιτροπή που περιλαμβάνει μέλη των εταιρειών IBM, Intel, Mellanox Technologies, Microsoft, Oracle, Broadcom, Gray και άλλων μεγάλων ονομάτων στο χώρο. Επίσης υποστηρίζεται από κορυφαίους

προμηθευτές όπως Cisco Systems, Hitachi, Fujitsu Siemens και άλλους [17], [18].

Το InfiniBand χρησιμοποιεί κανάλια επικοινωνίας από σημείο σε σημείο που μπορεί να είναι οπτική ίνα ή χαλκός. Παρέχει τη μεγαλύτερη απόδοση εφαρμογών και τη μικρότερη καθυστέρηση λόγω της αξιόπιστης ανταλλαγής μηνυμάτων και της χρήσης του πρωτοκόλλου Remote Direct Memory Access (RDMA) χωρίς να υπάρχει κάποιο ενδιάμεσο λογισμικό κατά την κίνηση των δεδομένων. Το υψηλό εύρος ζώνης και η χαμηλή καθυστέρηση δίνουν τα πλεονέκτημα χρήσης μίας μόνο σύνδεσης για τους διάφορους τύπους κίνησης. Είναι μία τεχνολογία που χρησιμοποιείται σε χιλιάδες κέντρα δεδομένων, σε συμπλέγματα υπολογιστών υψηλής απόδοσης και εφαρμογές που κλιμακώνονται [19].

Χρησιμοποιείται για τη διασύνδεση των περισσότερων γρηγορότερων υπερυπολογιστών του κόσμο. Παρέχει υψηλή απόδοση λόγω της υποστήριξης πρωτοκόλλων όπως το Remote Direct Memory Access το οποίο βελτιώνει την επεξεργασία του φορτίου εργασίας των πελατών. Έχει πάρα πολύ χαμηλή καθυστέρηση από άκρο σε άκρο 1μs καθιστώντας το έτσι κατάλληλο για υπολογιστές υψηλής απόδοσης και εφαρμογές κέντρων δεδομένων. Επιτρέπει τη χρησιμοποίηση πολλαπλών μονοπατιών με αυτόματη εναλλαγή διαδρομής σε περίπτωση διακοπής της σύνδεσης ενός τμήματος μεγιστοποιώντας τη διαθεσιμότητα και την αξιοπιστία. Προσφέρει μεταφορά φορτίου χωρίς απώλειες και ακεραιότητα των δεδομένων εφαρμόζοντας ελέγχους σε κάθε άλμα σε όλο το μήκος του δικτύου. Αξιοποιεί μόνο μία δικτυακή υποδομή για όλους τους τύπους κίνησης από επικοινωνίες, αποθήκευση, ομαδοποίηση και διαχείριση. Με τον τρόπο αυτό η κατανάλωση της συνολικής ισχύς και τα έξοδα διαχείρισης μειώνονται σημαντικά. Ακόμα οι διακόπτες και οι προσαρμογείς καναλιού των διακομιστών έχουν πολύ πιο ανταγωνιστικές τιμές σε σχέση με άλλες τεχνολογίες, προσφέροντας σημαντικό όφελος απόδοσης κόστους. Τέλος να αναφέρουμε ότι το InfiniBand είναι ένα διαλειτουργικό περιβάλλον, προσφέρει ανεξαρτησία από τους προμηθευτές βοηθώντας σημαντικά τους χρήστες να μην δεσμεύονται για την επιλογή των προϊόντων τους [19][17].

1.7.2 Converged Enhanced Ethernet

Οι οργανισμοί IEEE και IETF προσπάθησαν να βελτιώσουν το Ethernet ώστε να μπορεί να καλύψει τις απαιτήσεις των σύγχρονων κέντρων δεδομένων. Για να το πετύχουν αυτό υιοθέτησαν πολλές από τις ιδιότητες του InfiniBand. Οι νέες τεχνολογίες είναι γνωστές ως Data Center Ethernet (DCE) ή Converged Enhanced Ethernet (CEE)

Η Mellanox η οποία έχει πλέον αγοραστεί από την Nvidia είναι ένας από τους σημαντικότερους προμηθευτές των σύγχρονων κέντρων δεδομένων. Παρέχει προϊόντα και λύσεις διαχείρισης και για τα δύο πρότυπα Infiniband και Converged Enhanced Ethernet. Τα προϊόντα Mellanox Infiniband χρησιμοποιούνται εδώ και πολλά χρόνια για την διασύνδεση των μεγαλυτέρων υπερυπολογιστών στο κόσμο.

Η αρχιτεκτονική Converged Enhanced Ethernet της Mellanox προσφέρει τα παρακάτω πλεονεκτήματα. Αυξάνεται η απόδοση. Το κόστος και η πολυπλοκότητα του υλικού μειώνονται, η κατανάλωση ενέργειας μειώνεται και χρησιμοποιούνται διακόπτες υψηλότερης πυκνότητας σε σχέση με τους διακόπτες συνάθροισης. Η Mellanox προσφέρει λύσεις στις οποίες η απόδοση και το κόστος κυμαίνονται γραμμικά. Επίσης η διαχείριση της υποδομής γίνεται μέσω ενός διαχειριστεί ο οποίος παρακολουθεί και ρυθμίζει τους διακόπτες, φυσικούς και εικονικούς έτσι ώστε να ικανοποιούνται οι απαιτήσεις των εφαρμογών και των υπηρεσιών. Επίσης δημιουργούνται πολιτικές κίνησης για τους διακόπτες μέσω ενός διαχειριστή ώστε να εξασφαλίζεται η επιθυμητή απόδοση των εφαρμογών. Οι χρήστες μπορούν να καθορίσουν τις υπηρεσίες υποδομής που θέλουν και ο διαχειριστής της υποδομής εκτελεί τις απαραίτητες ενέργειες ώστε να λάβουν οι χρήστες τις επιθυμητές υπηρεσίες. Η Mellanox δίνει την δυνατότητα μέσω της αρχιτεκτονικής CEE να αποκτήσει ο πελάτης την κατάλληλη υποδομή χωρίς να θυσιάζεται η απόδοση και η επεκτασιμότητα [4].

1.8 Μετάβαση σε τεχνολογίες Ethernet 400G και 800G

Τα σημερινά κέντρα δεδομένων αλλάζουν και χρησιμοποιούν τεχνολογίες 400G με οπτικούς πομποδέκτες. Μερικοί από τους λόγους που οδήγησαν στη

δημιουργία και χρήση αυτή της τεχνολογίας νέας γενιάς είναι η υπολογιστική νέφους , η τεχνητή νοημοσύνη, υπηρεσίες 5G, ροής βίντεο, υπηρεσίες τηλεδιάσκεψης. Ακόμα ο τεράστιος όγκος δεδομένων που δημιουργούνται λόγω των μέσων κοινωνικής δικτύωσης, των έξυπνων συσκευών και του Διαδικτύου των Πραγμάτων οδηγούν επίσης προς αυτή την κατεύθυνση.

Οι παραπάνω παράγοντες δημιουργούν απαιτήσεις για πολύ υψηλές ταχύτητες μεταφοράς δεδομένων, μικρότερη καθυστέρηση και υψηλότερο εύρος ζώνης. Το Ethernet 400G καλύπτει αυτές τις απαιτήσεις και επιπλέον προσφέρει σημαντικά πλεονεκτήματα κόστους. Μία θύρα 400G κοστίζει πολύ λιγότερο από ότι κοστίζουν 4 θύρες 100G . Επιπλέον καταναλώνει λιγότερη ενέργεια σε σχέση με 4 θύρες 100G.

Ο τρόπος σχεδίασης και κατασκευής των κέντρων δεδομένων καθώς και των δικτύων διασύνδεσης των κέντρων δεδομένων αλλάξει σημαντικά με το Ethernet 400G. Υπάρχουν δύο τεχνολογίες που ανταγωνίζονται μεταξύ τους και είναι η QSFP-DD και η OSFP. Η τεχνολογία που συνήθως επιλέγεται είναι η QSFP-DD στην πλευρά του πελάτη γιατί προσφέρει περισσότερες επιλογές προσέγγισης και είναι συμβατή με παλαιού τύπου μονάδες [20][21][22].

Ακόμα μεγαλύτερες ταχύτητες 800G έχουν ξεκινήσει από το 2022 και αναμένεται η προτυποποίηση του Ethernet 800G από το IEEE μέσα στα επόμενα χρόνια.

ΚΕΦΑΛΑΙΟ 2 : Δικτύωση Οριζόμενη από το Λογισμικό και εικονικοποίηση λειτουργιών δικτύου

2.1 Βασικά χαρακτηριστικά της Δικτύωσης Οριζόμενης από το Λογισμικό

Στο κεφάλαιο αυτό θα ασχοληθούμε με τη Δικτύωση Οριζόμενη από το Λογισμικό (SDN) η οποία είναι μία αρχιτεκτονική με βασικό χαρακτηριστικό την αποσύνδεση του επιπέδου ελέγχου από το επίπεδο δεδομένων και τον προγραμματισμό το επιπέδου ελέγχου μέσω ενός κεντρικού σημείου που ονομάζεται ελεγκτής [23],[24]. Η αρχιτεκτονική αυτή μεταφέρει στον ελεγκτή την πολυπλοκότητα της διαχείρισης του δικτύου και δίνει μία αφαιρετική εικόνα της δικτυακής υποδομής.

Το χαρακτηριστικό του διαχωρισμού του επιπέδου ελέγχου και της προώθησης απλοποιεί τις δικτυακές συσκευές οι οποίες ελέγχονται από τον ελεγκτή. Ο ελεγκτής διαχειρίζεται το δίκτυο και παρέχει τις κατάλληλες οδηγίες στις απλοποιημένες συσκευές ώστε αυτές να μπορούν να πάρουν αποφάσεις προώθησης [23],[25].

Η διαχείριση του δικτύου και οι ρυθμίσεις γίνονται αυτόματα μέσω του κεντρικού ελεγκτή και διάφορων διεπαφών οι οποίες είναι γνωστές με τους όρους northbound και southbound APIs. Το ποιο γνωστό southbound API είναι το OpenFlow το οποίο χρησιμοποιείται για την επικοινωνία με τις συσκευές και στο οποίο θα αναφερθούμε παρακάτω. Μέσω των northbound APIs γίνεται η σύνδεση με εφαρμογές λογισμικού οι οποίες προσφέρουν πρωτόκολλα και αλγόριθμους για την αποδοτική λειτουργία του δικτύου. Αυτές οι εφαρμογές παρέχουν τη δυνατότητα για δυναμικές και γρήγορες αλλαγές στο δίκτυο, με βάση τις εκάστοτε ανάγκες. Ακόμα τα northbound APIs παρέχουν μια αφαιρετική εικόνα της τοπολογίας και των δικτυακών συσκευών και δίνουν τη δυνατότητα στις εφαρμογές λογισμικού να εκτελούν τις απαραίτητες ενέργειες χωρίς να γνωρίζουν τα χαρακτηριστικά των δικτυακών συσκευών. Αυτό έχει ως αποτέλεσμα την ανάπτυξη εφαρμογών οι οποίες μπορούν να λειτουργούν σε εξοπλισμό πολλών κατασκευαστών που μπορεί να διαφέρουν στις λεπτομέρειες εφαρμογής τους.

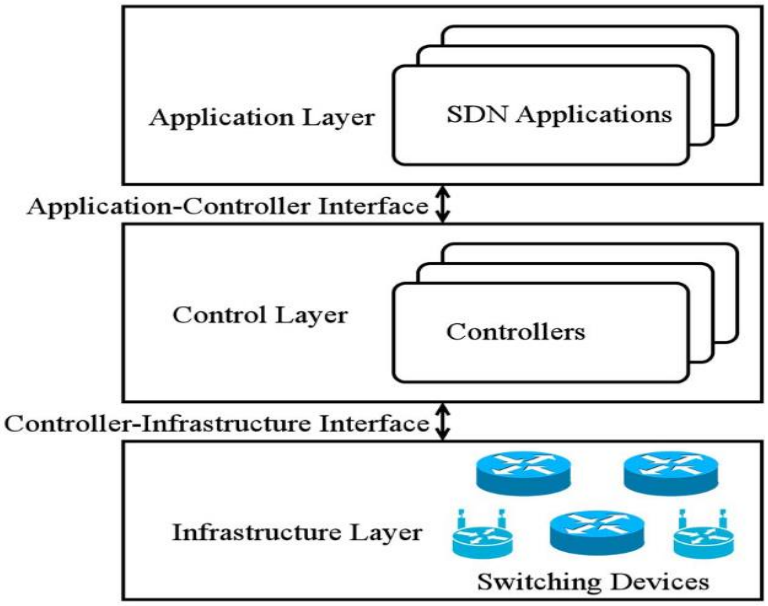
Το επίπεδο αφαίρεσης δίνει τη δυνατότητα εικονικοποίησης του δικτύου, αποσυνδέοντας τις φυσικές συσκευές από τις εικονικές υπηρεσίες. Οι διακομιστές που χρησιμοποιούν τους δικτυακούς πόρους δεν γνωρίζουν ότι χρησιμοποιούν εικονικούς και όχι φυσικούς δικτυακούς πόρους.

Ένα ακόμη χαρακτηριστικό του Open SDN είναι ότι οι διεπαφές δεν πρέπει να είναι ιδιόκτητες, πρέπει να είναι προτυποποιημένες. Τα APIs πρέπει να παρέχουν στο λογισμικό επαρκή έλεγχο ώστε να υπάρχει η δυνατότητα να πραγματοποιηθούν δοκιμές και έλεγχοι που θα οδηγήσουν στην καλύτερη και πιο γρήγορη τεχνολογική πρόοδο στη λειτουργία των δικτύων. Ακόμη οι διεπαφές ανοιχτού κώδικα επιτρέπουν τη διαλειτουργικότητα μεταξύ εξοπλισμού διαφορετικών προμηθευτών δημιουργώντας έτσι ένα ανταγωνιστικό περιβάλλον το οποίο συμβάλει στη μείωση του κόστους του δικτυακού εξοπλισμού για τους καταναλωτές [26].

2.2 Λειτουργία του SDN

Η αρχιτεκτονική SDN αποτελείται από τρία επίπεδα σύμφωνα με όσα προτείνει το Open Networking Foundation (ONF) , που είναι μία κοινοπραξία με μη κερδοσκοπικό χαρακτήρα και σκοπό την ανάπτυξη, τυποποίηση και εμπορευματοποίηση του SDN.Τα επίπεδα αυτά είναι τα εξής : το επίπεδο υποδομής, το επίπεδο ελέγχου και το επίπεδο εφαρμογών [27].

Το επίπεδο υποδομής ή αλλιώς επίπεδο δεδομένων περιέχει τις SDN συσκευές οι οποίες έχουν ως κύρια λειτουργία τους την προώθηση των εισερχόμενων πακέτων. Οι αποφάσεις προώθησης των πακέτων λαμβάνονται με βάση δεδομένα τα οποία έχουν οριστεί στις συσκευές από τον ελεγκτή. Για τα δεδομένα αυτά χρησιμοποιείται ο ορός ροές (flows). Κάθε ροή αφορά την μεταφορά ενός συνόλου πακέτων μεταξύ δύο τελικών σημείων του δικτύου. Τα τελικά σημεία μπορεί να προσδιορίζονται με εισερχόμενες θύρες, VLAN ID, IP διευθύνσεις, TCP/UDP θύρες, σήραγγες τρίτου επιπέδου και άλλα. Η κάθε ροή είναι μονής κατεύθυνσης. Αυτό σημαίνει ότι μπορεί να υπάρχουν δύο ροές για πακέτα που προωθούνται μεταξύ δύο τελικών σημείων, μία ροή για κάθε κατεύθυνση.



Εικόνα 2.1: Μοντέλο Αναφοράς SDN (“A Survey on Software-Defined Networking”, Xia el.al., 2015)

Οι ροές βρίσκονται ως καταχωρήσεις σε πίνακες ροών στις SDN συσκευές και περιέχουν τις αντίστοιχες ενέργειες που πρέπει να γίνουν όταν ένα εισερχόμενο πακέτο που φτάνει στην συσκευή ταιριάζει με κάποια από αυτές τις ροές. Οι πίνακες ροών έχουν δημιουργηθεί όπως αναφέραμε και παραπάνω με βάση τις ροές που στέλνει ο ελεγκτής στην κάθε συσκευή. Όταν ένα πακέτο που φτάνει στη συσκευή αντιστοιχηθεί με κάποια από τις ροές τότε εκτελούνται οι αντίστοιχες ενέργειες που συνήθως είναι η προώθηση του πακέτου. Οι καταχωρήσεις ροών προσδιορίζονται από προτεραιότητες. Το εισερχόμενο πακέτο ελέγχεται και επιλέγεται η ροή με την υψηλότερη προτεραιότητα. Αν δεν βρεθεί ταίριασμα με κάποια από τις ροές τότε ο διακόπτης στέλνει το πακέτο στον ελεγκτή ή πετάει το πακέτο. Αυτό εξαρτάται από τις ρυθμίσεις που έχουν γίνει στον διακόπτη και από την έκδοση του πρωτοκόλλου OpenFlow που χρησιμοποιείται [26].

Ο ελεγκτής συνδέει το επίπεδο υποδομής με το επίπεδο εφαρμογών με χρήση των κατάλληλων διεπαφών. Αφαιρεί τις λειτουργίες δικτύου από τις SDN συσκευές και παρέχει στις εφαρμογές που βρίσκονται στο πάνω επίπεδο μια αφαιρετική εικόνα των δικτυακών πόρων. Μεσολαβεί έτσι ώστε οι

εφαρμογές να μπορούν να ανταποκριθούν στα πακέτα που προωθούνται από τις SDN συσκευές στον ελεγκτή και να μπορούν να ορίσουν τις κατάλληλες ροές για αυτές τις συσκευές. Ο ελεγκτής έχοντας μία συνολική εικόνα του δικτύου μπορεί να παρέχει βέλτιστες αποφάσεις προώθησης με έναν προβλέψιμο τρόπο.

Ο ελεγκτής παρουσιάζεται ως ένα κεντρικό σημείο. Αυτό είναι ουσιαστικά ένα λογικό κεντρικό σημείο ενώ πρακτικά αποτελείται από πολλαπλούς διακομιστές έτσι ώστε να μην υπάρχει μοναδικό σημείο αστοχίας και να διασφαλίζεται υψηλή διαθεσιμότητα και κλιμακούμενη απόδοση [28][29]. Λόγω των πολλαπλών ελεγκτών απαιτείται η ύπαρξη και χρήση μίας διεπαφής για την επικοινωνία μεταξύ των ελεγκτών, ώστε να διασφαλίζεται ο συντονισμός και ο διαμοιρασμός της πληροφορίας του δικτύου για την επεξεργασία λήψης αποφάσεων [30][31].

Οι εφαρμογές διασυνδέονται με τον ελεγκτή και τον χρησιμοποιούν για τον ορισμό ροών στις SDN συσκευές. Οι ροές που μπορούν να ορίσουν χωρίζονται σε δύο κατηγορίες. Η πρώτη κατηγορία είναι οι προληπτικές ροές οι οποίες δημιουργούνται όταν ξεκινάει μία εφαρμογή. Οι ροές αυτές είναι στατικές και διατηρούνται μέχρι να γίνει κάποια αλλαγή στις ρυθμίσεις. Για παράδειγμα ο ελεγκτής μπορεί να αποφασίσει για λόγους εξισορρόπησης κίνησης να κάνει μία αλλαγή σε μία ροή.

Η δεύτερη κατηγορία είναι οι ροές αντίδρασης. Οι ροές αυτές δημιουργούνται ως απάντηση σε πακέτα τα οποία έχουν προωθήσει οι διακόπτες στον ελεγκτή. Οι εφαρμογές δείχνουν στον ελεγκτή πώς να ανταποκρίνεται σε τέτοιου τύπου πακέτα και ο ελεγκτής στέλνει μία νέα ροή στη συσκευή ως απάντηση. Με τον τρόπο αυτό ο διακόπτης μαθαίνει πώς να αντιδρά σε παρόμοια πακέτα στο μέλλον.

Παρέχεται με αυτό τον τρόπο η δυνατότητα δημιουργίας εφαρμογών οι οποίες υλοποιούν λειτουργίες όπως δρομολόγηση, προώθηση, έλεγχος πρόσβασης, εξισορρόπηση κίνησης και διάφορες άλλες λειτουργίες.

Επίσης ροές μπορεί να δημιουργηθούν ή να τροποποιηθούν ως αποτέλεσμα πακέτων που λαμβάνουν οι εφαρμογές όχι μόνο από τον ελεγκτή αλλά και

από άλλες πηγές όπως για παράδειγμα από ένα σύστημα ανίχνευσης εισβολών [26].

2.2.1 Επίπεδο Υποδομής - Επίπεδο Δεδομένων - SDN Συσκεύες

Μία SDN συσκευή αποτελείται από κάποια βασικά στοιχεία. Πρώτον μία διεπαφή για επικοινωνία με τον ελεγκτή που συνήθως είναι το OpenFlow. Δεύτερον από έναν ή περισσότερους πίνακες ροής οι οποίοι αποτελούν το λεγόμενο επίπεδο αφαίρεσης. Τρίτον από το κομμάτι που εκτελεί την επεξεργασία πακέτων. Το κομμάτι αυτό περιέχει κατάλληλους μηχανισμούς με τους οποίους όπως έχουμε αναφέρει και ποιο πάνω γίνεται η αναζήτηση ώστε να βρεθεί το ταίριασμα με την υψηλότερη προτεραιότητα για το εισερχόμενο πακέτο και να εκτελεστούν οι αντίστοιχες ενέργειες.

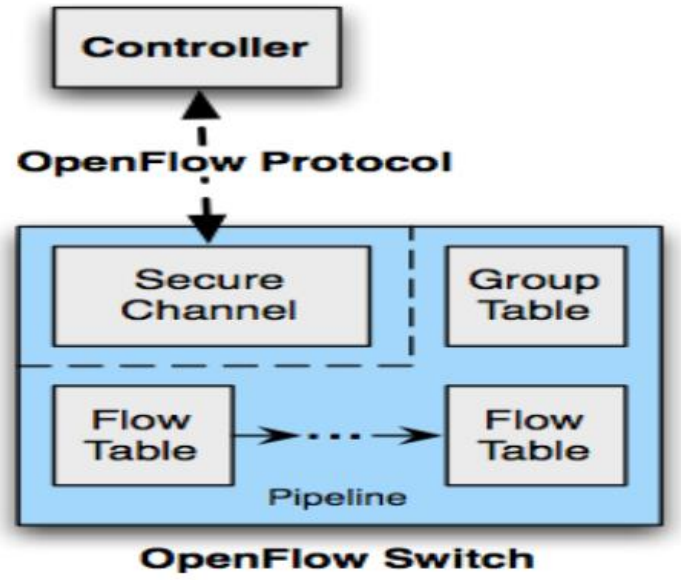
Μία ακόμα λειτουργία που εκτελούν οι συσκευές μεταγωγής είναι η συλλογή πληροφοριών που αφορούν την κατάσταση του δίκτυο, η προσωρινή τοπική αποθήκευσή τους και στην συνέχεια η αποστολή τους στον ελεγκτή. Οι πληροφορίες αυτές μπορεί να αφορούν για παράδειγμα την χρησιμοποίηση του δικτύου, την κίνηση του, καθώς επίσης και την τοπολογία του [24].

Οι συσκευές μεταγωγής περιέχουν επίσης ένα κανάλι ασφάλειας το οποίο μπορούν να χρησιμοποιήσουν για την επικοινωνία με τον ελεγκτή όταν χρειάζεται. Το κανάλι αυτό τις περισσότερες φορές είναι ένα TLS ή SSL κανάλι [32].

