ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ

Βελτιστοποίηση επιλογής κόμβων δεδομένων μέσω Μηχανικής Μάθησης για αποτελεσματική διαχείριση δεδομένων στο Pervasive Edge Computing

ΔΗΜΗΤΡΗΣ ΠΑΠΑΘΑΝΑΣΙΟΥ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων
Κωνσταντίνος Κολομβάτσος

Λαμία, Φεβρουάριος 2024

UNIVERSITY OF THESSALY

SCHOOL OF SCIENCE

INFORMATICS AND COMPUTATIONAL BIOMEDICINE

Optimizing Data Node Selection through Machine Learning for Efficient Data Management in Pervasive Edge Computing

Dimitris Papathanasiou

Master thesis

Supervisor
Konstantinos Kolomvatsos

Lamia, February 2024

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ
ΚΑΤΕΥΘΥΝΣΗ

«ΠΛΗΡΟΦΟΡΙΚΗ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΑΣΦΑΛΕΙΑ, ΔΙΑΧΕΙΡΙΣΗ ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΡΟΣΟΜΟΙΩΣΗ»

Βελτιστοποίηση επιλογής κόμβων δεδομένων μέσω Μηχανικής Μάθησης για αποτελεσματική διαχείριση δεδομένων στο Pervasive Edge Computing

Δημήτρης Παπαθανασίου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων
Κωνσταντίνος Κολομβάτσος

Λαμία, Φεβρουάριος 2024

«Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης»

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο [«τίτλος εργασίας»] αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο ΔΗΛΩΝ

5/2/2024

Υπογραφή

Βελτιστοποίηση επιλογής κόμβων δεδομένων μέσω Μηχανικής Μάθησης για αποτελεσματική διαχείριση δεδομένων στο Pervasive Edge Computing

Δημήτρης Παπαθανασίου

Τριμελής Επιτροπή:

Κων/νος Κολομβάτσος, Επίκουρος Καθηγητής ΠΘ (επιβλέπων)

Αθανάσιος Λουκόπουλος, Αναπληρωτής Καθηγητής ΠΘ

Νικόλαος Τζιρίτας, Επίκουρος Καθηγητής ΠΘ

ΠΕΡΙΛΗΨΗ

Η διαχείριση δεδομένων με βάση το πλαίσιο γίνεται ο κεντρικός πυρήνας πολλών ερευνητικών προσπαθειών. Η μέθοδος αυτή μπορεί να τοποθετηθεί ανάμεσα στον τομέα του Διαδικτύου των Πραγμάτων (IoT) και του Edge Computing (EC). Τεράστιοι όγκοι δεδομένων που καταγράφονται και συλλέγονται από συσκευές IoT επεξεργάζονται σε περιβάλλοντα EC. Ακόμα και αν οι κόμβοι στο Edge αναλαμβάνουν εργασίες διαχείρισης δεδομένων, χαρακτηρίζονται από περιορισμένους αποθηκευτικούς και υπολογιστικούς πόρους σε σύγκριση με το Cloud. Εμφανώς, αυτό ενεργοποιεί την εισαγωγή ευφυών μεθόδων επιλογής δεδομένων ικανών να αποφασίζουν ποια από τα συλλεγόμενα δεδομένα θα διατηρηθούν τοπικά βάσει αιτημάτων των τελικών χρηστών/εφαρμογών. Σε αυτό το άρθρο, δημιουργούμε έναν μηχανισμό όπου οι κόμβοι στο Edge εξάγουν γνώση χρησιμοποιώντας τα δικά τους φίλτρα επιλογής δεδομένων και αποφασίζουν τη διανεμημένη εκχώρηση νεότερων συλλεγόντων δεδομένων στους ομότιμους τους και/ή στο Cloud, όταν αυτά τα δεδομένα δεν συμμορφώνονται με τα τοπικά φίλτρα δεδομένων. Ο μηχανισμός μας σκοπεύει να αναβάλλει τις τελικές αποφάσεις για τη μεταφορά δεδομένων στο Cloud (π.χ., κέντρα δεδομένων) προκειμένου να διατηρήσει δεδομένα που είναι σχετικά με τους τελικούς χρήστες/εφαρμογές όσο το δυνατόν πιο κοντά. Ο μηχανισμός προκύπτει ένα χάρτη επιλογής δεδομένων σε κόμβους Edge μαθαίνοντας συγκεκριμένους υποχώρους δεδομένων, οι οποίοι ευνοούν τον προσδιορισμό των χώρων επεξεργασίας (π.χ., ερωτήματα ανάλυσης). Αξιολογούμε και αντιμετωπίζουμε την προσέγγισή μας

έναντι βασικών και σχεδίων που βρέθηκαν στη βιβλιογραφία, επιδεικνύοντας την εφαρμοσιμότητά της σε περιβάλλοντα πανταχού παρόντος Edge Computing.

Λέξεις Κλειδιά: Internet of Things, Pervasive Edge Computing, Tasks Management, Data Migration, Data Selectivity.

ABSTRACT

Abstract-Context-aware data management becomes the focus of several research efforts, which can be placed at the intersection between Internet of Things (IoT) and Edge Computing (EC). Huge volumes of data captured and collected by IoT devices are processed in EC environments. Even if edge nodes in EC environments undertake data management tasks, they are characterized by limited storage and computational resources compared to Cloud. Apparently, this mobilises the introduction of intelligent data selection methods capable of deciding which of the collected data should be kept locally based on end users/applications requests. In this paper, we devise a mechanism where edge nodes learn their own data selection filters and decide the distributed allocation of newly collected data to their peers and/or Cloud once these data are not conformed with the local data filters. Our mechanism intends to postpone final decisions on data transfer to Cloud (e.g., data centers) to pervasively keep relevant data as close to end users/applications as possible. The mechanism derives a data-selection map across edge nodes by learning specific data sub-spaces, which facilitates the placement of processing tasks (e.g., analytics queries). We evaluate and our approach against baselines and schemes found in the literature showcasing its applicability in pervasive edge computing environments.

Index Terms: Internet of Things, Pervasive Edge Computing, Tasks Management, Data Migration, Data Selectivity.

Table of Contents

ΠΕΡΙΛΗΨΗ I
ABSTRACT IV

ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ 2

CLOUD ΚΑΙ EDGE COMPUTING 1.1 2
ΑΝΑΓΚΗ ΔΙΑΧΕΙΡΗΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΠΕΡΙΒΑΛΛΟΝΤΑ ΧΑΜΗΛΩΝ ΠΟΡΩΝ 1.2 3

ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ 5

ΔΙΑΧΕΙΡΗΣΗ ΕΡΓΑΣΙΩΝ ΣΕ ΔΙΚΤΥΑ WSN 2.1 5
ΠΕΡΙΟΧΕΣ ΒΕΛΤΙΩΣΕΩΝ 2.1.Α 6
COLLABORATIVE LOCALIZATION 2.1.Β 7
SOURCE LOCALIZATION AND TRACKING 2.1.Γ 7
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ 2.2 8
ΣΥΝΔΙΑΣΜΟΣ EC ΚΑΙ AI 2.2.Α 9
ΚΑΤΑΝΕΜΗΜΕΝΗ ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΑΣΙΩΝ 2.3 10
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΡΓΑΣΙΩΝ ΣΤΟ MOBILE CROWDSENSING 2.3.Α 10
ΜΟΝΤΕΛΑ ΑΝΑΘΕΣΗΣ ΕΡΓΑΣΙΩΝ ΣΤΟ MOBILE CROWDSENSING 2.3.Β 11
ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ ΣΤΟ EC 2.4 12
ΚΕΝΤΡΟΠΟΙΗΜΕΝΗ ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ 2.4.Α 13
ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ BLOCKCHAIN ΣΤΟ EC 2.4.Β 14
PERVASIVE EDGE COMPUTING EC 2.5 14

ΚΕΦΑΛΑΙΟ 3 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΙΣΗ 17

ΚΕΦΑΛΑΙΟ 4 ΕΞΕΤΑΖΟΜΕΝΟ ΣΕΝΑΡΙΟ 21

ΚΕΦΑΛΑΙΟ 5 ΠΕΡΙΓΡΑΦΗ ΠΡΟΣΕΓΓΙΣΗΣ 23

ΦΙΛΤΡΑ ΚΑΙ ΡΟΗ ΕΝΕΡΓΕΙΩΝ 5.1 23
ΕΠΙΛΕΚΤΙΚΟΤΗΤΑ ΔΕΔΟΜΕΩΝ 5.2 26
ΔΙΑΧΕΙΡΗΣΗ ΦΙΛΤΡΩΝ ΔΕΔΟΜΕΝΩΝ 5.2.Α 27
ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ 5.2.Β 28

ΚΕΦΑΛΑΙΟ 6 ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ ΚΑΙ ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ 30

ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ 6.1 30
ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΜΑΘΗΣΗΣ 6.2 31

ΚΕΦΑΛΑΙΟ 7 ΠΕΙΡΑΜΑΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ 33

ΜΕΤΡΙΚΕΣ ΑΠΟΔΟΣΗΣ ΚΑΙ ΣΕΝΑΡΙΟ ΠΡΟΣΟΜΟΙΩΣΗΣ 7.1 33
ΑΞΙΟΛΟΓΗΣΗ 7.2 34
ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=3 7.2.Α 35
ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=4 7.2.Β 43
ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=5 7.2.Γ 47
ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ ΤΥΧΑΙΑ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ 7.2.Δ 50
ΠΕΡΙΟΡΙΣΜΟΙ 7.3 56

ΚΕΦΑΛΑΙΟ 8 ΣΥΜΠΕΡΑΣΜΑΤΑ 58

.ΒΙΒΛΙΟΓΡΑΦΙΑ 60

ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ

1.1 CLOUD ΚΑΙ EDGE COMPUTING

Με την ευρεία υιοθέτηση της νεφουπολογιστικής από πολλές εταιρίες κολοσσούς στον κλάδο του IT (Information Technologies), η διαδικασία της επεξεργασίας και αποθήκευσης δεδομένων έχει αλλάξει ριζικά και με τροπή προς μια πιο αποδοτική και κατανεμημένη διαδικασία. Διαθέτοντας ένα μεγάλο αριθμό απομακρυσμένων διακομιστών, το cloud computing μπορεί να παρέχει ευφυές υπολογιστικές, αποθηκευτικές και δικτυακές υπηρεσίες σε πραγματικό χρόνο ώστε να αυτές να συνάδουν με τις ανάγκες των χρηστών και τους πόρους που χρειάζονται [1].

Η ταχεία ανάπτυξη του Διαδικτύου των Πραγμάτων (Internet of Things - IoT) έχει οδηγήσει στην αύξηση του αριθμού των υλικών συσκευών και αισθητήρων που χρησιμοποιούνται παγκοσμίως. Οι συσκευές αυτές κατά κύριο λόγο, συλλέγουν δεδομένα από το περιβάλλον στο οποίο βρίσκονται και τα μεταδίδουν στο νέφος (Cloud) με σκοπό την επεξεργασία και αποθήκευση τους. Οι χρήστες μπορούν στη συνέχεια να αποκτούν πρόσβαση και να εξάγουν σχετικές πληροφορίες από αυτά τα δεδομένα, μέσο τεχνικών ανάλυσης και συσχέτισής τους [1]. Ωστόσο, η αυξανόμενη χρήση του IoT έχει αποκαλύψει προβλήματα με τον παραδοσιακό κεντροποιημένο υπολογισμό στο νέφος, όπως η αργή ταχύτητα επεξεργασίας, η υψηλή καθυστέρηση, πιθανή έλλειψη ιδιωτικότητας των δεδομένων και τέλος το επιπρόσθετο κόστος [1].

Η προσέγγιση του υπολογισμού στην ακμή (Edge Computing - EC) δημιουργήθηκε, έχοντας ως σκοπό την απόδοση λύσεων στις προκλήσεις αυτές. Η κεντρική ιδέα του EC είναι η κατανομή των εργασιών επεξεργασίας, αποθήκευσης και υπολογισμού από το κέντρο του νέφους στις ακμές του. Η χρήση αυτής της μεθόδου, μειώνει τους χρόνους μετάδοσης δεδομένων και ανταπόκρισης των συσκευών, αποφορτίζει ευρυζωνικά το δίκτυο και λιγοστεύει το κόστος μετάδοσης δεδομένων, ενισχύοντας παράλληλα την αποκεντροποίηση [1]. Το EC αναδύεται ως μια υποσχόμενη λύση για την αντιμετώπιση των περιορισμών των παραδοσιακών συστημάτων IoT που βασίζονται στο νέφος.

1.2 ΑΝΑΓΚΗ ΔΙΑΧΕΙΡΗΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΠΕΡΙΒΑΛΛΟΝΤΑ ΧΑΜΗΛΩΝ ΠΟΡΩΝ

Με την ταχεία ανάπτυξη και ευρεία διάδοση της τεχνολογίας της τεχνητής νοημοσύνης, εμφανίζονται όλο και περισσότερες καινοτόμες εφαρμογές που στοχεύουν στην αυτοματοποίηση εργασιών και πρόγνωση συμβάντων σε μεγάλες δικτυακές υποδομές όπως τα δίκτυα IoT. Μερικές από αυτές είναι η καταγραφή και αναγνώριση βιομετρικών δεδομένων, η καταγραφή και ανάλυση βίντεο σε πραγματικό χρόνο και η καταγραφή διάφορων περιβαλλοντικών δεδομένων [2]. Τέτοιου είδους έξυπνες εφαρμογές είναι συνήθως απαιτητικές στο κομμάτι των υπολογιστικών και αποθηκευτικών πόρων καθώς εκτελούν αλγόριθμους που απαιτούν έντονη υπολογιστική ισχύ (π.χ., βαθιά μάθηση), ενώ πολλές φορές χρειάζεται να επεξεργαστούν τα δεδομένα αυτά σε πραγματικό χρόνο. Παρόλα αυτά οι συσκευές που χρησιμοποιούνται στα δίκτυα αυτά (IoT) είναι κατά κύριο λόγο χαμηλής υπολογιστικής ισχύς. Αυτό συμβαίνει για 2 κύριους λόγους: 1) Το μέγεθος τους, καθώς οι περισσότεροι από αυτούς τους κόμβους είναι αισθητήρες και συνεπώς δεν μπορούν να κατέχουν δυνατές μονάδες επεξεργασίας λόγο όγκου και κόστους. 2) Η χρήση μπαταρίας για τη λειτουργία τους, η οποία έχει καθοριστικό ρόλο στη διαθεσιμότητα της συσκευής. Ως αποτέλεσμα της ανάγκης χρήση μπαταρίας για την λειτουργία τους, μια απαιτητική υπολογιστική διεργασία ή η χρήση ενός δυνατού chip επεξεργασίας θα κατανάλωνε υπερβολική ενέργεια που θα εξαντλούσε τη μπαταρία και θα επηρέαζε σε μεγάλο βαθμό το Quality of Service (QoS) της οποιασδήποτε εργασίας είναι υπεύθυνη η συσκευή [2].

Οι νέες αυτές, έξυπνες εφαρμογές περιλαμβάνουν συχνά την εκτέλεση εντατικών υπολογισμών σε δεδομένα αισθητήρων, όπως περιβαλλοντικά δεδομένα ή δεδομένα εικόνας, σε πραγματικό χρόνο με σκοπό την υλοποίηση γρήγορων αλληλεπιδράσεων με το φυσικό περιβάλλον που υπάγεται ο εκάστοτέ χρήστης. Με βάση την μελέτη της Cisco [3], είχε προβλεφθεί ότι οι συσκευές IoT θα αποτελούν το 50% ή 14.7 δισεκατομμύρια από όλες τις συσκευές που υπάγονται σε δικτυακές υποδομές το 2023. Εν τέλη, αυτό όχι μόνο πραγματοποιήθηκε, αλλά και ξεπέρασε τις προβλέψεις, φτάνοντας με τον αριθμό συσκευών να αγγίζει τις 15.14 δισεκατομμύρια. Επιπλέον στην μελέτη δηλώθηκε ότι οι μισές διεργασίες (workloads) θα εκτελούνται εκτός των εγκαταστάσεων των επιχειρήσεων και ότι η

εκτέλεσή τους θα μεταφερθεί σε κέντρα δεδομένων που υπάγονται ή όχι στο cloud ή στις παρυφές του δικτύου. Με βάση την ίδια μελέτη οι συσκευές αυτές αναμένεται να πράξουν 507.9 ZB δεδομένων μέσα στα επόμενα έτη. Παρόλα αυτά η ανάλυση και διαχείριση δεδομένων τόσο μεγάλου όγκου είναι δύσκολη για οργανισμούς που δεν έχουν λάβει ειδικά μέτρα για τη διαχείρισή τους [4]. Το edge computing γίνεται μια δημοφιλής επιλογή σε αυτό το πλαίσιο, καθώς το IoT έχει εδραιωθεί ως πρότυπο επεξεργασίας δεδομένων στις παρυφές του δικτύου [4].

Η διάθεση πόρων στο EC προσφέρει σημαντικό πλεονέκτημα σε σχέση με το παραδοσιακό cloud computing αφού εξαλείφει την ανάγκη μεταφοράς όλων των δεδομένων στο cloud για υπολογιστικές και αποθηκευτικές εργασίες. Αυτό απεγκλωβίζει σημαντικά την ευρυζωνικότητα του δικτύου και άλλους πόρους που καταλαμβάνονται από το cloud computing. Ωστόσο, καθώς οι εργασίες κατανέμονται σε κάθε κόμβο του Edge οι οποίοι χαρακτηρίζονται απο περιορισμένους πόρους, η αποτελεσματική και έξυπνη διαχείριση τόσο των πόρων του δικτύου όσο και τον εργασιών που υπάγονται σε αυτό, γίνεται ζωτικής σημασίας για την επιτυχία του Edge Computing [1].

ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΑΣΙΩΝ ΣΕ ΔΙΚΤΥΑ WSN

Τα δίκτυα WSN αποτελούνται από χιλιάδες μικρές, ασύρματες συσκευές που λειτουργούν με τη χρήση μπαταριών [5]. Η διαθεσιμότητα οικονομικών συσκευών αισθητήρων, έχει βοηθήσει στην ανάπτυξη μια νέας υποκλάσης δικτύων WSN, τα λεγόμενα Multimedia Wireless Sensor Networks (MWSN). Τα δίκτυα αυτά έχουν δημιουργήσει μια μεγάλη αύξηση σε εφαρμογές οι οποίες παρεμβάλουν στον φυσικό κόσμο, όπως εφαρμογές ελέγχου κυκλοφορίας και περιβαλλοντική παρακολούθηση [5]. Για την υποστήριξη εφαρμογών τέτοιας μορφής θα πρέπει να ληφθεί υπόψη, η υψηλή κατανάλωση ενέργειας που θα χρειάζεται για τη πραγματοποίηση του σκοπού της. Συνεπώς είναι αναγκαίο να προταθούν λύσεις με στόχο την καλύτερη διαχείριση της ενέργειας στοχεύοντας στο καλύτερο δυνατό QoS.

Οι κόμβοι σε ένα WSN δίκτυο πέρα από την ατομική χρήση τους, δηλαδή τη λειτουργία τους ως μεμονωμένες επεξεργαστικές μονάδες, μπορούν να χρησιμοποιηθούν και συνεργατικά με τους υπόλοιπους κόμβους του δικτύου χρησιμοποιώντας multi-hop επικοινωνία [6]. Κατά συνέπεια το δίκτυο WSN μπορεί να χρησιμοποιηθεί ως ένα σύστημα παράλληλου υπολογισμού μέσω της ανταλλαγής δεδομένων μεταξύ των κόμβων, δημιουργώντας με αυτόν το τρόπο ισχυρή υπολογιστική δύναμη μέσω του παράλληλου αυτού υπολογισμού.

Συνεπώς μια μεγάλη και απαιτητική ως προς πόρους διεργασία μπορεί να τμηματοποιηθεί σε μικρότερες υπο-διεργασίες, οι οποίες με τη σειρά τους θα εκτελεστούν στους ατομικούς κόμβους. Με τη χρήση αυτού του σεναρίου, κάθε κόμβος θα χρησιμοποιήσει συγκεκριμένους υπολογιστικούς πόρους για την επεξεργασία και την επικοινωνία στο δίκτυο. Για την εφαρμογή ενός τέτοιου τρόπου διαχείρισης των πληροφοριών του δικτύου πρέπει να υιοθετηθούν κατάλληλα μοντέλα και αρχιτεκτονικές συνεργασίας κόμβων, που θα αναλύουν και θα κατανέμουν τις διεργασίες στους καταλληλότερους κόμβους [6].

Ένα χαρακτηριστικό παράδειγμα εφαρμογής που χρησιμοποιεί τη συνεργατική επεξεργασία πληροφοριών σε WSN μπορεί εύκολα να θεωρηθεί ως συνεργατικός εντοπισμός και παρακολούθηση, το οποίο είναι ένα από τα βασικά πεδία εφαρμογής σε αυτόν τον τομέα. Οι αισθητήρες που είναι απομακρυσμένοι, συλλέγουν σήματα

από ένα κινούμενα αντικείμενα ή άλλες οντότητες εντός της περιοχής κάλυψής τους και μοιράζονται τις πληροφορίες για αυτό το αντικείμενο με άλλους αισθητήρες με σκοπό τον εντοπισμό του. Ανάλογα με τον τομέα εφαρμογής, τα αντικείμενα που παρακολουθούνται μπορεί να είναι οχήματα, άνθρωποι ή υλικά. Υπάρχουν πολλά ερευνητικά θέματα που προβληματίζουν τους ερευνητές στον τομέα του συνεργατικού εντοπισμού, τα οποία έχουν οδηγήσει στην ανάπτυξη διαφόρων αλγορίθμων και τεχνικών εντοπισμού για WSN [6].

2.1.α ΠΕΡΙΟΧΕΣ ΒΕΛΤΙΩΣΕΩΝ

Η συλλογική επεξεργασία σημάτων και πληροφοριών μέσω ενός δικτύου αντιπροσωπεύει έναν σχετικά πρόσφατο τομέα έρευνας, συνδεδεμένο με την αποκεντρωμένη σύγκλιση πληροφοριών. Κρίσιμα τεχνικά ζητήματα περιλαμβάνουν το ποσοστό διαμοιρασμού πληροφοριών μεταξύ των κόμβων και τη διαδικασία με την οποία οι κόμβοι συνοψίζουν πληροφορίες που προέρχονται από άλλους κόμβους. Η επεξεργασία δεδομένων από περισσότερους αισθητήρες συνήθως οδηγεί σε βελτιωμένη απόδοση, αλλά απαιτεί περισσότερους πόρους επικοινωνίας άρα και ενέργεια [6]. Λιγότερα δεδομένα χάνονται όταν οι πληροφορίες ανταλλάσσονται σε χαμηλότερο επίπεδο, όπως ατόφια σήματα, με το μειονέκτημα της μεγαλύτερης απαίτησης εύρους ζώνης. Συνεπώς, πρέπει να ληφθούν υπόψη οι πολλαπλές αντισταθμίσεις μεταξύ απόδοσης και εκμετάλλευσης πόρων κατά τη συλλογική επεξεργασία σήματος και πληροφοριών μέσω ασύρματων αισθητήρων [6].