Ακόμα θα πρέπει να αναφέρουμε ότι οι συσκευές μεταγωγής μπορεί να είναι εικονικές ή φυσικές συσκευές. Επίσης οι συσκευές μεταγωγής χωρίζονται σε δύο τύπους. Αυτές που είναι μόνο OpenFlow και αυτές που είναι υβριδικές. Οι συσκευές που υποστηρίζουν μόνο την OpenFlow λειτουργία επεξεργάζονται όλα τα πακέτα με βάση τις αποφάσεις προώθησης του ελεγκτή. Οι συσκευές που είναι υβριδικές μπορούν να υποστηρίξουν και την Open Flow λειτουργία και την παραδοσιακή λειτουργία [32]. Σημαντικοί κατασκευαστές δικτυακού εξοπλισμού όπως Cisco, IBM, HP, Juniper, Nec και Extreme έχουν ενσωματώσει σε κάποιες παραδοσιακές συσκευές μεταγωγής την υποστήριξη του πρωτοκόλλου OpenFlow παρέχοντας έτσι την δυνατότητα να μπορούν οι

συσκευές να λειτουργούν και στις δύο καταστάσεις ανάλογα με τις ανάγκες που υπάρχουν κάθε φορά.

Μία επιπλέον κατηγορία συσκευών είναι οι λεγόμενοι διακόπτες λευκού κουτιού “white-box switches”. Οι διακόπτες αυτοί κατασκευάζονται κυρίως από τσίπ μεταγωγής πυριτίου και μνήμη και CPU από κατασκευαστές που δεν είναι κάποιας γνωστής μάρκας. Το μεγαλύτερο μέρος του επιπέδου ελέγχου δεν υπάρχει σε αυτές τις συσκευές γιατί αναμένεται η λειτουργία αυτή να παρέχεται από τον ελεγκτή σε αντίθεση με τις παραδοσιακές συσκευές. Η κατηγορία αυτή καλύπτει μία από τις βασικές προϋποθέσεις του SDN η οποία είναι η μείωση του κόστους της φυσικής υποδομής χρησιμοποιώντας αντί για διακόπτες από γνωστούς κατασκευαστές δικτυακού εξοπλισμού, διακόπτες λευκού κουτιού με δυνατότητα Open Flow πολύ χαμηλότερου κόστους.



Εικόνα 2.2: Στοιχεία ενός διακόπτη που βασίζεται σε OpenFlow (“Considerations for Software Defined Networking (SDN): Approaches and Use Cases”, Bakshi, 2013)

2.2.2 SDN Ελεγκτής και Εφαρμογές

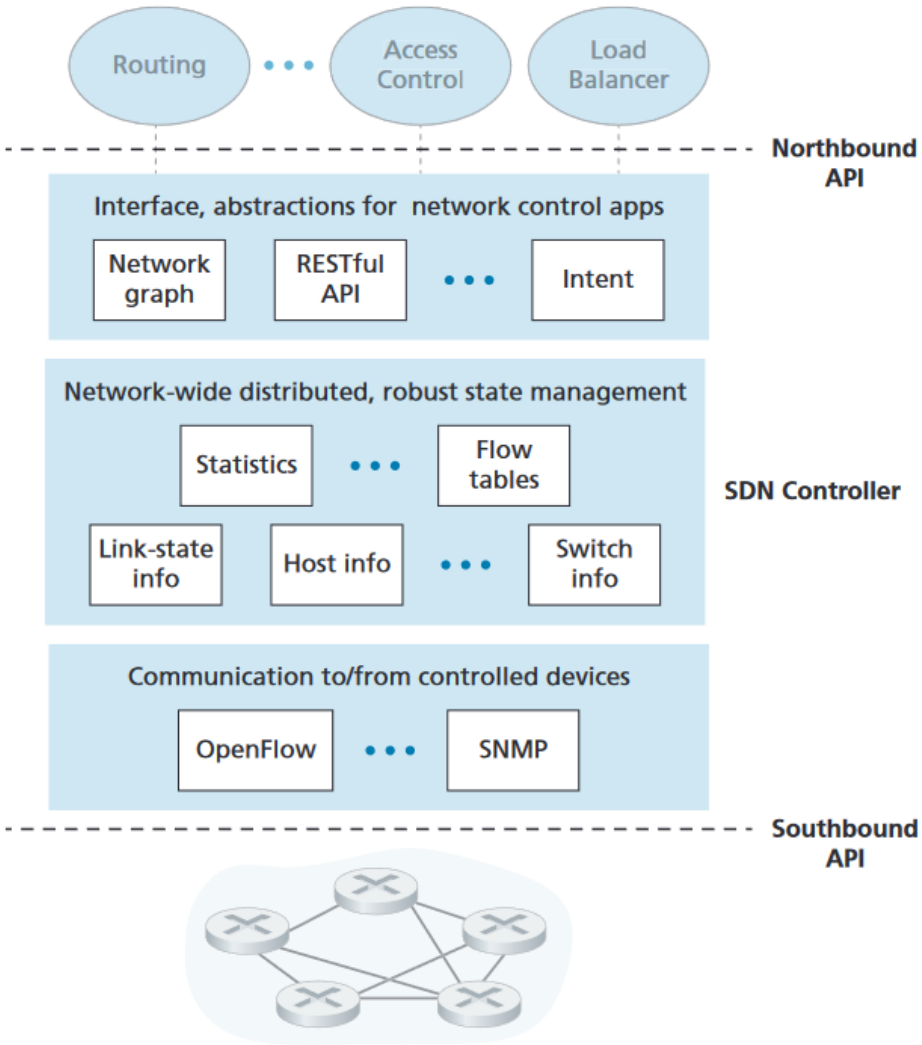
Ο SDN ελεγκτής όπως έχουμε αναφέρει και παραπάνω ελέγχει όλες τις SDN συσκευές, διατηρεί μία καθολική εικόνα ολόκληρου του δικτύου, εφαρμόζει αποφάσεις κάνοντας χρήση πολιτικών και χρησιμοποιεί διεπαφές για την επικοινωνία με τις συσκευές και τις εφαρμογές. Η εφαρμογή των πολιτικών που μπορεί για παράδειγμα να αφορούν την εξισορρόπηση κίνησης, την προώθηση, την δρομολόγηση, την ανακατεύθυνση και άλλες ενέργειες γίνεται από τον ελεγκτή και τις εφαρμογές. Οι ελεγκτές σε αρκετές περιπτώσεις έχουν δικές τους κοινές εφαρμογές.

Η επικοινωνία αυτή μεταξύ του ελεγκτή και των συσκευών γίνεται μέσω μίας διεπαφής “southbound”. Στην περίπτωση του Open SDN η διεπαφή που χρησιμοποιείται είναι το OpenFlow το οποίο εφαρμόζεται στην πλειοψηφία των ελεγκτών. Υλοποιούνται και κάποιες ιδιόκτητες λύσεις για εναλλακτικές SDN επιλογές. Υπάρχουν και περιπτώσεις που παρέχουν τη δυνατότητα τόσο OpenFlow όσο και κάποιου ιδιόκτητου πρωτοκόλλου στον ίδιο ελεγκτή. Το OpenFlow είναι η λύση που προτείνεται ωστόσο υπάρχουν και διάφορα άλλα πρότυπα όπως SNMP και Cisco CLI.

Η επικοινωνία προς το βορρά με τις εφαρμογές γίνεται μέσω northbound APIs. Για αυτό το κομμάτι δεν υπάρχει κάποιο πρότυπο αντίστοιχα με το OpenFlow που υπάρχει στο νότιο τμήμα επικοινωνίας. Υπάρχουν διάφοροι τύποι APIs όπως Java API, REST API, Python API και άλλα.

Οι βασικές λειτουργίες ενός ελεγκτής είναι οι εξής:

- Ανακάλυψη τελικών συσκευών όπως διακομιστές, εκτυπωτές και άλλα.
- Ανακάλυψη δικτυακών συσκευών όπως δρομολογητές, διακόπτες, ασύρματα σημεία πρόσβασης και άλλα.
- Γνώση του τρόπου διασύνδεσης των συσκευών μεταξύ τους.
- Διαχείριση ροών και συσκευών καθώς και παρακολούθηση διάφορων στατιστικών στοιχείων [26].



Εικόνα 2.3: Συστατικά ενός SDN ελεγκτή (“Computer Networking A Top Down Approach”, Kurose et. Al., 2022)

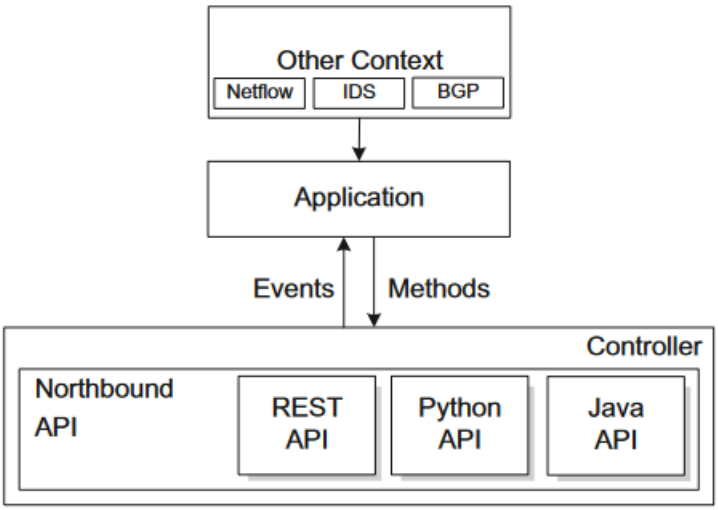
Ο ελεγκτής πρέπει να γνωρίζει τις SDN συσκευές της υποδομής, τόσο τις SDN δικτυακές συσκευές όσο και τις συσκευές τελικού χρήστη καθώς επίσης και τον τρόπο διασύνδεσης τους, να έχει δηλαδή εικόνα της τοπολογίας του δικτύου. Για να μπορεί να γίνει αυτό οι συσκευές στέλνουν τα διάφορα συμβάντα στον ελεγκτή. Για παράδειγμα μηνύματα τα οποία δηλώνουν ότι μία νέα συσκευή έχει συνδεθεί στο δίκτυο, ότι μία ζεύξη έχει διακοπή ή ότι μία νέα ζεύξη έχει δημιουργηθεί, μηνύματα τα οποία αποστέλλονται περιοδικά με σκοπό την επιβεβαίωση ότι μια συσκευή λειτουργεί κανονικά και άλλα. Με τον

τρόπο αυτό ο ελεγκτής έχει μια ενημερωμένη εικόνα της κατάστασης του δικτύου [28].

Καθώς το επιπέδου ελέγχου έχει ως τελικό στόχο να καθορίσει τους πίνακες ροής των διακοπτών που υπάρχουν στην υποδομή ο ελεγκτής διατηρεί αντίγραφα αυτών των πινάκων ροής τοπικά. Επίσης διατηρεί στατιστικά για κάθε ροή τα οποία συλλέγει από τους διακόπτες. Οι πίνακες ροής των διακοπτών έχουν μετρητές οι τιμές των οποίων αυξάνονται καθώς χρησιμοποιούνται οι ροές. Οι τιμές αυτές δίνονται στον ελεγκτή για στατιστικούς λόγους και μπορούν να αξιοποιηθούν από τις εφαρμογές στο αμέσως επόμενο επίπεδο. Η τοπολογία του δικτύου καθώς και τα στατιστικά διατηρούνται τοπικά σε βάσεις δεδομένων στον ελεγκτή.

Ο ελεγκτής όπως έχουμε αναφέρει είναι ένα λογικό κεντρικό σημείο. Πρακτικά οι βάσεις δεδομένων που διατηρεί και οι υπηρεσίες που προσφέρει είναι κατανεμημένες σε πολλούς διακομιστές για λόγους υψηλής απόδοσης, διαθεσιμότητας και ανοχής σε σφάλματα.

Οι εφαρμογές μέσω των διεπαφών που τους παρέχει ο ελεγκτής μπορούν να βλέπουν και να αλλάξουν την κατάσταση του δικτύου , τους πίνακες ροών και γενικά να πραγματοποιήσουν διάφορες ενέργειες. Υπάρχει η δυνατότητα να ειδοποιούνται όταν συμβαίνουν διάφορα γεγονότα ώστε να ανταποκρίνονται άμεσα και να εκτελούν κατάλληλες ενέργειες. Τα συμβάντα αυτά στέλνονται από τις SDN συσκευές στον ελεγκτή και από εκεί φτάνουν στις εφαρμογές.[28] Επίσης μπορεί να πραγματοποιούνται ενέργειες που να αλλάζουν την κατάσταση λόγω ειδοποιήσεων που φτάνουν στις εφαρμογές από άλλα σημεία πέρα από τον ελεγκτή όπως ένα σύστημα ανίχνευσης εισβολών [26].



Εικόνα 2.4: Διεπαφές SDN ελεγκτή προς το βορρά.(“Software defined networks: A comprehensive approach”, Goransson et.al., 2014)

Παρακάτω απεικονίζεται ένας πίνακας που περιέχει γνωστούς ελεγκτές [32],[39].

Controller	Language	Created by	Open source	OpenFlow Version
NOX	C++	Nicira	Yes	1.0
Beacon	Java	Stanford university	Yes	1.0
Floodlight, Floodlight-Plus	Java	Big switch networks	Yes	1.0,1.1, 1.3
Maestro	Java	Rice university	Yes	1.0
POX	Python	Nicira	Yes	1.0
Ryu	Python	NTT Labs	Yes	1.0,1.2,1.3,1.4
Trema	C, Ruby	NEC	Yes	1.0
OpenDaylight	Java	Linux Foundation	Yes	1.0,1.3

Πίνακας 2.1: SDN Ελεγκτές

2.3 Πρωτόκολλο OpenFlow

Το πρωτόκολλο OpenFlow χρησιμοποιείται όπως έχουμε αναφέρει για την επικοινωνία μεταξύ SDN ελεγκτή και SDN συσκευών. Λειτουργεί μέσω TCP με προεπιλεγμένο αριθμό θύρας 6653. Η επικοινωνία του ελεγκτή με τις συσκευές γίνεται με ανταλλαγή μηνυμάτων.

Παρακάτω αναφέρουμε ορισμένα από τα σημαντικά μηνύματα τα οποία στέλνει ο ελεγκτής σε έναν διακόπτη.

- Μηνύματα διαμόρφωσης. Τα μηνύματα αυτά χρησιμοποιούνται για να ρωτήσει ο ελεγκτής πληροφορίες σχετικά με το διακόπτη και στην συνέχεια να ρυθμίσει ορισμένες παραμέτρους διαμόρφωσης στον διακόπτη με βάση τις απαντήσεις που θα λάβει.
- Μηνύματα που αφορούν τροποποίηση κατάστασης. Είναι μηνύματα τα οποία αφορούν την εισαγωγή, τροποποίηση και διαγραφή ροών στους πίνακες ροών καθώς και τον ορισμό των ιδιοτήτων των θυρών.
- Μηνύματα που αφορούν ανάγνωση κατάστασης. Είναι μηνύματα που χρησιμοποίει ο ελεγκτής για να συλλέξει στατιστικά στοιχεία και να κάνει καταμέτρηση τιμών από τους πίνακες ροών.
- Μηνύματα που σχετίζονται με την αποστολή πακέτου. Είναι μηνύματα τα οποία χρησιμοποιεί ο ελεγκτής για να ορίσει την απευθείας αποστολή ενός συγκεκριμένου πακέτου σε μία συγκεκριμένη θύρα στον διακόπτη.

Στη συνέχεια αναφέρουμε ορισμένα από τα μηνύματα τα οποία στέλνει ένας διακόπτης σε έναν ελεγκτή.

- Μηνύματα εισαγωγής πακέτου. Όταν ο διακόπτης δεν βρει ένα ταίριασμα στον πίνακα ροών του για ένα εισερχόμενα πακέτο τότε ανάλογα με την έκδοση του OpenFlow και αν έχει ρυθμιστεί κατάλληλα μπορεί να στείλει ένα μήνυμα στον ελεγκτή με το οποίο να ζητάει πληροφορίες για την περαιτέρω επεξεργασία του πακέτου.
- Μηνύματα που αφορούν την διαγραφή μίας ροής. Τα μηνύματα αυτά χρησιμοποιούνται για να ενημερώσουν τον ελεγκτή για την αφαίρεση μίας καταχώρησης από τον πίνακα ροής. Ο λόγος για την αφαίρεση

μπορεί να είναι η λήψη ενός μηνύματος που απαιτούσε την συγκεκριμένη ενέργεια ή λήξη του χρονικού ορίου της συγκεκριμένης ροής.

- Μηνύματα που αφορούν την αλλαγή της κατάστασης μίας θύρας [28].

2.4 Εικονικοποίηση λειτουργιών δικτύου (Network Functions Virtualization - NFV)

Το NFV είναι η υλοποίηση σε λογισμικό των δικτυακών λειτουργιών και συσκευών. Η υλοποίηση μέσω λογισμικού της λειτουργικότητας μίας υπηρεσίας όπως για παράδειγμα ενός τείχους προστασίας(firewall) ή ενός εξισορροπητή φορτίου.

Η δυνατότητα αυτή δόθηκε λόγω της προόδου που έχει πραγματοποιηθεί στο υλικό διακομιστών γενικού σκοπού. Η πρόοδος αυτή είχε αρχικά υλοποιηθεί με σκοπό την υποστήριξη μεγάλου αριθμού εικονικών μηχανών και εικονικών διακοπτών που απαιτούνται για την υποστήριξη εικονικών δικτύων. Ωστόσο δίνει τη δυνατότητα πλήρους προγραμματιζόμενων συσκευών οι οποίες μπορούν να επαναπρογραμματίζονται σύμφωνα με τις απαιτήσεις ώστε να παρέχουν διαφορετικούς τύπους δικτυακών συσκευών.

Το SDN και το NFV είναι δύο έννοιες συμπληρωματικές και αλληλοεπικαλυπτόμενες [26].

Παρακάτω θα αναφερθούμε σε δύο παραδείγματα, ενός τείχους προστασίας και ενός εξισορροπητή φορτίου που υλοποιούνται μέσω λογισμικού ως εφαρμογές που τρέχουν πάνω από το επίπεδο του ελεγκτή.

2.4.1 Τείχος προστασίας

Το άρθρο [33] προτείνει τη χρήση ενός τείχους προστασίας με επίγνωση εφαρμογών.

Κάθε firewall χρησιμοποιεί μία βάση δεδομένων που περιέχει τις πολιτικές για την διαχείριση της κυκλοφορίας του δικτύου [34]. Τα firewalls μπορούν να λειτουργούν σε διαφορετικά επίπεδα της στοίβας του πρωτοκόλλου TCP/IP.

Ακόμα, υπάρχουν διαφορετικές κατηγορίες firewall: static packet filters, stateful firewalls, and application-level firewalls.

Τα παραδοσιακά κοινά firewall φιλτράρουν τα πακέτα με βάση την διεύθυνση αποστολέα και παραλήπτη και τις UDP/TCP πόρτες. Δεν μπορούν να καταλάβουν τις εφαρμογές. Αυτό έχει ως αποτέλεσμα να μπορούν να μπλοκάρουν μόνο συγκεκριμένες διευθύνσεις και πόρτες. Το τείχος προστασίας με επίγνωση εφαρμογών έχει το πλεονέκτημα ότι μπορεί να εντοπίσει και να απομονώσει συγκεκριμένες εφαρμογές.

Η αρχιτεκτονική του αποτελείται από τέσσερις συνεργαζόμενες μονάδες : την κύρια μονάδα, τη μονάδα αναγνώρισης εφαρμογής, τη μονάδα φιλτραρίσματος και τη μονάδα επιβολής ασφάλειας.

Η κύρια μονάδα λειτουργεί ως συντονιστής. Είναι τα σημείο επαφής για να μπορούν να αλληλεπιδρούν οι άλλες μονάδες μεταξύ τους. Αυτή η μονάδα ακούει στην πλευρά του ελεγκτή συμβάντα Packet_In.

Η μονάδα αναγνώρισης εφαρμογής ταξινομεί την κίνηση του δικτύου βασιζόμενη σε διάφορες πληροφορίες όπως στατιστικά δεδομένα, αριθμό θυρών ή πληροφορίες που λαμβάνει από το ωφέλιμο φορτίο εφαρμογών. Χρησιμοποιεί τρείς μεθόδους για να κάνει αναγνώριση κίνησης και ταξινόμηση σε επίπεδο σύνδεσης, πακέτου και ροής.

Η πρώτη προσέγγιση απαιτεί την εξέταση της κεφαλίδας του πακέτου στο επίπεδο μεταφοράς (αριθμός πόρτας TCP/UDP) για να μάθει τις θύρες προορισμού και προέλευσης ώστε να μπορέσει να γίνει ο προσδιορισμός της εφαρμογής. Χρησιμοποιεί γνωστές θύρες και τις συνδέει με εφαρμογές όπως ορίζονται από τον IANA. Είναι μια αποτελεσματική προσέγγιση για εφαρμογές που χρησιμοποιούν ένα σύνολο γνωστών θυρών ωστόσο, αποτυγχάνει σε εφαρμογές που χρησιμοποιούν δυναμική διαπραγμάτευση θυρών.

Η δεύτερη προσέγγιση αναζητάει γνωστές λέξεις - κλειδιά (υπογραφή) που προσδιορίζουν μοναδικά το πρωτόκολλο. Η αναζήτηση αυτή γίνεται στο ωφέλιμο φορτίου του πακέτου. Η υπογραφή είναι συνήθως μία κανονική έκφραση που βρίσκεται στο ωφέλιμο φορτίο του πακέτου του επιπέδου

εφαρμογής. Σε περίπτωση που εντοπιστεί η υπογραφή, γίνεται αντιστοίχιση με τη λίστα των διαθέσιμων υπογραφών.

Η τρίτη προσέγγιση έχει ως στόχο την αναγνώριση χαρακτηριστικών ροής. Στοχεύει στην ανάλυση της ροής με βάση την κατεύθυνση και το μέγεθος κάθε πακέτου από τα πρώτα N σε αριθμό πακέτα της ροής, τις IP διευθύνσεις καθώς και τις θύρες προορισμού και προέλευσης για τον εντοπισμό ταυτότητας εφαρμογής.

Η τρίτη μονάδα είναι η μονάδα φιλτραρίσματος. Χρησιμοποιείται για να παρέχει λεπτομερή έλεγχο της εισερχόμενης και εξερχόμενης δικτυακής κίνησης. Η μονάδα αυτή έχει έναν προκαθορισμένο πίνακα αναγνώρισης εφαρμογής τον οποίο χρησιμοποιεί για σύγκριση με την καθορισμένη εφαρμογή. Ο πίνακας αυτός είναι ένας πίνακας κατακερματισμού που περιέχει καταχωρήσεις αντιστοίχισης/ αναγνωριστικού. Οι γνωστές εφαρμογές αντιστοιχίζονται με συγκεκριμένα μοτίβα τα οποία περιλαμβάνουν γνωστές θύρες, προκαθορισμένες υπογραφές και αναγνωριστικά τα οποία θα χρησιμοποιηθούν από την πρώτη, δεύτερη και τρίτη προσέγγιση αντίστοιχα. Οι ροές για τις οποίες δεν μπορεί να γίνει ταξινόμηση με βάση κάποια από τις τρείς παραπάνω προσεγγίσεις κατηγοριοποιούνται ως άγνωστες [33].

Η τέταρτη και τελευταία μονάδα είναι η μονάδα επιβολής ασφάλειας. Χρησιμοποιείται για την επιβολή των απαιτούμενων υπηρεσιών του firewall σύμφωνα με τους κανόνες της πολιτικής που θέλουμε να εφαρμόσουμε. Χρησιμοποιείται για την εγκατάσταση της προκαθορισμένης ενέργειας που σχετίζεται με την αναγνωρισμένη εφαρμογή στον μεταγωγέα χρησιμοποιώντας ένα μήνυμα τροποποίησης OpenFlow.

Τέλος να αναφέρουμε ότι οι μονάδες του τείχους προστασίας συντηρούν δύο λίστες. Μία λίστα με κανόνες φιλτραρίσματος για τον προσδιορισμό των εφαρμογών και μία λίστα με κανόνες του firewall που αφορούν τις πολιτικές ασφάλειας ολόκληρου του δικτύου. Και οι δύο λίστες αποθηκεύονται σε έναν SDN ελεγκτή και είναι προσβάσιμες από τις μονάδες του firewall [33].

Κάποια από τα πλεονεκτήματα που μας προσφέρει η υλοποίηση του τείχους προστασίας ως εφαρμογή είναι τα εξής : Παρέχει καλύτερες δυνατότητες

καθώς μπορεί να αναγνωρίσει εφαρμογές. Ακόμα μπορεί να επιβάλλει δυναμικά κανόνες περιορισμού που αφορούν εφαρμογές οι οποίες μπορεί σε κάποια χρονική στιγμή να επηρεάσουν την δικτυακή απόδοση. Βελτιώνει την ασφάλεια του δικτύου και κάνει ποιο εύκολη τη διαχείριση της. Επίσης παρέχει υψηλή απόδοση, μειώνεται το κόστος καθώς δεν απαιτείται η αγορά συσκευής υλικού και δεν υπάρχει μοναδικό σημείο αστοχίας σε αντίθεση με ένα παραδοσιακό τείχος προστασίας.

2.4.2 Εξισορροπητής φορτίου

Μία ακόμα εφαρμογή που μελετήσαμε είναι ο εξισορροπητής φορτίου. Οι έξυπνες τεχνικές εξισορρόπησης φορτίου μεγιστοποιούν την απόδοση του δικτύου. Μοιράζουν την κίνηση σε πολλαπλές διαδρομές, ελαχιστοποιώντας με αυτό τον τρόπο την συμφόρηση [35]. Οι διαχειριστές δικτύων οι οποίοι χρησιμοποιούν έξυπνους εξισορροπητές φορτίου λογισμικού έχουν το σημαντικό πλεονέκτημα γνώσης προγνωστικών στοιχείων μέσω των οποίων είναι δυνατός ο εντοπισμός σημείων συμφόρησης κίνησης πριν αυτή δημιουργηθεί. Παρέχουν βελτιστοποιημένη κίνηση και ελαχιστοποίηση του χρόνου απόκρισης.

Στα παραδοσιακά δίκτυα δεν υπάρχει μεγάλη ευελιξία. Οι δρομολογητές χρησιμοποιούν κοινά πρωτόκολλα δρομολόγησης όπως το OSPF και το IS-IS για την απόκτηση μετρήσεων σύνδεσης και την τοπολογία του δικτύου. Λόγω της στατικής δρομολόγησης μπορεί να οδηγηθούμε σε συμφόρηση ορισμένων διαδρομών. Αν και θα μπορούσε κάποιος να υποστηρίξει ότι η αλλαγή του βάρους της σύνδεσης θα οδηγήσει σε αλλαγές διαδρομών αυτή δεν είναι η βέλτιστη λύση. Ο λόγος είναι ότι το μήνυμα αλλαγής κατάστασης δεν θα φτάσει σε όλους τους δρομολογητές την ίδια στιγμή, θα φτάσει σε διαφορετικό χρόνο. Επειδή δεν υπάρχει ακριβής συγχρονισμός μεταξύ των δρομολογητών όποτε υπάρχουν ενημερώσεις μετρικών είναι πολύ πιθανή η δημιουργία κάποιον παροδικών βρόχων κυκλοφορίας [36],[37]. Στο IP δίκτυο κυρίως λόγω του γεγονότος ότι δεν υπάρχει καθολική εικόνα του δικτύου είναι δύσκολο να διασφαλιστεί ότι η απόδοση εξισορρόπησης φορτίου είναι η βέλτιστη.

Επίσης οι παραδοσιακοί εξισορροπητές φορτίου δεν μπορούν να προγραμματιστούν επειδή είναι κλειδωμένοι από τον προμηθευτή. Αυτό έχει ως αποτέλεσμα οι διαχειριστές να μην έχουν την ευελιξία δημιουργίας των δικών τους αλγόριθμων για την εξισορρόπηση φορτίου. Το SDN επιτρέπει εξισορροπητές φορτίου λογισμικού οι οποίοι μπορούν να προγραμματιστούν δίνοντας έτσι το πλεονέκτημα σε κάποιον να σχεδιάζει και να εφαρμόζει τις δικές του στρατηγικές εξισορρόπησης φορτίου. Ένα ακόμα πλεονέκτημα των εξισορροπητών φορτίου λογισμικού είναι ότι μειώνεται το κόστος καθώς δεν απαιτείται η αγορά ξεχωριστής συσκευής υλικού. Επιπλέον στο SDN χρησιμοποιείται η δικτυακή τοπολογία για τη λήψη αποφάσεων δρομολόγησης έχοντας υπόψη τις απαιτήσεις εύρους ζώνης και προστασίας των υπηρεσιών που παρέχονται σε ολόκληρο το δίκτυο. Η εφαρμογή εξισορρόπησης φορτίου μαζί με τον ελεγκτή καθορίζουν τις διαδρομές δρομολόγησης χωρίς να χρειάζεται η χρήση αλγόριθμών που υπάρχουν σε κλασικές δικτυακές συσκευές [38].

2.5 Επιπλέον ανάγκες σύγχρονων κέντρων δεδομένων έναντι συμβατικών δικτύων.

Στα σύγχρονα κέντρα δεδομένων έχουν παρουσιαστεί ορισμένες ανάγκες τις οποίες θα εξετάσουμε. Οι ανάγκες αυτές είναι οι παρακάτω:

Πλήθος Διευθύνσεων Mac

Οι δικτυακές συσκευές , διακόπτες και δρομολογητές χρησιμοποιούν πίνακες MAC διευθύνσεων μέσω των οποίων καθορίζεται η πόρτα που θα χρησιμοποιήσουν για να στείλουν τα δεδομένα στον προορισμό τους. Οι πίνακες αυτοί υλοποιούνται με χρήση υλικού ώστε να εξασφαλίζεται μεγαλύτερη ταχύτητα. Ο αριθμός των εγγραφών που μπορεί να περιέχει ο κάθε πίνακας επηρεάζει ανάλογα και το κόστος του. Στα συμβατικά δίκτυα ο μέγιστος αριθμός των MAC διευθύνσεων ήταν διαχειρίσιμος.

Στα σύγχρονα κέντρα δεδομένων λόγω της εικονικοποίησης το μέγεθος δικτύων του δευτέρου επιπέδου έχει επεκταθεί πάρα πολύ. Ο αριθμός των εικονικών διακομιστών με δυνατότητα πολλαπλών εικονικών καρτών δικτύου έχει αυξηθεί σε σημαντικό βαθμό. Αυτό έχει ως αποτέλεσμα την ύπαρξη

τεράστιου αριθμού MAC διευθύνσεων. Οι πίνακες δεν έχουν σχεδιαστεί ώστε να μπορούν να περιέχουν τόσο μεγάλο αριθμό διευθύνσεων. Αυτό έχει ως αποτέλεσμα να γεμίζουν . Όταν δεν υπάρχει μία διεύθυνση στον πίνακα τότε τα δεδομένα προωθούνται προς όλες τις πόρτες. Αυτό επηρεάζει την απόδοση του δικτύου και γίνεται κακή χρήση του εύρους ζώνης.

Ακόμη τα VLAN τα οποία χρησιμοποιούνται πάρα πολύ λειτουργούν ως εξής: Όταν μία διεύθυνση προορισμού δεν υπάρχει στον πίνακα διευθύνσεων τότε γίνεται προώθηση του πακέτου μόνο προς όλες τις θύρες που ανήκουν στο ίδιο VLAN περιορίζοντας έτσι λίγο την κίνηση. Αντίθετα στην περίπτωση που ένας διακομιστής ανήκει σε πολλά VLAN τότε αυξάνεται ο αριθμός των εγγραφών του πίνακα έχοντας μία εγγραφή για κάθε VLAN.

Πλήθος Εικονικών Δικτύων VLANs

Τα εικονικά δίκτυα χρησιμοποιούνται για την υποστήριξη πολλών εικονικών δικτύων μέσα σε ένα φυσικό δίκτυο. Το μέγιστο πλήθος τους έχει οριστεί σε 4096 λόγω χρήσης πεδίου 12 bit για τον προσδιορισμό της ταυτότητας του VLAN. Το πλήθος αυτό είναι πολύ μεγάλο όταν πρόκειται για ένα κέντρο δεδομένων ενός μόνο πελάτη. Στα σύγχρονα κέντρα δεδομένων η φυσική εγκατάσταση χρησιμοποιείται από πολλούς πελάτες δημιουργώντας την ανάγκη για διαχωρισμό και ασφάλεια της κίνησης του κάθε πελάτη. Ο διαχωρισμός αυτός απαιτεί τη χρήση εικονικών δικτύων. Η συνεχής επέκταση των κέντρων δεδομένων με την αύξηση του αριθμού των πελατών και την εικονικοποίηση των διακομιστών δημιουργεί προβλήματα καθώς ο μέγιστος αριθμός των εικονικών δικτύων δεν επαρκή. Η δυνατότητα διαμοιρασμού των φυσικών πόρων όταν δεν υπάρχουν άλλα διαθέσιμα VLAN γίνεται πολύπλοκη [26],[40].

Spanning Tree Protocol

Το πρωτόκολλο αυτό χρησιμοποιείται για την δημιουργία μίας ιεραρχικής δομής χωρίς βρόγχους σε περιπτώσεις που υπάρχουν φυσική βρόγχοι στη δικτυακή τοπολογία. Κάθε φορά που συμβαίνει μία αλλαγή στο δίκτυο γίνονται νέοι υπολογισμοί ώστε να καθοριστεί η νέα ιεραρχική δομή. Η διαδικασία αυτή ονομάζεται σύγκληση. Το μειονέκτημα του πρωτοκόλλου αυτού είναι ότι

αφήνει ανεκμετάλλευτες κάποιες ζεύξεις και οδηγεί όλη την κίνηση στην κορυφή της ιεραρχικής δομής. Αυτή η διαδρομή δεν είναι η βέλτιστη για όλες τις περιπτώσεις. Στα σύγχρονα κέντρα δεδομένων όπου οι απαιτήσεις έχουν αυξηθεί πολύ χρειάζεται να αξιοποιείται όλο το διαθέσιμο εύρος ζώνης και να επιλέγεται κάθε φορά η βέλτιστη διαδρομή μεταξύ δύο κόμβων επικοινωνίας χωρίς να χρειάζεται απαραίτητα να υπάρχει μία ιεραρχική δομή. Ακόμα η αύξηση της εικονικοποίησης οδήγησε σε συχνές αλλαγές. Αυτό έχει ως αποτέλεσμα η διαδικασία της σύγκλησης να γίνεται συχνότερα . Η διαδικασία αυτή χρειάζεται ένα χρονικό διάστημα για να ολοκληρωθεί δημιουργώντας ακόμα ένα αρνητικό για τη χρήση του Spanning Tree Protocol στα σύγχρονα κέντρα δεδομένων [26].

Αλλαγές στους δικτυακούς πόρους

Ένα πολύ σημαντικό θέμα στα σύγχρονα κέντρα δεδομένων είναι ότι οι δικτυακές αλλαγές πρέπει να γίνονται γρήγορα και δυναμικά. Η προσθήκη, τροποποίηση και αφαίρεση δικτυακών πόρων πρέπει να γίνεται το ίδιο γρήγορα όπως γίνονται και οι αλλαγές στους εικονικούς διακομιστές και στις μονάδες αποθήκευσης. Στα συμβατικά δίκτυα μπορεί να χρειάζονται μέρες ή και εβδομάδες για σημαντικές αλλαγές, για παράδειγμα για αλλαγές σε VLANs. Αυτό γίνεται γιατί μία λάθος αλλαγή μπορεί να επηρεάσει όλους τους πόρους του κέντρου δεδομένων και όχι μόνο τους δικτυακούς. Δημιουργείται λοιπόν η ανάγκη οι αλλαγές αυτές να γίνονται αυτόματα χωρίς ανθρώπινη παρέμβαση. Για την παροχή μίας νέας υπηρεσίας, πρέπει πρώτα να δοθούν οι απαραίτητοι δικτυακοί πόροι, να γίνει δηλαδή η εικονική αλλαγή στο δίκτυο και μετά να γίνει οποιαδήποτε κίνηση σε διακομιστές για την παροχή της υπηρεσίας. Στα συμβατικά δίκτυα λόγω του τρόπου σχεδιασμού των πρωτοκόλλων οι ενέργειες γίνονται αντίστροφα [26],[40].

Αποκατάσταση αποτυχίας

Τα σύγχρονα κέντρα δεδομένων επεκτείνονται συνεχώς. Αυτό κάνει αρκετά πολύπλοκη την αποκατάσταση από μία αποτυχία. Χρειάζεται μία πλήρη εικόνα του δικτύου για να μπορεί να γίνει αντιληπτό το σημείο που υπάρχει το πρόβλημα και να γίνει μία γρήγορη αποκατάσταση. Η πρόβλεψη και οι

κατάλληλες αλλαγές μπορούν να βοηθήσουν σημαντικά στην αποφυγή μιας αποτυχίας [26].

Πολλαπλοί πελάτες σε ένα κέντρο δεδομένων

Τα σύγχρονα κέντρα δεδομένων αποτελούνται από μεγάλο αριθμό πελατών ο οποίος μπορεί να φτάσει τις εκατοντάδες ή και χιλιάδες. Αυτό δημιουργεί την ανάγκη οι πελάτες αυτοί να παραμένουν απομονωμένοι μεταξύ τους για λόγους ασφάλειας καθώς και για λόγους ποιότητας υπηρεσιών. Οι πελάτες αυτοί έχουν διαφορετικές ανάγκες οπότε χρειάζεται το κέντρο δεδομένων να εξασφαλίζει διαφορετικές ποιότητες υπηρεσιών στο ίδιο δίκτυο.

Μηχανική κίνησης και αποδοτικότητα μονοπατιού

Η μεγάλη κλιμάκωση των σύγχρονων κέντρων δεδομένων έχει οδηγήσει στη ανάγκη της αξιοποίησης των διαθέσιμων πόρων με το βέλτιστο τρόπο. Για να μπορεί να γίνει αυτό απαιτείται η χρήση κατάλληλων εργαλείων παρακολούθησης και μέτρησης της κίνησης του δικτύου. Οι ενέργειες αυτές θεωρούνταν πολυτέλεια στα συμβατικά δίκτυα. Ωστόσο στα σύγχρονα κέντρα δεδομένων είναι μία επιτακτική ανάγκη ώστε να έχουμε τη βέλτιστη χρήση του διαθέσιμου εύρους ζώνης και των ζεύξεων. Στα παραδοσιακά δίκτυα γίνεται χρήση της συντομότερης διαδρομής η οποία όμως δεν είναι πάντα η καλύτερη επιλογή διότι δεν λαμβάνει υπόψη της δεδομένα όπως το φορτίο κίνησης.

Ένας από τους λόγους που απαιτείται μεγάλη προσοχή στην επιλογή του αποδοτικότερου μονοπατιού είναι η μεγάλη αύξηση της κίνησης μεταξύ Ανατολής-Δύσης σε σχέση με την κίνηση μεταξύ Βορρά-Νότου. Όπως έχουμε αναφέρει και στο πρώτο κεφάλαιο στα παραδοσιακά δίκτυα η περισσότερη κίνηση υπήρχε στο δίκτυο κορμού δηλαδή από το κέντρο δεδομένων προς τον έξω κόσμο και αντίστροφα. Στα σύγχρονα κέντρα δεδομένων υπάρχει πολύ έντονη κίνηση μεταξύ των διακομιστών κατά μήκος όλου του δικτύου του κέντρου δεδομένων [26],[40].

2.6 Προσεγγίσεις SDN και τρόποι χρήσης

Για την καλύτερη αντιμετώπιση των αναγκών των σύγχρονων κέντρων δεδομένων έχουν αναπτυχθεί ορισμένες SDN προσεγγίσεις. Η προσέγγιση Open SDN στην οποία έχουμε αναφερθεί μέχρι στιγμής καλύπτει όλες τις απαιτήσεις των σύγχρονων κέντρων δεδομένων. Υπάρχουν δύο ακόμη προσεγγίσεις οι οποίες έχουν κερδίσει σημαντική προσοχή στην αγορά. Οι προσεγγίσεις αυτές είναι το SDN μέσω δικτύων επικάλυψης που βασίζονται σε hypervisors και το SDN μέσω API.

2.6.1 SDN μέσω δικτύων επικάλυψης που βασίζεται σε hypervisors

Στο SDN μέσω δικτύων επικάλυψης που βασίζεται σε hypervisors δημιουργούνται εικονικά δίκτυα πάνω από τη φυσική δικτυακή υποδομή. Στη περίπτωση αυτή η φυσική υποδομή και οι ρυθμίσεις των δικτυακών συσκευών δεν αλλάζουν. Οι SDN εφαρμογές χρησιμοποιούν τα εικονικά δίκτυα και τις θύρες τα οποία δεν σχετίζονται άμεσα με τα αντίστοιχα φυσικά. Η κίνηση του εικονικού δικτύου περνάει πάνω από τη φυσική υποδομή. Τα τελικά σημεία δεν γνωρίζουν λεπτομέρειες που σχετίζονται για παράδειγμα με το πώς γίνεται η δρομολόγηση μέσω της φυσική τοπολογίας. Τα εικονικά δίκτυα ελέγχονται συνήθως από τους hypervisors των εικονικών μηχανών που βρίσκονται στα άκρα του δικτύου. Χρησιμοποιείται ένας μηχανισμός σήραγγας (tunneling). Τα άκρα του εικονικού δικτύου ονομάζονται τελικά σημεία σήραγγας ή τελικά σημεία εικονικής σήραγγας (VTEPs). Ένα πακέτο στην πηγή στην άκρη του εικονικού δικτύου ενθυλακώνεται συνήθως από τον hypervisor μέσα σε ένα άλλο πλαίσιο. Ο hypervisor έχοντας τις κατάλληλες πληροφορίες από τον ελεγκτή στέλνει το ενθυλακωμένο πακέτο στο VTEP προορισμού. Το VTEP προορισμού όταν λαμβάνει το πακέτο αφαιρεί την ενθυλάκωση και προωθεί το πακέτο όπως ήταν στην αρχική του μορφή στον διακομιστή προορισμού.

Υπάρχουν διάφορες ιδιόκτητες μέθοδοι σήραγγας όπως για παράδειγμα το VXLAN της Cisco [40] , το STT της Nicira και το NVGRE της Microsoft.

Η προσέγγιση αυτή είναι κατάλληλη για τα κέντρα δεδομένων όπου υπάρχει ήδη κατάλληλο λογισμικό για την εικονικοποίηση υπολογιστών και

αποθήκευσης. Καλύπτει πολλές από τις ανάγκες των σύγχρονων κέντρων δεδομένων. Πρώτον καλύπτει το πρόβλημα του πλήθους των MAC διευθύνσεων καθώς είναι ορατές μέσω του φυσικού δικτύου μόνο οι MAC διευθύνσεις των τελικών σημείων της σήραγγας που βρίσκονται στους hypervisors. Επιπλέον για των διαχωρισμών και την απομόνωση της κίνησης μεταξύ των διαφορών πελατών του κέντρου δεδομένων χρησιμοποιείται τεχνολογία tunneling αντί για VLAN. Με αυτό τον τρόπο μπορούμε να έχουμε 16 εκατομμύρια ή και παραπάνω τμήματα δικτύων με σήραγγα χρησιμοποιώντας για παράδειγμα τεχνολογία σήραγγας VXLAN, STT ή NVGRE. Ακόμα λόγω της εικονικοποίησης μπορούν να γίνουν αυτόματα και πολύ εύκολα ενέργειες όπως δημιουργία, διαγραφή, μετακίνηση και αλλαγή εικονικών δικτύων. Οι ενέργειες αυτές γίνονται σε ελάχιστο χρόνο σε αντίθεση με το χρόνο που θα απαιτούνταν εάν γίνονταν οι αλλαγές στην φυσική δικτυακή υποδομή.

Ωστόσο έχει και κάποια μειονεκτήματα. Το λογικό δίκτυο δεν είναι συσχετισμένο με το φυσικό δίκτυο οπότε δεν μπορεί να αντιμετωπίσει προβλήματα που παρουσιάζονται στη φυσική υποδομή. Ακόμα λόγω της αλληλεπίδρασης μεταξύ του εικονικού και του φυσικού δικτύου στην περίπτωση μίας αστοχίας μπορεί να είναι δύσκολο να προσδιοριστεί το σημείο του προβλήματος. Επιπλέον δεν μπορεί να κάνει αποτελεσματική χρήση της δικτυακής υποδομής, να αντιμετωπίσει θέματα προτεραιότητας κίνησης και αποκλεισμένων μονοπατιών λόγω του Spanning Tree Protocol.

2.6.2 SDN μέσω API

Η επόμενη προσέγγιση που θα μελετήσουμε είναι το SDN μέσω API. Σε αυτή την περίπτωση ο ελεγκτής έχει πρόσβαση σε APIs τα οποία βρίσκονται στις δικτυακές συσκευές και μπορεί μέσω αυτών να διαμορφώσει και να προγραμματίσει το επίπεδο ελέγχου που βρίσκεται στις δικτυακές συσκευές. Υπάρχουν APIs παλαιού τύπου τα οποία χρησιμοποιούν μεθόδους όπως SNMP και CLI τα οποία δεν είναι για συχνή χρήση και επιτρέπουν ενέργειες όπως στατική διαχείριση. Για τα σύγχρονα κέντρα δεδομένων όπου απαιτούνται συχνές, αυτοματοποιημένες και δυναμικές εργασίες χρησιμοποιούνται άλλου τύπου APIs ώστε να είναι δυνατή η απομακρυσμένη

πραγματοποίηση αλλαγών. Το πιο γνωστό API νέου τύπου είναι το RESTful API το οποίο χρησιμοποιεί το πρωτόκολλο HTTP ή HTTPS και τυπική θύρα TCP με αποτέλεσμα να μην απαιτείται ειδική ρύθμιση στο τείχος προστασίας για τις κλήσεις API.

Ένα πλεονεκτήματα χρήσης αυτής της προσέγγισης είναι ότι χρησιμοποιεί κλασικούς διακόπτες και έτσι δεν απαιτείται η αναβάθμιση ή η αντικατάσταση των διακοπτών με νεότερους που να υποστηρίζουν το πρωτόκολλο OpenFlow. Αυτό θα μπορούσε να είναι μια καλή λύση στη περίπτωση που μία εταιρία έχει ήδη επενδύσει σε ακριβές κλασικού τύπου συσκευές και θα ήταν επιπλέον κόστος η αντικατάσταση τους με λιγότερο ακριβές και πιο απλές συσκευές. Ακόμα επιτρέπει σε κάποιο βαθμό τον έλεγχο των δικτυακών συσκευών από ένα κεντρικό σημείο, τον ελεγκτή. Επιπλέον παρέχει τη δυνατότητα γρήγορων και αυτόματων αλλαγών. Τα Restful APIs κάνουν δυνατή τη δημιουργία λογισμικού το οποίο μπορεί να χρησιμοποιηθεί για συντονισμό ώστε να γίνονται αυτόματα και γρήγορα δικτυακές αλλαγές**.**

Ωστόσο αυτή η προσέγγιση δεν δίνει λύση στο πρόβλημα των MAC διευθύνσεων και των VLAN ούτε μπορεί να προσφέρει κάτι στο ζήτημα των πολλαπλών πελατών. Η χρήση ενός ελεγκτή και η αξιοποίηση νέου τύπου βελτιωμένων APIs για την αυτόματη ενημέρωση των συσκευών μπορούν να προσφέρουν μία βελτίωση στην περίπτωση αποτυχίας. Εάν οι αποφάσεις για τις διαδρομές λαμβάνονται από τον ελεγκτή τότε τα βελτιωμένα APIs μπορούν να βοηθήσουν σε περιπτώσεις αστοχιών. Ωστόσο αυτή δεν αποτελεί μία κλασική περίπτωση SDN μέσω API. Ακόμα το SDN μέσω API μπορεί να προσφέρει σε κάποιο βαθμό λύση σε θέματα αποδοτικότητας διαδρομών.