Άλλα ζητήματα περιλαμβάνουν τους τρόπους με τους οποίους θα πληρούνται οι απαιτήσεις σε χρονική καθυστέρηση και αξιοπιστία κατά την μετάδοση πληροφοριών. Ένα πυκνό δίκτυο οικονομικών αισθητήρων μπορεί να επιτρέψει τη χωρική δειγματοληψία χωρίς την ανάγκη για απαιτητικούς αλγόριθμους. Αυτοί οι αλγόριθμοι πρέπει να λειτουργούν ασύγχρονα, δεδομένου ότι οι ταχύτητες των επεξεργαστών και οι δυνατότητες επικοινωνίας μπορεί να διαφέρουν ή ακόμη και να αλλάζουν δραματικά. Η έρευνα και η πείρα έχουν αποδείξει ότι η βέλτιστη συνεργασία μεταξύ των αισθητήρων μπορεί να ενισχύσει σημαντικά την αποτελεσματικότητα της ανίχνευσης και της επεξεργασίας σε δίκτυα αισθητήρων [6].

2.1.β COLLABORATIVE LOCALIZATION

Ο εντοπισμός (localization) στα WSN αποτελεί μια τεχνολογία κλειδί. Πολλές προκλήσεις στον αυτόματο καθορισμό της τοποθεσίας των αισθητήρων υπάρχουν λόγω των περιορισμένων πόρων του δικτύου καθώς και του υψηλού ποσοστού αστοχίας τους. Η έρευνα στο τομέα αυτό επικεντρώνεται στις απαιτήσεις που δημιουργούνται από τις διάφορες εφαρμογές, όπως είναι η επεκτασιμότητά τους, η ενεργειακή αποδοτικότητα, η ασφάλεια και η ακρίβεια.

Σε διάφορα σενάρια, ιδίως σε εφαρμογές όπως ο εντοπισμός αντικειμένων, ο αυτόματος εντοπισμός των αισθητήρων είναι ζωτικής σημασίας για ένα δίκτυο WSN. Ένας κόμβος είναι αναγκαίο να γνωρίζει τη θέση του στο περιβάλλον που βρίσκεται έτσι ώστε να μπορεί να συλλέξει ουσιαστικά δεδομένα από τη τοποθεσία του [8]. Τα παραδοσιακά συστήματα GPS, δεν είναι πάντα κατάλληλα για χρήση σε αυτό το πλαίσιο καθώς υπάρχουν σημαντικοί περιορισμοί κόστους και ενσωμάτωσης [9]. Για αυτό το λόγω έχουν αναπτυχθεί τεχνικές εντοπισμού που χρησιμοποιούν κόμβους beacon ή anchor, οι οποίοι έχουν γνωστές θέσεις στο δίκτυο και με αυτό τον τρόπο είναι εύκολο να δημιουργηθεί ένα τοπικό σύστημα συντεταγμένων.

Αλγόριθμοι εκτιμούν τις θέσεις άλλων αισθητήρων με βάση τις μετρήσεις μεταξύ των αισθητήρων, επιτρέποντας κατανεμημένα ή κεντροειδή σχήματα για τον καθορισμό της θέσης. Οι συνεργατικές τεχνικές εντοπισμού επιτρέπουν την εκτίμηση της θέσης για κόμβους που δεν είναι anchor ή beacon εκμεταλλευόμενες τις μετρήσεις απόστασης μεταξύ των αισθητήρων, με ακόμη πιο ακριβή αποτελέσματα, ειδικά σε πολύπλοκες τοπολογίες δικτύων.

Αυτές οι τεχνικές ενισχύουν την αποδοτικότητα των WSN, προσφέροντας γνώση της θέσης στους αισθητήρες, κάτι που είναι ουσιαστικό για λειτουργίες όπως η δρομολόγηση και η κάλυψη ανίχνευσης. Οι ερευνητές έχουν εξερευνήσει μεθόδους, όπως η πολυδιάστατη κλιμάκωση (MDS), για τον υπολογισμό των θέσεων των αισθητήρων με βάση τις εκτιμώμενες αποστάσεις και την κανονικοποίηση των συντεταγμένων με βάση τις γνωστές θέσεις των κόμβων, προσφέροντας έτσι μεγαλύτερη ακρίβεια στην εντοπισμό των αισθητήρων.

2.1.γ SOURCE LOCALIZATION AND TRACKING

Μια τυπική εφαρμογή των WSN είναι ο εντοπισμός κινούμενων στόχων χρησιμοποιώντας στρατηγικά τοποθετημένους ασύρματους κόμβους. Η κύρια

πρόκληση σε αυτό το πλαίσιο είναι η διατήρηση της περιορισμένης ενέργειας των μπαταριών των αισθητήρων, καθώς η αντικατάστασή τους είναι αρκετά δύσκολη. Για να αντιμετωπιστεί αυτό, έχουν εισαχθεί συνεργατικές τεχνικές και αλγόριθμοι οι οποίοι αποσκοπούν στο διαμοιρασμό του φορτίου άρα και στη μείωση κατανάλωσης ενέργειας στους μεμονωμένους κόμβους. Οι προσεγγίσεις αυτές υποθέτουν ότι κάθε αισθητήρας μπορεί να χρησιμοποιήσει τα ήδη ληφθέντα δεδομένα για να βελτιστοποιήσει τις μελλοντικές ενέργειες ανίχνευσης, βελτιώνοντας έτι τη χρήση των πόρων και την αποδοτικότητα της επικοινωνίας [6].

2.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Τα τελευταία χρόνια, η μηχανική μάθηση έχει αναδειχθεί ως ένα επαναστατικό εργαλείο στον τομέα της διαχείρισης δεδομένων, προσφέροντας νέες προσεγγίσεις για την βελτιστοποίηση διαφόρων πτυχών της χειρισμού και αποθήκευσης δεδομένων. Η μηχανική μάθηση, δίνει τη δυνατότητα να αξιοποιήσουμε τη δύναμη της λήψης αποφάσεων με βάση τα δεδομένα αλλά επίσης αυξάνει τις δυνατότητες παραμετροποίησης και βελτιστοποίησης των στρατηγικών που χρησιμοποιούνται σε πραγματικό χρόνο, προσαρμόζοντάς τις στη δυναμική φύση του ΕΟ.

Η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI) αποτελεί μια τεχνολογία που προσδίδει στις μηχανές ένα επίπεδο νοημοσύνης, επιτρέποντάς τους να εκτελούν εργασίες με τρόπο παρόμοιο με τους ανθρώπους [1,10]. Αντίθετα οι αλγόριθμοι βασισμένοι σε ευρητικές μεθόδους και η εξόρυξη δεδομένων (Data Mining) [11] έχουν παίξει σημαντικό ρόλο στις λύσεις της τεχνητής νοημοσύνης για το IoT τις τελευταίες δεκαετίες, επικεντρωνόμαστε κυρίως στην Μηχανική Μάθηση (Machine Learning - ML), έναν πρόσφατα αναδυόμενο τομέα του AI. Είναι σημαντικό να σημειώσουμε ότι, ενώ το data mining και το machine learning έχουν κοινές χρήσεις όσον αφορά την αξιοποίηση μαζικών δεδομένων, το ML επικεντρώνεται στην απομίμηση της διαδικασίας μάθησης του ανθρώπου, ενώ η ΕΔ σχεδιάστηκε για την εξαγωγή κανόνων από τα δεδομένα [1,11]. Σε αντίθεση με το DM, το ML αντιπροσωπεύει ένα υψηλότερο επίπεδο νοημοσύνης και αποτελεί ένδειξη της μελλοντικής κατεύθυνσης του AI.

Η εκτεταμένη εφαρμογή της τεχνητής νοημοσύνης, ειδικότερα της μηχανικής μάθησης, έχει αναμφισβήτητα εξελιχθεί σε μια αναπόφευκτη τάση στη "εποχή των

μεγάλων δεδομένων" (Big Data Era) που φέρνει το IoT. Είναι σημαντικό να τονιστεί ότι πολλές από αυτές τις εφαρμογές του AI επιβάλλουν αυστηρές απαιτήσεις για τον χρόνο απόκρισης και την αξιοπιστία του δικτύου, απαιτήσεις που συχνά δεν ικανοποιούνται από τον καθιερωμένο υπολογισμό στο νέφος. Αντιθέτως, το μοντέλο EC μπορεί να ανταποκριθεί σε αυτές τις απαιτήσεις, χρησιμοποιώντας το AI στον ακριβή προσδιορισμό και κατανέμοντας πόρους υπολογισμού και αποθήκευσης σε συσκευές στο άμεσο περιβάλλον των σημείων πρόσβασης. Παρότι το EC φέρνει οφέλη, όπως η μείωση του χρόνου απόκρισης, η βελτίωση της ασφάλειας των δεδομένων και ενίσχυση της συνολικής ασφάλειας, οι περιορισμένες δυνατότητες υπολογισμού και αποθήκευσης στις συσκευές στο edge έχουν προκαλέσει νέα προβλήματα. Η χρήση της μηχανικής μάθησης για τη βελτιστοποίηση του EC και την επίλυση των προβλημάτων που αντιμετωπίζει έχει γίνει μια νέα τάση στη σύγχρονη έρευνα [1,12].

2.2.α ΣΥΝΔΥΑΣΜΟΣ EC ΚΑΙ AI

Οι προσπάθειες για τον συνδυασμό του AI και του EC σε πρόσφατες έρευνες έχουν ενθαρρυνθεί από δύο κύριους λόγους, οι οποίοι αντικατοπτρίζουν πλήρως το αμοιβαίο όφελος μεταξύ του AI και του EC:

1. Η ανάπτυξη του EC αντιμετωπίζει ακόμα πολλές προκλήσεις, όπως η προγραμματισμός εργασιών, η κατανομή πόρων, η βελτιστοποίηση καθυστέρησης, η βελτιστοποίηση κατανάλωσης ενέργειας και τα θέματα ιδιωτικότητας και ασφαλείας. Ως απάντηση σε αυτές τις προκλήσεις, πολλοί ερευνητές έχουν υιοθετήσει λύσεις βασισμένες σε τεχνητή νοημοσύνη για την προώθηση της ανάπτυξης του EC [1].
2. Παρά τη γρήγορη εξέλιξη του AI, η αποτελεσματική εφαρμογή του εξαρτάται από ισχυρούς υπολογιστικούς πόρους. Το παραδοσιακό cloud computing παρέχει επαρκείς υπολογιστικούς και αποθηκευτικούς πόρους, αλλά η χρήση της τεχνητής νοημοσύνης στο νέφος για εξαγωγή γνώσης και εκπαίδευσης μπορεί να οδηγήσει σε σημαντικές καθυστερήσεις, καθώς και σε ζητήματα ιδιωτικότητας και ασφάλειας δεδομένων. Με την εκτέλεση των εργασιών τεχνητής νοημοσύνης σε κόμβους στο άμεσο περιβάλλον του τελικού χρήστη, το EC μπορεί

αποτελεσματικά να αντιμετωπίσει τα παραπάνω ζητήματα με βελτιωμένη σταθερότητα, αξιοπιστία καθώς και την εμπειρία του χρήστη [1].

Οι ερευνητές έχουν δημιουργήσει σημαντικά άλματα για την αντιμετώπιση των προβλημάτων αυτών με κύρια επικέντρωση στη ενίσχυση της αποδοτικότητας των αλγόριθμων τεχνίτης νοημοσύνης που διαχειρίζονται περιορισμένους υπολογιστικούς και ενεργειακούς πόρους όπως παρουσιάζεται στο [1].

2.3 ΚΑΤΑΝΕΜΗΜΕΝΗ ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΑΣΙΩΝ

Με την εδραίωση των WSN ως ένα επαναστατικό μέσω επικοινωνίας μεταξύ συσκευών πολύπλοκων δικτυακών υποδομών, όπου μια πληθώρα αισθητήρων συλλέγουν και διακινούν δεδομένα συνεργατικά, η αποτελεσματική οργάνωση των εργασιών αποτελεί ένα μείζον ζήτημα. Οι προκλήσεις που προκύπτουν από τους περιορισμένους πόρους και τη δυναμική φύση των WSN υπογραμμίζουν την κρίσιμη σημασία της επιδεξιότητας στη διαχείριση κατανεμημένων εργασιών με σκοπό την βελτιστοποίηση της αποτελεσματικότητας των δικτύων αυτών. Καθώς το τοπίο των WSN συνεχίζει να εξελίσσεται, η ανάγκη για μηχανισμούς έξυπνης και προσαρμοστικής διαχείρισης εργασιών γίνεται ολοένα και πιο εμφανής καθώς τα δίκτυα WSN υιοθετούνται σε κρίσιμους τομείς όπως οι στρατιωτικές εφαρμογές, περιβαλλοντική επίβλεψη, εφαρμογές εντοπισμού κ.α. [5].

2.3.α ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΡΓΑΣΙΩΝ ΣΤΟ MOBILE CROWDSENSING

Στον τομέα του mobile crowdsensing, η διαχείριση των εργασιών αποτελείται από 2 τύπους: 1) Τη συλλογή δεδομένων από αισθητήρες, 2) Την επεξεργασία και ανάλυση των δεδομένων που συλλέχθηκαν. Οι βασικοί φορείς στην συλλογή δεδομένων, μέσω αισθητήρων, είναι οι φορείς (carriers) και οι οντότητες εκδήλωσης εργασίας [13]. Οι φορείς χρησιμοποιούν αισθητήρες για τη συλλογή πληροφοριών, καθιστώντας τις διαθέσιμες για μελλοντική χρήση. Οι οντότητες εκδήλωσης εργασίας διαδραματίζουν κρίσιμο ρόλο στην ανάθεση εργασιών στους φορείς και στον έλεγχο της ανταλλαγής δεδομένων μεταξύ φορέων και οντοτήτων επεξεργασίας δεδομένων [13]. Μεταξύ αυτών, οι πάροχοι υπηρεσιών (Service Providers - SP) ξεχωρίζουν ως οντότητες που ενδιαφέρονται για τις πληροφορίες που συλλέγονται από τους φορείς. Οι αισθητηριακές εργασίες κατηγοριοποιούνται συστηματικά με βάση τα χαρακτηριστικά τους και τη συμμετοχή των οντοτήτων εκδηλώσεων

εργασίας, συνεισφέροντας στην πλήρη κατανόηση του τοπίου του mobile crowdsensing.

Οι εργασίες που βασίζονται σε συγκεκριμένα γεγονότα (Event-based tasks), περιλαμβάνουν την αισθητηριακή αντίδραση κατά συγκεκριμένων συνθηκών ή συμβάντων, όπως η καταγραφή εικόνων φυσικών καταστροφών. Οι περιοδικές εργασίες (Periodic tasks) περιλαμβάνουν τη δημιουργία δεδομένων από τους φορείς σε τακτικά χρονικά διαστήματα, και έχουν εύρος από την ανάλυση συμπεριφοράς σε αίθουσα μαθημάτων έως την παρακολούθηση των σημάτων καρδιάς ασθενών. Οι χωρικές εργασίες (Spatial tasks) εκμεταλλεύονται το GPS για εργασίες που βασίζονται στην τοποθεσία, όπως η ανίχνευση ταχύτητας αυτοκινήτων ή η αναφορά καταστάσεων δρόμων. Οι χρονικές εργασίες (Temporal tasks) δεν απαιτούν συγκεκριμένη τοποθεσία, αλλά προϋποθέτουν την ανίχνευση και μετάδοση δεδομένων σε συγκεκριμένες χρονικές στιγμές, με διάκριση ανάμεσα σε εργασίες κατά και μετά τη διάρκεια του αιτήματος. Οι χωροχρονικές εργασίες (Spatiotemporal tasks) απαιτούν από έναν φορέα να συλλέγει δεδομένα σε μια συγκεκριμένη περιοχή κατά τη διάρκεια συγκεκριμένης χρονικής στιγμής, όπως μελέτες συμπεριφοράς ατόμων. Στον τομέα της ανάλυσης δεδομένων, οι φορείς σε ένα σύστημα mobile crowdsensing συλλέγουν δεδομένα και τα μεταδίδουν στον SP. Ο SP εξάγει χρήσιμες πληροφορίες από τα δεδομένα που συλλέχθηκαν, χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Οι εργασίες χωρίζονται σε δύο κατηγορίες σύμφωνα με την επεξεργασία δεδομένων: Επεξεργασία πακέτων (Batch processing), όπου οι φορείς μεταβιβάζουν τα δεδομένα τους στον SP για ανάλυση, και Πραγματικού χρόνου επεξεργασία (Real-time processing), όπου η ανάλυση των δεδομένων γίνεται περιοδικά [13].

2.3.β ΜΟΝΤΕΛΑ ΑΝΑΘΕΣΗΣ ΕΡΓΑΣΙΩΝ ΣΤΟ MOBILE CROWDSENSING

Στο mobile crowdsensing, οι οντότητες εργασιών διαδραματίζουν έναν πρωταγωνιστικό ρόλο στην ανάθεση εργασιών στους φορείς, μέσω μοντέλων ανάθεσης εργασιών. Αυτά τα μοντέλα λειτουργούν βασισμένα σε κριτήρια όπως η Συμμετοχή του Ανθρώπου (Human Involvement), η Γνώση που είναι διαθέσιμη στους Παρόχους Υπηρεσιών (Knowledge available to Service Providers) και η Χωρική Διανομή (Spatial Distribution). Τα μοντέλα ανάθεσης εργασιών κατηγοριοποιούνται ευρέως σε μοντέλα επιλογής χρήστη και μοντέλα εκχώρησης

εργασιών από τον διακομιστή. Στα μοντέλα επιλογής χρήστη, οι φορείς έχουν την ευελιξία να επιλέγουν και να εκτελούν εργασίες χωρίς να αποκαλύπτουν την ταυτότητά τους, χρησιμοποιώντας κεντρικά ή συντονισμένα μοντέλα. Στα μοντέλα εκχώρησης εργασιών από τον διακομιστή, οι φορείς στέλνουν πληροφορίες τοποθεσίας στον SP, με μοντέλα όπως τα κατανεμημένα ή αυτόνομα να καθορίζουν τις εργασίες [13].

Σε συγκεκριμένες μελέτες, αλγόριθμοι όπως οι NTA (oNline Task Asignment) και FTA (oFline Task Assignment) έχουν εισαχθεί για την αντιμετώπιση ζητημάτων κινητικότητας αλλά είναι ευάλωτοι σε επιθέσεις [13,14]. Προβλήματα όπως το inimum-Average-Makespan και το Minimum-Largest-Makespan έχουν επιλυθεί με τη χρήση μοντέλων όπως το AOTA και το LOTA [13,15]. Άλλες έρευνες εξετάζουν προκλήσεις όπως η ποιότητα των εργασιών, τα κόστη και οι μηχανισμοί της ευκαιριακής και συμμετοχικής ανίχνευσης. Πλαίσια όπως ο ασφαλής διπλασιασμός δεδομένων με υποστήριξη Fog και ανοικτού κώδικα λύσεις όπως το Mew [16], στοχεύουν στον βέλτιστο διαμοιρασμό εργασιών και την ευελιξία των προγραμματιστών. Οι επιθυμίες για διατήρηση του απορρήτου στην ανάθεση εργασιών, η πραγματοποίηση συλλογής δεδομένων σε πραγματικό χρόνο και η ποιότητα της πληροφορίας εξετάζονται σε διάφορα σχήματα, αναδεικνύοντας τον εξελισσόμενο χώρο της διαχείρισης εργασιών στο mobile crowdsensing.

2.4 ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ ΣΤΟ EC

Στο δυναμικά μεταβαλλόμενο πεδίο του cloud computing, η έλευση του edge computing έχει εισαγάγει νέες προοπτικές και προκλήσεις. Καθώς η εξάρτησή από την αποκεντρωμένη επεξεργασία ενισχύεται, έτσι αυξάνεται και η ανάγκη για εξειδικευμένες στρατηγικές διαχείρισης δεδομένων στον περιβάλλοντα χώρο. Η αποτελεσματική διαχείριση δεδομένων στο πεδίο του edge computing αναδεικνύεται κρίσιμη για ποικίλους λόγους. Αρχικά, προσφέρει λύσεις στον περιορισμό της τοπικής χωρητικότητας αποθήκευσης. Οι κόμβοι στον περιβάλλοντα χώρο, συχνά υπόκεινται σε περιορισμούς πόρων, αδυνατούν να αποθηκεύσουν όλα τα παραγόμενα δεδομένα εντός των τοπικών τους πόρων. Επιπλέον, ελαχιστοποιεί την χρονική καθυστέρηση που συνδέεται με τη λήψη δεδομένων από

απομακρυσμένους κεντρικούς εξυπηρετητές, βελτιώνοντας, κατά συνέπεια, τη συνολική απόδοση των συστημάτων edge computing.

Η διαχείριση δεδομένων αναφέρεται στη διαδικασία αποτελεσματικής συλλογής, αποθήκευσης, επεξεργασίας και ανάλυσης δεδομένων στο άκρο του δικτύου, κοντά στην πηγή παραγωγής των δεδομένων, έναντι της αποκλειστικής εξάρτησης από κεντρικούς διακομιστές στο νέφος. Αυτή η αποκεντρωμένη προσέγγιση είναι κρίσιμη για την ελαχιστοποίηση της καθυστέρησης, την ενίσχυση της λήψης αποφάσεων πραγματικού χρόνου και την διαχείριση μεγάλου όγκου δεδομένων (Big Data) που παράγονται στην πηγή, καθιστώντας το edge computing ένα ζωτικό συστατικό σε περιπτώσεις όπου η έγκαιρη και αποτελεσματική επεξεργασία δεδομένων είναι πρωταρχική, όπως σε περιβάλλοντα IoT, συστήματα υγείας και άλλα σενάρια που χαρακτηρίζονται από περιορισμένους πόρους [17].

2.4.α ΚΕΝΤΡΟΠΟΙΗΜΕΝΗ ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ

Στα συστήματα που χαρακτηρίζονται απο δημιουργία μεγάλου όγκου δεδομένων, τα όρια της κεντροποιημένης διαχείρισης δεδομένων έρχονται στο προσκήνιο. Οι προκλήσεις γίνονται εμφανείς, καθώς τα συστήματα αυτά αντιμετωπίζουν τη συνεχή ροή σημαντικών συνόλων δεδομένων, οδηγώντας σε κρίσιμα ζητήματα όπως τα overload states και οι σημαντικές καθυστερήσεις στην επεξεργασία. Οι δομικοί περιορισμοί που είναι ενσωματωμένοι στη κεντροποιημένη διαχείριση τη καθιστούν συχνά μη αποδοτική, εμποδίζοντας τη δυνατότητά των συστημάτων να επεξεργάζονται εκτεταμένα σύνολα δεδομένων σε πραγματικό χρόνου αποδοτικά]17,18]. Αυτός ο περιορισμός υπονομεύει τη γρήγορη ανάλυση και διάδοση κρίσιμων πληροφοριών, δυσχεραίνοντας τις διαδικασίες λήψης αποφάσεων. Ως αποτέλεσμα, η επεκτασιμότητα των κεντροποιημένων μοντέλων έρχεται υπό αμφισβήτηση, αναγκάζοντας σε επανεκτίμηση της αποτελεσματικότητάς τους στην αντιμετώπιση των αυξανόμενων απαιτήσεων των σύγχρονων αναγκών [18].

Επιπλέον, οι απειλές ασφάλειας επιδεινώνουν τις ευπάθειες που είναι προφανείς στα κεντρικά συστήματα διαχείρισης δεδομένων. Η έννοια του μοναδικού σημείου αποτυχίας (Single Point of Failure) εμφανίζεται έντονα, αντιπροσωπεύοντας ένα σημαντικό κίνδυνο για την ακεραιότητα και τη διαθεσιμότητα των δεδομένων. Σε περίπτωση αποτυχίας της κεντρικής αρχής, το σύστημα δέχεται επίθεση στο σύνολό του, με δυνητική απώλεια ή μη διαθεσιμότητα

των δεδομένων. Οι επιθέσεις Κατανεμημένης Άρνησης Εξυπηρέτησης (DDoS) εισάγει ένα επιπλέον επίπεδο ανησυχίας, καθώς κακόβουλες οντότητες μπορεί να εκμεταλλευτούν την περιορισμένη χωρητικότητα της κεντρικής υποδομής για να εισάγουν πολλαπλά αιτήματα (floding), διαταράσσοντας τη φυσιολογική ροή επεξεργασίας των δεδομένων. Αυτές οι αναφορές στην ασφάλεια υποστηρίζουν μια μετάβαση προς αποκεντρωμένες και κατανεμημένες στρατηγικές διαχείρισης δεδομένων στον τομέα του μεγάλου όγκου δεδομένων, με στόχο τον περιορισμό των κινδύνων και την ενδυνάμωση της ανθεκτικότητας των συστημάτων έναντι των εξελισσόμενων απειλών [18].

2.4.β ΔΙΑΧΕΙΡΗΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ BLOCKCHAIN ΣΤΟ EC

Πρόσφατες έρευνες εξετάζουν την ενσωμάτωση της τεχνολογίας blockchain για τη δημιουργία αποκεντρωμένης διαχείρισης δεδομένων σε κινητά συστήματα που επικεντρώνονται στη δημιουργία μεγάλου όγκου δεδομένων. Η διαφάνεια και η δυνατότητα κοινής χρήσης που προσφέρει η τεχνολογία blockchain χρησιμοποιούνται για να διασφαλίσουν την ακεραιότητα, την ανιχνευσιμότητα, την ασφαλή αποκεντρωμένη πρόσβαση, κοινοποίηση και αποθήκευση δεδομένων [18, 19].

Έχουν αναληφθεί πρωτοβουλίες για την αντιμετώπιση της ομοιογενούς φύσης των μεγάλων δεδομένων με τον συνδυασμό της τεχνολογίας blockchain με το cloud computing, εξασφαλίζοντας ασφαλή αποθήκευση και κοινοποίηση χωρίς τη συμμετοχή εξωτερικών φορέων. Παρά τις πρωτότυπες υλοποιήσεις όπως αυτές που αναφέρονται στο [18, 19, 20], οι προκλήσεις εξακολουθούν να υφίστανται στη διευκόλυνση της άμεσης μεταφοράς δεδομένων και της πρόσβασης σε πραγματικό χρόνο λόγω της περίπλοκης φύσης και των πιθανών κινδύνων που σχετίζονται με την εγκατάσταση της τεχνολογίας blockchain.

2.5 PERVASIVE EDGE COMPUTING EC

Όπως προαναφέρθηκε, είναι γεγονός ότι ο συνδυασμός του IoT και του Edge Computing προσφέρει μια πληθώρα ευκαιριών για τη δημιουργία έξυπνων εφαρμογών που θα διευκολύνουν τις δραστηριότητες των τελικών χρηστών. Το IoT και το EC παρέχουν δύο τεράστιες συνδεδεμένες υποδομές, όπου πολλοί κόμβοι μπορούν να είναι ενεργοί και να συνεργάζονται μεταξύ τους. Οι συσκευές IoT

αναλαμβάνουν την ευθύνη της αλληλεπίδρασης με το περιβάλλον και τους τελικούς χρήστες για τη συλλογή δεδομένων και την υποστήριξη της διεπαφής των παρεχόμενων εφαρμογών. Παράλληλα, οι κόμβοι EC τοποθετούνται κοντά στους τελικούς χρήστες και τις συσκευές IoT [4], διαμορφώνοντας το νέο σχήμα του ούτως Pervasive Edge Computing (PEC). Το PEC είναι η μετασχηματισμός του παλαιού Pervasive Computing ενισχυμένου με μια επιπλέον υποδομή (το EC), που εντάχθηκε με σκοπό να υποστηρίξει τις επεξεργαστικές δραστηριότητες κοντά στην πηγή των δεδομένων. Το PEC γίνεται ο μεσολαβητής μεταξύ του οικοσυστήματος του IoT και του Cloud back end, προσφέροντας ένα επιπλέον επίπεδο όπου η νοημοσύνη μπορεί να υιοθετηθεί και να παραδοθεί με μεγάλα οφέλη για τους τελικούς χρήστες. Προφανώς, η σύνδεση αυτών των τριών επιπέδων (IoT, EC, Cloud) και η παρουσία πολλών συσκευών με εντελώς διαφορετικά χαρακτηριστικά υπολογιστικής ισχύος και ζήτηση για επεξεργασία καθιστούν την υιοθέτηση έξυπνων μοντέλων για τη διαχείριση των δεδομένων απαραίτητη.

Ένα κρίσιμο ζήτημα στο PEC είναι η διαχείριση των μεγάλου όγκου δεδομένων που συλλέγονται από το οικοσύστημα του IoT και μεταφέρονται προς το Cloud. Ακόμη και αν ενσωματωθούν μαζικά πόροι στο EC ώστε να εξελιχθεί σε ένα μεγάλο κατανεμημένο περιβάλλον υπολογισμού, παραμένει ένα αφιλόξενο μέρος για την επεξεργασία μεγάλου όγκου δεδομένων [4]. Λαμβάνοντας υπόψη ότι οι κόμβοι του EC εμφανίζουν περιορισμένες υπολογιστικές δυνατότητες σε σύγκριση με τους πόρους του Cloud, γίνεται προφανές ότι η συλλογή μεγάλου όγκου δεδομένων στο EC δεν είναι εφικτή. Το PEC στοχεύει στο να παρέχει εφαρμογές on-node, κυρίως σχετιζόμενες με την πραγματοποίηση επεξεργασίας δεδομένων σε πραγματικό χρόνο. Επομένως, είναι απαραίτητη η προετοιμασία κάθε κόμβου του EC για τη φιλοξενία κατανεμημένων συνόλων δεδομένων υπό τους περιορισμούς των τοπικών υπολογιστικών δυνατοτήτων. Αυτού του είδους η προετοιμασία, μπορεί να σχετίζεται με την απόφαση σχετικά με τις υπηρεσίες και τα δεδομένα που πρέπει να διατηρηθούν τοπικά, ενώ τα υπόλοιπα μεταφέρονται προς το Cloud back end ή τους ομότιμους κόμβους. Η προαναφερθείσα απόφαση επηρεάζεται φυσικά από τα εισερχόμενα αιτήματα για επεξεργασία, συνήθως σε μορφή εργασιών. Κάθε εργασία απαιτεί μια συγκεκριμένη επεξεργαστική δραστηριότητα που καθορίζει ένα σύνολο περιορισμών που σχετίζονται με τις υπηρεσίες και τα δεδομένα που πρέπει να

αποτελέσουν μέρος της απαιτούμενης επεξεργασίας. Ωστόσο, εάν τα απαραίτητα δεδομένα και υπηρεσίες που πρέπει να υιοθετηθούν για την εκτέλεση των αιτηθέντων εργασιών λείπουν, οι κόμβοι μπορούν να υιοθετήσουν δύο στρατηγικές: εκκίνηση εργασιών ή μετανάστευση των κατάλληλων υπηρεσιών/δεδομένων τοπικά [21]. Η ενεργή μετανάστευση υπηρεσιών ή δεδομένων ανάμεσα σε κόμβους με διαφορετικά χαρακτηριστικά αποτελεί ένα θεμελιώδες πρόβλημα στο να υποστηρίζεται η πανταχού παρούσα εφαρμογή κοντά στους τελικούς χρήστες [22].

ΚΕΦΑΛΑΙΟ 3 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

Προς το παρόν, το IoT έχει εξελιχθεί σε μια κύρια υποδομή, όπου ένα μεγάλο εύρος συσκευών μπορεί να αλληλοεπιδρά με το περιβάλλον και να συλλέγει τεράστιο όγκο δεδομένων. Τα δεδομένα συλλέγονται/αποστέλλονται σε ροές και συνήθως, αντιπροσωπεύουν την εξέλιξη ενός φαινομένου υπό εξέταση (π.χ., πυρκαγιά, ατμοσφαιρική ρύπανση [23], [24], [25]). Οι ροές γίνονται αντικείμενο επεξεργαστικών δραστηριοτήτων όταν χρειάζονται απαντήσεις σε πραγματικό χρόνο, ενώ τα υπόλοιπα δεδομένα χρησιμοποιούνται στη μακροπρόθεσμη λήψη αποφάσεων, μέσω της διάφορων τεχνικών μηχανικής μάθησης. Η διαχυτικότητα των εφαρμογών απαιτεί και τις δύο προαναφερθέντες καταστάσεις για την προώθηση των υποστηριζόμενων υπηρεσιών με σαφή οφέλη για τους τελικούς χρήστες. Κάθε απάντηση παρέχεται υπό τη μορφή επεξεργαστικών αποτελεσμάτων που απαιτούνται από τις τελούμενες εργασίες. Παραδείγματα εργασιών είναι ερωτήσεις ανάκτησης δεδομένων (data retrieval queries) [26], [27], εκπαίδευση μοντέλων Μηχανικής Μάθησης (ML) [28], [29], επιλογή δεδομένων για την προσωρινή αποθήκευση [30], προ επεξεργασία βίντεο [31]. Εμφανώς, οι εργασίες εκτελούνται πάνω στα δεδομένα που συλλέγονται στην υποδομή EC, ενώ οι κόμβοι εκδηλώνουν συγκεκριμένους υπολογιστικούς πόρους που επηρεάζουν την ταχύτητα της επεξεργασίας. Οι κόμβοι μπορούν να εκτελέσουν μόνο ένα υποσύνολο των αρχικά καθορισμένων εργασιών, εάν το στόχος είναι να αποφευχθεί η υπερφόρτωση. Έτσι, μερικές από αυτές τις εργασίες γίνονται ο λόγος λήψης αποφάσεων εκφόρτωσης, με σκοπό την εύρεση του κατάλληλου κόμβου για να τις φιλοξενήσει [29], [32]. Επιπλέον, οι αποφάσεις εκφόρτωσης εργασιών, επιλογής δεδομένων και μεταφοράς δεδομένων μπορούν να συνδυαστούν για να ενισχύσουν τον αυτόνομο χαρακτήρα των κόμβων με οφέλη στην απόδοσή τους. Διάφορες στρατηγικές, έχουν εισαχθεί για όλα τα προαναφερθέντα ερευνητικά προβλήματα που στοχεύουν κυρίως στη συγκέντρωση τοπικών δεδομένων [33], στη συνεργασία μεταξύ κόμβων [34] ή στον καθορισμό ενός κόμβου εξαγωγής ως του μοναδικού εκπροσώπου μιας ομάδας [35], όπου τα δεδομένα αποθηκεύονται και λαμβάνονται οι δραστηριότητες επεξεργασίας [36]. Ανεξάρτητα από την επιλεγείσα στρατηγική, το συζητούμενο πρόβλημα είναι πολύπλοκο εάν θέλουμε να έχουν οι κόμβοι ανεξάρτητη δράση προσπαθώντας να

ανταποκριθούν στις προκλήσεις απόδοσης που επιβάλλουν δυναμικά και κατανεμημένα περιβάλλοντα.

Η αποστολή εργασιών και η μεταφορά δεδομένων αποτελούν διαφορετικές απόψεις του ίδιου νομίσματος, δηλαδή προσπαθούν να λύσουν το πρόβλημα της αδυναμίας (για οποιονδήποτε λόγο) τοπικής εκτέλεσης των εισερχόμενων εργασιών. Η αποστολή εργασιών επιβάλλει τη μεταφορά των επεξεργαστικών δραστηριοτήτων σε έναν ομότιμο κόμβο (peer node), όπου τα κατάλληλα δεδομένα και το φορτίο αφήνουν ελεύθερο χώρο για την εκτέλεση, ενώ η μεταφορά δεδομένων επιβάλλει τον διαμοιρασμό των κατάλληλων τμημάτων πληροφορίας στο τοπικό σύνολο δεδομένων. Διάφορες προκλήσεις προκύπτουν πριν από την εκτέλεση της αποστολής εργασιών, όπως η επιβάρυνση της επικοινωνίας [37]. Οι ερευνητές προσπαθούν να επιτύχουν την καλύτερη δυνατή απόδοση και να παρέχουν μοντέλα πλήρως ευθυγραμμισμένα με τις δυναμικές ανάγκες των κόμβων, λαμβάνοντας υπόψη τις περιβαλλοντικές πληροφορίες, όπως η κατάσταση κάθε κόμβου στην ομάδα ή η απόδοση του δικτύου. Εμφανώς, η θεμελιώδης επίγνωση βρίσκεται στα διαθέσιμα και σχετικά δεδομένα στους κόμβους, προκειμένου να ενεργοποιηθούν και οι δύο αυτές ενέργειες [38]. Ένας «χάρτης» της τοποθέτησης των δεδομένων στο οικοσύστημα [39] μπορεί να βοηθήσει στον εντοπισμό του κόμβου όπου μια εργασία θα μπορούσε να ανατεθεί ή όπου τα δεδομένα μπορούν να μετακινηθούν. Ο τελικός στόχος είναι να αποφευχθεί η υπερφόρτωση των κόμβων που θα μπορούσε να θέσει σε κίνδυνο την ικανότητά τους να εκτελέσουν τις εισερχόμενες αιτήσεις [40]. Η εκτέλεση των εργασιών αποτελεί ένα πρόβλημα βελτιστοποίησης δύσκολο να λυθεί λόγω του υψηλού αριθμού παραμέτρων που πρέπει να ενσωματωθούν και να επεξεργαστούν στα προτεινόμενα μοντέλα. Είναι απαραίτητο να αποφεύγονται σενάρια όπου η απόδοση θα κινδυνεύσει να μειωθεί, επομένως, Συνεργατικά σχήματα [5], γραμμικός προγραμματισμός [41], μοντέλα νοημοσύνης σμηνών [42], [43], γενετικοί αλγόριθμοι [44], ο έξυπνος διαχειρισμός σχημάτων βασισμένων σε γράφους [45] ή η θεωρία βέλτιστου σταματήματος [2] μπορούν να προσφέρουν κάποιες λύσεις στο προαναφερθέν πρόβλημα. Ωστόσο, ο υπολογιστικός φόρτος στους κόμβους για τον προσδιορισμό των δεδομένων που πρέπει να αποθηκεύονται και να διατηρούνται τοπικά, σε συνδυασμό με τη ζήτηση για επεξεργασία, επηρεάζει τη συνολική τους απόδοση [29], [46]. Οι κόμβοι θα επωφελούνταν εάν ήταν ενήμεροι για τη διανομή,

τις τάσεις και τα πρότυπα πρόσβασης των εισερχόμενων εργασιών πάνω στα τοπικά τους δεδομένα [28], [47]. Αυτό μπορεί να οδηγήσει σε ένα αποτελεσματικό μοντέλο επιλεκτικότητας δεδομένων για το φιλτράρισμα των συλλεγμένων δεδομένων (λόγω των περιορισμένων δυνατοτήτων αποθήκευσης στους κόμβους EC) και τη διατήρηση τοπικά μόνο εκείνων που συμφωνούν με τις ζητούμενες εργασίες.

Μια στρατηγική επιλεκτικότητας δεδομένων ή φιλτραρίσματος δεδομένων σχεδιάζεται για να βοηθήσει στον προσδιορισμό των εξαρτήσεων μεταξύ των κόμβων, των συνόλων δεδομένων τους και των εισερχόμενων εργασιών, απορρίπτοντας τα άσχετα δεδομένα και επιτρέποντας τη μείωση περιττών μεταδόσεων μέσω του δικτύου [48], [49]. Το θέμα του [50] είναι να παρουσιάσει το μοντέλο DTFilter, το οποίο είναι ικανό να αξιολογεί τις εισερχόμενες εργασίες βασιζόμενο σε ένα αλγόριθμο ισορροπημένου δέντρου αναζήτησης, που αποφεύγει τη μετάδοση διπλότυπων αντικειμένων στα ρεύματα δεδομένων. Η προσέγγιση φιλτραρίσματος δεδομένων που συζητείται στο [51] υιοθετεί ετικετοποίηση δεδομένων και εκφράσεις κανόνων πολιτικής. Οι συγγραφείς του [52] παρουσιάζουν κατανεμημένα φίλτρα Bloom που εκμεταλλεύονται τον διαμερισμό των δεδομένων βασισμένο σε συμβάντα, λαμβάνοντας υπόψη την αντιμετώπιση ροών συμβάντων χρησιμοποιώντας παράθυρα ολίσθησης. Το θέμα του [53] είναι να προτείνει ένα σχήμα φιλτραρίσματος δεδομένων που αξιολογεί μια μεθοδολογία βασισμένη σε μηχανική μάθηση για τον εντοπισμό ανωμαλιών. Τα φιλτραρισμένα σύνολα δεδομένων υποστηρίζουν την εκπαίδευση των τεχνητών νευρωνικών δικτύων για την πρόβλεψη της απόδοσης εφαρμογών. Στο [54], οι συγγραφείς παρουσιάζουν το πρότυπο IEEE P2805.2 για την υποστήριξη γενικών πρωτοκόλλων απόκτησης, φιλτραρίσματος και επικάλυψης δεδομένων μεταξύ των υποδομών Cloud και EC. Το θέμα του [55] είναι να συζητήσει μια μέθοδο ανίχνευσης για το EC που χρησιμοποιεί υπάρχοντα μοντέλα ML για την ταξινόμηση ενός ύποπτου αρχείου στις κατηγορίες ασφαλούς, κακόβουλου είτε απρόβλεπτου τύπου. Πρέπει να σημειωθεί ότι οι προαναφερθείσες στρατηγικές δεν λαμβάνουν υπόψη τα πρότυπα που απεικονίζουν τα αιτήματα επεξεργασίας όπως αυτά εκτίθενται από τις εισερχόμενες εργασίες.

Σε αυτή την εργασία, αποκλίνουμε σημαντικά από τις νέες τεχνολογίες αιχμής, και επικεντρωνόμαστε στη διαχείριση των κατανεμημένων φίλτρων δεδομένων που ανήκουν στους κόμβους EC. Στόχος μας είναι η ενίσχυση της συμπεριφοράς των

κόμβων EC και η ανίχνευση των κατάλληλων ομότιμων κόμβων που μπορούν να φιλοξενήσουν δεδομένα που βασίζονται πέρα από τα τοπικά φίλτρα δεδομένων. Αυτή η στρατηγική στοχεύει στο να 'αναβάλει' τη μεταφορά δεδομένων στο Cloud, κρατώντας τα κοντά στις πηγές τους και στους τελικούς χρήστες. Λαμβάνοντας υπόψη ότι τα τοπικά φίλτρα δεδομένων παρέχονται κατά την επεξεργασία των εισερχόμενων εργασιών, αποκαλύπτουμε τις δυνατότητες ενός πλήρως αυτόνομου EC όπου οι κόμβοι μεταναστεύουν ενεργά δεδομένα στις σωστές τοποθεσίες (δηλαδή, στους ομότιμους κόμβους) και ενισχύουν τα τοπικά μοντέλα ML που μπορούν να χτιστούν πάνω σε αυτά. Δεδομένου της εξονυχιστικής διερεύνησης της βιβλιογραφίας, αυτή είναι μία από τις πρώτες προσπάθειες για την επεξεργασία στρατηγικών επιλεκτικότητας και τον έξυπνο καθορισμό της κατανομής των δεδομένων. Η εστίασή μας είναι στην επεξεργασία της ανάθεσης context-aware δεδομένων, καθορίζοντας ένα καινοτόμο 'χάρτη' πάνω στα δεδομένα των κόμβων που μπορεί να υποστηρίξει τη λήψη αποφάσεων και να εξυπηρετήσει οποιαδήποτε καθορισμένη εργασία. Η μετανάστευση των δεδομένων θα ενισχύσει τα κατανεμημένα σύνολα δεδομένων ενσωματώνοντας επιπλέον κομμάτια πληροφοριών με μεγάλα οφέλη για την πιθανή εκπαίδευση των τοπικών μοντέλων ML. Εκτός από αυτό, είμαστε σε θέση να γνωρίζουμε ενεργά την ακριβή τοποθεσία των δεδομένων και την αποδοτική κατανομή των εργασιών εξαλείφοντας την καθυστέρηση στην αντίστοιχη διαδικασία λήψης αποφάσεων.

ΚΕΦΑΛΑΙΟ 4 ΕΞΕΤΑΖΟΜΕΝΟ ΣΕΝΑΡΙΟ

Υποθέτουμε ένα σενάριο όπου γίνεται χρήση Federated Learning, με σκοπό τη συγκέντρωση ενός υποσύνολο μοντέλων Μηχανικής Μάθησης που παράγονται από κατανεμημένους κόμβους. Ο κόμβος που συλλέγει τα κοινόχρηστα σχήματα ML αναλαμβάνει την ευθύνη για τον σχηματισμό του γενικού ML μοντέλου που απεικονίζει τη γνώση των ατομικών κόμβων με βάση τα τοπικά τους δεδομένα. Είναι προφανές ότι η διαδικασία εκπαίδευσης δεν εκτίθεται στον κεντρικό κόμβο και δεν αποκαλύπτει πρόσβαση/πρότυπα/ενδιαφέροντα πάνω στα τοπικά δεδομένα. Οι κόμβοι βασίζονται σε μια μέθοδο επιλεκτικότητας δεδομένων, δηλαδή ένα φίλτρο που παραδίδεται με τα εισερχόμενα αιτήματα για να απορρίπτει νέα δεδομένα που δεν ταιριάζουν με αυτά, εξοικονομώντας έτσι τοπικούς πόρους. Με αυτόν τον τρόπο, οι κόμβοι γίνονται 'ειδικοί' σε δεδομένα με συγκεκριμένη στατιστική περιγραφή και είναι αποτελεσματικοί ως προς την αποθήκευση, πλήρως προσαρμοσμένοι στις ανάγκες των εφαρμογών. Σε αυτήν την περίπτωση, όπου κάθε ML μοντέλο προέρχεται από ένα συγκεκριμένο σύνολο δεδομένων, οι πληροφορίες για τα τοπικά φίλτρα δεδομένων θα μπορούσαν να είναι χρήσιμες για τον κεντρικό κόμβο κατά την εκτέλεση της συγκέντρωσης των κοινόχρηστων τοπικών σχημάτων. Φυσικά, ο κεντρικός κόμβος μπορεί να βασιστεί σε ένα πιο πολύπλοκο σχήμα, όπου υποσύνολα ML μοντέλων μπορούν να συνδυαστούν δυναμικά καθώς τα σύνολα δεδομένων εξελίσσονται με τον χρόνο. Στο παραπάνω σενάριο, μπορούμε εύκολα να αναγνωρίσουμε την ανάγκη για τον διαμοιρασμό των νεοφερμένων δεδομένων όταν δεν ταιριάζουν με το σχήμα επιλεκτικότητας τοπικών δεδομένων (δηλαδή το φίλτρο δεδομένων).