2.6.3 Open SDN

Η προσέγγιση του Open SDN όπως αναφέραμε και παραπάνω αντιμετωπίζει όλες τις ανάγκες των σύγχρονων κέντρων δεδομένων. Ο ελεγκτής έχει τη δυνατότητα να δημιουργήσει σήραγγες μεταξύ τελικών σημείων και χρησιμοποιώντας το πρωτόκολλο OpenFlow να ορίσει κανόνες μέσω των οποίων θα γίνει η προώθηση της κίνησης στα κατάλληλα tunnels. Η χρήση συσκευών που υποστηρίζουν τη λειτουργία σήραγγας έχει ως αποτέλεσμα την καλύτερη απόδοση λόγω υλικού. Έτσι μπορεί να αντιμετωπίσει τα

προβλήματα περιορισμών των MAC διευθύνσεων και των VLANs με παρόμοιο τρόπο με την προσέγγιση μέσω δικτύων επικάλυψης. Επίσης αν χρησιμοποιήσουμε το Open SDN για τη δημιουργία εικονικών δικτύων μπορούν και σε αυτή την περίπτωση να γίνουν αυτόματα και πολύ εύκολα ενέργειες όπως δημιουργία, διαγραφή, μετακίνηση και αλλαγή των δικτύων. Το πλεονέκτημα του Open SDN σε σχέση με την προσέγγιση μέσω δικτύων επικάλυψης είναι ότι μπορεί να πραγματοποιήσει αλλαγές και στο φυσικό δίκτυο που υπάρχει από κάτω. Ακόμα μπορεί να προσφέρει τη δυνατότητα χρήσης σήραγγας και στο επίπεδο δύο αυξάνοντας τον αριθμό των πολλαπλών πελατών που μπορούν να υπάρχουν σε ένα κέντρο δεδομένων.

Στο Open SDN ο ελεγκτής έχει μία πλήρη εικόνα της τοπολογίας ολόκληρου του δικτύου. Αυτό του δίνει τη δυνατότητα να αντιλαμβάνεται που υπάρχει πρόβλημα και να μπορεί να αλλάξει τη δρομολόγηση της κίνησης σε περίπτωση εμφάνισης κάποιας αποτυχίας. Μπορεί να προβλέψει καταστάσεις και να πάρει κατάλληλες αποφάσεις δρομολόγησης το οποίο είναι πολύ σημαντικό. Μπορεί να λάβει υπόψη παράγοντες όπως το εύρος ζώνης και το φορτίο κίνησης σε κάθε διαδρομή. Έχει την εικόνα των στατιστικών τα οποία λαμβάνει από τις δικτυακές συσκευές και αντιλαμβάνεται πιο μονοπάτι είναι λιγότερο επιφορτισμένο. Έτσι μπορεί να ανοίγει εναλλακτικά μονοπάτια και να προσφέρει πολύ καλύτερη απόδοση [26].

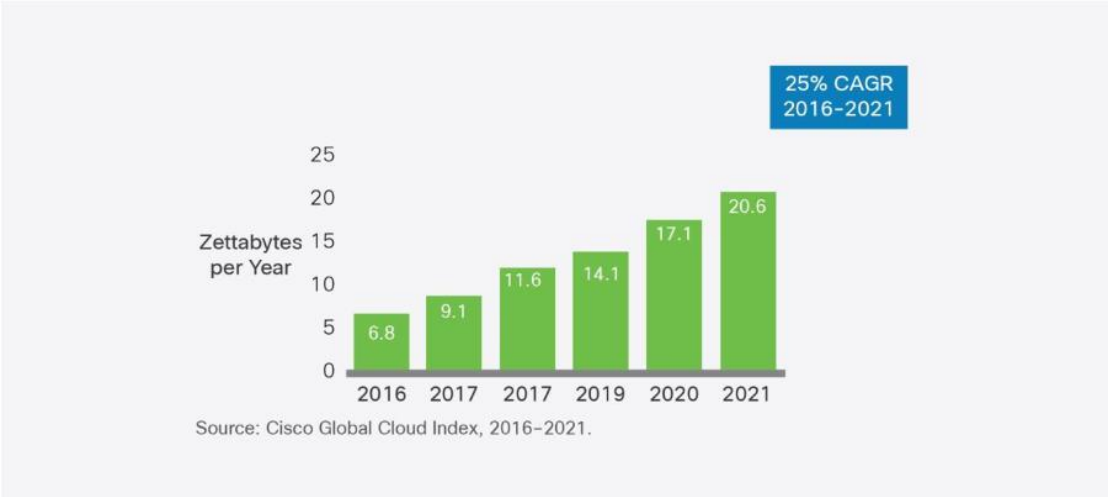
ΚΕΦΑΛΑΙΟ 3 : Προκλήσεις στα σύγχρονα δίκτυα κέντρων δεδομένων

Παρακάτω θα αναφερθούμε στις προκλήσεις και τις απαιτήσεις που δημιουργούνται στα σύγχρονα δίκτυα των κέντρων δεδομένων λόγω των τάσεων που επικρατούν.

3.1 Κλιμάκωση κίνησης κέντρων δεδομένων

Τα σύγχρονα κέντρα δεδομένων χρησιμοποιούνται για την παροχή εφαρμογών και υπηρεσιών όπως υπηρεσίες υπολογιστικής νέφους, εφαρμογές μηχανικής μάθησης, ροής πολυμέσων υψηλής ευκρίνειας, 5G κινητές επικοινωνίες, διαδίκτυο των πραγμάτων, υπηρεσίες κοινωνικής δικτύωσης και άλλα. Η κλιμάκωση αυτών των υπηρεσιών και εφαρμογών έχει αυξήσει σε μεγάλο βαθμό την κίνηση μέσα στα κέντρα δεδομένων [41],[42],[43].

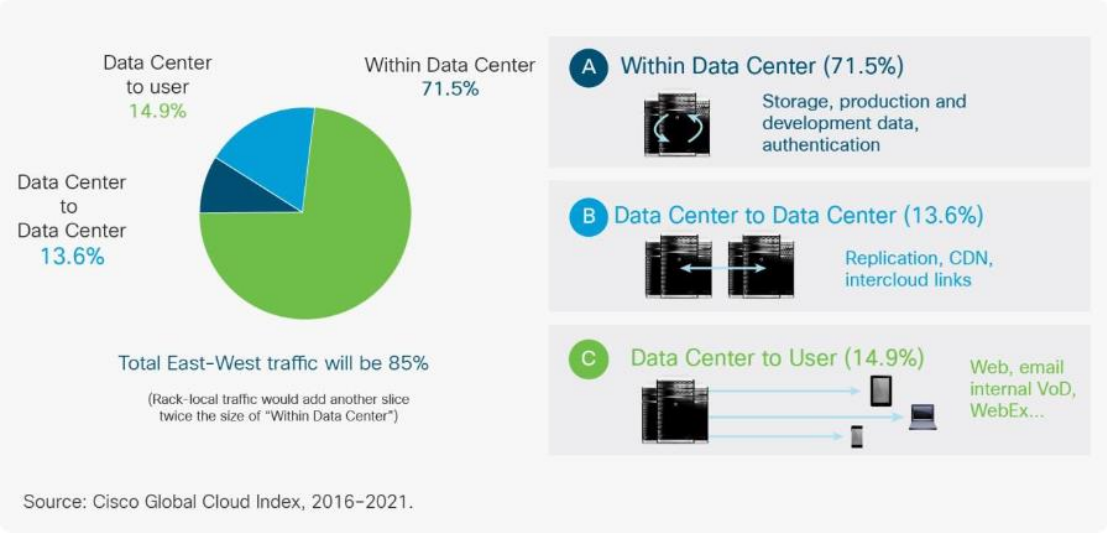
Παρακάτω βλέπουμε σύμφωνα με στοιχεία από την εταιρία Cisco ότι η κλιμάκωση αυτών των εφαρμογών και υπηρεσιών οδηγεί σε μία ετήσια αύξηση της κίνησης στα κέντρα δεδομένων της τάξης του 25%. Η παγκόσμια κίνηση των κέντρων δεδομένων το 2021 υπολογίζεται γύρο στο 20,6ZB το 2021 και είναι τριπλάσια σε σχέση με το 2016 όπου ήταν 6,8ZB .



Εικόνα 3.1: Παγκόσμια αύξηση κίνησης κέντρου δεδομένων (“Cisco Global Cloud Index: Forecast and Methodology, 2016-2021”)

Η κίνηση αυτή είναι πολύ μεγαλύτερη σε σχέση με την κίνηση του διαδικτύου και των WAN δικτύων.

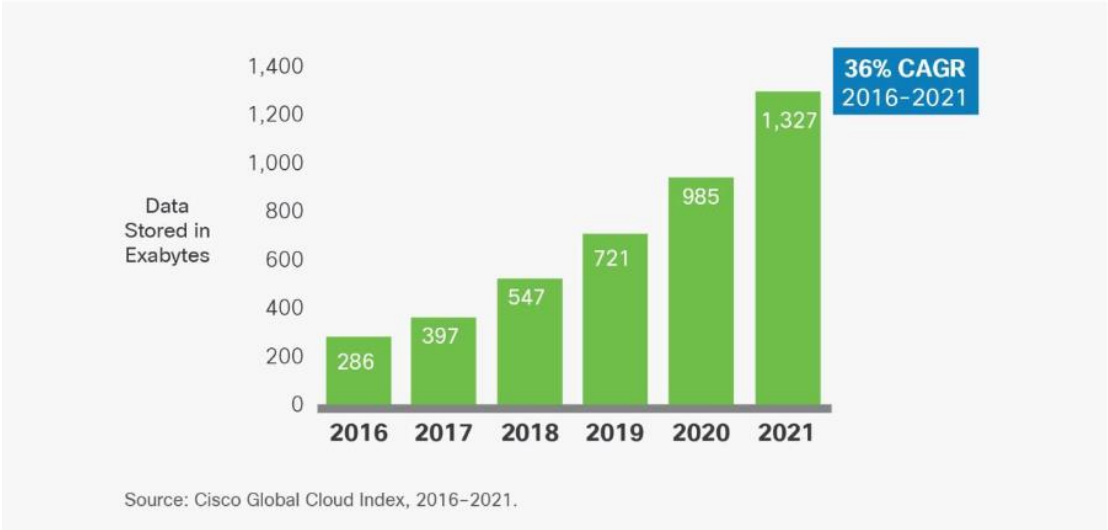
Όπως βλέπουμε και στην παρακάτω εικόνα το μεγαλύτερο ποσοστό της κίνησης είναι μέσα στα κέντρα δεδομένων.



Εικόνα 3.2: Παγκόσμια κίνηση κέντρου δεδομένων (“Cisco Global Cloud Index: Forecast and Methodology, 2016-2021”)

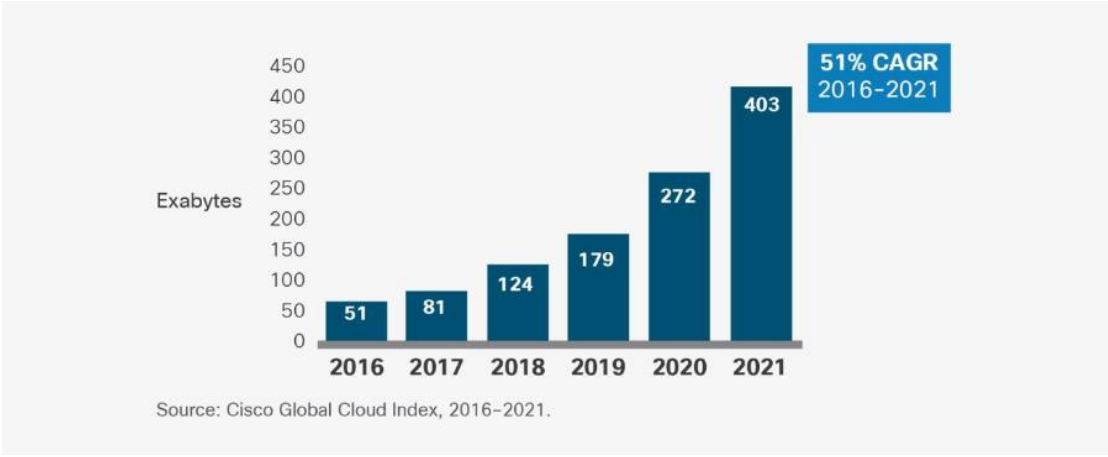
Είναι η λεγόμενη κίνηση μεταξύ Ανατολής – Δύσης. Περιλαμβάνει την κίνηση εντός του κέντρου δεδομένων καθώς και την κίνηση μεταξύ των κέντρων δεδομένων και υπολογίζεται γύρο στο 85% το 2021. Η κίνηση από το κέντρο δεδομένων προς το διαδίκτυο ή δίκτυα WAN αποτελεί μόνο το 15%. Είναι η λεγόμενη κίνηση μεταξύ Βορρά – Νότου. Η κίνηση αυτή είναι μεταξύ του κέντρου δεδομένων και του χρήστη, κίνηση μεταξύ πελάτη και εξυπηρετητή για διάφορες υπηρεσίες όπως email, Web, WebEx και διάφορα άλλα.

Ακόμα η αύξηση των δεδομένων αποθήκευσης είναι μεγάλη και διαφέρει ανάλογα με τον τύπο. Γενικά είναι μεταξύ 30 και 70 τοις εκατό. Παρακάτω βλέπουμε σύμφωνα με τα στοιχεία που μας παρέχει η εταιρία Cisco ότι τα δεδομένα τα οποία είναι αποθηκευμένα στα κέντρα δεδομένων παγκόσμια υπολογίζονται στα 1,3ZB το 2021. Βλέπουμε ότι υπάρχει μία αύξηση της τάξης του 36% περίπου ετησίως.



Εικόνα 3.3: Δεδομένα που είναι αποθηκευμένα σε κέντρα δεδομένων (“Cisco Global Cloud Index: Forecast and Methodology, 2016-2021”)

Τα μεγάλα δεδομένα (big data) για παράδειγμα τα οποία δημιουργούν μεγάλες απαιτήσεις κίνησης παίζουν πολύ σημαντικό ρόλο στην αύξηση των συνολικών αποθηκευμένων δεδομένων. Υπολογίζονται γύρο στα 403EB το 2021 αποτελώντας το 30% του συνόλου των αποθηκευμένων δεδομένων. Όπως βλέπουμε και στην παρακάτω εικόνα έχουν ένα αρκετά μεγάλο ετήσιο ρυθμό αύξησης που φτάνει το 51% περίπου [41].



Εικόνα 3.4: Όγκος μεγάλων δεδομένων (“Cisco Global Cloud Index: Forecast and Methodology, 2016-2021”)

Για να μπορέσουν τα δίκτυα κέντρων δεδομένων να ανταποκριθούν και να καλύψουν τις υψηλές αυτές ανάγκες χρειάζεται να επεκταθούν και να γίνουν πιο ευέλικτα [43],[44],[45].

3.2 Υψηλό εύρος ζώνης και χαμηλή καθυστέρηση

Η κλιμάκωση της κίνησης των δικτύων των κέντρων δεδομένων δημιουργεί πρόσθετες ανάγκες καθώς οι παραδοσιακοί ηλεκτρικοί διακόπτες που χρησιμοποιούνται για την προώθηση της κίνησης χρησιμοποιούν ολοκληρωμένο κύκλωμα ειδικής εφαρμογής (ASIC) το οποίο έχει περιορισμένο εύρος ζώνης εισόδου/εξόδου. Η αύξηση του εύρους ζώνης θα μπορούσε να γίνει χρησιμοποιώντας μία δομή πολλαπλών επιπέδων, αξιοποιώντας μία στοίβα πολλαπλών ASICs και αυξάνοντας με αυτό τον τρόπο το διαθέσιμο εύρος ζώνης. Ωστόσο η λύση αυτή εισάγει επιπλέον καθυστέρηση στο δίκτυο καθώς επίσης και σημαντικό κόστος και μεγάλη κατανάλωση ενέργειας για τις επιπλέον διασυνδέσεις υψηλής χωρητικότητας [46][47].

Οι αρχιτεκτονικές δικτύων κέντρων δεδομένων οι οποίες χρησιμοποιούν πολλαπλά επίπεδα μεταγωγής με ηλεκτρικούς διακόπτες όπως για παράδειγμα η αρχιτεκτονική Fat-tree δεν μπορούν να προσφέρουν την απαιτούμενη ευελιξία και απόδοση για να καλύψουν τις υψηλές ανάγκες που δημιουργεί η κλιμάκωση της κίνησης των δικτύων [48],[49],[50]. Οι κλασικές ηλεκτρικές διασυνδέσεις δεν μπορούν να προσαρμόσουν το εύρος ζώνης και την τοπολογία του δικτύου. Η σταθερή δικτυακή τοπολογία δύσκολα καλύπτει τις σύγχρονες ανάγκες όπως για παράδειγμα τις ανάγκες κίνησης των νευρωνικών δικτύων. Η κλιμάκωση των δικτύων δεν μπορεί να ανταποκριθεί στα δεδομένα εκπαίδευσης τα οποία αυξάνονται με πολύ υψηλούς ρυθμούς [43].

Τα οπτικά δίκτυα μπορούν να ανταποκριθούν στις απαιτήσεις των σύγχρονων κέντρων δεδομένων. Οι οπτικοί διακόπτες και οι οπτικές διασυνδέσεις μπορούν να προσφέρουν πολύ υψηλότερο εύρος ζώνης και χαμηλή καθυστέρηση. Ακόμα έχουν τη δυνατότητα αναδιαμόρφωσης. Λόγω του υψηλού εύρους ζώνης δεν χρειάζεται η χρήση ιεραρχικής δομής. Αντίθετα

επιτρέπουν τη χρήση μίας περισσότερο επίπεδης αρχιτεκτονικής η οποία βοηθάει στην χαμηλή καθυστέρηση [43],[51],[52].

Σε δίκτυα μεγάλης κλίμακας για επικοινωνία σε μεγάλες αποστάσεις οι οπτικές διασυνδέσεις είναι πολύ καλύτερες σε σχέση με τις ηλεκτρικές διότι δεν εισάγουν καθυστέρηση σε σχέση με την απόσταση σε αντίθεση με τις ηλεκτρικές διασυνδέσεις.

Επίσης μπορούμε να έχουμε οπτικές διασυνδέσεις σε ένα μήκος κύματος με μετάδοση μέχρι και 800 Gb/s, σε αντίθεση με τις ηλεκτρικές διασυνδέσεις όπου για μικρές αποστάσεις ένα καλώδιο φτάνει μόνο δεκάδες Gb/s [43].

3.3 Πολλαπλοί πελάτες στην ίδια δικτυακή υποδομή

Μια ακόμη πρόκληση που παρουσιάζεται στα σύγχρονα δίκτυα κέντρων δεδομένων είναι η διαχείριση υπηρεσιών και εφαρμογών από πολλαπλούς πελάτες οι οποίοι υπάρχουν ταυτόχρονα στην ίδια δικτυακή υποδομή. Οι εφαρμογές για πολλαπλούς πελάτες αναπτύσσονται με πολύ γρήγορους ρυθμούς, έχουν διάφορες ροές κίνησης και διαφορετικές απαιτήσεις για την καθυστέρηση και τις απώλειες πακέτων. Το γεγονός αυτό δημιουργεί την ανάγκη για παροχή διαφορετικής ποιότητας υπηρεσίας σε κάθε πελάτη δυναμικά [53],[54].

Τα κλασσικά ιεραρχικά δίκτυα κέντρων δεδομένων που αποτελούνται μόνο από ηλεκτρικούς διακόπτες δεν είναι κατάλληλα για να φιλοξενούν με ασφάλεια κρίσιμες εφαρμογές για πολλαπλούς πελάτες. Για τις απαιτήσεις αυτές είναι κατάλληλη η χρήση οπτικών δικτύων κέντρων δεδομένων με δυναμική παροχή ποιότητας υπηρεσίας.

Μια καλή τακτική είναι ο ευέλικτος τεμαχισμός της υποδομής με ένα διαχειρίσιμο και λειτουργικό τρόπο και η παροχή διαφορετικής ποιότητας υπηρεσιών σύμφωνα με τις ανάγκες και τις απαιτήσεις του κάθε πελάτη. Η OPSquare όπως βλέπουμε στο άρθρο [55] είναι μία αρχιτεκτονική οπτικών δικτύων κέντρων δεδομένων η οποία ευέλικτα και αυτόματα παρέχει και διαμορφώνει το κάθε τεμάχιο της υποδομής του κέντρου δεδομένων μέσω ενός εκτεταμένου SDN επιπέδου ελέγχου για να είναι δυνατή η παροχή διαφορετικής ποιότητας υπηρεσίας μεταξύ των πολλαπλών πελατών [55].

3.4 Κατανάλωση ενέργειας, κόστος και πολυπλοκότητα

Τα κλασσικά δίκτυα κέντρων δεδομένων με ηλεκτρικές διασυνδέσεις όσο κλιμακώνονται δημιουργούν ορισμένα προβλήματα. Όσο μεγαλώνει η διάμετρος του δικτύου μεγαλώνει σε σημαντικό βαθμό το κόστος του και η πολυπλοκότητα του [43] . Όπως αναφέραμε και παραπάνω για να μπορέσει να αυξηθεί το εύρος ζώνης και να καλύπτει τις σύγχρονες ανάγκες μπορούν να χρησιμοποιηθούν πολλαπλά ASICs χρησιμοποιώντας μια δομή πολλαπλών επιπέδων. Αυτό ωστόσο αυξάνει σημαντικά το κόστος και την πολυπλοκότητα λόγω της ανάγκης χρήσης επιπλέον διασυνδέσεων υψηλής χωρητικότητας οι οποίες καταναλώνουν μεγάλα ποσά ενέργειας [46],[47].

Οι ηλεκτρικές διασυνδέσεις καταναλώνουν σημαντικά ποσά ενέργειας και το κόστος αυτό είναι ένας παράγοντας ο οποίος περιορίζει την επέκταση των κέντρων δεδομένων [56],[57]. Οι οπτικές διασυνδέσεις από την άλλη καταναλώνουν πολύ μικρότερη ενέργεια. Επιπλέον η χρήση οπτικών διακοπτών μειώνει ακόμη περισσότερο την κατανάλωση ενέργειας διότι δεν απαιτούνται μετατροπές από οπτικό σήμα σε ηλεκτρικό και ξανά οπτικό. Αυτό μας προσφέρει ένα σημαντικό πλεονέκτημα απόδοσης από ενεργειακή και οικονομική άποψη [58],[59],[60]. Επίσης μειώνεται ο χρόνος μεταγωγής. Οι αμιγώς οπτικές αρχιτεκτονικές δικτύων κέντρων δεδομένων μπορούν να πετύχουν μία μείωση στην κατανάλωση ενέργειας που μπορεί να φτάσει μέχρι και το 75% μειώνοντας έτσι σε τεράστιο βαθμό και το κόστος [43].