Η πρότασή μας είναι τα δεδομένα να διατηρηθούν στο οικοσύστημα EC όσο το δυνατόν περισσότερο, προκειμένου να ελαχιστοποιήσουμε την καθυστέρηση στην παροχή απαντήσεων. Υποστηρίζουμε ότι η διατήρηση των τοπικά απορριφθέντων δεδομένων στο οικοσύστημα EC, δηλαδή σε ομότιμους που κατέχουν ένα παρόμοιο σύνολο δεδομένων, επιβάλλει λιγότερη καθυστέρηση σε σύγκριση με την εξάρτηση από το Cloud back end λόγω της σύντομης απόστασης με τους συνομήλικους κόμβους. Στην εργασία αυτή, αντιμετωπίζουμε την πρόκληση της επιλεκτικότητας δεδομένων μαθαίνοντας τον ομότιμο κόμβο που είναι ο “host” παρόμοιων συνόλων δεδομένων με τα τοπικά απορριφθέντα δεδομένα. Τα μοτίβα δεδομένων, στη μορφή

των υλοποιημένων φίλτρων δεδομένων, κοινοποιούνται στο δίκτυο, βοηθώντας στον καθορισμό των κατάλληλων τοποθεσιών όπου πρέπει να διατεθούν τα δεδομένα. Η καθιέρωση μιας αποτελεσματικής μεθοδολογίας μετανάστευσης δεδομένων (στατικής ή δυναμικής) αποτελεί το κλειδί των ποιοτικών εφαρμογών στο Edge [56]. Στη προσέγγισή μας, εφαρμόζουμε ένα στατιστικό μοντέλο μάθησης για τον εντοπισμό των πλέον κατάλληλων ομότιμων κόμβων για κάθε εισερχόμενο πακέτο δεδομένων και εφαρμόζουμε ένα adapted weighted KNN σχήμα [57] για να αποφασιστεί η τελική λίστα των ομότιμων κόμβων που επιλέγονται για να φιλοξενήσουν τα τοπικά απορριφθέντα δεδομένα. Εκτός από αυτό, προτείνουμε ένα δεύτερο μοντέλο, δηλαδή μια ευρητική προσέγγιση για τον εντοπισμό ομότιμων κόμβων που υποστηρίζουν ένα φίλτρο δεδομένων που είναι στατιστικά παρόμοιο με τα απορριφθέντα δεδομένα. Οι τεχνικές συνεισφορές που γίνονται στην εργασία είναι:

- Ένα σχήμα για τη διαχείριση των κοινόχρηστων φίλτρων δεδομένων.
- Ένα μοντέλο για τον εντοπισμό της στατιστικής αντιστοιχίας μεταξύ των απορριφθέντων δεδομένων και των κοινόχρηστων τοπικών φίλτρων δεδομένων
- Ένα ML μοντέλο για τον εντοπισμό των καλύτερων αντιστοιχιών μεταξύ των απορριφθέντων δεδομένων και των φίλτρων δεδομένων που κοινοποιούνται από ομότιμους κόμβους.
- Ένα ευρετικό μοντέλο για την αποκάλυψη της στατιστικής σχέσης μεταξύ των τοπικά απορριφθέντων δεδομένων και των κοινόχρηστων φίλτρων δεδομένων.
- Μια σφαιρική αξιολόγηση των μηχανισμών μας έναντι συγκεκριμένων βάσεων, παρουσιάζοντας τα οφέλη και την εφαρμογή του.

ΚΕΦΑΛΑΙΟ 5 ΠΕΡΙΓΡΑΦΗ ΠΡΟΣΕΓΓΙΣΗΣ

5.1 ΦΙΛΤΡΑ ΚΑΙ ΡΟΗ ΕΝΕΡΓΕΙΩΝ

Στον **Πίνακα 1** φαίνεται η επεξήγηση των γραφικών συμβόλων που χρησιμοποιούνται στην εργασία. Λαμβάνουμε υπόψη ένα περιβάλλον PEC με ένα σύνολο από N κόμβους Edge Computing. Οι κόμβοι είναι άμεσα συνδεδεμένοι με ένα σύνολο συσκευών IoT, ενώ οι άμεσες αλληλεπιδράσεις μεταξύ τους ενισχύουν την υποστήριξη της συνεργατικής επίλυσης προβλημάτων. Οι κόμβοι μπορούν να λαμβάνουν και να αποθηκεύουν τα δεδομένα που συλλέγονται από τις συσκευές IoT και να εκτελούν τοπικές εργασίες επεξεργασίας (π.χ., προβλέψεις από ανάλυση δεδομένων) για την υποστήριξη των εφαρμογών των χρηστών. Δεδομένου ότι οι υπολογιστικές και αποθηκευτικές δυνατότητες των κόμβων δεν είναι υπερβολικά επαρκείς (είναι πολύ μικρότερες από τους πόρους του Cloud), μαθαίνουν να επιλέγουν τα δεδομένα που θα φιλοξενήσουν, ενώ τα υπόλοιπα μπορούν να μεταφερθούν σε άλλους κόμβους ή σε κέντρα Cloud/δεδομένων. Αυτή η συμπεριφορά επιλεκτικότητας δεδομένων επηρεάζεται από εισερχόμενες εργασίες που ζητούν επεξεργασία σε τοπικά αποθηκευμένα δεδομένα. Σε τέτοια περιβάλλοντα, οι εργασίες περιλαμβάνουν συνήθως την εξαγωγή γνώσης (π.χ., data mining, εντοπισμός outliers), προβλέψεις αναλυτικών (π.χ., εκπαίδευση μοντέλων ML), λίστες top-k αντικειμένων από ροές δεδομένων για προτάσεις [4], [29], [40], [58], και κατανεμημένες εργασίες μάθησης (Federated Learning [59], [60]). Οι στόχοι της επιλογής δεδομένων είναι να κρατούν (σχετικά) δεδομένα όσο το δυνατόν πιο κοντά στους τελικούς χρήστες διανέμοντας τα σε κόμβους PEC. Το πλεονέκτημα είναι η υποστήριξη γρήγορης επεξεργασίας εργασιών, χαμηλής καθυστέρησης και μειωμένου εύρους ζώνης [61].

Αναμένεται ότι η επιλεκτικότητα των δεδομένων θα επηρεαστεί από συγκεκριμένους περιορισμούς των εισερχόμενων αιτημάτων που υποδεικνύουν τη ζήτηση για συγκεκριμένες εργασίες επεξεργασίας πάνω στα τοπικά σύνολα δεδομένων των κόμβων. Αν μια εργασία απαιτεί την επεξεργασία δεδομένων που δεν είναι διαθέσιμα τοπικά, ο κόμβος θα πρέπει είτε να εκτελέσει αυτήν την εργασία είτε σε έναν ή περισσότερους ομότιμους (peer) κόμβους που διαθέτουν τα κατάλληλα δεδομένα, είτε στο Cloud, είτε ακόμη και να εκτελέσει τη μεταφορά δεδομένων από

άλλες πηγές για να καλύψει τις ανάγκες του αιτήματος. Στη περίπτωση που εξετάζουμε, θεωρούμε ότι όλοι οι κόμβοι διαθέτουν τα τοπικά τους φίλτρα δεδομένων μετά την επεξεργασία των εισερχόμενων αιτημάτων. Ο στόχος των φίλτρων δεδομένων είναι να ορίσουν συνθήκες για την επιλεκτικότητα των δεδομένων και να απορρίπτουν δεδομένα που βρίσκονται πέρα από αυτές τις συνθήκες. Εάν τα φίλτρα δεδομένων είναι ενεργά, οι κόμβοι θα πρέπει να μεταφέρουν τα εισερχόμενα δεδομένα που αποκλίνουν από το τοπικό φίλτρο, δηλαδή να μεταναστεύσουν δεδομένα στον πιο κατάλληλο κόμβο. Όταν ένα εισερχόμενα δεδομένα παραβιάζουν το τοπικό φίλτρο, τότε οι κόμβοι, κατά σειρά, ελέγχουν εάν τα δεδομένα μπορούν να μεταναστεύσουν σε (i) ομότιμους κόμβους που διαθέτουν παρόμοια σύνολα δεδομένων, ή (ii) στο Cloud σε περίπτωση έλλειψης σχετικού ομότιμου κόμβου. Με αυτόν τον τρόπο, αναβάλλουμε τη μεταφορά των δεδομένων στο Cloud, κρατώντας τα κοντά στους τελικούς χρήστες και δημιουργούμε έναν χάρτη δεδομένων του οικοσυστήματος γνωρίζοντας πού αποθηκεύεται κάθε δεδομένο.

Τα συλλεγμένα δεδομένα είναι πολυδιάστατοι διανυσματικοί χώροι, δηλαδή, $x = [x_1, x_2, \ldots, x_m] \in \mathbb{R}^M$ όπου M ο αριθμός των διαστάσεων. Στα κατανεμημένα σύνολα δεδομένων, θα εκτελεστεί ένα σύνολο εργασιών που αναφέρονται σε ροές. Λαμβάνουμε υπόψη ότι

$$T = \{[y_j^l, y_j^h]\} \tag{1}$$

Για παράδειγμα, ένα αίτημα εργασίας, T = [5,18] με $y_j^l = 5, y_j^h = 18$, απαιτεί τιμές δεδομένων από την παραπάνω επιλογή εύρους για συγκεκριμένη επεξεργασία, π.χ., επιλογή δεδομένων για εκπαίδευση μοντέλου μηχανικής μάθησης. Στη περιγραφή που δόθηκε, η επικέντρωση ήταν στη j_{th} διάσταση. Η ροή του T παρουσιάζει ένα σύνολο εύρους τιμών για κάθε αίτηση. Επομένως, το τοπικό φίλτρο δεδομένων πρέπει να τα 'συγκεντρώνει' στο τελικό διάστημα εύρους για την j-οστή διάσταση, δηλαδή $[y_j^{\sim l}, y_j^{\sim h}]$. Εμφανώς, κάθε διάνυσμα εισερχόμενων δεδομένων x πρέπει να συγκριθεί με το $[y_j^{\sim l}, y_j^{\sim h}]$ για να αποφασιστεί εάν θα διατηρηθεί τοπικά ή όχι. Το τελικό φίλτρο δεδομένων για τον i-οστό κόμβο υπολογίζεται από

$$\widehat{T_\iota} = \{[y_j^{\sim l}, y_j^{\sim h}]\}, j = 1,2, \ldots, M \tag{2}$$

καθορίζοντας τις συνθήκες ταιριάσματος για όλες τις πιθανές διαστάσεις. Αυτά τα φίλτρα $\hat{T}$ κατανέμονται στο οικοσύστημα PEC σε προκαθορισμένα διαστήματα για

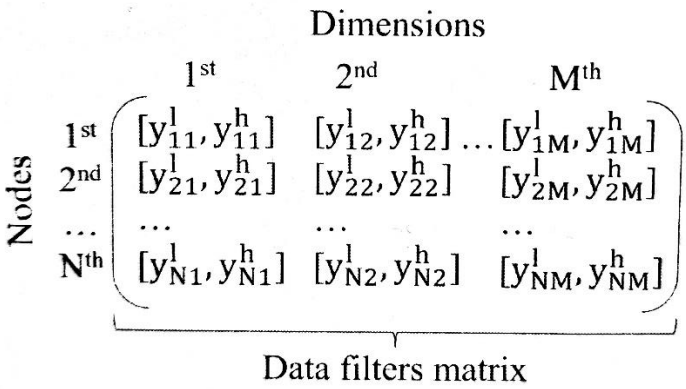
να κρατούν τους ομότιμους κόμβους ενημερωμένους σχετικά με τα 'στατιστικά' των τοπικών συνόλων δεδομένων και να υποστηρίξουν τον μηχανισμό λειτουργίας. Υπάρχει μια φυσική αβεβαιότητα μεταξύ δύο διαδοχικών χρονικών στιγμών αναφοράς σχετικά με τις αλλαγές στα φίλτρα δεδομένων.

Εάν επικεντρωθούμε στον κόμβο n_i, μπορούμε εύκολα να διακρίνουμε το σύνολο των ληφθέντων φίλτρων δεδομένων που τοποθετούνται σε μια ειδική δομή δεδομένων. Στην Εικόνα 1 παρουσιάζεται ένα παράδειγμα των δομών δεδομένων που χρησιμοποιούνται για την αποθήκευση των εισερχομένων φίλτρων δεδομένων. Χωρίς απώλεια γενικότητας και για να διευκολύνουμε τη σημειογραφία μας, θεωρούμε ότι ο n_i αποθηκεύει N φίλτρα δεδομένων (αντί για N−1). Όταν ένα νέο δεδομένο x φτάνει στο n_i, ο κόμβος πρέπει να ελέγξει το ταίριασμα με το τοπικό φίλτρο δεδομένων $\widehat{T_i}$ και να ακολουθήσει μία από τις ακόλουθες ενέργειες/αποφάσεις:

- **Ενέργεια 1:** Αν το x ταιριάζει με το $\widehat{T_i}$, το x αποθηκεύεται στο τοπικό σύνολο δεδομένων.
- **Ενέργεια 2:** Αν το x δεν ταιριάζει με το $\widehat{T_i}$, n_i πρέπει να επιλέξει τον καλύτερο ομότιμο κόμβο για να προωθήσει το x. Αυτή η διαδικασία γίνεται μέσω των φίλτρων δεδομένων που υπάρχουν στο $\widehat{T}$ με τη χρήση μηχανισμών ελέγχου ομοιότητας.
- **Ενέργεια 3:** Αν το x δεν ταιριάζει με το $\widehat{T_i}$, ενώ δεν έχει βρεθεί κάποιος κατάλληλος ομότιμος κόμβος για την αποθήκευση του x, το x μεταφέρεται στο Cloud back end.

Εμφανώς, η **Ενέργεια 2** λαμβάνεται υπό την **Ενέργεια 1** και η **Ενέργεια 3** λαμβάνεται υπό την **Ενέργεια 2**, δημιουργώντας ένα σειριακό μοντέλο λήψης αποφάσεων που προσπαθεί να 'αναβάλει' τη μεταφορά του x στο Cloud εντοπίζοντας κόμβους στο οικοσύστημα PEC με παρόμοια δεδομένα που μπορούν να το φιλοξενήσουν. Η βασική ερώτηση είναι πώς μπορούμε να ορίσουμε το ταίριασμα μεταξύ του x και των διαθέσιμων φίλτρων δεδομένων για να υποστηρίξουμε μια τέτοια λήψη αποφάσεων. Στην εργασία αυτή, προτείνουμε τη χρήση ενός έξυπνου σχήματος που συνδυάζει μια στατιστική προσέγγιση με ένα μοντέλο μηχανικής μάθησης με σκοπό την παραγωγή της κατάλληλης ενέργειας για τα εισερχόμενα δεδομένα. Ορίζουμε την έννοια του Δείκτη Ταιριάσματος Φίλτρου (Filter Matching

Indicator - FMI) αναφερόμενο ως φ_i, που αντιπροσωπεύει εάν το x ταιριάζει με τα ατομικά φίλτρα του i-οστού κόμβου για ένα υψηλό αριθμό διαστάσεων. Πρακτικά, ορίζουμε ένα κατώφλι θ πάνω από το οποίο το φ_i θεωρείται αποδεκτό και ο αντίστοιχος κόμβος συμμετέχει στη λογική μας, που επιλέγει τη λίστα των κόμβων όπου το x πρέπει να εκχωρηθεί. Το κύριο πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορούμε ενεργητικά να γνωρίζουμε τη φυσική τοποθεσία που βρίσκονται τα ζητούμενα δεδομένα, έτσι, μπορεί να παρέχεται αποτελεσματικός προγραμματισμός των κατανομών εργασιών σε όλους τους κόμβους.



Εικόνα 1. Παράδειγμα δομής των δεδομένων οπού τα συλλεγόμενα φίλτρα δεδομένων είναι αποθηκευμένα.

Πίνακας 1. Ονοματολογία

Σύμβολο	**Περιγραφή**
N	Αριθμός κόμβων
N	Σύνολο κόμβων
n_i	Ο i_{th} κόμβος
x	Πολυδιάστατο διάνυσμα δεδομένων
M	Αριθμός διαστάσεων στο πολυδιάστατο διάνυσμα
T	Το ερώτημα προς επεξεργασία
$[y_j^l, y_j^h]$	Ζητούμενο διάστημα για επεξεργασία
$y_j^{\sim l}$	Το χαμηλότερο όριο στο τελευταίο φίλτρο
$y_j^{\sim h}$	Το υψηλότερο όριο στο τελευταίο φίλτρο
$\widehat{T_i}$	Το τελευταίο τοπικό φίλτρο δεδομένων για τον i_{th} κόμβο
φ_i	Ο δείκτης ταιριάσματος για τον i_{th} κόμβο
ϑ	Κατώφλι ταιριάσματος για τα διανύσματα δεδομένων

5.2 ΕΠΙΛΕΚΤΙΚΟΤΗΤΑ ΔΕΔΟΜΕΩΝ

Σε αυτό το υπο κεφάλαιο, παρέχεται μια συζήτηση για την προτεινόμενη προσέγγιση παρουσιάζοντας τα μέρη του μηχανισμού λήψης αποφάσεων που δημιουργήσαμε. Για κάθε μέρος, αναλύουμε το μοντέλο και παρουσιάζουμε τις αντίστοιχες διατυπώσεις.

5.2.α ΔΙΑΧΕΙΡΗΣΗ ΦΙΛΤΡΩΝ ΔΕΔΟΜΕΝΩΝ

Η υιοθέτηση του $\hat{T}$ μας δίνει τη δυνατότητα να εντοπίσουμε εάν τα εισερχόμενα δεδομένα που αναφέρονται από το xx ταιριάζουν με τα καταγεγραμμένα φίλτρα δεδομένων. Αξίζει να σημειωθεί ότι τα δεδομένα και τα φίλτρα είναι πολυδιάστατα διανύσματα, και επομένως, πρέπει να προσδιοριστεί εάν η αντιστοίχιση πραγματοποιείται για τις υιοθετημένες διαστάσεις. Για παράδειγμα, δεν μπορούμε να βασιστούμε σε μια αντιστοίχιση που αναφέρεται σε ένα περιορισμένο αριθμό διαστάσεων (π.χ., λιγότερες από θ). Θεωρούμε ότι το φ_i, δηλαδή το φίλτρο δεδομένων για τον i-οστό κόμβο, είναι η συγκέντρωση ενός συνόλου δεικτών που εκθέτουν εάν έχουμε αντιστοίχιση μεταξύ του x και του $\hat{T}$. Ορίζουμε την ακόλουθη συνάρτηση δείκτη:

$$I(x_j) = \begin{cases} 1, & Αν\ \tilde{y_j}^l \leq x_j \leq \tilde{y_j}^h \\ 0, & Διαφορετικά \end{cases} \tag{3}$$

Το $I(x_i)$ υποδηλώνει εάν υπάρχει αντιστοιχία μεταξύ των εισερχόμενων δεδομένων και του i-οστού φίλτρου για τη διάσταση j. Η συγκέντρωση των M δεικτών θα οδηγήσει στο τελικό φ_i, $\varphi_i = \sum_{j=1}^{M} I(x_j)$. Κατά την εξέταση της αντιστοιχίας του x με το $\hat{T}_i$, πρέπει να παρατηρήσουμε έναν αυξημένο αριθμό ατομικών αντιστοιχίσεων, δηλαδή $\varphi_i \geq \theta$. Το θ, επιλέγεται στρατηγικά για να απεικονίσει την πρόθεση ενσωμάτωσης του βαθμού αντιστοιχίας που θα μπορούσε να οδηγήσει στην υιοθέτηση του i-οστού κόμβου ως υποψήφιο για τη φιλοξενία του x.

Θεωρούμε ότι η διαδικασία αντιστοίχισης για μια μεμονωμένη διάσταση είναι το αποτέλεσμα μιας δοκιμής Bernoulli με πιθανότητα επιτυχίας p_j. Βάσει αυτής της παρατήρησης, η τυχαία μεταβλητή Φ_i καθορίζεται από τη φ_i, ακολουθεί μια Binomial κατανομή όπου πραγματοποιούνται M δοκιμές Bernoulli. Αξίζει να σημειωθεί ότι οιδιαστάσεις που υιοθετούνται είναι ανεξάρτητες μεταξύ τους και ακολουθούν μια συγκεκριμένη κατανομή πιθανοτήτων. Επιπλέον, διακρίνουμε δύο διαφορετικές περιπτώσεις που αφορούν τις δοκιμές Bernoulli που υποδεικνύουν την παρουσία μιας αντιστοίχισης στις υιοθετημένες διαστάσεις.

- Περίπτωση A: Οι προαναφερθείσες δοκιμές Bernoulli έχουν την ίδια πιθανότητα επιτυχίας, δηλαδή $p_j = \mathrm{p}, \forall\, \mathrm{j}$.

- Περίπτωση Β: Οι υιοθετημένες δοκιμές Bernoulli έχουν διαφορετικές πιθανότητες επιτυχίας p_j.

Εμφανώς, όταν το $I(x_j)$ είναι μια διαδικασία Bernoulli, μπορούμε εύκολα να διατυπώσουμε την κατανομή του Φ_i , δηλαδή $\Phi_i \sim B(M,p)$ (για απλότητα, χρησιμοποιούμε το p για να σημειώσουμε την πιθανότητα επιτυχίας στην Περίπτωση Α). Η συνάρτηση μάζας πιθανοτήτων (pmf) του Φ_i δίνεται από τη συνάρτηση

$$f(k) = \binom{M}{k} p^k (1-p)^{M-k} \tag{4}$$

και αντιπροσωπεύει την πιθανότητα να έχουμε ακριβώς k αντιστοιχίσεις εκ των M διαστάσεων. Επιπλέον, η αθροιστική συνάρτηση κατανομής (cdf) δίνεται από τη συνάρτηση

$$F(k) = \sum_{r=0}^{k} \binom{M}{k} p^k (1-p)^{M-k} \tag{5}$$

Στη περίπτωση οπού οι δείκτες ατομικού ταιριάσματος έχουν διαφορετικές πιθανότητες επιτυχίας p_j (Περίπτωση Β), το Φ_i ακολουθεί τη διωνυμική κατανομή Poisson. Στη περίπτωση αυτή, η pmf και η cdf δίνονται από τις ακόλουθες συναρτήσεις:

$$f(k) = \sum_{A \in G_\upsilon} \prod_{r \in A} p_r \prod_{t \in A^c} (1-p_l) \tag{6}$$

$$f(k) = \sum_{\upsilon=0}^{k} \sum_{A \in G_\upsilon} \prod_{r \in A} p_r \prod_{t \in A^c} (1-p_l) \tag{7}$$

Οπού το G_υ είναι το σύνολο των υποσυνόλων των υ ακαίρων που μπορούν να επιλεχθούν από $\{1,2,\dots,M\}$.

5.2.β ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

Όπως αναφέρεται το $I(x_j)$ υποδηλώνει αν το x_j ταιριάζει με το φίλτρο της j-οστής διάστασης $[y_j^{\sim l}, y_j^{\sim h}]$. Η αναμενόμενη τιμή της I() παρέχεται από τη συνάρτηση:

$$E(I) = \int_{y_j^{\sim l}}^{y_j^{\sim \eta}} fx_j(x_j)\, dx_j \tag{8}$$

Δεδομένου ότι το x_j είναι η πραγματοποίηση της τυχαίας μεταβλητής X, η οποία ακολουθεί πιθανοτική κατανομή με συνάρτηση πυκνότητας πιθανότητας (pdf) $fx_j(x)$ και cdf $Fx_j(x)$. Γενικότερα μπορούμε να λάβουμε την αναμενόμενη τιμή της $I(x_j)$ μέσω της υιοθέτησης της παρακάτω συνάρτησης:

$$E(I) = Fx_j(y_j^{\sim h}) - Fx_j(y_j^{\sim l}) \tag{9}$$

Πρόταση 1. Η αναμενόμενη τιμή του $I(x_j)$ για μια μεμονωμένη διάσταση των εισερχόμενων δεδομένων, που παρουσιάζονται από το x, δίνεται από:

- $\frac{\tilde{y}_j^{l} - \tilde{y}_j^{h}}{b-a}$, αν $x_j \sim \cup(a,b)$;
- $e^{-\lambda \tilde{y}_j^{l}} - e^{-\lambda \tilde{y}_j^{h}}$, αν $x_j \sim Exp(\lambda)$;
- $\frac{1}{2}\left(\mathrm{erf}\left(\frac{\tilde{y}_j^{h}-\mu}{\sigma\sqrt{2}}\right) - \mathrm{erf}\left(\frac{\tilde{y}_j^{l}-\mu}{\sigma\sqrt{2}}\right)\right)$, αν $x_j \sim N(\mu,\sigma)$.

Απόδειξη. Η απόδειξη μπορεί να ανακτηθεί εύκολα αν εφαρμόσουμε τη συνάρτηση cdf για κάθε κατανομή στην συνάρτηση E(I).

Ας επικεντρωθούμε στην Περίπτωση Α, όπου όλες οι όμοιες δραστηριότητες για τις σχεδιαζόμενες διαστάσεις εκδηλώνουν την ίδια πιθανότητα επιτυχίας p. Η αποδοχή ενός υποψηφίου κόμβου πρέπει να βασίζεται στο εάν το φίλτρο του ταιριάζει με το x για έναν αριθμό διαστάσεων πάνω από θ. Με άλλα λόγια, πρέπει να παρατηρήσουμε τουλάχιστον θ επιτυχίες από το M για να θεωρήσουμε τον συγκεκριμένο κόμβο ως υποψήφιο για το x. Σε αυτήν την περίπτωση, η πιθανότητα του Φ_i να είναι πάνω από θ δίνεται από την παρακάτω εξίσωση:

$$P(\Phi_i \geq \vartheta) = \sum_{k=\vartheta}^{M} \binom{M}{k} p^k (1-p)^{M-k} \qquad (10)$$

Αν επικεντρωθούμε στη Περίπτωση Β, οπού το Φ_i ακολουθεί τη διωνυμική κατανομή Poisson, μπορούμε να λάβουμε τη πιθανότητα τουλάχιστον θ ταιριαστών ενεργειών, εκ των M, όπως φαίνεται στην εξίσωση:

$$P(\Phi_i \geq \vartheta) = 1 - \sum_{\upsilon=0}^{\theta} \sum_{A \in G_\upsilon} \prod_{r \in A} p_r \prod_{l \in A^c} (1-\pi_\lambda) \qquad (11)$$

ΚΕΦΑΛΑΙΟ 6 ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ ΚΑΙ ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ

Σε αυτό το κεφάλαιο, γίνεται ανάλυση των προτεινόμενων μοντέλων που έχουν υιοθετηθεί για την υποστήριξη της λήψης αποφάσεων σχετικά με την κατανομή του x σε περίπτωση που δεν αυτό δεν ταιριάζει με το τοπικό φίλτρο δεδομένων. Τα μοντέλα που δημιουργήθηκαν προσπαθούν να ανιχνεύσουν τον πιο όμοιο κόμβο πριν το x οδηγηθεί στην τελική του θέση. Η συζητούμενη λήψη αποφάσεων βασίζεται στη διαχείριση του πίνακα φίλτρων δεδομένων $\tilde{T}$ και των ατομικών φίλτρων $[y_j^l, y_j^h]$ για κάθε διάσταση j. Προτείνουμε τη χρήση δύο μοντέλων, (i) Κατανομή δεδομένων με στατιστικές μεθόδους και (ii) Κατανομή δεδομένων βασισμένη στη μάθηση.

6.1 ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ

Αρχικά, βασιζόμαστε σε μια μέθοδο μη παραμετρικής επίβλεψης, δηλαδή το μοντέλο weighted KNN classifier [6]. Όπως ήδη αναφέρθηκε, η πληροφορία που παρέχεται από τον $\tilde{T}$ αναφέρεται στα φίλτρα δεδομένων που συλλέγονται από τους κόμβους, δηλαδή $\{[y_j^l, y_j^h]\}$, $j = 1,2, \dots, M$. Επομένως, πρέπει να συγκρίνουμε το x_j με τα αντίστοιχα διαστήματα $[y_j^l, y_j^h]$ και να υπολογίσουμε την απόσταση του x με το $\tilde{T}_\iota$ εξάγοντας τους k Κοντινότερους Γείτονες (k-NN) για την κατανομή των εισερχομένων δεδομένων. Δεδομένης μιας νόρμας $||.||$ στο R^M και το x, πρέπει να ταξινομήσουμε τα $\tilde{T}_\iota, i = 1,2, \dots, N$ έτσι ώστε $\left||\tilde{T}_1 - x|\right| \leq \left||\tilde{T}_2 - x|\right| \leq \cdots \leq ||\widetilde{T_N} - x||$ επιλέγοντας τους πρώτους k ομότιμους κόμβους για το x.

Στο Μοντέλο Α εισάγουμε μια συνάρτηση εξομάλυνσης που λαμβάνει υπόψιν της τη θέση του x_j σε κάθε αντίστοιχο διάστημα. Για να επιτύχουμε τη λειτουργία αυτή λαμβάνουμε υπόψη μας ότι ακόμα και αν ένας κόμβος είναι στους K κοντινότερους πιθανόν να υπάρχουν κάποιες διαστάσεις του x_j, στις οποίες τα δεδομένα του να είναι εκτός των φίλτρων ή το x_j να παρατηρείται στα όρια της j-οστης διάστασης του φίλτρου. Υποθέτουμε ότι το μέσο των δεδομένων ακολουθεί κανονική κατανομή. Αξίζει να σημειωθεί ότι δεν υποθέτουμε ότι οι καταγεγραμμένες τιμές είναι κανονικά κατανεμημένες αλλά ότι το μέσο κατά μήκος των δειγμάτων, είναι

κανονικά κατανεμημένο. Η υπόθεση αυτή γίνεται, επειδή οι περισσότεροι πληθυσμοί έχουν μέσο μ και τυπική απόκλιση σ, διαλέγοντας από αυτούς του πληθυσμούς μια μεγάλη τιμή δειγμάτων z, θα οδηγήσει στη κανονική τους κατανομή με μέσο μ και τυπική απόκλιση $\frac{\sigma}{\sqrt{z}}$ [62], [63]. Φυσικά, ο μέσος όρος των δειγμάτων σημαίνει ότι ο το μέσο του πληθυσμού θα είναι μ, όμως η τυπική απόκλιση των δειγμάτων θα είναι χαμηλότερη από τη τυπική απόκλισή του πληθυσμού. Η υπόθεση αυτή εφαρμόζεται στα εισερχόμενα αιτήματα, για να θεωρήσουμε ότι κάθε φίλτρο δεδομένων ακολουθεί κανονική κατανομή. Το κύριο στοιχείο της λογικής μας είναι ότι οι εισερχόμενες εργασίες διαμορφώνουν τα φίλτρα δεδομένων με τιμές που δεσμεύονται κοντά στο μέσο $c_j = \frac{\tilde{y_j}^l + \tilde{y_j}^h}{2}$. Αυτή η υπόθεση οδηγεί στο συμπέρασμα ότι οι εισερχόμενες εργασίες, που καθορίζουν τα φίλτρα δεδομένων, ζητούν τιμές που είναι συγκεντρωμένες στο κέντρο του διαστήματος, c_j . Για αυτό, ορίζουμε το βάρος για κάθε ομότιμο κόμβο που συμπεριλαμβάνεται στους κ κοντινότερους στην εξής συνάρτηση:

$$D' = \sum_{j=1}^{M} \frac{1}{e^{\alpha' \cdot |x_j - c_j| + \beta'}} \qquad (12)$$

Όπου α'$\in R^+$ και β'$\in R^-$ οι παράμετροι εξομάλυνσης. Όσο μεγαλύτερη η απόσταση από το κέντρο ενός φίλτρου είναι το x_j, αυξάνεται η διαφορά μεταξύ τους. Οι ατομικές διαφορές (πχ. Η διαφορά για κάθε διάσταση) προστίθενται για να επιφέρουν τη τελική τιμή του D'. Η μέθοδος αυτή γίνεται η βάση για την ταξινόμηση των ομότιμων κόμβων και τη τελική επιλογή του συνόλου για τη αποθήκευση του x.

6.2 ΚΑΤΑΝΟΜΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΜΑΘΗΣΗΣ

Εκτός από το πρώτο μοντέλο, προτείνουμε ένα ακόμα (Μοντέλο Β), το οποίο βελτιώνει τη τελική αποθήκευση του x. Το Μοντέλο Β βελτιώνει τη προαναφερθείσα πρόταση της συνάρτησης εξομάλυνσης και επιλογής των ομότιμων κόμβων. Στο Μοντέλο Β ο κανόνας που υιοθετούμε είναι ένας υπολογισμός απόστασης μεταξύ των συσσωρευμένων διαφορών του $\tilde{T}_i$ και x.

$$D_i = \left\| \tilde{T}_i - x \right\| = \sum_{j=1}^{M} d_j \qquad (13)$$

όπου

$$d_j = \begin{cases} 0, & Αν\ y_j^{\sim l} \leq x_j \leq y_j^{\sim h} \\ \left|x_j - y_j^{\sim l}\right|, & Αν\ x_j < y_j^{\sim l} \\ \left|x_j - y_j^{\sim h}\right|, & Αν\ x_j > y_j^{\sim h} \end{cases} \quad (14)$$

Ταξινομώντας τη λίστα $\{D_i\}_{i=1}^{N}$ σε αύξουσα σειρά, μπορούμε να διακρίνουμε τους Κ κοντινότερους γείτονες του x υποθέτοντας ότι το βάρος για κάθε ομότιμο είναι $w_{n_i} = \frac{1}{k}$. Παρόλα αυτά, προχωρώντας ένα βήμα μπροστά και καθορίζοντας το βάρος κάθε ομότιμου κόμβου να είναι αντίστροφος ανάλογο της σωρευμένης διαφοράς ανάμεσα στο x και το $\tilde{T}_i$, εφαρμόζοντάς τι στη βελτιωμένη συνάρτηση εξομάλυνσης, με σκοπό την απόρριψη κόμβων που παρουσιάζουν μεγάλη διαφορά με το x. Αυτή η προσέγγιση εφαρμόζεται μόνο για τους κ γείτονες $n_{(i)}, i = 1,2,\cdots,k$ που διαδέχθηκαν από τον υπολογισμό συσσωρευμένων διαφορών D_i. Ισχύει η επόμενη συνάρτηση:

$$\widehat{w}_{n_i} = \frac{a}{\max\left\{1, \left|x_j \notin \left[y_j^{\sim l}, y_j^{\sim h}\right]\right|\right\} + e^{\beta \cdot D_i + \gamma}} \quad (15)$$

όπου α,β $\in R^+$ και γ$\in R^-$ οι παράμετροι εξομάλυνσης. Στη παραπάνω συνάρτηση, το $\left|x_j \notin \left[y_j^{\sim l}, y_j^{\sim h}\right]\right|$ αντιπροσωπεύει το μέγεθος της συλλογής που παρουσιάζει τις τιμές του x_j που δεν είναι στο αντίστοιχο φίλτρο δεδομένων. Εμφανώς, το $\left|x_j \notin \left[y_j^{\sim l}, y_j^{\sim h}\right]\right| < \theta$. Ακολουθώντας αυτή τη προσέγγιση, δίνουμε μεγαλύτερη έμφαση στα φίλτρα με τα οποία το το x έχει περισσότερα ταιριάσματα. Τέλος, η λίστα ομότιμων κόμβων ταξινομείται ξανά σε φθίνουσα σειρά με βάση το $\widehat{w}_{n_i}$, και το εισερχόμενο διάνυσμα x δεσμεύεται στους $k' < k$ κοντινότερους κόμβους.

ΚΕΦΑΛΑΙΟ 7 ΠΕΙΡΑΜΑΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

7.1 ΜΕΤΡΙΚΕΣ ΑΠΟΔΟΣΗΣ ΚΑΙ ΣΕΝΑΡΙΟ ΠΡΟΣΟΜΟΙΩΣΗΣ

Τα προτεινόμενα μοντέλα, εφαρμόστηκαν σε ένα περιβάλλον όπου εισερχόμενα δεδομένα πολλαπλών μεταβλητών (M) εισέρχονταν στο δίκτυο με σκοπό τη κατηγοριοποίησή τους. Κάθε μεταβλητή αντιπρόσωπε διάφορα είδη δεδομένων, όπως (υγρασία, θερμοκρασία). Για το σενάριο μας χρησιμοποιήθηκαν διανύσματα εισόδου 6 μεταβλητών (M=6). Για κάθε διάσταση, καθορίζουμε ένα y_j^l, y_j^h, βασισμένοι στη κανονική κατανομή. Ο σκοπός μας είναι να διακρίνουμε εάν τα μοντέλα, μπορούν να ανταποκριθούν και να αποθηκεύσουν τα εισερχόμενα δεδομένα στους πιο ταιριαστούς κόμβους του δικτύου με βάση τα άνω και κάτω όρια των φίλτρων στους κόμβους.

Για να αξιολογήσουμε την απόδοση των μοντέλων, χρησιμοποιούμε μερικές μετρικές που στοχεύουν στην απεικόνιση της επιτυχίας κατηγοριοποίησης των εισερχόμενων πακέτων καθώς και το χρόνο που χρειάζονται τα μοντέλα για να κάνουν τη κατηγοριοποίηση αυτή. Για αυτό οι μετρικές που χρησιμοποιήθηκαν είναι οι εξής:

- **ε:** Η μετρική αυτή χρησιμοποιείται για δείξει την ακρίβεια του μοντέλου, για παράδειγμα $\varepsilon = \frac{|J|}{W}$, όπου $|J|$ είναι το cardinality της συλλογής αιτημάτων που μπορεί να διαχειριστεί από τα τοπικά δεδομένα. Για ε=1, έχουμε τη καλύτερη δυνατή απόδοση.
- **τ:** Το **τ,** επιλέχθηκε για να αναδείξει το χρόνο που χρειάζεται για την εκτέλεση των υπολογισμών και την τελική κατάληξη του εισερχόμενου διανύσματος με τη χρήση των προτεινόμενων μοντέλων. Προφανώς το **τ** πρέπει να είναι όσο το δυνατόν μικρότερο έτσι ώστε, τα μοντέλα να μπορούν να ανταποκριθούν σε πραγματικού χρόνου διαχείριση των εισερχόμενων δεδομένων.
- **ρ:** Με τη μετρική αυτή μετράμε τα εισερχόμενα διανύσματα τα οποία ταιριάζουν απόλυτα με τον κόμβο που τα μοντέλα επέλεξαν να τα ταιριάξουν. Αυτό σημαίνει ότι όλα τα δεδομένα τα οποία περιέχονταν στο

εισερχόμενο διάνυσμα, ήταν μέσα στα όρια που έθεταν τα φίλτρα του κόμβου.

- **q:** Η μετρική αυτή, μετράει όλα τα εισερχόμενα διανύσματα τα οποία κατηγοριοποιήθηκαν σε κόμβο με τον οποίο τα φίλτρα του δεν ταίριαζαν με καμία από τις τιμές του x. Αυτό σημαίνει ότι όλες οι τιμές του εισερχόμενου διανύσματος x ήταν εκτός των y_j^l, y_j^h του εκάστοτέ επιλεγμένου κόμβου.
- **r:** Η μεταβλητή **r** αποτελεί έναν δείκτη που μετρά τη μισή αντιστοίχιση των μοντέλων. Αυτό σημαίνει ότι περιλαμβάνει τα διανύσματα που ταξινομήθηκαν σε κόμβους όπου ο αριθμός των μεταβλητών στο διάνυσμα εισόδου x ήταν μικρότερος από όλους τους κόμβους, αλλά μεγαλύτερος από το ήμισυ $\frac{M}{2} \leq matches < M$, όπου M ο αριθμός των διαστάσεων του x.
- **s:** Αντίστοιχα με το r, το **s**, αντιπροσωπεύει τα εισερχόμενα πακέτα που ήταν λιγότερο από ήμισυ ταιριαστά αλλά όχι καθόλου, με τον κόμβο που τα μοντέλα επέλεξαν να τα ταιριάξουν $0 \leq matches < \frac{M}{2}$.

7.2 ΑΞΙΟΛΟΓΗΣΗ

Για την αξιολόγηση των μοντέλων διεξήχθησαν προσομοιώσεις με τη δημιουργία τυχαίων πακέτων εισόδου, με σκοπό της αναπαράστασης της δυναμικής φύσης του IoT. Αυτή η προσέγγιση επιτρέπει την αποτύπωση της πραγματικής διακύμανσης και της απρόβλεπτης φύσης των σεναρίων στα οποία τα μοντέλα πρέπει να ανταποκρίνονται. Για κάθε μέτρηση των στατιστικών χρησιμοποιήθηκαν 100 απεσταλμένα πακέτα. Κάθε πακέτο που δημιουργούνταν ταίριαζε απόλυτα με τα φίλτρα ενός από τους κόμβους του δικτύου, γεγονός που έγινε έτσι ώστε να μπορούμε να μετρήσουμε με ακρίβεια την αποτελεσματικότητα των μοντέλων. Όπως προαναφέρθηκε τα πακέτα εισόδου υπάγονταν από 6 διαστάσεις, και δοκιμές έγιναν με διάφορες τιμές των:

- θ, $\theta = [3, 4, 5]$
- k, $k = [2, 5, 10]$
- N, N = [5, 10, 20, 50, 100]

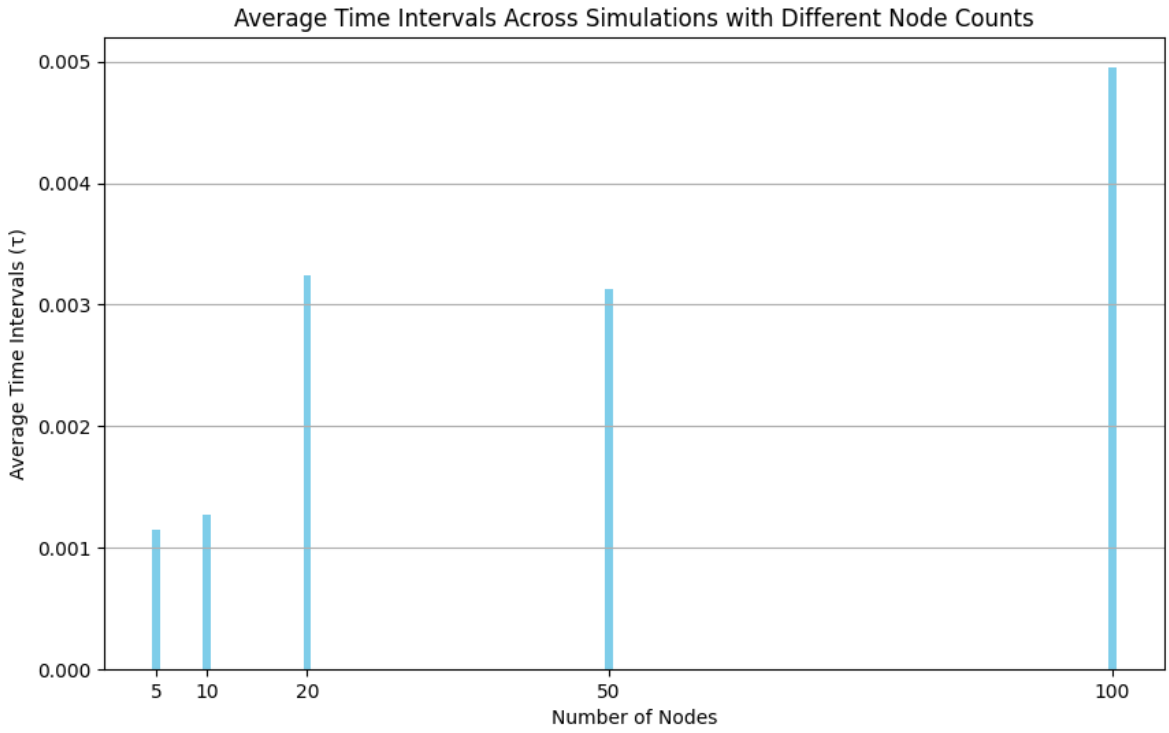
Για τις προσομοιώσεις των διαφορετικών τιμών του θ, τα μοντέλα εξετάστηκαν σε όλες τις τιμές του N,k. Για να το πετύχουμε αυτό οι προσομοιώσεις χωρίστηκαν σε 2 τμήματα.

1. Σταθερό k=3, για την εξέταση των διαφορετικών τιμών του N.
2. Σταθερό N=20, για την εξέταση των διαφορετικών τιμών του k.

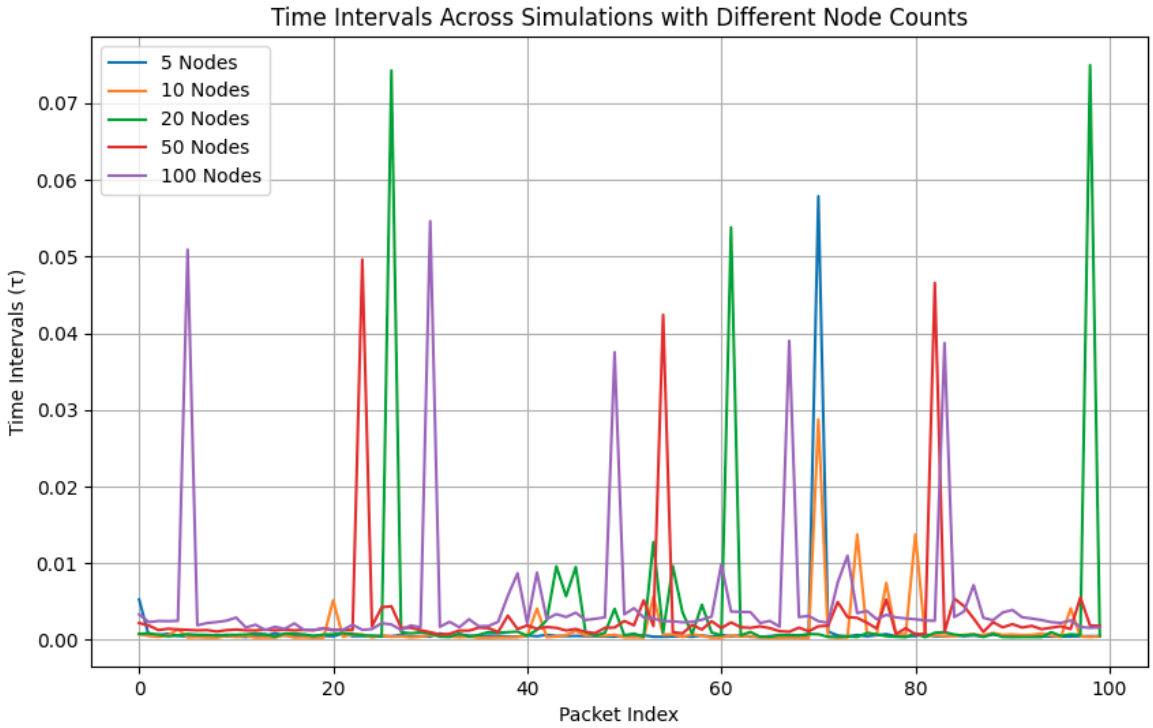
Κρατώντας σταθερό το k για μεταβλητά N, στοχεύουμε στη παρατήρηση της απόδοσης του μοντέλου στην εξαγωγή προτύπων για τα εισερχόμενα δεδομένα ανάλογα με το μέγεθος των κόμβων του δικτύου. Αυτή η ανάλυση μας βοηθά να αναγνωρίσουμε εάν μια μέση επιλεγμένη τιμή του k είναι κατάλληλη για σύνολα δεδομένων διαφορετικών μεγεθών. Με αυτόν τον τρόπο, εξετάζουμε την ευαισθησία του μοντέλου στο μέγεθος του dataset και την ικανότητά του να εξάγει σημαντικές πληροφορίες ανεξάρτητα από τον αριθμό των κοντινότερων γειτόνων. Αυτό βοηθά στην αξιολόγηση των μοντέλων με άξονα τη κλιμακούμενη χρήση. Η διατήρηση σταθερού N, κατά την εξέταση διαφορετικών τιμών k μας επιτρέπει να εξετάσουμε την ευαισθησία των κοντινότερων γειτόνων με ξεκάθαρους άξονες σε ένα ομοιογενές περιβάλλον. Η διαδικασία δοκιμής διάφορων τιμών του k επιτρέπει την επιλογή βέλτιστου k, προσφέροντας μια μέθοδο για τον προσδιορισμό του ιδανικού αριθμού γειτόνων που επιτυγχάνει την καλύτερη ισορροπία μεταξύ προκατάληψης και διακύμανσης (Bias-Variance Tradeoff) [64] στο μοντέλο.