3.5 Ευέλικτη κατανομή των πόρων και βελτίωση της αξιοποίησης τους

Η τεράστια κλίμακα των σύγχρονων κέντρων δεδομένων έχει δημιουργήσει την ανάγκη για βελτίωση της αποτελεσματικότητας τους [61]. Το φορτίο εργασίας στα κέντρα δεδομένων παγκόσμια υπολογίζεται γύρω στα 500 εκατομμύρια το 2021 σύμφωνα με στοιχεία της εταιρίας Cisco [41] και παρουσιάζει έναν ετήσιο ρυθμό αύξησης γύρο στο 19%.

Εικόνα 3.5: Παγκόσμιο φορτίο εργασίας κέντρων δεδομένων (“Cisco Global Cloud Index: Forecast and Methodology, 2016-2021”)

Για να μπορέσουν λοιπόν τα κέντρα δεδομένων να ανταποκριθούν στις μεγάλες ανάγκες λόγω του υψηλού φορτίου εργασίας πρέπει οι χειριστές τους να αυξήσουν τη συνολική χωρητικότητα των διαθέσιμων πόρων δηλαδή τη χωρητικότητα των δικτυακών πόρων, των υπολογιστικών πόρων και των πόρων αποθήκευσης. Ωστόσο η χρησιμοποίηση της μνήμης και των κεντρικών μονάδων επεξεργασίας δεν είναι μεγάλη. Τα συμπλέγματα χρησιμοποιούν περίπου το 50 τοις εκατό της μνήμης για το 55 τοις εκατό του χρόνου και από 10 έως 30 τοις εκατό της χωρητικότητας της κεντρικής μονάδας επεξεργασίας το 80 τοις εκατό του χρόνου [62].

Παρατηρείται λοιπόν ότι ένα μεγάλο ποσοστό των διαθέσιμων πόρων δεν μπορεί να αξιοποιηθεί. Για να μπορέσουν τα σύγχρονα κέντρα δεδομένων να ανταποκριθούν στη μεγάλη αύξηση του φορτίου εργασίας θα πρέπει να τοποθετηθεί επιπλέον υλικό πράγμα το οποίο αυξάνει σημαντικά το κόστος και την κατανάλωση ενέργειας.

Το χαμηλό ποσοστό χρήσης των διαθέσιμων πόρων εμφανίζεται γιατί οι απαιτήσεις για πόρους από τις εφαρμογές και τις υπηρεσίες δεν ταιριάζουν με τις ποσότητες των πόρων οι οποίες βρίσκονται μέσα σε λεπίδες φυσικών διακομιστών (blade servers) αλλιώς γνωστών και ως ολοκληρωμένων διακομιστών. Οι ποσότητες των πόρων μέσα σε αυτούς τους ολοκληρωμένους διακομιστές είναι σταθερές [63],[64].

Οι παραπάνω διακομιστές είναι τοποθετημένοι μέσα σε διαφορετικά ικριώματα, συνδέονται μέσω καρτών δικτύου με τους ToR διακόπτες και επικοινωνούν μέσω κίνησης Ethernet/IP. Η ποσότητα των πόρων είναι σταθερή σε κάθε διακομιστή. Λόγω αυτής της σταθερής διαμόρφωσης του υλικού δημιουργείται εξάρτηση πόρων. Αυτό σημαίνει ότι όταν ένα τύπος από τους πόρους έχει εξαντληθεί σε έναν διακομιστή τότε αυτός ο διακομιστής αν και μπορεί να έχει διαθέσιμο μεγάλο ποσοστό από τους υπόλοιπους τύπους των πόρων του δεν μπορεί να εκτελέσει άλλο φορτίο εργασίας. Αντίστοιχα στην περίπτωση αποτυχίας ενός από τους τύπους των πόρων του διακομιστή επηρεάζεται η διαθεσιμότητα όλων των πόρων πράγμα το οποίο οδηγεί στην αποτυχία ολόκληρου του διακομιστή.

Ακόμα η τοποθέτηση όλων των τύπων πόρων μέσα στο ίδιο πλαίσιο διακομιστή δεν επιτρέπει την αναβάθμιση ή την αλλαγή μόνο σε ορισμένους από τους τύπους των πόρων. Στην περίπτωση αυτή απαιτείται η αντικατάσταση ολόκληρου του διακομιστή με κάποιον άλλο. Αυτό μπορεί να οδηγήσει στην αναβολή χρήσης υλικών νεότερης γενιάς λόγω του κόστους που απαιτείται για την ολική αναβάθμιση [63].

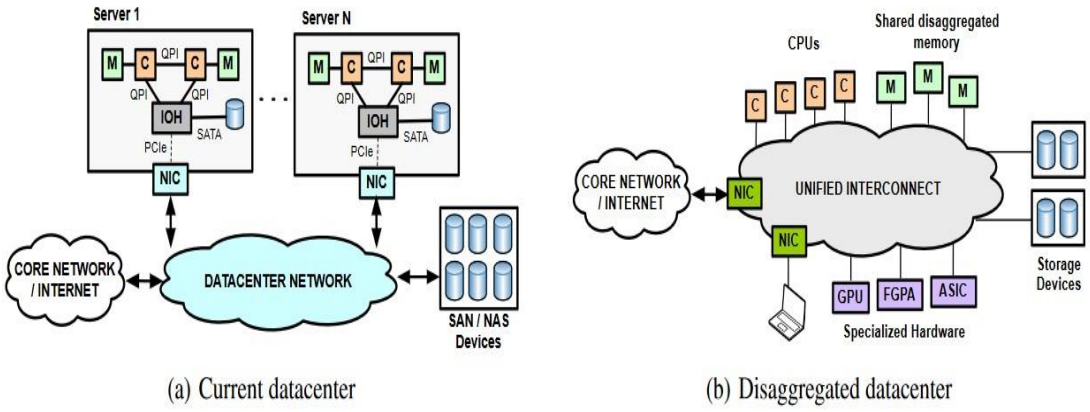
3.5.1 Διαχωρισμένα κέντρα δεδομένων

Ένας πιθανός τρόπος για την αποφυγή της μη σωστής αξιοποίησης των πόρων είναι ο διαχωρισμός τους. Οι διαφορετικοί τύποι πόρων αποσυνδέονται μεταξύ τους δίνοντας έτσι τη δυνατότητα να μπορούν να κατανεμηθούν ξεχωριστά στις διάφορες υπηρεσίες και εφαρμογές σε αντίθεση με τους ολοκληρωμένους διακομιστές. Η ευέλικτη κατανομή των πόρων αυξάνει σημαντικά την διαθεσιμότητα και την αξιοποίηση τους.

Τα διαχωρισμένα κέντρα δεδομένων χωρίζονται σε δύο κατηγορίες με βάση το επίπεδο διαχωρισμού. Τα μερικώς διαχωρισμένα και τα πλήρως διαχωρισμένα κέντρα δεδομένων. Τα μερικώς διαχωρισμένα κέντρα δεδομένων έχουν χρησιμοποιηθεί ευρέως τα τελευταία χρόνια. Ο διαχωρισμός των μονάδων αποθήκευσης ανήκει σε αυτή την περίπτωση. Οι υπόλοιποι υπολογιστικοί πόροι εξακολουθούν να υπάρχουν μέσα σε ολοκληρωμένους διακομιστές και διασυνδέονται με τις μονάδες αποθήκευσης μέσω ξεχωριστής δικτυακής υποδομής. Μπορεί να χρησιμοποιείται μία ειδική

κάρτα δικτύου για παράδειγμα InfiniBand για την επικοινωνία των μονάδων αποθήκευσης με τους υπολογιστικούς πόρους. Σε αυτή την περίπτωση η χρησιμοποίηση της μνήμης και της κεντρικής μονάδας επεξεργασίας εξακολουθεί να είναι εξαρτημένη.

Στα πλήρως διαχωρισμένα κέντρα δεν υπάρχουν ολοκληρωμένοι διακομιστές που να περιέχουν διάφορους τύπους πόρων αντίθετα υπάρχουν διάφορες μονάδες οι οποίες περιέχουν πόρους ίδιου τύπου. Οι μονάδες μπορεί να είναι μία λεπίδα (blade), ένα ικρίωμα, ή ακόμα και ένα ολόκληρο σύμπλεγμα. Αυτές οι μονάδες διασυνδέονται για να επικοινωνούν μεταξύ τους. Τα πλήρως διαχωρισμένα κέντρα δεδομένων ενισχύουν σημαντικά τη χρησιμοποίηση και τη διαθεσιμότητα των πόρων. Επίσης επιτρέπουν την αναβάθμιση και την αντικατάσταση μεμονωμένων πόρων όποτε είναι αναγκαίο [64],[65].



Εικόνα 3.6: Αρχιτεκτονικές διαφορές μεταξύ κέντρων δεδομένων με επίκεντρο το διακομιστή και τους πόρους. (“Network Requirements for Resource Disaggregation”, Gao et.al., 2016)

Ωστόσο η επικοινωνία μεταξύ διαφορετικών τύπων πόρων έχει αυστηρές απαιτήσεις για πολύ υψηλό εύρος ζώνης μετάδοσης και εξαιρετικά μικρή καθυστέρηση. Οι απαιτήσεις αυτές διαφέρουν ανάλογα με τον τύπο των πόρων. Το απαιτούμενο εύρος ζώνης μπορεί να κυμαίνεται από λίγα gigabit μέχρι αρκετές εκατοντάδες gigabit το δευτερόλεπτο. Επίσης οι απαιτήσεις καθυστέρησης για επικοινωνία μεταξύ κεντρικής μονάδας επεξεργασίας και

μονάδας αποθήκευσης είναι της κλίμακας των milliseconds, ενώ μεταξύ μνήμης και κεντρικής μονάδας επεξεργασίας είναι της κλίμακας των nanoseconds. Αν οι παραπάνω απαιτήσεις δεν μπορούν να ικανοποιηθούν αυτό θα έχει ως αποτέλεσμα τη σημαντική μείωση στην απόδοση των εφαρμογών και υπηρεσιών. Οι οπτικές επικοινωνίες μπορούν να προσφέρουν χαμηλή καθυστέρηση και υψηλό εύρος ζώνης αλλά όχι απεριόριστο [65].

Το δίκτυο αποτελεί έναν βασικό παράγοντα για τη χρήση ή τον αποκλεισμό του διαχωρισμού των πόρων καθώς η επικοινωνία που πριν γινόταν μέσα σε έναν διακομιστή πρέπει στην περίπτωση διαχωρισμένων πόρων να διασχίσει το δίκτυο. Αυτό αυξάνει τη δικτυακή κίνηση. Θα πρέπει το δίκτυο να είναι σε θέση να υποστηρίξει εξαιρετικά χαμηλή καθυστέρηση για το αυξημένο φορτίο. Οι οπτικές τεχνολογίες αν και προσφέρουν όπως είπαμε και παραπάνω χαμηλή καθυστέρηση και υψηλό εύρος ζώνης δεν μπορούν πάντα να ικανοποιήσουν τις υψηλές απαιτήσεις εύρους ζώνης που χρειάζονται τα πλήρως διαχωρισένα κέντρα δεδομένων. Αποτελεί πρόκληση η πρόοδος στις οπτικές επικοινωνίες ώστε να είναι δυνατή η πλήρης αξιοποίηση των πλεονεκτημάτων των πλήρως διαχωρισμένων κέντρων δεδομένων [61],[65],[66].

ΚΕΦΑΛΑΙΟ 4 : Αμιγώς οπτικές αρχιτεκτονικές

4.1 Σύγχρονες απαιτήσεις

Τα κλασικά πολυεπίπεδα δίκτυα κέντρων δεδομένων δεν μπορούν να ανταποκριθούν στην μεγάλη ζήτηση που υπάρχει για κέντρα δεδομένων υψηλής απόδοσης και εφαρμογών νέφους στα οποία το φορτίο εργασίας αυξάνεται συνεχώς [69]. Οι σύγχρονες εφαρμογές όπως για παράδειγμα η μηχανική μάθηση δημιουργούν μεγαλύτερη ανάγκη για ευέλικτα και κλιμακούμενα δίκτυα. Η κλίμακα των δικτύων αυξάνεται συνεχώς λόγω της γρήγορης ανάπτυξης του μεγέθους του μοντέλου, των δεδομένων εκπαίδευσης και των απαιτήσεων που δημιουργούνται για υπολογιστικούς πόρους. Η ευρεία χρήση της τεχνητής νοημοσύνης, ο όγκος των μεγάλων δεδομένων, η υπολογιστική νέφους και άλλα έχουν αλλάξει σημαντικά τις ανάγκες των δικτύων κέντρων δεδομένων. Οι απαιτήσεις κίνησης έχουν αυξηθεί σημαντικά.

Τα παραδοσιακά ηλεκτρικά δίκτυα δεν μπορούν να προσαρμόσουν την τοπολογία του δικτύου και το διαθέσιμο εύρος ζώνης. Αυτό έχει ως αποτέλεσμα να γίνεται σπατάλη των διαθέσιμων πόρων και να απαιτείται υψηλό κόστος για την επέκταση των δικτύων. Η σταθερή δικτυακή τοπολογία είναι δύσκολο να προσαρμοστεί με τις σύγχρονες απαιτήσεις κίνησης. Οι απαιτήσεις εύρους ζώνης και καθυστέρησης καθώς επίσης και η μεγάλη κατανάλωση ενέργειας καθιστούν ακατάλληλα τα παραδοσιακά ηλεκτρικά δίκτυα για την αντιμετώπιση των σύγχρονων απαιτήσεων. Οι οπτικές διασυνδέσεις από την άλλη παρέχουν υψηλό εύρος ζώνης, έχουν τη δυνατότητα επαναδιαμόρφωσης και καταναλώνουν πολύ λιγότερη ενέργεια. Για το λόγο αυτό αποτελούν καλύτερη επιλογή για την αντιμετώπιση αυτών των αναγκών [43].

Όπως έχουμε αναφέρει και νωρίτερα τα κέντρα δεδομένων αποτελούν μέρος στο οποίο υπάρχουν οι υπολογιστικοί και αποθηκευτικοί πόροι στο νέφος. Οι υπηρεσίες νέφους δίνουν τη δυνατότητα ανάπτυξης όταν απαιτείται νέων εφαρμογών πολύ εύκολα, γρήγορα και αποδοτικά από άποψη κόστους. Οι εφαρμογές εκτελούνται μέσα σε εικονικές μηχανές και μπορούν να

δημιουργηθούν πολύ γρήγορα. Ωστόσο για την χρησιμοποίηση των εφαρμογών πρέπει να δημιουργηθούν αντίστοιχα και οι δικτυακές υπηρεσίες. Στα κλασσικά δίκτυα λόγω της ανθρώπινης διαμόρφωσής τους υπάρχει μεγάλη καθυστέρηση και αύξηση του λειτουργικού κόστους. Για την αντιμετώπιση αυτών των ζητημάτων υπάρχουν κατάλληλες αρχιτεκτονικές. Ένα παράδειγμα μίας τέτοιας αρχιτεκτονικής θα αναλυθεί παρακάτω στην ενότητα 4.2 .

Οι οπτικές τεχνολογίες μπορούν να ανταποκριθούν αποτελεσματικά στις ανάγκες υψηλής απόδοσης. Επιπλέον, η χρήση ενός επιπέδου ελέγχου που βασίζεται στο SDN βοηθάει ακόμα περισσότερο στην αξιοποίηση των πλεονεκτημάτων των οπτικών δικτύων κέντρων δεδομένων παρέχοντας ευέλικτες, δυναμικές και ανθεκτικές υπηρεσίες δικτύου. Για την αντιμετώπιση των σύγχρονων απαιτήσεων δημιουργείται η ανάγκη χρησιμοποίησης οπτικών αρχιτεκτονικών.

Υπάρχουν υβριδικές αρχιτεκτονικές ηλεκτρο–οπτικών κέντρων δεδομένων όπως η Helios [67], η C-through [68] και άλλες. Ωστόσο αυτές οι αρχιτεκτονικές δεν μπορούν να αντιμετωπίσουν εφαρμογές διαφορετικού τύπου [69].

Το έργο EC FP7 LIGHTNESS προτείνει μια επίπεδη οπτική αρχιτεκτονική η οποία μπορεί να παρέχει προγραμματιζόμενες, δυναμικές, ευέλικτες και εξαιρετικά διαθέσιμες υπηρεσίες συνδεσιμότητας δικτύων κέντρων δεδομένων για την αντιμετώπιση των απαιτήσεων των σύγχρονων εφαρμογών κέντρων δεδομένων και νέφους [70],[71],[72],[73].

4.2 LIGHTNESS: Δίκτυο κέντρου δεδομένων το οποίο βασίζεται αποκλειστικά σε οπτική μεταγωγή κυκλώματος και πακέτου με στόχο τη βελτίωση της επεκτασιμότητας, της καθυστέρησης και της απόδοσης.

Στη συνέχεια θα αναφερθούμε στη αρχιτεκτονική η οποία προτείνεται στο έργο LIGHTNESS. Η αρχιτεκτονική αυτή βασίζεται στη χρήση υβριδικών αρχών οπτικής μεταγωγής κυκλώματος και οπτικής μεταγωγής πακέτου σε όλα τα επίπεδα του δικτύου τα οποία περιλαμβάνουν τις δικτυακές κάρτες, τους οπτικούς διακόπτες πάνω από τα ικριώματα (ToR) και τους οπτικούς

διακόπτες πάνω από τα συμπλέγματα (ToC). Ένας SDN ελεγκτής χρησιμοποιείται για τη λήψη αποφάσεων δρομολόγησης, για την προώθηση τους στις δικτυακές συσκευές καθώς και για την κατανομή των δικτυακών πόρων παρέχοντας έτσι ένα προγραμματιζόμενο επίπεδο δεδομένων [74].

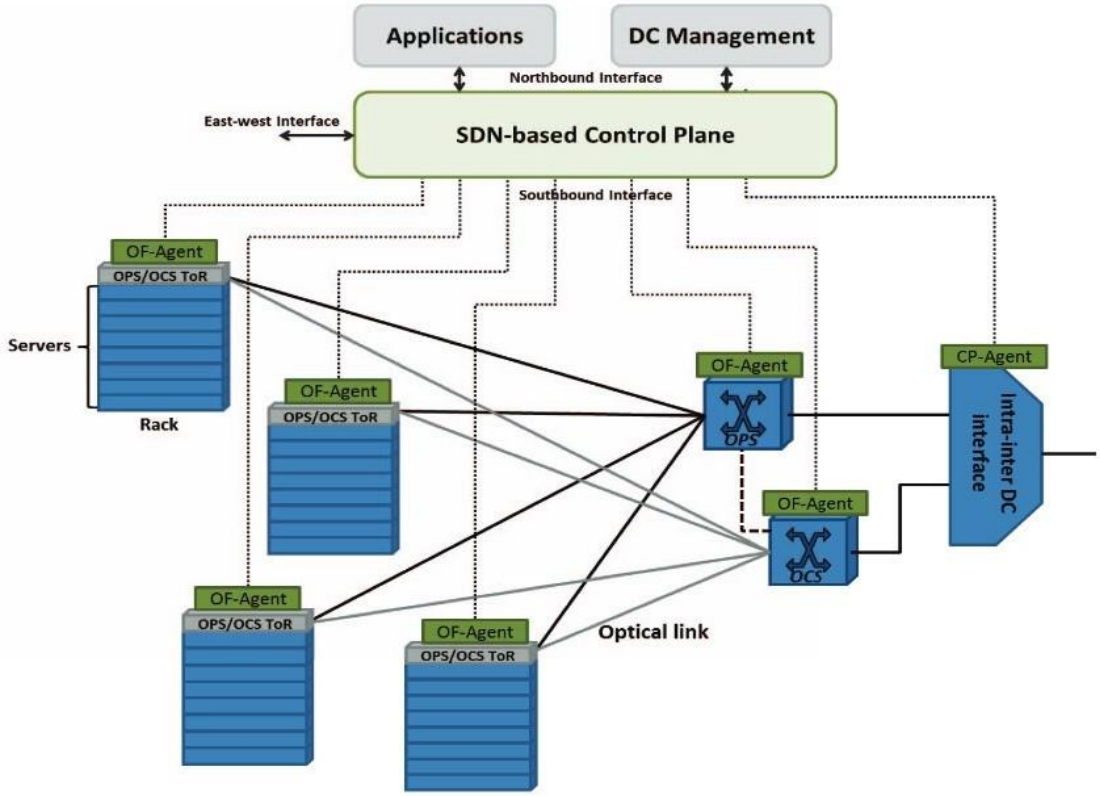
4.2.1 LIGHTNESS Αρχιτεκτονική Επιπέδου Δεδομένων

Η αρχιτεκτονική αυτή εισάγει ένα επίπεδο δεδομένων το οποίο βασίζεται στην υβριδική μεταγωγή η οποία αναφέρεται σε οπτική μεταγωγή κυκλώματος (OCS) και οπτική μεταγωγή πακέτου (OPS). Δημιουργείται μία πιο επίπεδη αρχιτεκτονική σε σχέση με την κλασσική αρχιτεκτονική πολλαπλών επιπέδων των ηλεκτρικών δικτύων κέντρων δεδομένων η οποία προσφέρει μεγαλύτερη απόδοση, μικρότερη καθυστέρηση και βελτιωμένη επεκτασιμότητα [71].

Μέσα στα κέντρα δεδομένων υπάρχουν εφαρμογές οι οποίες δημιουργούν ροές δεδομένων μεταξύ διακομιστών μικρής διάρκειας και αυστηρών απαιτήσεων όσον αφορά την καθυστέρηση η οποία συνήθως απαιτείται να είναι της τάξης μικρότερης του 1μs. Επίσης υπάρχουν εφαρμογές οι οποίες δημιουργούν σταθερές ροές δεδομένων μεγάλης διάρκειας. Η χρήση μίας ενιαίας τεχνολογίας οπτικής μεταγωγής και για τα δύο είδη εφαρμογών μπορεί να δημιουργήσει προβλήματα καθυστέρησης, απόδοσης, προβλήματα με το μέγεθος της μνήμης και απώλεια πακέτων [67]. Οι οπτικοί διακόπτες χαμηλότερης ταχύτητας μπορούν να χειριστούν ικανοποιητικά τις σταθερές ροές δεδομένων μεγάλης διάρκειας. Αντίθετα οι ροές μικρής διάρκειας με αυστηρές απαιτήσεις καθυστέρησης χρειάζονται διακόπτες γρήγορης προώθησης. Στην περίπτωση που χρησιμοποιούνται οι ίδιοι διακόπτες και για τα δύο είδη κίνησης υπάρχει το πρόβλημα της μνήμης που μπορεί να δημιουργήσει σημαντική καθυστέρηση στις ροές μικρής διάρκειας.

Οι διακομιστές που βρίσκονται στα ικριώματα συνδέονται στους οπτικούς διακόπτες ToR. Ο οπτικός διακόπτης ToR μπορεί να είναι ένα παθητικό οπτικό στοιχείο όπως ένα πλέγμα κυματοδηγών (Arrayed Waveguide Grating - AWG) ή ένα ενεργό οπτικό στοιχείο όπως ένας επιλεκτικός διακόπτης μήκους κύματος (Wavelength Selective Switch - WSS) ή ένα επιλεκτικός διακόπτης φάσματος (Spectrum Selective Switch - SSS). Με τον τρόπο αυτό

όπως βλέπουμε και στην παρακάτω εικόνα γίνεται η σύνδεση των διακομιστών στο υβριδικό OCS/OPS δίκτυο [74].