7.2.α ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=3

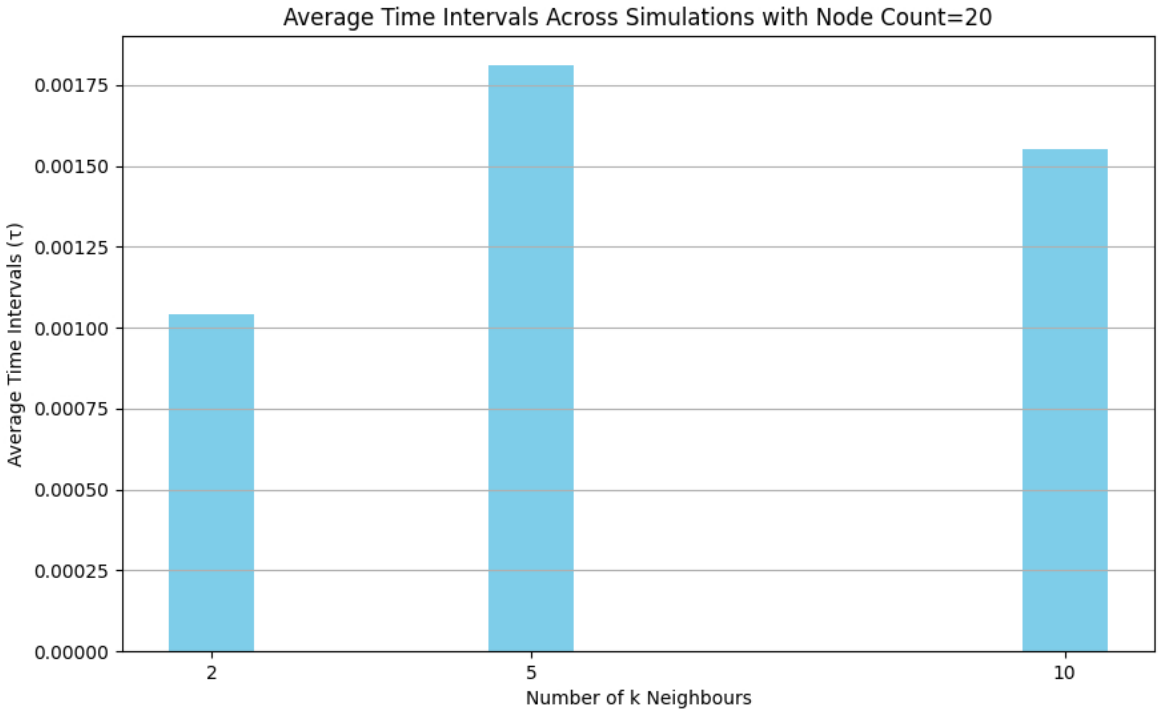
Με την επιλογή μια χαμηλής τιμής ταιριάσματος (θ) στα μοντέλα, μπορούμε να αναμένουμε ορισμένα πρότυπα στις μετρικές μας. Αρχικά ο χρόνος τ, ενδέχεται να μειωθεί καθώς τα μοντέλα επεξεργάζονται λιγότερα δεδομένα λόγω της πιο αυστηρής διαλογής. Αυτή η ρύθμιση αναμένεται να ενισχύσει τη συνολική αποδοτικότητα, ιδίως σε σενάρια που απαιτούν διαχείριση δεδομένων πραγματικού χρόνου. Ωστόσο αντίστοιχα είναι πολύ πιθανό να μειωθούν τα τέλεια ταιριάσματα ρ, καθώς το επίπεδο ομοιότητας των εισερχόμενων δεδομένων πρέπει να είναι αρκετά μεγάλο για να επιτευχθεί το ταίριασμα. Επιπλέον, και οι δύο μετρικές κατηγορίας (q) και μερικών ταιριασμάτων (r και s) πιθανόν να αυξηθούν ως αποτέλεσμα της μείωσης του ρ. Σε γενικές γραμμές, ένα χαμηλό κατώφλι ταιριάσματος υπογραμμίζει την ακρίβεια και την ειδικότητα, με πιθανές βελτιστοποιήσεις στην αποδοτικότητα.



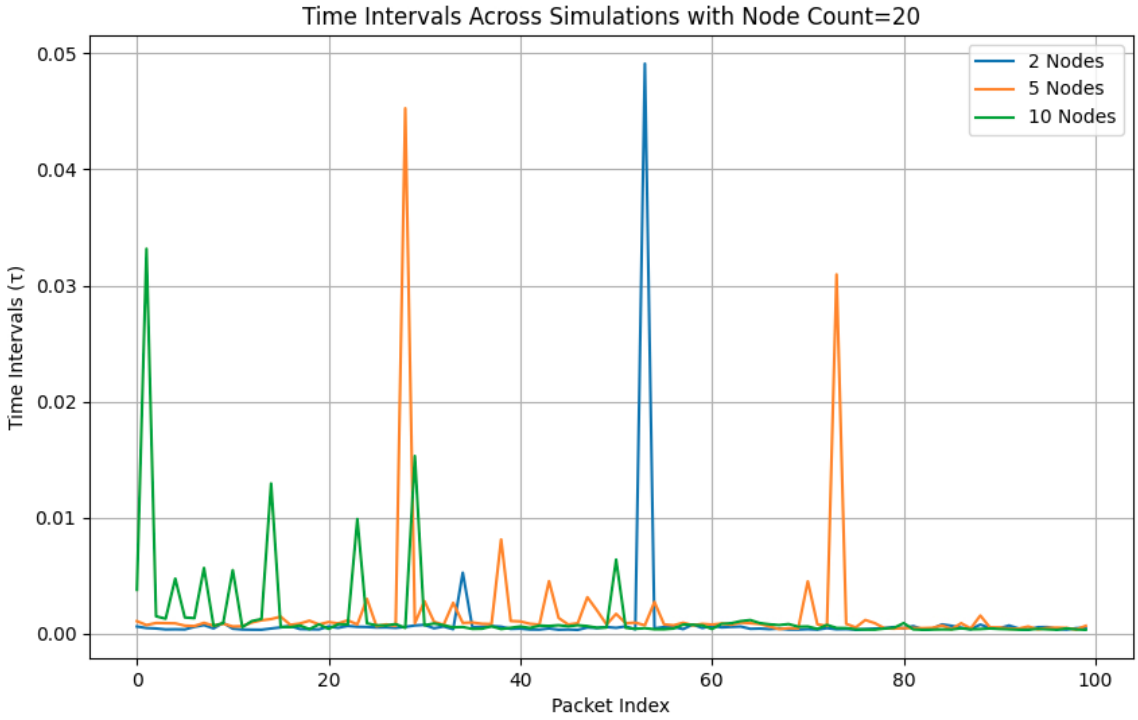
Εικόνα 2. Μέσος χρόνος διαχείρισης πακέτου με k=3 και μεταβλητό N για το πρώτο μοντέλο.



Εικόνα 3. Χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το πρώτο μοντέλο.



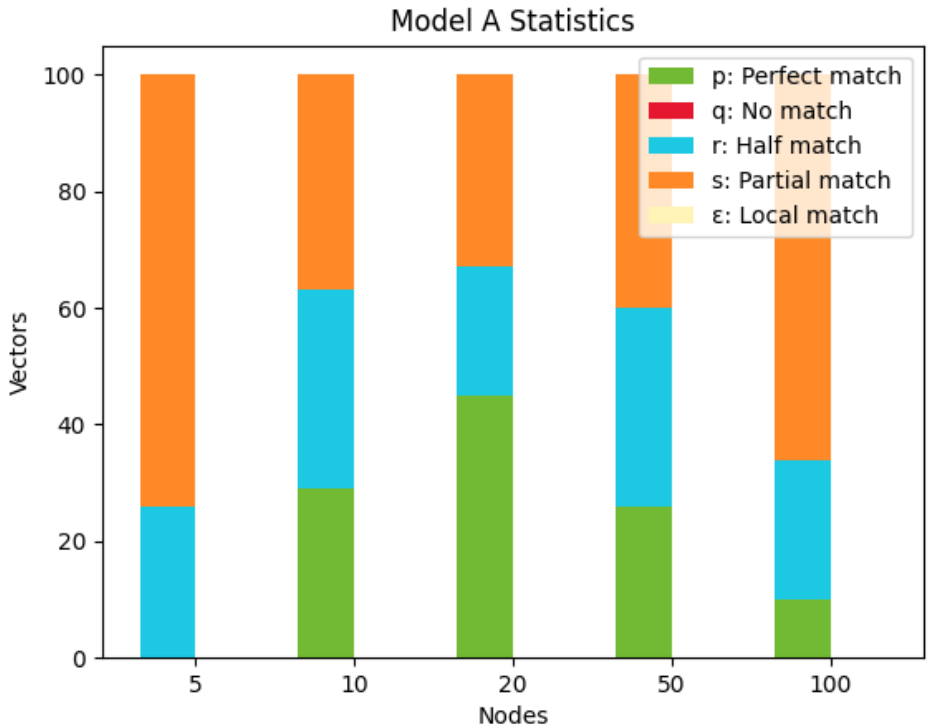
Εικόνα 4. Μέσος χρόνος διαχείρισης πακέτου με N=20 και μεταβλητό k, για το πρώτο μοντέλο.



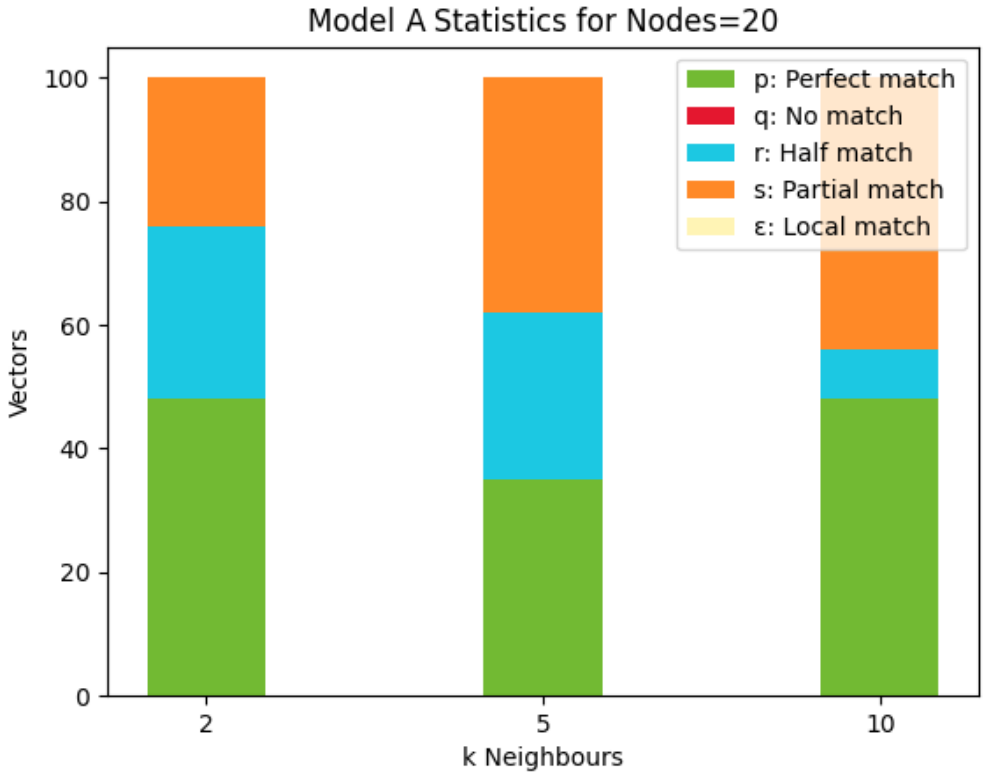
Εικόνα 5. Χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το πρώτο μοντέλο.

Όπως φαίνεται από τις μετρήσεις στην **Εικόνα 2** ο χρόνος διαχείρισης και κατηγοριοποίησης του πακέτου αυξάνεται, όταν αυξάνονται και οι κόμβοι του δικτύου, με εξαίρεση τη περίπτωση όπου N=20 στην οποία ο μέσος χρόνος διαχείρισης πακέτου ήταν μεγαλύτερος από το αντίστοιχο δίκτυο 50 κόμβων. Αυτή

η μέτρηση πιθανών να προκλήθηκε από πακέτα outliers που δυσκόλεψαν τη πρόβλεψη του μοντέλου. Το γεγονός αυτό μπορεί να παρατηρηθεί από την **Εικόνα 3** στην οποία διακρίνονται απότομες αιχμές (spikes) σε διαφορετικές περιπτώσεις στη κατηγοριοποίηση των πακέτων. Παρόλο που οι παρουσίαση τέτοιου είδους αιχμών είναι φυσιολογική, καθώς παρουσιάζεται μικρό αριθμό του δείγματος αλλά και από όλες τις προσομοιώσεις, η κλίμακα της καθυστέρησης που παρουσιάζεται στο δίκτυο με 20 κόμβους δημιουργεί μια εσφαλμένη αντίληψη πτώσης απόδοσης του μοντέλου. Συνεπώς μπορούμε να υποθέσουμε ότι η αύξηση της χρονικής καθυστέρησης στη κατηγοριοποίηση εισερχόμενων πακέτων έχει γραμμική σχέση με την αύξηση του όγκου του δικτύου. Κατά τα άλλα οι χρόνοι διαχείρισης που παρουσίασε το πρώτο μοντέλο ήταν αρκετά ικανοποιητικοί για περιβάλλοντα διαχείρισης δεδομένων πραγματικού χρόνου. Στις **Εικόνα 4, Εικόνα 5** φαίνονται οι αντίστοιχοι χρόνοι που καταγράφτηκαν για το πρώτο μοντέλο με σταθερό αριθμό κόμβων και μεταβλητές τιμές k. Εδώ τα αποτελέσματα έδειξαν ότι ο χρόνος μπορεί να μειωθεί σημαντικά ρίχνοντας τον αριθμό των κοντινότερων γειτόνων, χωρίς μάλιστα σημαντικές επιπτώσεις στα αποτελέσματα του μοντέλου, όπως θα συζητηθεί παρακάτω.



Εικόνα 6. Αποτελέσματα μετρικών για το πρώτο μοντέλο, με k=3 και μεταβλητό N.

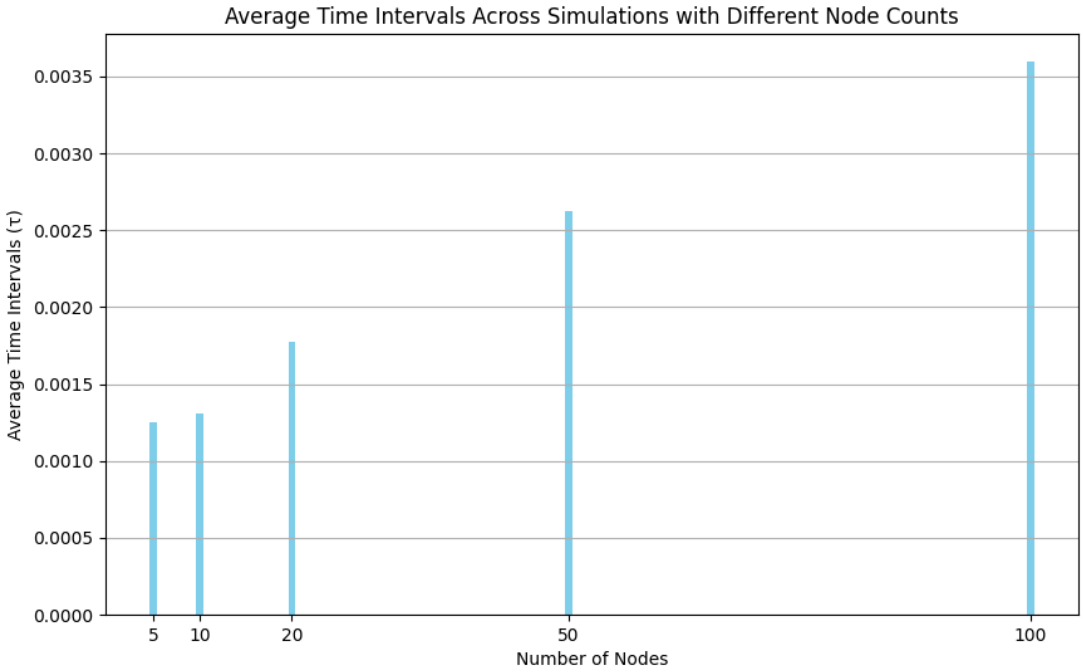


Εικόνα 7. Αποτελέσματα μετρικών για το πρώτο μοντέλο, με N=20 και μεταβλητό k.

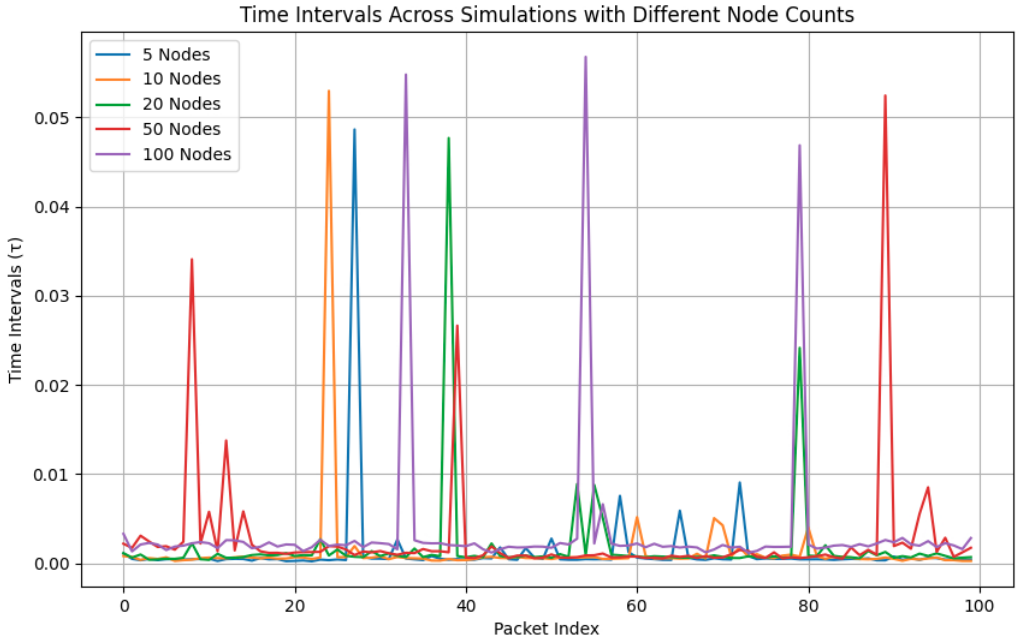
Στα αποτελέσματα, μπορούμε να διακρίνουμε ότι το πρώτο μοντέλο δίνει τα καλύτερα αποτελέσματα ταιριάσματος για εύρος κόμβων από 10-50, με την απόδοση στα άκρα να είναι αρκετά μικρότερη. Πιο συγκεκριμένα όταν το μοντέλο κλήθηκε να κατηγοριοποιήσει τα πακέτα σε δίκτυο 5 κόμβων κανένα από τα εισερχόμενα πακέτα δεν ταίριαξε με τον κόμβο που επιλέχθηκε, ενώ στην αντίπερα όχθη, δηλαδή το δίκτυο με τους 100 κόμβους, τα ποσοστά τέλειου ταιριάσματος ήταν αντικειμενικά υποδεέστερα από αυτά των μικρότερων δικτύων. Στα μεσαία δίκτυα των 10, 20 και 50 κόμβων τα αποτελέσματα ήταν πιο υποσχόμενα με το μοντέλο να δίνει τη καλύτερη πρόβλεψη κοντά στο 45% τέλειου ταιριάσματος στο δίκτυο που αποτελούνται από 20 κόμβους, και με ποσοστό κοντά στο 30% στα δίκτυα με 10 και 50. Ακόμα σημαντικό είναι ότι το μεγαλύτερο ποσοστό των πακέτων στα δίκτυα αυτά (60%>) ανατέθηκαν σε κόμβους που είτε ταίριαζαν απόλυτα είτε πάνω από το μισό. Επιπλέον κανένα πακέτο δεν κατηγοριοποιήθηκε σε κόμβο που δεν ταίριαζε καθόλου. Στις προσομοιώσεις τώρα με διαφορετικές τιμές k, η απόδοση φαίνεται να είναι καλύτερη με μικρό αριθμό γειτόνων αφού για k=2, το μοντέλο παρουσίασε ποσοστό τέλειας ή καλύτερης από το ήμισυ κατηγοριοποίησης κοντά στο 80%. Αντίθετα για k=5 και k=10, τα ποσοστά «καλής» κατηγοριοποίησης πέφτουν στο

~60%, που συνεχίζει να είναι ένα ικανοποιητικό ποσοστό αν λάβουμε ξανά υπόψη ότι κανένα πακέτο δεν κατηγοριοποιήθηκε σε τέλειος άσχετο κόμβο. Αξίζει να σημειωθεί ότι παρόλο το παρόμοιο ποσοστό «καλής» πρόβλεψης, το μοντέλο για k=10, παρουσίασε αρκετά μεγαλύτερο ποσοστό τέλειου ταιριάσματος σε σχέση με το μοντέλο που χρησιμοποιούσε k=5.

Όσο αφορά το δεύτερο μοντέλο, που αποτελεί και μια βελτιωμένη έκδοση του πρώτου τα αποτελέσματα ήταν αρκετά καλύτερα, τόσο στις χρονικές όσο και στις ποιοτικές μετρήσεις. Συγκεκριμένα, όπως φαίνεται στην **Εικόνα 8,** το Μοντέλο Β παρουσίασε μείωση χρόνου στο μέσο υπολογισμό καταλληλότερου κόμβου για πλήθος κόμβων 20, 50 και 100 με ποσοστό μείωσης 46.87%, 14% και 26% αντίστοιχα. Για πλήθος κόμβων 5 και 10 τα αποτελέσματα ήταν σχεδόν πανομοιότυπα με κανένα μοντέλο να μην εμφανίζει σημαντικό χρονικό πλεονέκτημα έναντι του άλλου. Ακόμα από την **Εικόνα 9,** μπορούμε να διακρίνουμε ότι παρόλες τις φυσιολογικές διακυμάνσεις στους χρόνος εκτέλεσης, αυτές δεν ήταν τόσο σημαντικές ώστε να επηρεάσουν τη γραμμική αύξηση του χρόνου εκτέλεσης του μοντέλου με διαφορετικά N, όπως διακρίθηκε στο Μοντέλο Α. Από αυτό το γεγονός μπορούμε να θεωρήσουμε ότι το Μοντέλο Β είναι πιο σταθερό από το Μοντέλο Α.