Εικόνα 4.1: Αρχιτεκτονική κέντρου δεδομένων όπως προτείνεται στο έργο LIGHTNESS. (“A Novel SDN enabled Hybrid Optical Packet/Circuit Switched Data Center Network: the LIGHTNESS approach”, Peng et.al., 2014)

Οι διακόπτες OPS έχουν ως στόχο υψηλό αριθμό θυρών οι οποίες λειτουργούν στα 40 Gb/s με μικρή καθυστέρηση από άκρο σε άκρο (μικρότερη από 1 μs) και επιλέγονται για ροές πακέτων μικρής διάρκειας. Από την άλλη οι διακόπτες OCS επιλέγονται για ροές δεδομένων μεγάλης διάρκειας και λειτουργούν στα 100 Gb/s. Το οπτικό ToR με επίγνωση εφαρμογών κάνει ταξινόμηση της κίνησης σε μικρής και μεγάλης διάρκειας και εκτελεί συνάθροιση κυκλοφορίας [71]. Αυτό δίνει τη δυνατότητα προγραμματισμού και δυναμικότητας πράγμα πολύ σημαντικό για τα σύγχρονα κέντρα

δεδομένων [74]. Ακόμα οι διακόπτες OPS και OCS συνδέονται με μία διεπαφή έτσι ώστε όταν παρουσιάζεται ανάγκη να γίνεται η σύνδεση μεταξύ κέντρων δεδομένων.

Η δυνατότητα επεκτασιμότητας παρέχεται μέσω της χρησιμοποίησης πολλαπλών διακοπτών OCS και OPS οι οποίοι διασύνδεουν τα ικριώματα επιτρέποντας με αυτό τον τρόπο να αυξηθεί ο αριθμός των μηκών κύματος εισόδου/εξόδου μεταξύ των διακοπτών ToR στην περίπτωση που υπάρχει περιορισμός λόγω του αριθμού των θυρών των διακοπτών OCS ή OPS [71]. Για παροχή ακόμα μεγαλύτερης επεκτασιμότητας υπάρχει η δυνατότητα δημιουργίας συμπλεγμάτων με σταθερό αριθμό ικριωμάτων. Ο κάθε διακόπτης συμπλέγματος μπορεί να διασυνδεθεί με τα άλλα συμπλέγματα μέσω ενός οπτικό διακόπτη [74].

Για την βελτίωση της επικοινωνίας οι διακόπτες ToR έχουν σχεδιαστεί και υλοποιηθεί χρησιμοποιώντας πλατφόρμες προγραμματιζόμενης πύλης υψηλής ταχύτητας (Field Programmable Gate Array - FPGA) [75], οπτοηλεκτρονικούς πομποδέκτες και οπτικά συστήματα. Ο προγραμματισμός του υλικού και η χρήση κατάλληλων τεχνικών επεξεργασίας και πλαισίωσης συμβάλουν στην επεξεργασία με εξαιρετικά μικρή καθυστέρηση. Επιπλέον κατάλληλες τεχνικές συγκέντρωσης κίνησης συμβάλουν στην προώθηση μέγιστης χωρητικότητας. Η κίνηση των διακομιστών αναλύεται μέσω πρωτοκόλλων και αντιστοιχίζεται σε οπτικά πακέτα προσπαθώντας να διασφαλίσουμε όσο το δυνατόν μικρότερη καθυστέρηση.

Οι οπτικοί διακόπτες έχουν διεπαφές μέσω των οποίων συνδέονται στους OPS και OCS κόμβους. Η προγραμματιζόμενη πλατφόρμα προσφέρει τη δυνατότητα δυναμικών και ευέλικτων αλλαγών στην επεξεργασία και εναλλαγή της κίνησης σύμφωνα με τις απαιτήσεις που υπάρχουν κάθε στιγμή.

Με τη χρήση των οπτικών διακοπτών αποφεύγονται οι μετατροπές σήματος από οπτικό σε ηλεκτρικό και ξανά οπτικό. Αυτό έχει ως αποτέλεσμα τη μείωση της ενέργειας που καταναλώνεται και του κόστους καθώς επίσης και τη δημιουργία σταθερών συνδέσεων χαμηλής καθυστέρησης [71].

Όπως αναφέραμε παραπάνω ο οπτικός διακόπτης μπορεί να είναι ένα παθητικό οπτικό στοιχείο ή ένα ενεργό οπτικό στοιχείο. Αυτό που προτιμάται είναι ένα ενεργό οπτικό στοιχείο γιατί είναι ευέλικτο, μπορεί να προσαρμόσει το εύρος ζώνης και να αναδιαμορφωθεί. Παρέχει έτσι τη δυνατότητα ικανοποίησης πιθανών μελλοντικών αναγκών στην κατανομή του εύρους ζώνης με μικρότερη απόσταση καναλιών και υψηλότερη φασματική απόδοση σε σχέση με τα σημερινά δεδομένα.

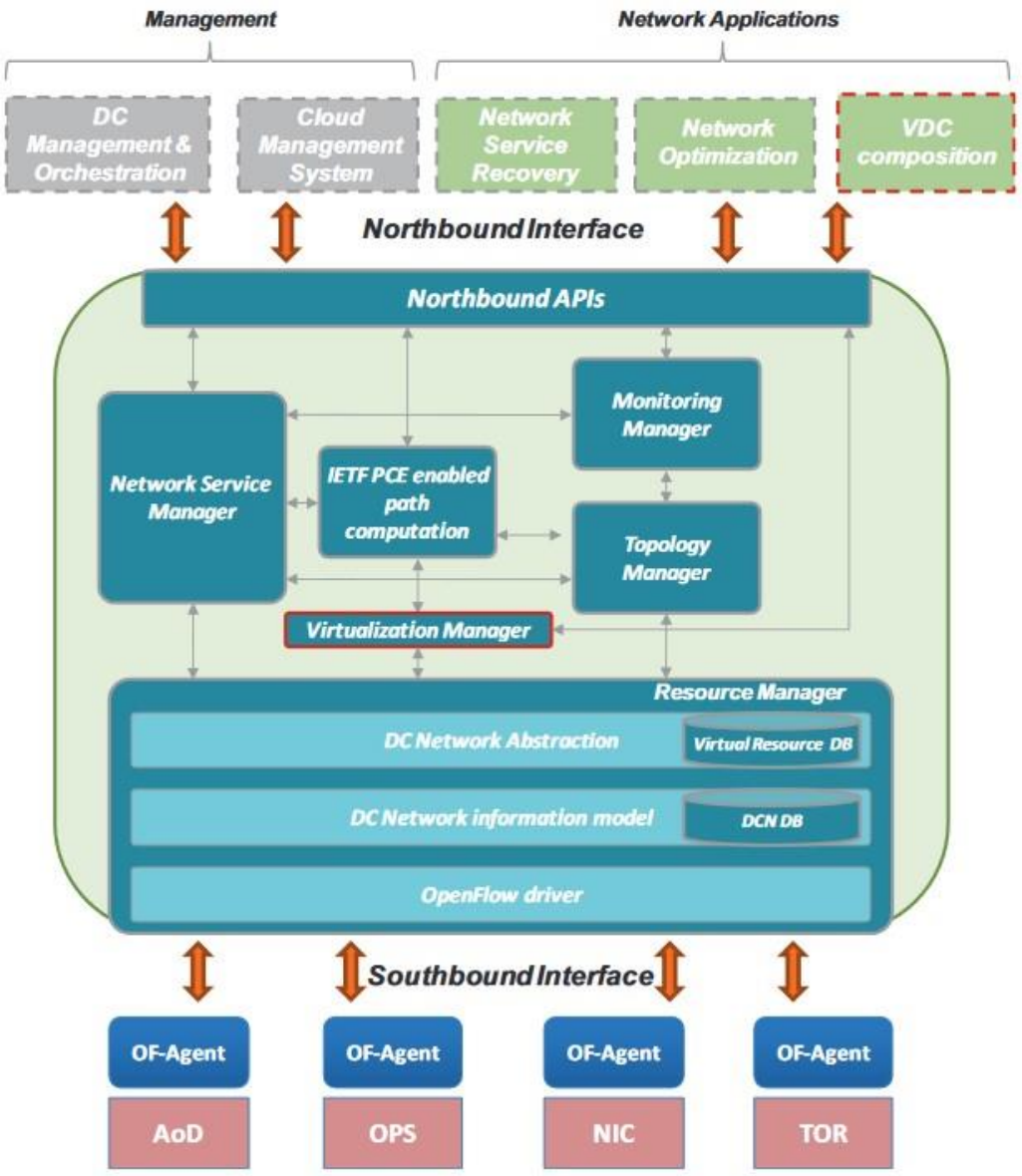
Τα πλεονεκτήματα που προσφέρονται από αυτή την υποδομή και τα οποία χρειάζονται για τα σύγχρονα κέντρα δεδομένων είναι η χαμηλή καθυστέρηση και η παροχή εύρους ζώνης με βάση τη ζήτηση που υπάρχει. Επίσης κάθε πομποδέκτης μπορεί να συνδεθεί σε OCS ή OPS κόμβο. Η επιλογή γίνεται με βάση τις απαιτήσεις κίνησης και παρέχεται η δυνατότητα προγραμματισμού και αναδιαμόρφωσης των συνδέσεων.

4.2.2 LIGHTNESS Αρχιτεκτονική Επιπέδου Ελέγχου

Στη συνέχεια θα αναφερθούμε στο επίπεδο ελέγχου το οποίο χρησιμοποιείται για την παροχή των δικτυακών πόρων χρησιμοποιώντας ένα σύνολο από κατάλληλες διαδικασίες και λειτουργίες. Φροντίζει για την εγκατάσταση των υπηρεσιών συνδεσιμότητας και για την αποσύνδεση τους, για τον υπολογισμό των κατάλληλων μονοπατιών και των ροών, για την παρακολούθηση του δικτύου, για την δυναμική τροποποίηση υπηρεσιών καθώς και για την βελτιστοποίηση των πόρων με αυτόματο και δυναμικό τρόπο. Ακόμα έχει “northbound” διεπαφές για σύνδεση στο ανώτερο επίπεδο με εφαρμογές χρηστών, με εργαλεία διαχείρισης και ενορχήστρωσης, με κατάλληλα εργαλεία για σχεδίαση εικονικών κέντρων δεδομένων και άλλα. Επίσης χρησιμοποιεί μία διεπαφή “southbound” για την επικοινωνία με τις δικτυακές συσκευές που βρίσκονται στο χαμηλότερο επίπεδο, για την διαμόρφωση τους και για την παρακολούθηση των συσκευών και της δικτυακής κίνηση. Η διεπαφή αυτή που προτείνεται στο έργο LIGHTNESS είναι το OpenFlow που όπως έχουμε αναφέρει και σε προηγούμενο κεφάλαιο είναι ανοιχτό πρωτόκολλο και ανεξάρτητο από τον προμηθευτή.

Το επίπεδο ελέγχου αποτελείται από έναν SDN ελεγκτή που στην περίπτωση αυτή είναι ο OpenDaylight. Ο ελεγκτής OpenDaylight υλοποιεί κάποιες

βασικές λειτουργίες και πρωτόκολλα για τον έλεγχο του δικτύου έτσι ώστε να μπορεί να καλύψει τις απαιτήσεις των εφαρμογών και υπηρεσιών του κέντρου δεδομένων. Στην παρακάτω εικόνα απεικονίζεται η αρχιτεκτονική του επιπέδου ελέγχου.



Εικόνα 4.2: Αρχιτεκτονική επιπέδου ελέγχου όπως προτείνεται στο έργο LIGHTNESS. (“LIGHTNESS : All-Optical SDN-enabled Intra-DSN with Optical Circuit and Packet Switching”, Saridis et.al., 2018)

Η αρχιτεκτονική αυτή διαφέρει σε σχέση με άλλες SDN αρχιτεκτονικές γιατί όπως βλέπουμε και στην παραπάνω εικόνα χρησιμοποιεί έναν διαχειριστή εικονικοποίησης (Virtualization Manager). Το στοιχείο αυτό χρησιμοποιείται για την παροχή εικονικών δικτύων κέντρων δεδομένων δίνοντας έτσι τη δυνατότητα ύπαρξης πολλαπλών πελατών μέσα στο ίδιο φυσικό κέντρο δεδομένων. Βρίσκεται πάνω από το διαχειριστή πόρων και αξιοποιεί την αφαιρετική άποψη που του παρέχει αυτός για το υβριδικό οπτικό δίκτυο. Έτσι ο κάθε πελάτης του κέντρου δεδομένων μπορεί με βάση τις απαιτήσεις ποιότητας υπηρεσιών που έχει να φτιάξει το δικό τους εικονικό δίκτυο [69],[74].

Οι λειτουργίες και τα χαρακτηριστική που παρέχονται από το επίπεδο ελέγχου μπορούν να επεκταθούν και να δημιουργηθούν εφαρμογές στο αμέσως ανώτερο επίπεδο, το επίπεδο εφαρμογών, οι οποίες υποστηρίζουν αυτές τις λειτουργίες. Ένα τέτοιο παράδειγμα είναι η εφαρμογή VDC composition που βλέπουμε στην εικόνα 4.2 . Μέσω της εφαρμογής αυτής οι πελάτες του κέντρου δεδομένων μπορούν να δημιουργούν και να τροποποιούν δυναμικά τα δίκτυα τους με βάση τις εκάστοτε ανάγκες και απαιτήσεις τους.

Η εφαρμογή VDC composition διασυνδέεται με το διαχειριστή εικονικοποίησης. Οι αλγόριθμοι και οι διαδικασίες εκτελούνται στην εφαρμογή VDC και στη συνέχεια ο διαχειριστής εικονικοποίησης παρέχει τους πόρους και φροντίζει για τη διασφάλιση της ποιότητας των υπηρεσιών. Υπάρχουν διάφοροι αλγόριθμοι όπως για παράδειγμα ένας αλγόριθμος ο οποίος κατανέμει δυναμικά εικονικά τμήματα δικτύου ή ένας αλγόριθμος ο οποίος κατανέμει στατικά εικονικά δίκτυα.

Όλες οι οπτικές συσκευές, οπτικά ToR, OPS, OCS στο επίπεδο δεδομένων ελέγχονται από τον SDN ελεγκτή με τον οποίο επικοινωνούν μέσω της διεπαφής OpenFlow. Οι οπτικοί διακόπτες OPS χρησιμοποιούνται όπως έχουμε αναφέρει για γρήγορη προώθηση της κλίμακα των nanoseconds. Παρακολουθούν τη λήψη πακέτων , τις πιθανές διενέξεις μεταξύ των πακέτων και φροντίζουν για την προώθηση σε έναν ή πολλαπλούς παραλήπτες. Ο SDN ελεγκτής είναι υπεύθυνος για τη δημιουργία των αποφάσεων προώθησης και για την παροχή των πινάκων προώθησης στις δικτυακές

συσκευές. Οι OPS διακόπτες λαμβάνουν τα οπτικά πακέτα και με βάση την οπτική ετικέτα τα προωθούν. Με τον τρόπο αυτό αποσυνδέεται η γρήγορη λειτουργία του επιπέδου δεδομένων το οποίο λειτουργεί σε nanoseconds από τις αργές λειτουργίες του επιπέδου ελέγχου το οποίο λειτουργεί σε milliseconds. Πετυχαίνουμε έτσι γρήγορη προώθηση η οποία έχει αποσυνδεθεί από το επίπεδου ελέγχου το οποίο φροντίζει για τις λειτουργίες εικονικοποίησης δικτύου και για την εφαρμογή του σχεδιασμού του εικονικού κέντρου δεδομένων [74].

4.3 Πλήρως διαχωρισμένα κέντρα δεδομένων

Η μεγάλη κλίμακα και το υψηλό φορτίο εργασίας στο οποίο καλούνται να ανταποκριθούν τα σύγχρονα κέντρα δεδομένων δημιουργούν την ανάγκη αξιοποίησης με τον καλύτερο δυνατό τρόπο των πόρων που διαθέτουν. Τα διαχωρισμένα κέντρα δεδομένων όπως έχουμε αναφέρει και στο προηγούμενο κεφάλαιο βελτιώνουν σημαντικά τη χρήση των διαθέσιμων πόρων. Αυτό είναι αποτέλεσμα της δυνατότητας που έχουν για ευέλικτη κατανομή τους. Ωστόσο η επικοινωνία μεταξύ της μνήμης και της κεντρικής μονάδας επεξεργασίας απαιτεί εξαιρετικά μικρή καθυστέρηση και πολύ υψηλό εύρος ζώνης. Για το λόγο αυτό η χρήση οπτικών διασυνδέσεων είναι απαραίτητη για την μεταξύ τους επικοινωνία.

Τα κέντρα δεδομένων μπορούν να διαχωρίσουν τους πόρους σε διάφορα επίπεδα. Μπορεί να γίνει διαχωρισμός σε επίπεδο ικριώματος, δηλαδή αντί να υπάρχουν ολοκληρωμένοι διακομιστές σε λεπίδες μέσα σε ένα ικρίωμα υπάρχουν λεπίδες όπου η κάθε μία έχει ένα τύπο πόρου. Ακόμα μπορεί να γίνει διαχωρισμός σε επίπεδο συμπλέγματος. Αυτό σημαίνει ότι διαφορετικοί τύποι πόρων μπορούν να βρίσκονται σε διαφορετικά ικριώματα μέσα σε ένα σύμπλεγμα. Και τέλος μπορεί να γίνει διαχωρισμός σε επίπεδο κέντρου δεδομένων, δηλαδή να υπάρχουν συμπλέγματα που να περιέχουν ένα τύπο πόρου και να δημιουργούνται συνδέσεις μεταξύ συμπλεγμάτων για την επικοινωνία των πόρων μεταξύ τους.

Συνήθως επιλέγεται η πρώτη περίπτωση επειδή σε αυτή την κλίμακα λόγω της μικρής απόστασης μπορούμε να αντιληφθούμε πολύ πιο εύκολα τη χαμηλή καθυστέρηση και το υψηλό εύρος ζώνης που απαιτείται για την

επικοινωνία. Στις άλλες δύο περιπτώσεις οι προκλήσεις που υπάρχουν για την επικοινωνία μεταξύ των πόρων είναι μεγαλύτερες λόγω των μεγαλύτερων αποστάσεων που δημιουργούνται [65].

Στη συνέχεια θα αναφερθούμε σε μία αρχιτεκτονική για ένα πλήρως διαχωρισμένο κέντρο δεδομένων η οποία αφορά την πρώτη περίπτωση.

4.4 Αρχιτεκτονική πλήρως διαχωρισμένου κέντρου δεδομένων με οπτικές διασυνδέσεις

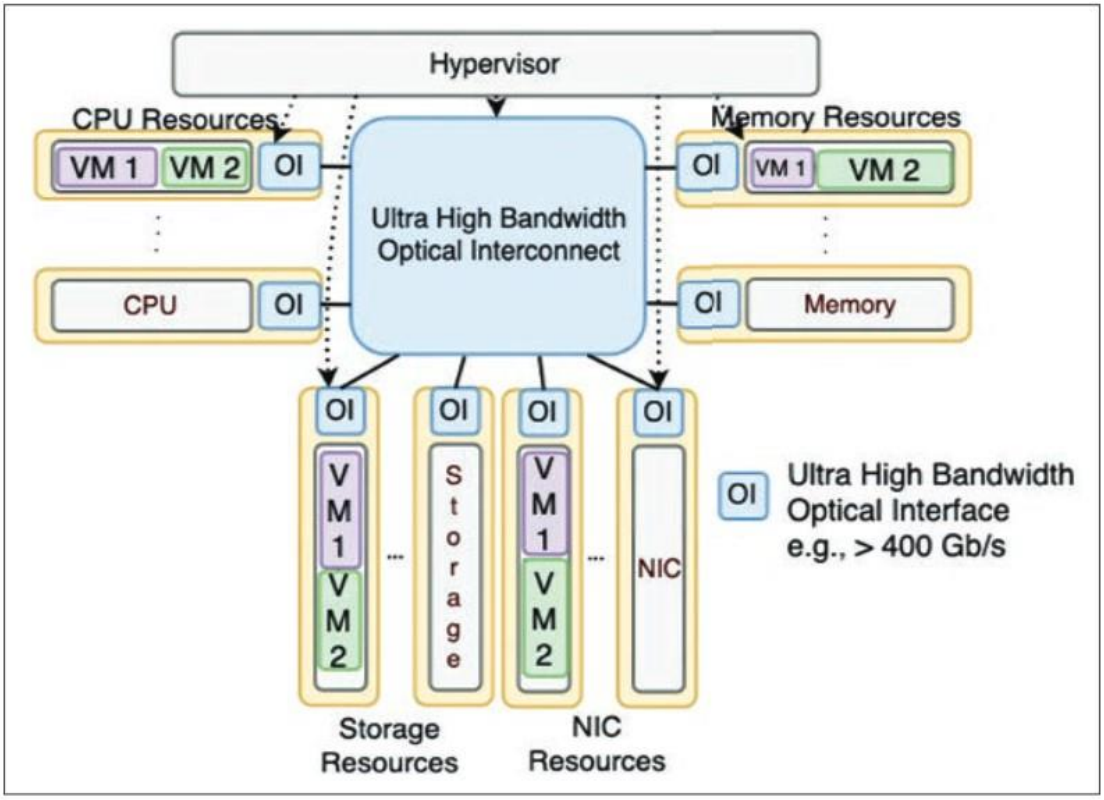
Η αρχιτεκτονική αυτή αφορά ένα πλήρως διαχωρισμένο κέντρο δεδομένων σε επίπεδο ικριώματος. Μέσα στα ικριώματα υπάρχουν λεπίδες όπου η κάθε λεπίδα περιέχει μόνο ένα τύπο πόρου σε αντίθεση με την κλασική περίπτωση όπου μία λεπίδα περιέχει έναν ολοκληρωμένο διακομιστή. Με αυτό τον τρόπο πετυχαίνουμε την πλήρη αποσύνδεση των πόρων. Οι λεπίδες διασυνδέονται στο ικρίωμα μέσω οπτικών διεπαφών.

Όλοι η επικοινωνία η οποία γινόταν στην κλασική περίπτωση των ολοκληρωμένων διακομιστών μέσα στους δίαυλους της μητρικής πλακέτας μεταφέρεται τώρα στις εξωτερικές οπτικές ζεύξεις μέσω των οποίων γίνεται η επικοινωνία των διαφόρων τύπων πόρων. Δηλαδή η επικοινωνία της μνήμης με την κεντρική μονάδα επεξεργασίας, της μνήμης με τις μονάδες αποθήκευσης, της μνήμης με την κάρτα δικτύου και άλλα μεταφέρεται στις εξωτερικές οπτικές ζεύξεις. Για το λόγο αυτό θα πρέπει οι οπτικές διεπαφές να είναι σε θέση να ανταπεξέλθουν στις αυστηρές απαιτήσεις σχετικά με το εύρος ζώνης και την καθυστέρηση στην επικοινωνία των πόρων έτσι ώστε να μην δημιουργούνται προβλήματα στην απόδοση των εφαρμογών οι οποίες τους χρησιμοποιούν.

Κάθε λεπίδα στην αρχιτεκτονική αυτή έχει μία οπτική διασύνδεση η οποία υποστηρίζει όλους τους τύπους επικοινωνίας ακόμα και την πιο απαιτητική που είναι ανάμεσα στη μνήμη νέας γενιάς και την κεντρική μονάδα επεξεργασίας. Η επικοινωνία αυτή σε περιπτώσεις αιχμής έχει συνήθως απαιτήσεις υψηλότερες από 400 Gb/s.

Η χρησιμοποίηση μίας διασύνδεσης για όλους τους τύπους επικοινωνίας μειώνει την πολυπλοκότητα όσον αφορά την καλωδίωση. Ωστόσο το γεγονός

ότι όλη η επικοινωνία διέρχεται από την ίδια διεπαφή κάνει πιο δύσκολο τον συντονισμό της επικοινωνίας. Για παράδειγμα η επικοινωνία της μνήμης με την κεντρική μονάδα επεξεργασίας απαιτεί υψηλό εύρος ζώνης. Χρειάζεται προσοχή ώστε να μην καταναλώνεται διαρκώς όλο το διαθέσιμο εύρος ζώνης για την μεταξύ τους επικοινωνία γιατί αυτό θα δημιουργούσε πρόβλημα στην επικοινωνία με τους άλλους πόρους. Θα μπορούσε να εξασθενήσει ή ακόμα και να διακόψει την επικοινωνία της μνήμης με τις μονάδες αποθήκευσης και την επικοινωνία της μνήμης με την κάρτα δικτύου [63],[64].



Εικόνα 4.3: Αρχιτεκτονική πλήρως διαχωρισμένου κέντρου δεδομένων σε επίπεδο ικριώματος με αμιγώς οπτικές διασυνδέσεις. (“Disaggregated Data Center: Challenges and Trade-offs”, Lin et.al., 2020)

4.4.1 Διαχείριση Πόρων

Για την διαχείριση των πόρων εφαρμόζεται μία τεχνική η οποία χρησιμοποιείται στην υπολογιστική νέφους. Η χρήση εικονικών μηχανών προσφέρει το πλεονέκτημα χρησιμοποίησης οποιουδήποτε λειτουργικού συστήματος απαιτείται για τη σωστή λειτουργία των εφαρμογών ανεξάρτητα από το υλικό που χρησιμοποιείται. Οι αλλαγές που μπορεί να γίνονται στο υλικό δεν πρέπει να επηρεάζουν τις εφαρμογές που τρέχουν στις εικονικές μηχανές. Στη συγκεκριμένη αρχιτεκτονική χρησιμοποιείται ένας hypervisor ο οποίος προσφέρει μία αφαιρετική άποψη του υλικού στις εικονικές μηχανές και αποκρύπτει όλες τις αλλαγές που γίνονται στο υλικό.

Ο hypervisor όπως βλέπουμε και στην εικόνα 4.3 τρέχει πάνω από το ικρίωμα. Παρακολουθεί τη χρήση των πόρων σε όλες τις λεπίδες και σε κάθε νέο αίτημα που γίνεται για μία εικονική μηχανή κατανέμει πόρους με βάση τη διαθεσιμότητα που υπάρχει. Αυτό που θα πρέπει να λαμβάνεται υπόψη με ιδιαίτερη προσοχή είναι το γεγονός ότι το εύρος ζώνης της οπτικής διεπαφής της κάθε λεπίδας που περιέχει τους διαθέσιμους πόρους είναι περιορισμένο. Σε περίπτωση μίας αποτυχίας στην οπτική διεπαφή θα επηρεαστούν όλες οι εικονικές μηχανές στο διαχωρισμένο ικρίωμα. Τέτοιες καταστάσεις μπορούν να αντιμετωπιστούν διατηρώντας αντίγραφα ασφαλείας των εικονικών μηχανών σε άλλα ικριώματα ή συμπλέγματα ή άλλα κέντρα δεδομένων [76] ελαχιστοποιώντας το επιπλέον κόστος [65].

4.4.2 Απαιτήσεις επικοινωνίας μεταξύ των πόρων

Οι απαιτήσεις για την επικοινωνία μεταξύ των πόρων στα διαχωρισμένα κέντρα δεδομένων διαφέρουν ανάλογα με τον τύπο των πόρων. Οι απαιτήσεις καθυστέρησης για την επικοινωνία μεταξύ της κεντρικής μονάδας επεξεργασίας και των μονάδων αποθήκευσης είναι της κλίμακα των microseconds και οι απαιτήσεις για εύρος ζώνης είναι της κλίμακας των λίγων Gb/s.

Αντίθετα οι απαιτήσεις για την επικοινωνία μεταξύ της μνήμης και της κεντρικής μονάδας επεξεργασίας είναι πολύ μεγαλύτερες. Η απαίτηση για την καθυστέρηση είναι της κλίμακας <100 ns. Η απόδοση της μνήμης και της

κεντρικής μονάδας επεξεργασίας επηρεάζουν σε μεγάλο βαθμό το ζητούμενο εύρος ζώνης. Ο υπολογισμός του γίνεται πολλαπλασιάζοντας τον αριθμό των ελεγκτών μνήμης που υπάρχουν στην κεντρική μονάδα επεξεργασίας, την ταχύτητα του ρολογιού μνήμης και το μέγεθος της λέξης του επεξεργαστή. Για παράδειγμα για μνήμη 4ης γενιάς με διπλό ρυθμό δεδομένων με τρείς ελεγκτές μνήμης σε έναν επεξεργαστή 64 bit με ταχύτητα ρολογιού 2133 Mhz απαιτείται ένα εύρος ζώνης περίπου 400 Gb/s. Το συνολικό εύρος ζώνης που απαιτείται σε μια οπτική διεπαφή μίας λεπίδας η οποία περιέχει πόρους επεξεργαστών αυξάνεται όσο αυξάνεται ο αριθμός των επεξεργαστών και ο αριθμός των πυρήνων που υπάρχουν στην λεπίδα. Τα προϊόντα τα οποία είναι διαθέσιμα στην αγορά αυτή τη στιγμή είναι πάρα πολύ δύσκολο να καλύψουν τέτοιες ανάγκες υψηλού εύρους ζώνης [66].

4.4.3 Χρήση οπτικών διασυνδέσεων για επικοινωνία μεταξύ των πόρων

Οι οπτικές διασυνδέσεις λόγω της χαμηλής καθυστέρησης και του μεγάλου εύρους ζώνης που μπορούν να προσφέρουν είναι η μόνη τεχνολογία η οποία μπορεί να χρησιμοποιηθεί για την αντιμετώπιση των αναγκών επικοινωνίας μεταξύ των πόρων.

Με βάση τις τεχνικές που χρησιμοποιούνται για την ανίχνευση κατηγοριοποιούνται σε δύο βασικούς τύπους οι οποίοι είναι οι εξής :

- Το συνεκτικό σύστημα
- Διαμόρφωση έντασης και σύστημα άμεσης ανίχνευσης (IM/DD)

Το συνεκτικό σύστημα χρησιμοποιείται σε μεγάλο βαθμό για μετάδοση σε μεγάλες αποστάσεις. Έχει υψηλό κόστος και μεγάλη πολυπλοκότητα τα οποία συμβάλουν στην αποφυγή χρησιμοποίησης του σε μικρές αποστάσεις. Επιπλέον η επεξεργασία σήματος που γίνεται στους αναμεταδότες λόγω πολυπλοκότητας εισάγει μεγάλη καθυστέρηση. Αυτό μπορεί να έχει ως αποτέλεσμα να μην καλύπτονται οι ανάγκες για εξαιρετικά χαμηλή καθυστέρηση που υπάρχουν στα πλήρως διαχωρισμένα περιβάλλοντα και είναι ένας λόγος ο οποίος αποτρέπει τη χρήση του σε αυτά.

Αντίθετα η οπτική μετάδοση η οποία χρησιμοποιεί διαμόρφωση έντασης και συστήματα άμεσης ανίχνευσης προσφέρει υψηλό εύρος ζώνης και έχει απλή

εγκατάσταση. Είναι η τεχνική η οποία προτιμάται στα πλήρως διαχωρισμένα κέντρα δεδομένων για τα οποία μιλάμε στην συγκεκριμένη περίπτωση.

Στον παρακάτω πίνακα βλέπουμε στοιχεία για επικοινωνία με χρήση τελευταίας τεχνολογίας οπτικής μετάδοσης για ταχύτητες 400 Gb/s και μεγαλύτερες σε μικρές αποστάσεις.

Modulation	Wavelength band (nm)	Data rate per fiber	Multiplexing	Reach	Optical link	Transceiver	Pre-FEC BER
PAM4 [7]	850	4 × 100 Gb/s	WDM	105 m	MMF	Silicon photonics	1e-6
NRZ/EDB [8]	1550	7 × 100 Gb/s	SDM	10 km	MCF	EAM	5e-5
NRZ [9]	1310	8 × 4 × 25 Gb/s	SDM/WDM	1.1 km	MCF	VCSEL	1e-12
PAM4 [10]	1550	7 × 149 Gb/s	SDM	1 km	MCF	VCSEL	3.8e-3

WDM: wavelength-division multiplexing; NRZ: non-return-to-zero; SDM: spatial-division multiplexing;
MMF: multi-mode fiber; VCSEL: vertical-cavity surface emitting laser; EDB: electrical duo-binary; MCF: multi-core fiber;
EAM: electro-absorption modulator; PAM4: 4-level pulse amplitude modulation.

Πίνακας 4.1: Οπτική μετάδοση μικρής απόστασης (“Disaggregated Data Centers: Challenges and Trade-offs”, Lin et.al., 2020)

Υπάρχει όπως βλέπουμε η δυνατότητα χρήσης διαφορετικών μορφών διαμόρφωσης, διαφορετικών τύπων πομποδεκτών, διαφορετικών τύπων πολυπλεξίας, διαφορετικών τεχνικών για την διόρθωση σφαλμάτων και την επεξεργασία των σημάτων. Όλα τα παραπάνω αποτελούν τεχνικές που μπορούν πιθανόν να χρησιμοποιηθούν σε διαχωρισμένα κέντρα δεδομένων. Για να μπορέσουμε να πετύχουμε χαμηλή κατανάλωση ενέργειας και να μειώσουμε το κόστος γίνεται χρήση υψηλού ρυθμού μετάδοσης των δεδομένων. Η χρήση απλών μορφών διαμόρφωσης όπως η non-return-to zero on-off-keying NRZ-OOK και η EDB [77] πετυχαίνουν μετάδοση 100 Gb/s και μεγαλύτερη σε πραγματικό χρόνο. Ωστόσο η βασική επιλογή είναι η διαμόρφωση πλάτους με τέσσερα επίπεδα (PAM4) [78],[79] η οποία πετυχαίνει μεγαλύτερη απόδοση του εύρους ζώνης.

Οι διάφορες τεχνικές πολυπλεξίας όπως η πολυπλεξία χωρικής διαίρεσης (SDM) [79] , η πολυπλεξία μήκους κύματος (WDM) [78] καθώς επίσης και ο συνδυασμός διαφόρων τεχνικών πολυπλεξίας όπως αυτές οι δύο για παράδειγμα [80] έχουν ως αποτέλεσμα την αύξηση της χωρητικότητας των οπτικών ινών. Με αυτούς τους τρόπους βελτιώνεται η δυνατότητα κάλυψης του ιδιαίτερα μεγάλου εύρους ζώνης που είναι αναγκαίο για την επικοινωνία στα διαχωρισμένα κέντρα δεδομένων. Η χρήση της πολυπλεξίας χωρικής διαίρεσης απαιτεί την χρησιμοποίηση προηγμένης τεχνολογίας οπτικών ινών αλλά έχει μικρότερο κόστος σε αντίθεση με την πολυπλεξία μήκους κύματος για την οποία χρησιμοποιούνται πομποδέκτες υψηλού κόστους.

Ακόμα να αναφέρουμε ότι μία μονότροπη οπτική ίνα (SMF) θα πρέπει να επιλέγεται σε περιπτώσεις μεγάλων αποστάσεων. Αυτό σημαίνει ότι είναι κατάλληλη για διαχωρισμένα κέντρα μεγαλύτερης κλίμακας και όχι για διαχωρισμένα κέντρα σε επίπεδο ικριώματος. Από την άλλη οι πολύτροπες οπτικές ίνες (MMF) μπορούν να χρησιμοποιηθούν μαζί με πομποδέκτες χαμηλού κόστους. Μπορούν να καλύψουν μικρές αποστάσεις και προσφέρουν περιορισμένο εύρος ζώνης. Χρειάζεται και σε αυτή την περίπτωση να γίνει απλή επεξεργασία σήματος ωστόσο ο χρόνος που απαιτείται για την επεξεργασία είναι πολύ λιγότερος [78].

Οι πομποδέκτες στα διαχωρισμένα κέντρα δεδομένων πρέπει να είναι οικονομικά αποδοτικοί, μικροί και η ενσωμάτωση και εφαρμογή τους στη μονάδα των πόρων να είναι εύκολη. Υπάρχουν δύο τύποι πομποδεκτών που χρησιμοποιούνται κυρίως και αυτοί είναι τα laser VCSEL [80],[81] και τα ολοκληρωμένα κυκλώματα φωτονικού πυριτίου (SiP) [78]. Και οι δυο τύποι είναι οικονομικά αποδοτικοί και έχουν μικρό ενεργειακό αποτύπωμα.

Τα VCSEL σε συνδυασμό με πολυπλεξία χωρικής διαίρεσης χρησιμοποιούνται σε μεγάλο βαθμό για οπτική επικοινωνία σε μικρές αποστάσεις. Με σωστή σχεδίαση μπορούν να λειτουργούν σε μεγάλο εύρος θερμοκρασιών χωρίς να επηρεάζεται ιδιαίτερα η απόδοση τους και χωρίς να χρειάζονται επιπλέον παρακολούθηση. Αυτό τα καθιστά κατάλληλα για λειτουργία στα κέντρα δεδομένων όπου το φορτίο εργασίας που υπάρχει την κάθε στιγμή επηρεάζει σημαντικά την αλλαγή στην θερμοκρασία.

Τα SiP σε συνδυασμό με πολυπλεξία μήκους κύματος χρησιμοποιούνται για μεγάλο ρυθμό μετάδοσης δεδομένων. Εταιρίες όπως η Intel, η Acacia και η Luxtera υποστηρίζουν τα διαχωρισμένα κέντρα δεδομένων προσφέροντας λύσεις 400G με χρήση SiP [65].

4.4.4 Τεχνολογίες οπτικής μεταγωγής οι οποίες χρησιμοποιούνται για επικοινωνία μεταξύ των πόρων

Όλη η επικοινωνία μεταξύ των πόρων στα πλήρως διαχωρισμένα κέντρα δεδομένων γίνεται μέσω των διασυνδέσεων τους. Για το λόγο αυτό πρέπει οι κόμβοι διασύνδεσης να μπορούν να ανταποκριθούν στις ιδιαίτερα αυστηρές απαιτήσεις για ελάχιστη καθυστέρηση και υψηλό εύρος ζώνης.

Υπάρχουν τεχνολογίες όπως για παράδειγμα η Infiniband η οποία θα μπορούσε να χρησιμοποιηθεί για την επικοινωνία μεταξύ πόρων των οποίων οι απαιτήσεις εύρους ζώνης κυμαίνονται σε ένα μέτριο επίπεδο. Η χρήση ενός ηλεκτρικού διακόπτης όπως για παράδειγμα ενός Cisco Nexus 9316D ή ενός Exablaze FastMux παρέχει μία καθυστέρηση 50+ ns και ένα εύρος ζώνης 400 Gb/s σε κάθε θύρα. Παρατηρούμε λοιπόν ότι μπορεί να καλύψει τις ιδιαίτερα αυξημένες απαιτήσεις σε περιπτώσεις αιχμής για την επικοινωνία μεταξύ της κεντρικής μονάδας επεξεργασίας και της μνήμης. Ωστόσο την ίδια στιγμή θα πρέπει να είναι σε θέση να καλύψει τις απαιτήσεις για επικοινωνία ανάμεσα σε πολλές λεπίδες πόρων και αυτό αυξάνει ακόμη περισσότερο τη ζήτηση για χωρητικότητα. Ακόμα θα πρέπει να λάβουμε υπόψη μας ότι οι τεχνολογίες της μνήμης και της κεντρικής μονάδας επεξεργασίας βελτιώνονται και αυτό οδηγεί σε αύξηση των απαιτήσεων για μικρότερη καθυστέρηση και μεγαλύτερο εύρος ζώνης πράγμα το οποίο δημιουργεί περιορισμούς με τη χρήση των ηλεκτρικών διακοπτών. Επιπλέον η χρήση ενός ηλεκτρονικού διακόπτη εισάγει πρόσθετη καθυστέρηση λόγω των μετατροπών του κάθε σήματος από οπτικό σε ηλεκτρικό και ξανά οπτικό, αυξάνει την κατανάλωση ενέργειας και το λειτουργικό κόστος. Με βάση αυτά θα μπορούσαμε να πούμε ότι οι οπτικοί διακόπτες είναι πιο κατάλληλοι για χρήση σε διαχωρισμένα κέντρα δεδομένων.

Μπορούμε να χωρίσουμε τους οπτικούς διακόπτες σε δύο βασικές κατηγορίες. Η πρώτη κατηγορία αφορά διακόπτες οι οποίοι για την αναδιαμόρφωση τους χρειάζονται περισσότερο χρόνο ο οποίος είναι της τάξης των microsecond και θεωρούνται αργοί διακόπτες. Ένα κλασικό παράδειγμα αυτής της κατηγοράς είναι οι οπτικοί διακόπτες οι οποίοι βασίζονται σε μικρο-ηλεκτρομηχανικά συστήματα (MEMS). Χρησιμοποιούνται συνήθως για οπτική μεταγωγή κυκλώματος και ο χρόνος αναδιαμόρφωσης τους είναι μεγαλύτερος από δεκάδες microseconds [81]. Για τη χρήση τους στα διαχωρισμένα περιβάλλοντα θα πρέπει η διαμόρφωση τους να γίνεται στην αρχή όταν γίνεται η κατανομή των πόρων στα εικονικά μηχανήματα. Δημιουργούνται έτσι τα κανάλια για την επικοινωνία μεταξύ των πόρων και κατά την λειτουργία των εικονικών μηχανών δεν χρειάζεται πρόσθετος χρόνος για την αναδιαμόρφωση τους.

Αντίθετα η δεύτερη κατηγορία αφορά διακόπτες των οποίων ο χρόνος αναδιαμόρφωσης είναι της τάξης των nanosecond και θεωρούνται γρήγοροι διακόπτες. Σε αυτή την κατηγορία ανήκουν τα laser γρήγορου συντονισμού με ένα πλέγμα κυματοδηγών [82] και οι διακόπτες φωτονικού πυριτίου υψηλής ακτίνας βασισμένους σε μικροδακτύλιο [83]. Οι διακόπτες αυτοί χρησιμοποιούνται για οπτική μεταγωγή πακέτων. Αν χρησιμοποιηθεί αυτός ο τύπος διακοπτών στα διαχωρισμένα κέντρα δεδομένων τότε η καθυστέρηση στην επικοινωνία των πόρων εξαρτάται από το χρόνο που χρειάζεται για να γίνει η μεταγωγή. Όσο μικρότερος θα είναι ο χρόνος μεταγωγής τόσο μικρότερη θα είναι και η καθυστέρηση. Επειδή όμως δεν υπάρχει οπτική μνήμη για την αποθήκευση των πακέτων σε περιπτώσεις διένεξης πακέτων θα υπάρχει μεγαλύτερη καθυστέρηση. Για το λόγο αυτό η τεχνική η οποία προτιμάτε στα διαχωρισμένα κέντρα δεδομένων είναι η οπτική μεταγωγή κυκλώματος με χρήση διακοπτών αργής μεταγωγής [65].

4.4.5 Αξιολόγηση απόδοσης στα πλήρως διαχωρισμένα κέντρα δεδομένων

Η χρήση των πλήρως διαχωρισμένων κέντρων δεδομένων έχει ως στόχο την πιο αποδοτική αξιοποίηση των διαθέσιμων πόρων σε σύγκριση με τα κέντρα δεδομένων όπου γίνεται χρήση ολοκληρωμένων διακομιστών. Με βάση

στοιχεία τα οποία βλέπουμε στο άρθρο [65] τα οποία αποκτήθηκαν με μετρήσεις που έγιναν χρησιμοποιώντας έναν προσομοιωτή βασισμένο σε Python προκύπτει το συμπέρασμα ότι το διαθέσιμο εύρος ζώνης για την επικοινωνία μεταξύ των πόρων στα πλήρως διαχωρισμένα κέντρα δεδομένων δεν μπορεί να θεωρηθεί απεριόριστο ακόμα και στην περίπτωση χρήσης οπτικής μετάδοσης με πολύ υψηλές ταχύτητες.

Έγιναν μετρήσεις χρησιμοποιώντας δύο τύπους μνήμης στα εικονικά μηχανήματα. Μνήμη τύπου DDR3 – 1600 MHz η οποία για την επικοινωνία με την κεντρική μονάδα επεξεργασίας έχει μέγιστη απαίτηση εύρους ζώνης 200 Gb/s και μνήμη τύπου DDR4 – 3200 MHz της οποίας οι απαιτήσεις για επικοινωνία με την κεντρική μονάδα επεξεργασίας σε στιγμές αιχμής φτάνουν τα 400 Gb/s. Επίσης για τα πειράματα χρησιμοποιήθηκαν οπτικές διεπαφές με ρυθμούς δεδομένων 400 Gb/s και 800 Gb/s. Όπως έχουμε αναφέρει και νωρίτερα πρόσφατα έχει γίνει η τυποποίηση του 400G από τον IEEE και έχουν ξεκινήσει από το 2022 ακόμα μεγαλύτερες ταχύτητες 800G. Μέσα στα επόμενα χρόνια αναμένεται και η τυποποίηση του 800G από τον IEEE.

Το συμπέρασμα που βγήκε από τις πειραματικές μετρήσεις είναι ότι για να μπορέσουμε να αξιοποιήσουμε τα οφέλη που προσφέρουν τα διαχωρισμένα κέντρα δεδομένων θα πρέπει το διαθέσιμο εύρος ζώνης στις οπτικές διασυνδέσεις να είναι αποδοτικό. Χρησιμοποιώντας οπτικές διεπαφές 400 Gb/s με τις πιο εξελιγμένες μνήμες τύπου DDR4 έχει ως αποτέλεσμα πολλά αιτήματα για εικονικές μηχανές να μην μπορούν να εξυπηρετηθούν γιατί το εύρος ζώνης δεν επαρκεί. Έτσι οδηγούμαστε σε μία κατάσταση στην οποία γίνεται μικρή χρήση των πόρων μνήμης και κεντρικής μονάδας επεξεργασίας λόγω της έλλειψης εύρους ζώνης. Με την αύξηση του εύρους ζώνης στα 800 Gb/s πετυχαίνουμε σημαντική βελτίωση στη χρήση των διαθέσιμων πόρων και γενικότερα βελτίωση της συνολικής εικόνας. Ωστόσο για να μπορέσουμε να πετύχουμε μεγαλύτερη απόδοση των πλήρως διαχωρισμένων κέντρων δεδομένων σε σχέση με τα μερικώς διαχωρισμένα κέντρα και τα κέντρα ολοκληρωμένων διακομιστών χρειάζεται το εύρος ζώνης να είναι μεγαλύτερο από 800 Gb/s. Αυτό φανερώνει την ανάγκη για περαιτέρω πρόοδο στις οπτικές επικοινωνίες ώστε να πετύχουμε μεγαλύτερο εύρος ζώνης σε μικρές αποστάσεις για παράδειγμα πάνω από 1 Tb/s [65].

ΚΕΦΑΛΑΙΟ 5 : Συμπεράσματα

Η συνεχής κλιμάκωση των κέντρων δεδομένων αυξάνει ολοένα και περισσότερο το φορτίο εργασίας δημιουργώντας την ανάγκη για χρησιμοποίηση κατάλληλων αρχιτεκτονικών έτσι ώστε να έχουμε την καλύτερη δυνατή αξιοποίηση των διαθέσιμων υπολογιστικών και δικτυακών πόρων καθώς και των πόρων αποθήκευσης. Η μεγάλη ζήτηση για υψηλή απόδοση στα σύγχρονα κέντρα δεδομένων αυξάνει τις απαιτήσεις για υψηλότερο εύρος ζώνης και χαμηλότερη καθυστέρηση. Οι αρχιτεκτονικές δικτύων με ηλεκτρικούς διακόπτες οι οποίες χρησιμοποιούν πολλαπλά επίπεδα μεταγωγής δεν μπορούν να προσφέρουν την απαιτούμενη ευελιξία και απόδοση ώστε να ικανοποιήσουν τις σύγχρονες απαιτήσεις. Τα κλασσικά ηλεκτρικά δίκτυα που βασίζονται σε μία ιεραρχική δομή δεν μπορούν να προσαρμόσουν το διαθέσιμο εύρος ζώνης και την τοπολογία του δικτύου με αποτέλεσμα να γίνεται σπατάλη των διαθέσιμων πόρων. Η επέκταση τους απαιτεί υψηλό κόστος τόσο για την αγορά υλικού όσο και για την κατανάλωση ενέργειας και αυξάνει σημαντικά την πολυπλοκότητα.