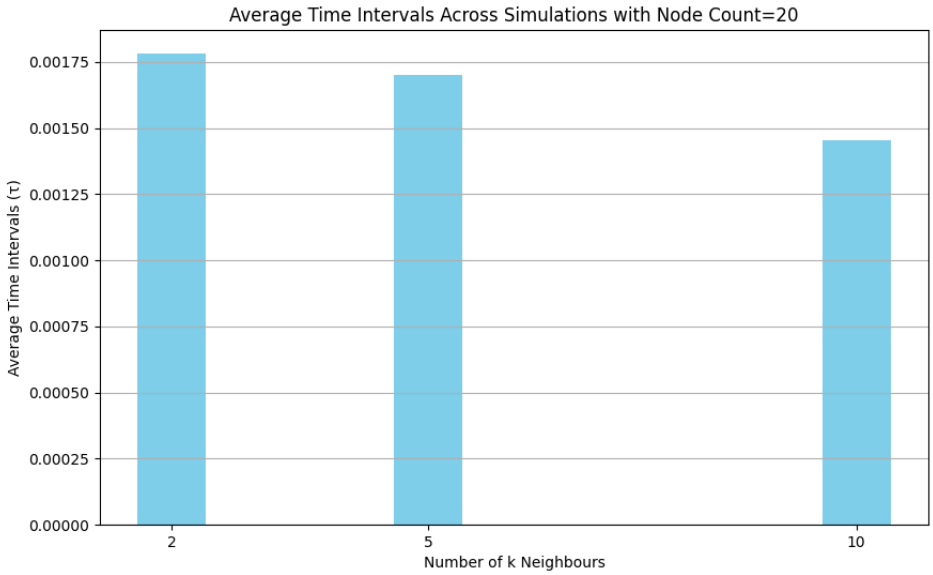


Εικόνα 8. Μέσος χρόνος διαχείρισης πακέτου με k=3 και μεταβλητό N για το δεύτερο μοντέλο.

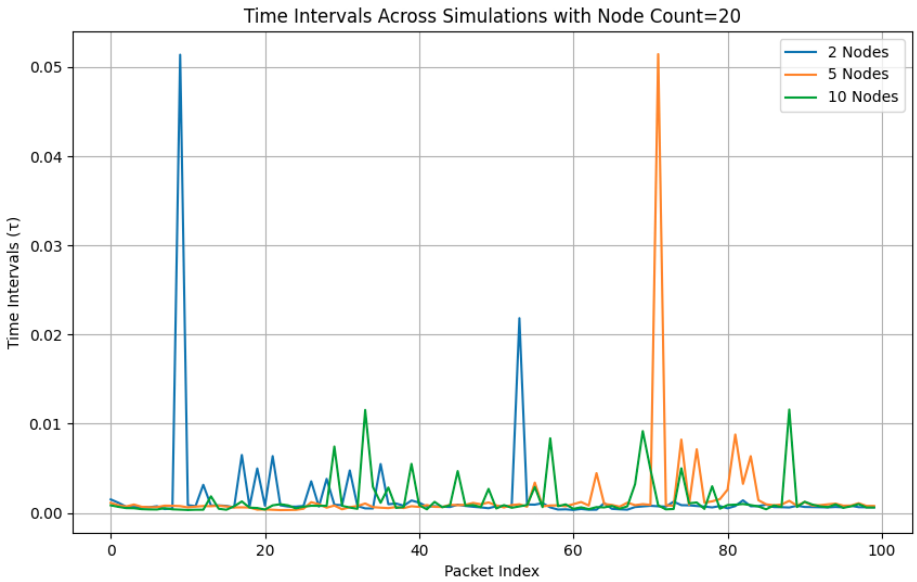


Εικόνα 9. Χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το δεύτερο μοντέλο.

Για τις διαφορετικές τιμές του κ, η μέση τιμή του χρόνου υπολογισμού που σημειώθηκε για το Μοντέλο Β ήταν μεγαλύτερη για k=2 ενώ ελάχιστα μικρότερη για τις υπόλοιπες τιμές **Εικόνα 10**. Αξίζει να σημειωθεί ότι η μεγαλύτερος μέσος χρόνος διαχείρισης πακέτου σημειώθηκε με το μικρότερο κ (k=2) ενώ ο μικρότερος μέσος χρόνος διαχείρισης παρουσιάστηκε για k=10. Σημαντικό είναι ακόμα να αναφερθεί ότι παρόλο που οι μέσοι χρόνοι για k=2 και k=5 ήταν υψηλότεροι, στις προσομοιώσεις τους εμφανίστηκαν μεγάλες ακμές σε αντίθεση με τη προσομοίωση για k=10, όπου η διαχείριση των πακέτων ήταν πιο ομαλή **Εικόνα 11**.

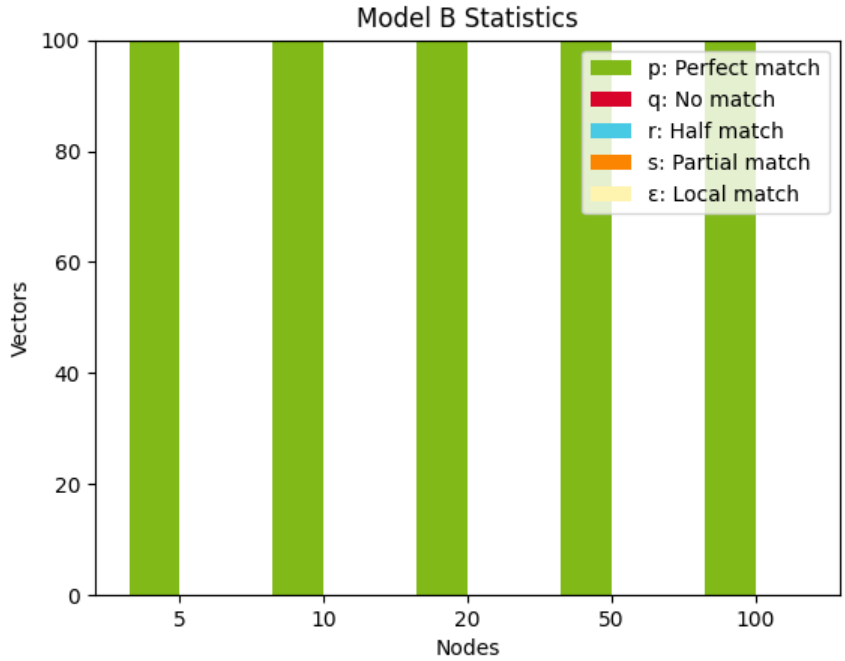


Εικόνα 10. Μέσος χρόνος διαχείρισης πακέτου με N=20 και μεταβλητό k, για το δεύτερο μοντέλο.

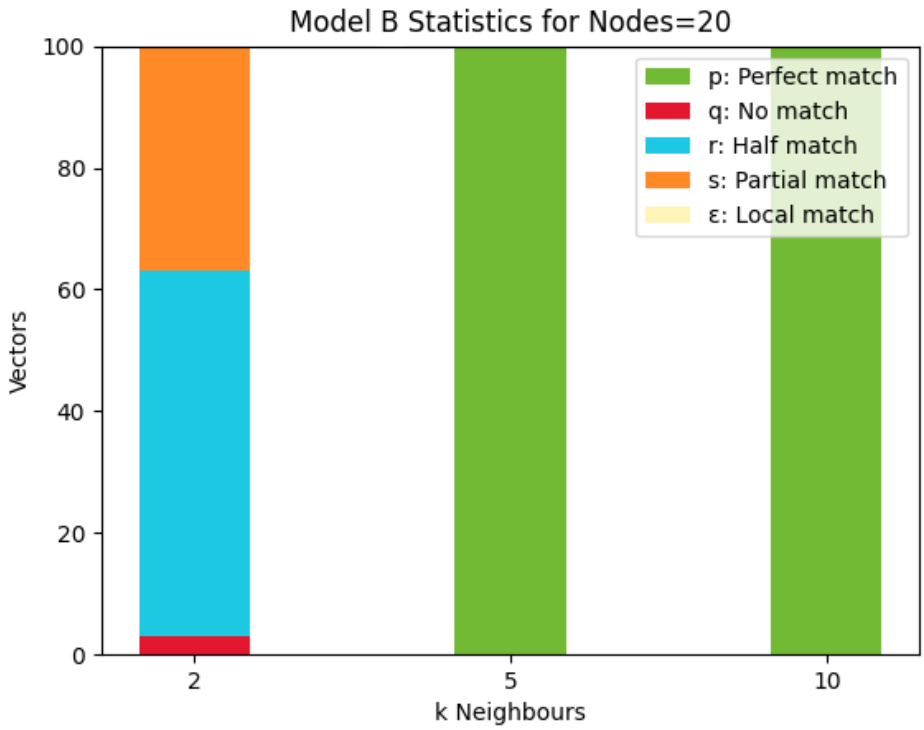


Εικόνα 11. Χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το δεύτερο μοντέλο.

Όσο αφορά τα αποτελέσματα, το Μοντέλο Β ήταν αρκετά ακριβέστερο, σε σχέση με το Μοντέλο Α, στην απόδοση σωστού κόμβου για κάθε πακέτο. Το μοντέλο σημείωσε τέλεια κατηγοριοποίηση των πακέτων ανεξάρτητα το πλήθος των κόμβων του δικτύου, όπως φαίνεται στην **Εικόνα 12**. Για τις διαφορετικές τιμές του κ τα αποτελέσματα ήταν χειρότερα για k=2 σε σχέση με αυτά του μοντέλου Α, καθώς το Μοντέλο Β κατηγοριοποίησε λανθασμένα ή λιγότερο σωστά τα εισερχόμενα πακέτα. Όμως με την αύξηση του k σε 5 και 10 η ακρίβεια του μοντέλου αυξήθηκε ξανά στο 100% επιτυχής για όλα τα εισερχόμενα πακέτα **Εικόνα 13**. Μπορούμε λοιπόν να συμπεράνουμε ότι για k≥3 το Μοντέλο Β παράγει πολύ πιο σταθερή κατηγοριοποίηση σε σχέση με το Μοντέλο Α.



Εικόνα 12. Αποτελέσματα μετρικών για το δεύτερο μοντέλο, με k=3 και μεταβλητό N.

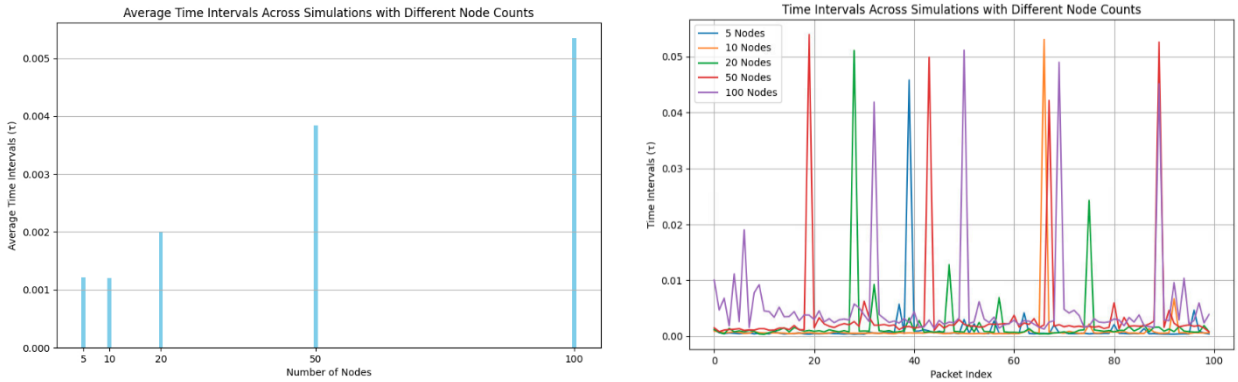


Εικόνα 13. Αποτελέσματα μετρικών για το δεύτερο μοντέλο, με N=20 και μεταβλητό k.

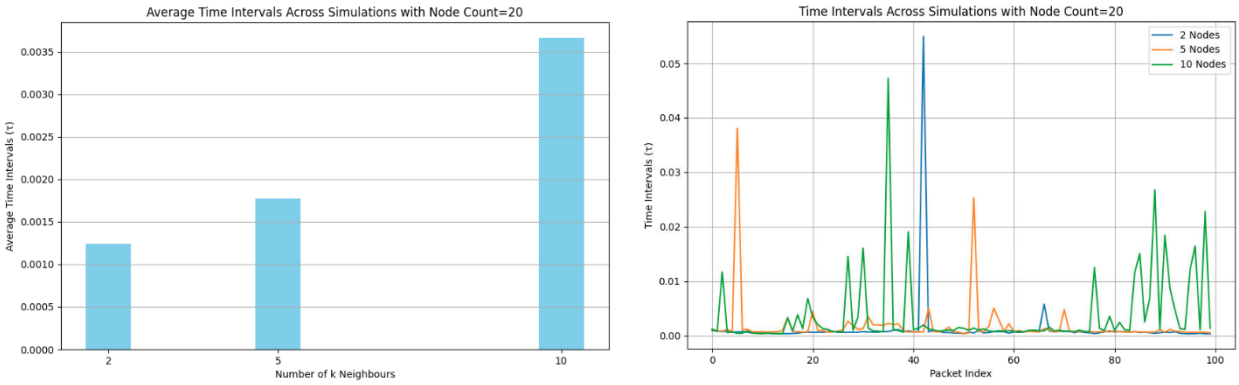
7.2.β ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=4

Με την αύξηση του θ, δίνεται στα μοντέλα ένας μεγαλύτερος βαθμός ελευθερίας στη διαλογή των κόμβων επιτρέποντάς το να μεγαλώσει το εύρος των διαστάσεων ομοιότητας των φίλτρων του κόμβου και του εισερχόμενου πακέτου x.

Στις χρονικές μετρήσεις, οι χρόνοι διαχείρισης των πακέτων είναι σχεδόν ίδιοι με μικρές αυξομειώσεις κυρίως στις προσομοιώσεις με 100 και 20 κόμβους **Εικόνα 14**. Σε αυτές τις περιπτώσεις παρουσιάστηκε άνοδος 10% και μείωση 39% στους μέσους χρόνους διαχείρισης αντίστοιχα, όταν η τιμή του θ αυξήθηκε από 3 σε 4. Ακόμα για τις μετρήσεις με τα διαφορετικά k που φαίνονται στην **Εικόνα 15**, έδειξαν αυξήσεις 30% και 140% για τις τιμές του k=2 και k=10 αντίστοιχα. Η μεγάλη αύξηση του χρόνου διαχείρισης των πακέτων για τη τιμή του k=10 και θ=4, υποδηλώνει τη δυσκολία ανταπόκρισης του μοντέλου Α στην επεκτασιμότητα του δικτύου πέρα από συγκεκριμένα όρια αφού η χρονική του απόδοση στις υπόλοιπες τιμές του k ήταν κανονική σε σχέση με τις προηγούμενες προσομοιώσεις.



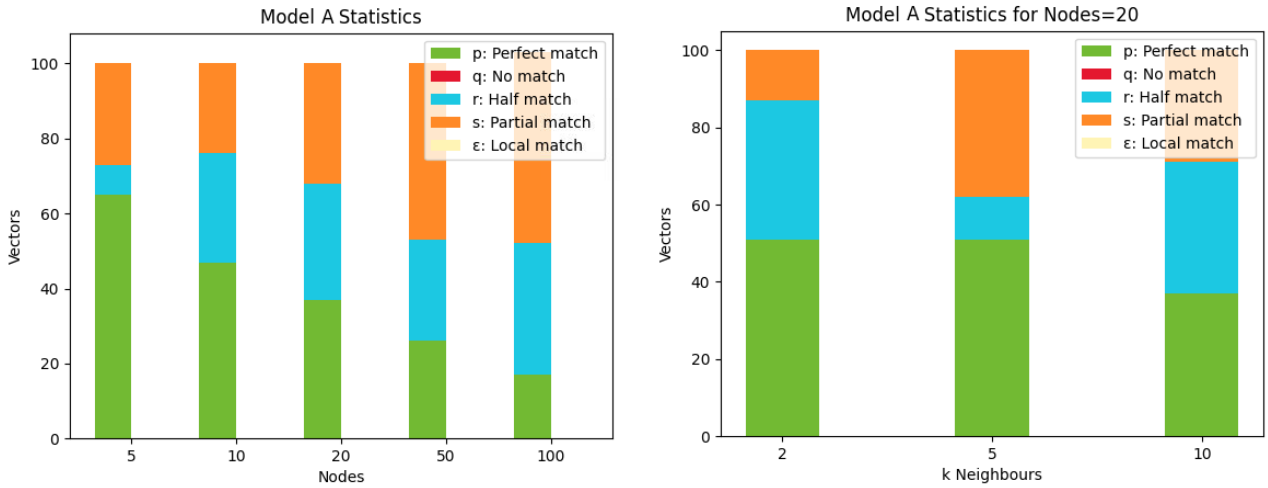
Εικόνα 14. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το πρώτο μοντέλο.



Εικόνα 15. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το πρώτο μοντέλο.

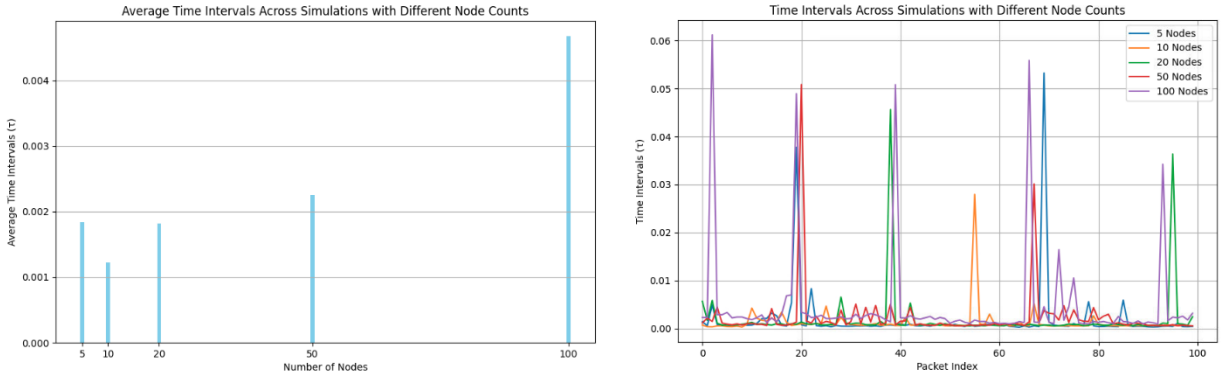
Τα αποτελέσματα ανάθεσης πακέτων με την αύξηση του θ δείχνουν ότι το Μοντέλο Α έχει αυξήσει τη δυνατό τα πλήρους αντιστοίχισης ενός πακέτου με τους κόμβους του δικτύου **Εικόνα 16**. Κυρίως στις προσομοιώσεις μικρών δικτύων με N=5 και N=10, διακρίνεται μια μεγάλη αύξηση των τέλειων ταιριασμάτων των

πακέτων. Παρόλα αυτά η ακρίβεια του μοντέλου Α φαίνεται να πέφτει ανάλογα με την αύξηση του μεγέθους του δικτύου, συνεχίζοντας παρόλη τη πτώση αυτή να παρέχει καλύτερα αποτελέσματα με την αύξηση της μεταβλητής ευαισθησίας. Για τις διαφορετικές τιμές του k, το μοντέλο φαίνεται να παραμένει σταθερό με τις μικρές και μεσαίες τιμές του k αλλά η ακρίβειά του να πέφτει όταν η τιμή φτάσει το 10. Η πτώση αυτή της ακρίβειας του μοντέλου σε συνδυασμό με την αύξηση του μέσου όρου διαχείρισης πακέτων κάνει το μοντέλο καταλληλότερο για εφαρμογή σε δίκτυα μικρού μεγέθους, τουλάχιστον για τη τιμή θ=4.

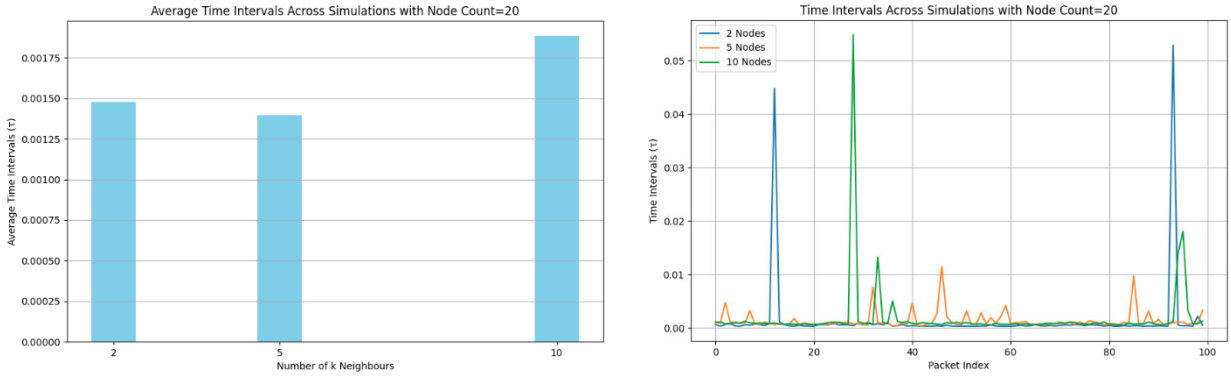


Εικόνα 16. Αποτελέσματα μετρικών για το πρώτο μοντέλο, για διαφορετικά N και διαφορετικά k.

Για το βελτιωμένο Μοντέλο Β τα αποτελέσματα στις χρονικές επιδόσεις διαχείρισης πακέτου είχαν ελάχιστες διαφορές από αυτά με τη τιμή του θ ορισμένη σε 3 **Εικόνα 17**. Η κύρια και σημαντικότερη διαφορά καταγράφηκε στη προσομοίωση με N=100, κατά την οποία ο μέσος χρόνος διαχείρισης αυξήθηκε κατά 45% με σχέση τον αντίστοιχο χρόνο για τη μικρότερη τιμή του θ. Στις υπόλοιπες περιπτώσεις οι χρόνοι εμφάνισαν μικρές αυξομειώσεις, η εμφάνιση των οποίων ήταν αναμενόμενη με την αλλαγή του θ. Επιπλέον στις μετρήσεις χρόνου για τις διαφορετικές τιμές του k, το Μοντέλο Β εμφάνισε την ακριβώς αντίθετη συμπεριφορά με τη προσομοίωση με θ=3, αφού όπως φαίνεται στην **Εικόνα 18,** οι χρόνοι διαχείρισης για k=2 και k=5 μειώθηκαν κατά περίπου 15% υπήρξε αντίστοιχη αύξηση στη προσομοίωση για k=10. Παρόλα αυτά το Μοντέλο Β εμφάνισε ταχύτερους χρόνους σε σχέση με το Μοντέλο Α και σε αυτή τη περίπτωση.