Τα οπτικά δίκτυα μπορούν να ανταποκριθούν στις σύγχρονες απαιτήσεις. Προσφέρουν αρκετά υψηλότερο εύρος ζώνης και χαμηλή καθυστέρηση. Λόγω του υψηλού εύρους ζώνης μπορεί να χρησιμοποιηθεί μία πιο επίπεδη αρχιτεκτονική η οποία συμβάλει στη μείωση της καθυστέρησης. Επίσης οι οπτικές διασυνδέσεις καταναλώνουν πολύ μικρότερα ποσά ενέργειας και η χρήση αποκλειστικά οπτικών διακοπτών μειώνει ακόμα περισσότερο την απαιτούμενη ενέργεια μέχρι και 75% καθώς και το χρόνο μεταγωγής εφόσον δεν χρειάζεται να γίνονται μετατροπές του σήματος από οπτικό σε ηλεκτρικό και ξανά οπτικό. Από τα παραπάνω συμπεραίνουμε ότι η χρήση αμιγώς οπτικών αρχιτεκτονικών προσφέρει τεράστια οφέλη και από άποψη κόστους.

Τα πλεονεκτήματα των οπτικών δικτύων κέντρων δεδομένων μπορούν να αξιοποιηθούν ακόμα καλύτερα χρησιμοποιώντας μία SDN αρχιτεκτονική η οποία παρέχει ευέλικτες, προγραμματιζόμενες, δυναμικές και εξαιρετικά ανθεκτικές υπηρεσίες δικτύου οι οποίες μπορούν να ικανοποιήσουν τις σύγχρονες απαιτήσεις των κέντρων δεδομένων. Η SDN αρχιτεκτονική προσφέρει βελτιωμένη επεκτασιμότητα, μικρότερη καθυστέρηση και

μεγαλύτερη απόδοση. Έτσι ο συνδυασμός της με χρήση οπτικών δικτύων βοηθάει στη βελτίωση της αποτελεσματικότητας τους.

Όπως αναφέραμε και παραπάνω το υψηλό φορτίο εργασίας στα σύγχρονα κέντρα δεδομένων έχει δημιουργήσει την ανάγκη για αξιοποίηση των διαθέσιμων πόρων με τον καλύτερο δυνατό τρόπο. Αυτό οδηγεί στη χρήση αρχιτεκτονικών διαχωρισμένων κέντρων δεδομένων. Υπάρχουν όπως είδαμε δύο τύποι διαχωρισμένων κέντρων δεδομένων. Τα μερικώς διαχωρισμένα κέντρα και τα πλήρως διαχωρισμένα κέντρα δεδομένων. Στα πλήρως διαχωρισμένα κέντρα δεδομένων όλοι οι πόροι είναι διαχωρισμένοι μεταξύ τους. Αυτό βελτιώνει σημαντικά τη διαθεσιμότητα και τη χρησιμοποίηση τους και επιτρέπει την αντικατάσταση ή την αναβάθμιση όποτε είναι αναγκαίο μεμονωμένων πόρων.

Ωστόσο ο διαχωρισμός των πόρων δημιουργεί αυστηρές απαιτήσεις για την επικοινωνία μεταξύ τους. Οι απαιτήσεις αυτές διαφέρουν ανάλογα με τον τύπο των πόρων που θέλουν να επικοινωνήσουν. Οι απαιτήσεις εύρους ζώνης για την επικοινωνία μεταξύ της κεντρικής μονάδας επεξεργασίας και των μονάδων αποθήκευσης είναι της τάξης των λίγων Gb/s και οι απαιτήσεις καθυστέρησης είναι της τάξης των microseconds. Αντίθετα για την επικοινωνία μεταξύ της κεντρικής μονάδας επεξεργασίας και της μνήμης οι απαιτήσεις είναι πολύ μεγαλύτερες και το επιθυμητό εύρος ζώνης εξαρτάται από την απόδοση τους. Νεότερες τεχνολογίες έχουν υψηλότερες απαιτήσεις. Παράδειγμα μια μνήμη τύπου DDR4 – 3200 MHz έχει μέγιστη απαίτηση εύρους ζώνης 400 Gb/s. Ακόμα το ζητούμενο εύρος ζώνης αυξάνεται όσο μεγαλώνει ο αριθμός των επεξεργαστών και ο αριθμός των πυρήνων. Η ζητούμενη καθυστέρηση είναι της κλίμακας <100 ns.

Προκύπτει λοιπόν το συμπέρασμα ότι το δίκτυο θα πρέπει να είναι σε θέση να υποστηρίξει το απαιτούμενο υψηλό εύρος ζώνης και την εξαιρετικά χαμηλή καθυστέρηση. Διαφορετικά δεν μπορούμε να αξιοποιήσουμε τα οφέλη που μπορούν να προσφέρουν τα διαχωρισμένα κέντρα δεδομένων. Οι οπτικές τεχνολογίες δεν μπορούν πάντα να ικανοποιήσουν τις απαιτήσεις εύρους ζώνης και η περαιτέρω πρόοδος τους αποτελεί πρόκληση για να είναι δυνατή

η καλύτερη αξιοποίηση των πλεονεκτημάτων των διαχωρισμένων κέντρων δεδομένων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] *What is a data center?* (no date a) *IBM*. Available at: https://www.ibm.com/topics/data-centers (Accessed: 07 June 2023).

[2] *What is a data center?* (2023) *Cisco*. Available at: https://www.cisco.com/c/en/us/solutions/data-center-virtualization/what-is-a-data-center.html (Accessed: 07 June 2023).

[3] Singh, A. *et al.* (2015) 'Jupiter rising', *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication* [Preprint]. doi:10.1145/2785956.2787508.

[4] *Scaling-out ethernet for the data center* Available at: https://network.nvidia.com/pdf/whitepapers/WP-ethernet%20scaleout-WEB.pdf

[5] *Data Center Architecture Overview* (2015) *Cisco*. Cisco. Available at: https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_1.html (Accessed: May 6, 2023).

[6] Hammadi, A. and Mhamdi, L. (2014) "A survey on architectures and energy efficiency in Data Center Networks," *Computer Communications*, 40, pp. 1–21. Available at: https://doi.org/10.1016/j.comcom.2013.11.005.

[7] Zhang, Y. and Ansari, N. (2013) "On architecture design, congestion notification, TCP INCAST and power consumption in data centers," *IEEE Communications Surveys & Tutorials*, 15(1), pp. 39–64. Available at: https://doi.org/10.1109/surv.2011.122211.00017.

[8] Chen, T., Gao, X. and Chen, G. (2016) "The features, hardware, and architectures of Data Center Networks: A survey," *Journal of Parallel and Distributed Computing*, 96, pp. 45–74. Available at: https://doi.org/10.1016/j.jpdc.2016.05.009.

[9] Barroso, L.A. and Hölzle, U. (2007) "The case for energy-proportional computing," *Computer*, 40(12), pp. 33–37. Available at: https://doi.org/10.1109/mc.2007.443.

[10] Al-Fares, M., Loukissas, A. and Vahdat, A. (2008) "A scalable, Commodity Data Center Network Architecture," *Proceedings of the ACM SIGCOMM 2008 conference on Data communication* [Preprint]. Available at: https://doi.org/10.1145/1402958.1402967.

[11] Cisco Global Cloud Index: Forecast and Methodology, 2011–2016," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns1175/Cloud Index White Paper.pdf

[12] V. Vasudevan, A. Phanishayee, H. Shah, E. Krevat, D. G. Andersen, G. R. Ganger, G. A. Gibson, and B. Mueller, "Safe and effective fine-grained TCP re-

transmissions for datacenter communication," in Proc. of ACM SIGCOMM'09, 2009

[13] M. Alizadeh, A. Greenberg, D. A. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, and M. Sridharan, "Data center TCP (DCTCP)," in Proc. of ACM SIGCOMM'10, 2010

[14] Alizadeh, M. and Edsall, T. (2013a) 'On the data path performance of leaf-spine datacenter fabrics', *2013 IEEE 21st Annual Symposium on High-Performance Interconnects* [Preprint]. doi:10.1109/hoti.2013.23.

[15] *How to choose data center spine and leaf switches?: FS Community* (no date) *Knowledge*. Available at: https://community.fs.com/blog/how-to-choose-data-center-spine-and-leaf-switches.html (Accessed: 17 May 2023).

[16] *What is spine-leaf architecture and how to design it: FS Community* (no date) *Knowledge*. Available at: https://community.fs.com/blog/leaf-spine-with-fs-com-switches.html (Accessed: 17 May 2023).

[17] Rajkumar Buyya; Toni Cortes; Hai Jin, "An Introduction to the InfiniBand Architecture," in *High Performance Mass Storage and Parallel I/O: Technologies and Applications* , IEEE, 2002, pp.616-632, doi: 10.1109/9780470544839.ch42

[18] Sadmin (2017) *Infiniband accelerates the world's fastest supercomputer, two of the top five supercomputers, and 77 percent of new HPC systems on the TOP500 list, InfiniBand Trade Association*. Available at: https://www.infinibandta.org/infiniband-accelerates-the-worlds-fastest-supercomputer-two-of-the-top-five-supercomputers-and-77-percent-of-new-hpc-systems-on-the-top500-list/ (Accessed: 07 June 2023).

[19] *Infiniband - a low-latency, high-bandwidth interconnect* (2019) *InfiniBand Trade Association*. Available at: https://www.infinibandta.org/about-infiniband/ (Accessed: 07 June 2023).

[20] Pincemin, E. *et al.* (2023) 'End-to-end interoperable 400-GBE optical communications through 2-km 400GBASE-FR4, 8 × 100-km 400G-openroadm and 125-km 400-zr fiber lines', *Journal of Lightwave Technology*, 41(4), pp. 1250–1257. doi:10.1109/jlt.2022.3204731.

[21] *HOW 400G has transformed Data Centers: FS Community* (no date) *Knowledge*. Available at: https://community.fs.com/blog/how-400g-has-transformed-data-centers.html (Accessed: 07 June 2023).

[22] *400g optics in Hyperscale Data Centers: Coping with the ever-increasing bandwidth requirements: FS Community* (no date) *Knowledge*. Available at: https://community.fs.com/blog/400g-optics-in-hyperscale-data-centers-coping-with-the-ever-increasing-bandwidth-requirements.html (Accessed: 07 June 2023).

[23] Open Networking Fundation , "Software-defined Networking: The New norm for Networks.", ONF White Paper, (2012) , available online:https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf

[24] Xia, W. *et al.* (2015) 'A survey on software-defined networking', *IEEE Communications Surveys & Tutorials*, 17(1), pp. 27–51. doi:10.1109/comst.2014.2330903.

[25] T. D. Nadeau and K. Gray, Software Defined Networks, CA: O' Reilly, (2013)

[26] Göransson, P. and Black, C. (2014) *Software defined networks: A comprehensive approach*. Amsterdam: Elsevier, Morgan Kaufmann.

[27] *Open Networking Foundation*. Available at: https://opennetworking.org/ (Accessed: 11 August 2023)

[28] Kurose, J.F. and Ross, K.W. (2022) *Computer networking: A top-down approach*. Harlow, United Kingdom: Pearson Education Limited.

[29] Blial, O., Ben Mamoun, M. and Benaini, R. (2016) 'An overview on SDN architectures with multiple controllers', *Journal of Computer Networks and Communications*, 2016, pp. 1–8. doi:10.1155/2016/9396525.

[30] *SDNI: A message exchange protocol for software defined networks (sdns ...* Available at: https://www.semanticscholar.org/paper/SDNi%3A-A-Message-Exchange-Protocol-for-Software-Tsou-Aranda/c8b854a110ca1d929dcfc9ccce2027a99a6193a1 (Accessed: 20 August 2023).

[31] H. Xie et al., "Software-defined networking efforts debuted at IETF 84," IETF J., Oct. 2012. [Online]. Available: http://www.internetsociety.org/fr/node/45708

[32] Rowshanrad, S. *et al.* (2014) 'A survey on SDN, the future of Networking', *Journal of Advanced Computer Science & Technology*, 3(2), p. 232. doi:10.14419/jacst.v3i2.3754.

[33] Nife, F.N. and Kotulski, Z. (2020) 'Application-aware firewall mechanism for software defined networks', *Journal of Network and Systems Management*, 28(3), pp. 605–626. doi:10.1007/s10922-020-09518-z.

[34] Cheng, Y. *et al.* (2019) 'FPC: A new approach to firewall policies compression', *Tsinghua Science and Technology*, 24(1), pp. 65–76. doi:10.26599/tst.2018.9010003.

[35] Lin, W. and Zhang, L. (2016) 'The load balancing research of SDN based on Ant Colony algorithm with Job Classification', *Proceedings of the 2016 2nd Workshop on Advanced Research and Technology in Industry Applications* [Preprint]. doi:10.2991/wartia-16.2016.95.

[36] Barreto, F. (2012) 'Fast emergency paths schema to overcome transient link failures in OSPF routing', *International journal of Computer Networks & Communications*, 4(2), pp. 17–34. doi:10.5121/ijcnc.2012.4202.

[37] Nelakuditi, S. *et al.* (2007) 'Fast local rerouting for handling transient link failures', *IEEE/ACM Transactions on Networking*, 15(2), pp. 359–372. doi:10.1109/tnet.2007.892851.

[38] Semong, T. *et al.* (2020) 'Intelligent load balancing techniques in software defined networks: A survey', *Electronics*, 9(7), p. 1091. doi:10.3390/electronics9071091.

[39] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," in *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14-76, Jan. 2015, doi: 10.1109/JPROC.2014.2371999.

[40] Bakshi, K. (2013) 'Considerations for Software Defined Networking (SDN): Approaches and use cases', *2013 IEEE Aerospace Conference* [Preprint]. doi:10.1109/aero.2013.6496914.

[41] *Cisco Global Cloud Index: Forecast and methodology, 2016 2021 White Paper*. Available at: https://virtualization.network/Resources/Whitepapers/0b75cf2e-0c53-4891-918e-b542a5d364c5_white-paper-c11-738085.pdf (Accessed: 21 September 2023).

[42] Report, 2016 Internet Trends (2016) *2016 internet trends report, Kleiner Perkins | Make History*. Available at: https://www.kleinerperkins.com/perspectives/2016-internet-trends-report (Accessed: 21 September 2023).

[43] Lu, Y. and Gu, H. (2019) 'Flexible and scalable optical interconnects for Data Centers: Trends and challenges', *IEEE Communications Magazine*, 57(10), pp. 27–33. doi:10.1109/mcom.001.1900326.

[44] Emara, T.Z. and Huang, J.Z. (2019) 'A distributed data management system to support large-scale data analysis', *Journal of Systems and Software*, 148, pp. 105–115. doi:10.1016/j.jss.2018.11.007.

[45] K. Bonawitz et al., "Towards federated learning at scale: System design,"arXiv: 1902.01046, 2019

[46] Ghiasi, A. (2015) 'Large data centers interconnect bottlenecks', *Optics Express*, 23(3), p. 2085. doi:10.1364/oe.23.002085.

[47] Ghiasi, A. (2012) 'Is there a need for on-chip photonic integration for large data warehouse switches', *The 9th International Conference on Group IV Photonics (GFP)* [Preprint]. doi:10.1109/group4.2012.6324075.

[48] Quttoum, A.N. (2018) 'Interconnection structures, management and routing challenges in cloud-service data center networks: A survey', *International Journal of Interactive Mobile Technologies (iJIM)*, 12(1), p. 36. doi:10.3991/ijim.v12i1.7573.

[49] Imran, M. and Haleem, S. (2018) 'Optical interconnects for cloud computing data centers: Recent advances and future challenges', *Proceedings of International Symposium on Grids and Clouds 2018 in conjunction with Frontiers in Computational Drug Discovery — PoS(ISGC 2018 & FCDD)* [Preprint]. doi:10.22323/1.327.0017.

[50] Miao, W., Yan, F. and Calabretta, N. (2016) 'Towards petabit/s all-optical flat data center networks based on WDM optical cross-connect switches with flow control', *Journal of Lightwave Technology*, 34(17), pp. 4066–4075. doi:10.1109/jlt.2016.2593040.

[51] Z. Chai, X. Hu, F. Wang, X. Niu, J. Xie, and Q. Gong, "Ultrafast all-optical switching," Adv. Opt. Mater., vol. 5, no. 7, 2017, Art. no. 1600665

[52] Testa, F. and Pavesi, L. (no date) *Optical switching in Next Generation Data Centers*. Cham: Springer 2017.

[53] Amiri, M. *et al.* (2017) 'Sdn-enabled game-aware routing for Cloud Gaming Datacenter Network', *IEEE Access*, 5, pp. 18633–18645. doi:10.1109/access.2017.2752643.

[54] Pagès, A. *et al.* (2019) 'Orchestrating virtual slices in data centre infrastructures with optical DCN', *Optical Fiber Technology*, 50, pp. 36–49. doi:10.1016/j.yofte.2019.02.011.

[55] Xue, X. *et al.* (2020) 'SDN-controlled and orchestrated opsquare DCN enabling automatic network slicing with differentiated QoS provisioning', *Journal of Lightwave Technology*, 38(6), pp. 1103–1112. doi:10.1109/jlt.2020.2965640.

[56] Rumley, S. *et al.* (2017) 'Optical interconnects for Extreme Scale Computing Systems', *Parallel Computing*, 64, pp. 65–80. doi:10.1016/j.parco.2017.02.001.

[57] Chen, K. *et al.* (2014) 'OSA: An optical switching architecture for data center networks with unprecedented flexibility', *IEEE/ACM Transactions on Networking*, 22(2), pp. 498–511. doi:10.1109/tnet.2013.2253120.

[58] Ballani, H. *et al.* (2018) 'Bridging the last mile for optical switching in data centers', *Optical Fiber Communication Conference* [Preprint]. doi:10.1364/ofc.2018.w1c.3.

[59] Shukla, V., Srivastava, R. and Choubey, D.K. (2019) 'Optical switching in next-generation data centers', *Contemporary Developments in High-Frequency Photonic Devices*, pp. 164–193. doi:10.4018/978-1-5225-8531-2.ch008.

[60] Cheng, Q. *et al.* (2018) 'Photonic switching in high performance datacenters [invited]', *Optics Express*, 26(12), p. 16022. doi:10.1364/oe.26.016022.

[61] Han, S. *et al.* (2013) 'Network support for resource disaggregation in next-generation datacenters', *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks* [Preprint]. doi:10.1145/2535771.2535778.

[62] Cheng, Y., Chai, Z. and Anwar, A. (2018) 'Characterizing co-located datacenter workloads', *Proceedings of the 9th Asia-Pacific Workshop on Systems* [Preprint]. doi:10.1145/3265723.3265742.

[63] Roozbeh, A. *et al.* (2018) 'Software-defined "hardware" infrastructures: A survey on enabling technologies and open research directions', *IEEE Communications Surveys & Tutorials*, 20(3), pp. 2454–2485. doi:10.1109/comst.2018.2834731.

[64] Li, C.-S. *et al.* (2017) 'Composable architecture for Rack Scale Big Data Computing', *Future Generation Computer Systems*, 67, pp. 180–193. doi:10.1016/j.future.2016.07.014.

[65] Lin, R. *et al.* (2020) 'Disaggregated Data Centers: Challenges and trade-offs', *IEEE Communications Magazine*, 58(2), pp. 20–26. doi:10.1109/mcom.001.1900612.

[66] Gao, X. et al.(2016) 'Network Requirements for Resource Disaggregation', 12th UNISEX symposium on operating systems design and implementation (OSDI 16)

[67] N. Farrington, et al., "Helios: a hybrid electrical/optical switch architecture for modular data centers," in ACM SIGCOMM 2010

[68] G. Wang, et al, "c-Through: Part-time Optics in Data centers," in ACM SIGCOMM 2010

[69] Peng, S. *et al.* (2014) 'A novel sdn enabled hybrid optical packet/circuit switched data centre network: The lightness approach', *2014 European Conference on Networks and Communications (EuCNC)* [Preprint]. doi:10.1109/eucnc.2014.6882622.

[70] 'Lightness EU FP7 project'. [Online]. Available: http://www.ict-lightness.eu/

[71] Perello, J. *et al.* (2013) 'All-optical packet/circuit switching-based data center network for enhanced scalability, latency, and throughput', *IEEE Network*, 27(6), pp. 14–22. doi:10.1109/mnet.2013.6678922.

[72] Saridis, G.M. *et al.* (2016) 'Lightness: A function-virtualizable software defined data center network with all-Optical Circuit/packet switching', *Journal of Lightwave Technology*, 34(7), pp. 1618–1627. doi:10.1109/jlt.2015.2509476.

[73] Saridis, G.M. *et al.* (2015) 'Lightness: A deeply-programmable SDN-Enabled Data Centre Network with OCS/OPS Multicast/unicast switch-over', *2015 European Conference on Optical Communication (ECOC)* [Preprint]. doi:10.1109/ecoc.2015.7341690.

[74] Saridis, G.M. *et al.* (2017) 'Lightness: All-optical SDN-enabled intra-dcn with Optical Circuit and packet switching', *Optical Switching in Next Generation Data Centers*, pp. 147–165. doi:10.1007/978-3-319-61052-8_8.

[75] Guo, B. *et al.* (2015) 'SDN-enabled Programmable optical packet/circuit switched Intra Data Centre Network', *Optical Fiber Communication Conference* [Preprint]. doi:10.1364/ofc.2015.th4g.5.

[76] Lu, P. *et al.* (2015) 'Highly efficient data migration and backup for big data applications in Elastic Optical inter-data-center Networks', *IEEE Network*, 29(5), pp. 36–42. doi:10.1109/mnet.2015.7293303.

[77] Lin, R. *et al.* (2018) 'Real-time 100 gbps/λ/core NRZ and EDB IM/DD transmission over multicore fiber for intra-datacenter Communication Networks', *Optics Express*, 26(8), p. 10519. doi:10.1364/oe.26.010519.

[78] Lavrencik, J. *et al.* (2017) '4Λ × 100gbps Vcsel pam-4 transmission over 105m of wide band Multimode Fiber', *Optical Fiber Communication Conference* [Preprint]. doi:10.1364/ofc.2017.tu2b.6.

[79] Ozolins, O. *et al.* (2018) '7×149 gbit/s PAM4 transmission over 1 km multicore fiber for short-reach optical interconnects', *Conference on Lasers and Electro-Optics* [Preprint]. doi:10.1364/cleo_si.2018.sm4c.4.

[80] Hayashi, T. *et al.* (2016) '125-μm-cladding eight-core multi-core fiber realizing ultra-high-density cable suitable for O-band short-reach optical interconnects', *Journal of Lightwave Technology*, 34(1), pp. 85–92. doi:10.1109/jlt.2015.2470078.

[81] Porter, G. *et al.* (2013) 'Integrating microsecond circuit switching into the Data Center', *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM* [Preprint]. doi:10.1145/2486001.2486007.

[82] Werner, S. *et al.* (2018) 'AWGR-based optical processor-to-memory communication for low-latency, low-energy vault accesses', *Proceedings of the International Symposium on Memory Systems* [Preprint]. doi:10.1145/3240302.3240318.

[83] Bergman, K. *et al.* (2014) 'Silicon Photonics for exascale systems', *Optical Fiber Communication Conference* [Preprint]. doi:10.1364/ofc.2014.m3e.1.