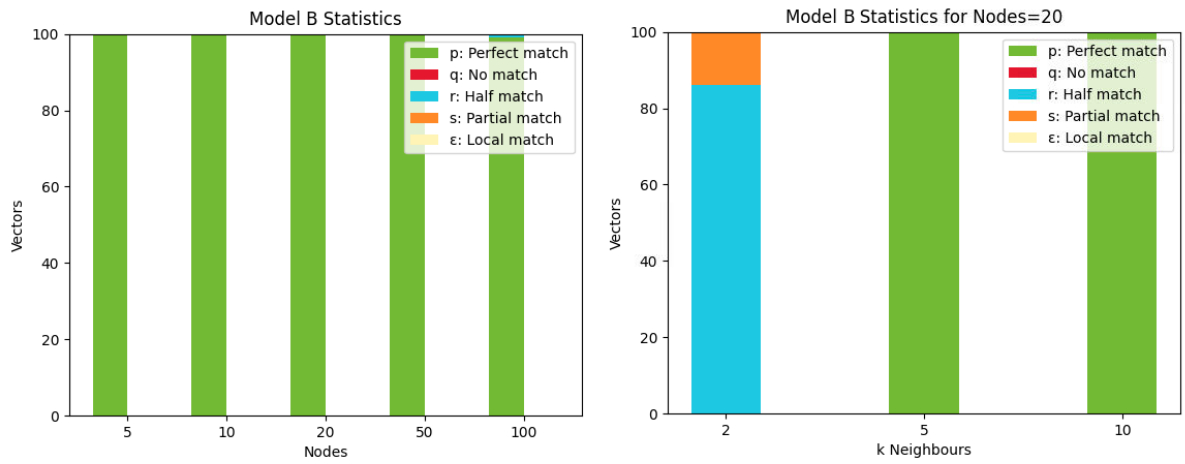


Εικόνα 17. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το δεύτερο μοντέλο.



Εικόνα 18. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το δεύτερο μοντέλο.

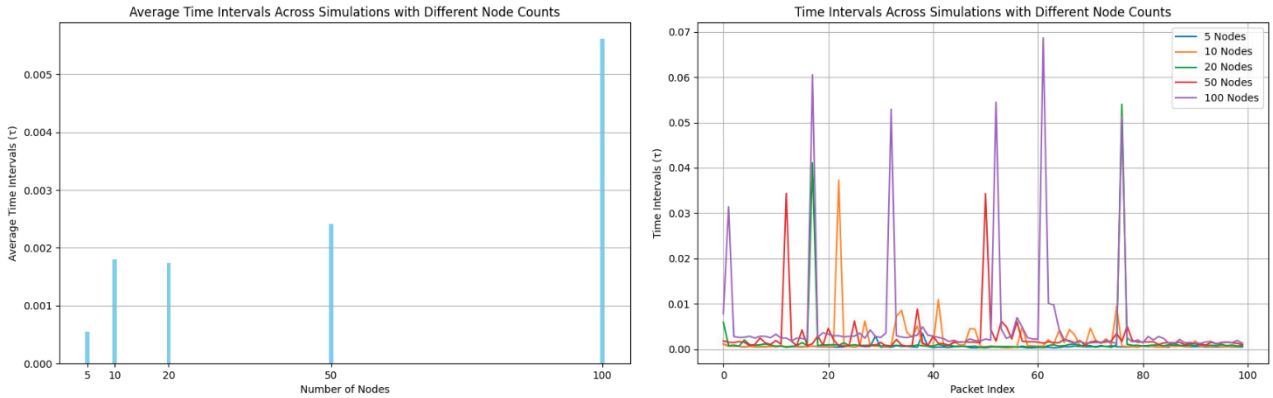
Η απόδοση του μοντέλου Β τώρα όσο αφορά την ακρίβεια, έδειξε όμοια αποτελέσματα ακρίβειας σε 4εις από τις 5 περιπτώσεις όπως φαίνεται στην **Εικόνα 19**. Η μόνη διαφορά ήταν το 99% τέλειου ταιριάσματος που σημειώθηκε στο δίκτυο των 100 κόμβων σε σχέση με το απόλυτο 100% που καταγράφηκε στη προηγούμενη προσομοίωση. Για τις μετρήσεις με τις διαφορετικές τιμές του k, όπως και στη προηγούμενη προσομοίωση το Μοντέλο Β δεν έφερε καλά αποτελέσματα για τη τιμή k=2, ενώ παρουσίασε τη μέγιστη δυνατή αντιστοίχιση διανύσματος-κόμβου με τις μεγαλύτερες τιμές. Παρόλα αυτά με την αύξηση της ευαισθησίας τα ταιριάσματα που ήταν άνω του ήμισυ ή ακριβώς στη μέση αυξήθηκαν κατά 43%. Το Μοντέλο Β παρόλα αυτά ήταν ξανά χειρότερο στη σύγκρισή του με το Μοντέλο Α για k=2, αφού το Μοντέλο Α εμφάνισε περισσότερα τέλεια ταιριάσματα έχοντας σχεδόν το ίδιο ποσοστό κατανομής διανυσμάτων σε κόμβους που ταίριαζαν κάτω του ήμισυ αλλά όχι καθόλου. Αυτό υποδηλώνει ότι το Μοντέλο Β δεν ενδείκνυται να χρησιμοποιείται με τιμές του k μικρότερες από 3 αφού η απόδοσή του πέφτει σημαντικά.



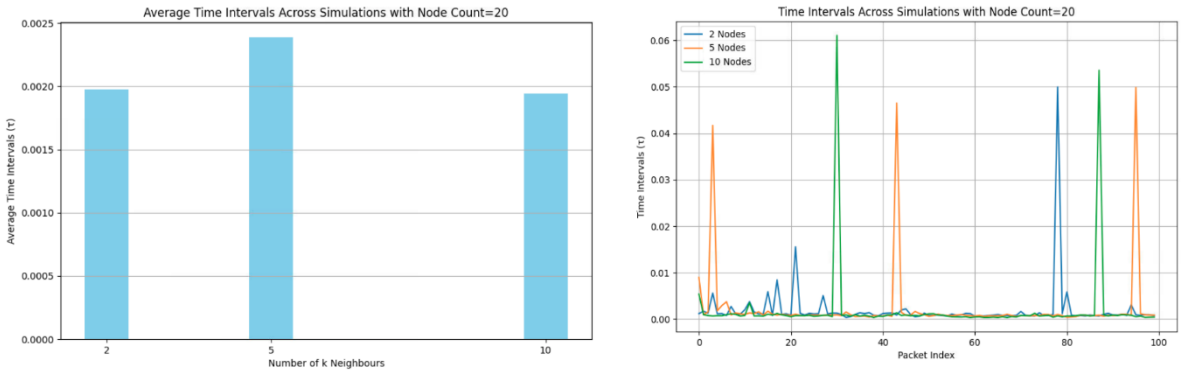
Εικόνα 19. Αποτελέσματα μετρικών για το δεύτερο μοντέλο, για διαφορετικά N και διαφορετικά k.

7.2.γ ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ Θ=5

Τέλος, έγιναν προσομοιώσεις και για τιμή του θ=5, η οποία αφήνει μεγάλο περιθώριο ελευθερίας στο ταίριασμα των διαστάσεων του εισερχόμενου διανύσματος x και τους N κόμβους του δικτύου.



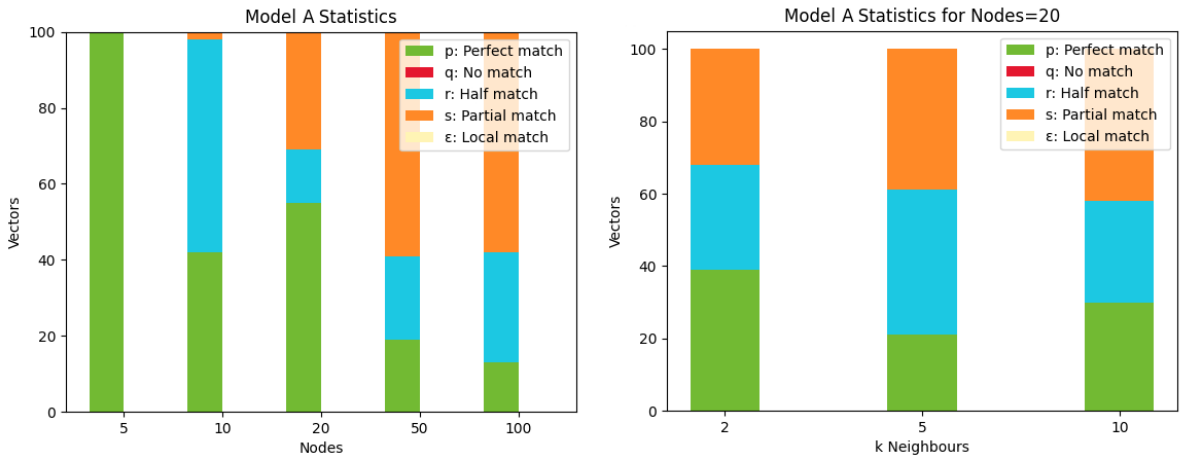
Εικόνα 20. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το πρώτο μοντέλο.



Εικόνα 21. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το πρώτο μοντέλο.

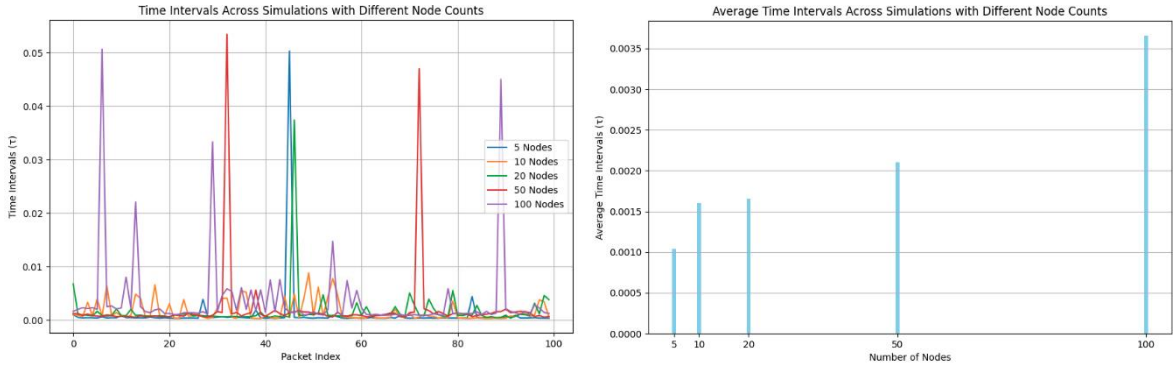
Στην **Εικόνα 20** φαίνονται οι μέση χρόνοι διαχείρισης πακέτου, οι οποίοι δε παρουσιάζουν ιδιαίτερες διαφορές σε σχέση με τις προηγούμενες προσομοιώσεις του μοντέλου, για τα διαφορετικά θ, για το μέγεθος του δικτύου με N= [10, 20, 50, 100]. Για τη τιμή N=5 όμως ο μέσος χρόνος εκτέλεσης έχει μειωθεί περισσότερο από 50% με βάση τις 2 προηγούμενες προσομοιώσεις. Η μείωση χρόνου αυτή είναι αρκετά σημαντική καθώς όπως θα αναφέρουμε αργότερα η απόδοση του μοντέλου Α ήταν πολύ αποδοτική στην ανάθεση πακέτων στους όμοιους κόμβους του δικτύου. Έτσι το Μοντέλο Α παρέχει γρήγορες ταχύτητες χωρίς να μειώνεται η αποδοτικότητα του μοντέλου. Όσο αφορά τις διαφορετικές τιμές του k που αναγράφονται στην **Εικόνα 21**, οι μέσοι χρόνοι είναι αυξημένοι σε σχέση με τις τιμές τους στις προσομοιώσεις με θ=3 και θ=4. Η μόνη εξαίρεση είναι η περίπτωση του k=10 στη προσομοίωση θ=4, όπου οι χρόνοι που παρουσιάστηκαν για θ=5 είναι μικρότεροι.

Τα αποτελέσματα της προσομοίωσης ήταν αρκετά ικανοποιητικά για σε σχέση με τις προηγούμενες προσομοιώσεις στα μικρά δίκτυα (N = [5, 10]), **Εικόνα 22**. Η απόδοση του μοντέλου για 5 κόμβους έφτασε το 100% τέλειου ταιριάσματος σε συνδυασμό όπως προαναφέρθηκε με μείωση του μέσου χρόνου διαχείρισης. Ακόμα στο δίκτυο N=10, παρόλο που τα τέλεια ταιριάσματα ήταν στο 40%, η κατηγοριοποίηση των πακέτων έγινε σε κόμβους που ήταν σε μεγάλο βαθμό όμοιοι με το εισερχόμενο διάνυσμα σε ένα πολύ μεγάλο ποσοστό 55%. Το γεγονός αυτό δείχνει ένα σημαντικό βαθμό αποδοτικότητας του μοντέλου Α στα μικρά δίκτυα. Για τις μεσαίες και μεγάλες τιμές μεγέθους δικτύων (N = [20, 50, 100]), το Μοντέλο Α, εμφάνισε αποτελέσματα παρόμοια με τις προηγούμενες προσομοιώσεις όπου η απόδοση για N=20 ήταν 55% για τα τέλεια ταιριάσματα και 14% για τα πάνω από το ήμισυ όμοια. Όσο το δίκτυο μεγάλωνε σε μέγεθος αυτή έπεφτε κατακόρυφα. Οι διαφορετικές τιμές του k, έριξαν τα τέλεια ταιριάσματα κατά 29%, 66.8%, 45.45% για τις αντίστοιχες τιμές του k 2, 5, 10. Παρόλα αυτά με την αύξηση των μερικών ταιριασμάτων η ακρίβεια της άνω από ήμισυ αντιστοιχίας των εισερχόμενων διανυσμάτων με τους κόμβους να φτάνει περίπου το 60% των εισερχόμενων διανυσμάτων.

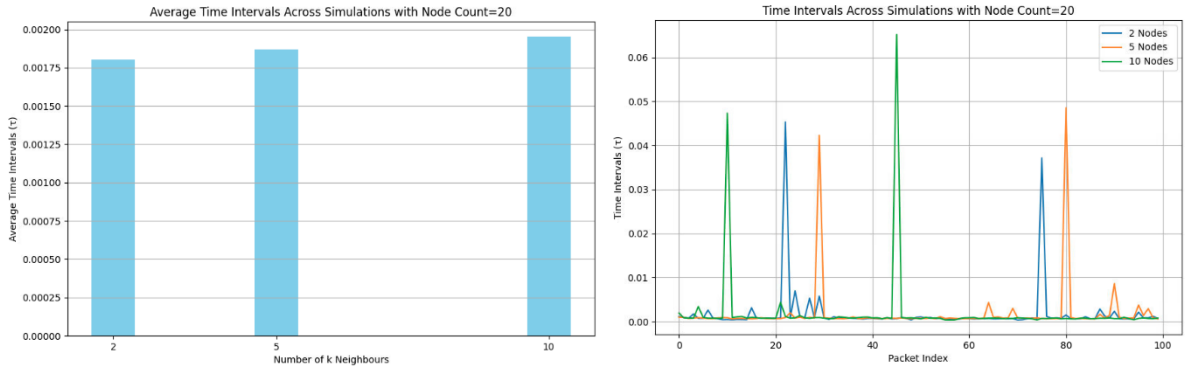


Εικόνα 22. Αποτελέσματα μετρικών για το πρώτο μοντέλο, για διαφορετικά N και διαφορετικά k.

Για το Μοντέλο Β, οι μέσοι χρόνοι διαχείρισης πακέτου ήταν ξανά μικρότεροι από αυτούς που σημειώθηκαν στο Μοντέλο Α σε όλες τις περιπτώσεις εκτός από το δίκτυο με N=5 όπου το Μοντέλο Α φάνηκε να έχει το πλεονέκτημα **Εικόνα 23**. Οι τιμές του μέσου χρόνου διαχείρισης ήταν ανταγωνιστικοί και με τις προηγούμενες προσομοιώσεις του μοντέλου Β, σημειώνοντας καλύτερες επιδόσεις από τη προσομοίωση για θ=4, ενώ για θ=5 οι χρόνοι διαχείρισης εισερχόμενου διανύσματος ήταν καλύτεροι για δίκτυα με N >= 20. Για τις διαφορετικές τιμές του k, το Μοντέλο Β και εδώ είχε το πλεονέκτημα σε σχέση με το Μοντέλο Α αλλά οι χρόνοι του ήταν μεγαλύτεροι σε σχέση με τις αντίστοιχες μετρικές για θ=3 και θ=4 αντίστοιχα.

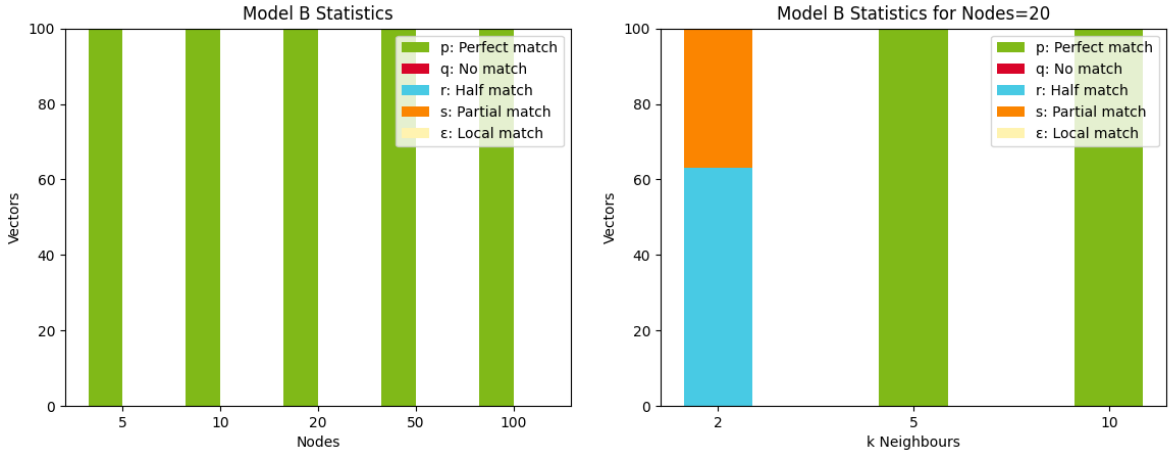


Εικόνα 23. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το δεύτερο μοντέλο.



Εικόνα 24. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το δεύτερο μοντέλο.

Τα αποτελέσματα ήταν για άλλη μια φορά τέλεια για όλες τις διαφορετικές τιμές του N, όπως φαίνεται στην **Εικόνα 25**. Η απόδοση του μοντέλου παραμένει η ίδια παρόλες τις αλλαγές τόσο της ελευθερίας όσο και του μεγέθους του δικτύου, γεγονός που το καθιστά απόλυτα σταθερό στις εξωτερικές αλλαγές. Για τις διαφορετικές τιμές του k, το μοντέλο εμφάνισε ξανά τέλεια ακρίβεια για k=5 και k=10 αλλά συνέχισε το μοτίβο αστοχιών σε τέλεια ταιριάσματα για k=2. Παρόλο που οι τιμές ταιριασμάτων των εισερχόμενων διανυσμάτων ήταν κατά κύριο λόγο σε κόμβους που αυτά ταιριάζανε περισσότερο ή κατά το ήμισυ, τα αποτελέσματα του μοντέλου A φαίνεται για άλλη μια φορά να υπερτερούν για τις μικρότερες τιμές του k πετυχαίνοντας ένα ικανοποιητικό ποσοστό τέλειων ταιριασμάτων σε αντίθεση με το Μοντέλο B.



Εικόνα 25. Αποτελέσματα μετρικών για το δεύτερο μοντέλο, για διαφορετικά N και διαφορετικά k.

7.2.δ ΠΡΟΣΟΜΟΙΩΣΕΙΣ ΜΕ ΤΥΧΑΙΑ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ

Για να αντιμετωπιστεί η χαμηλή απόδοση του μοντέλου B για τις χαμηλές τιμές του k, διεξήχθησαν μετρήσεις για κατηγοριοποίηση των πακέτων με βάση τυχαία

επιλογή των 2 μοντέλων για κάθε πακέτο. Τα 2 μοντέλα είχαν ισοπίθανο ποσοστό επιλογής για τη κατηγοριοποίηση του πακέτου. Οι προσομοιώσεις με την μέθοδο της τυχαίας επιλογής έγιναν και για διαφορετικές τιμές του θ όπως και στις μετρήσεις απόδοσης των μοντέλων ξεχωριστά.

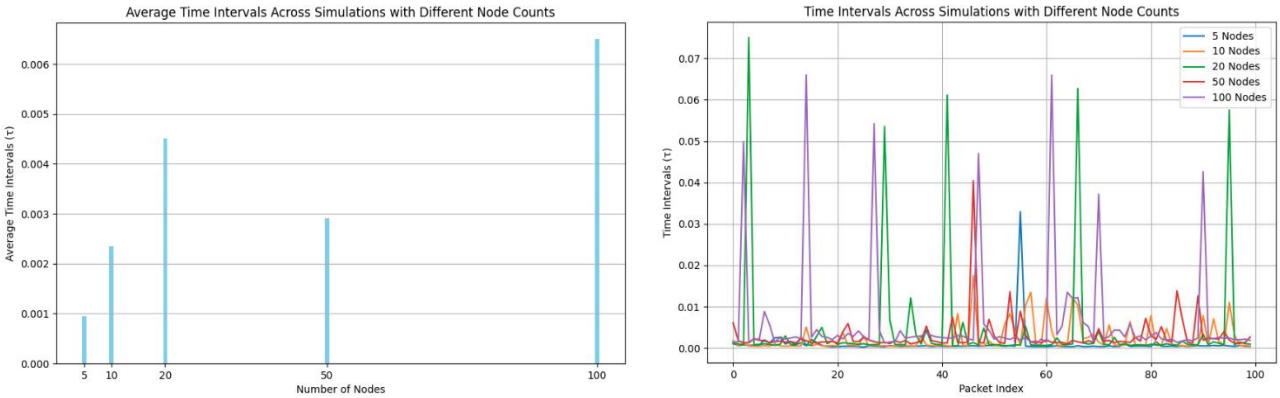
Για θ=3, το η τεχνική τυχαίας επιλογής, όπως φαίνεται στον **Πίνακα 2**, υστέρησε σε όλες τις περιπτώσεις πέραν του δικτύου με μέγεθος N=5. Στις περισσότερες περιπτώσεις τα μοντέλα ξεχωριστά παρουσίασαν πολύ καλύτερους μέσους χρόνους διαχείρισης των εισερχόμενων πακέτων. Επιπλέον όπως φαίνεται στην **Εικόνα 26**, οι χρόνοι διαχείρισης παρουσιάζουν αρκετές ακμές γεγονός που υποδηλώνει μια αστάθεια στη διαχείριση των εισερχόμενων πακέτων με την τακτική εναλλαγή του μοντέλου κατηγοριοποίησης. Αυτό ίσως να είναι και το γεγονός που οδηγεί τη τυχαία επιλογή μοντέλου διαχείρισης να εμφανίζει μεγαλύτερους συγκριτικά μέσους χρόνους. Για τις διαφορετικές τιμές του k, όπως φαίνεται στον **Πίνακα 3**, η χρήση της τυχαίας επιλογής μοντέλου έριξε τους χρόνους διαχείρισης με τις μεγαλύτερες τιμές του k ενώ υστερούσε για k=2. Σημαντική ήταν η μείωση του χρόνου για k=5, όπου η χρήση της τυχαίας επιλογής επέφερε αρκετά ταχύτερα αποτελέσματα σε σχέση με την αποκλειστική χρήση ενός εκ των 2 μοντέλων. Ακόμα οι προσομοιώσεις παρουσιάζονται αρκετά ομαλές με ελάχιστες ακμές στα διάφορα πακέτα, όπως παρουσιάζεται στην **Εικόνα 27**. Τέλος για τα διάφορα N, τα αποτελέσματα παρόλο που ήταν αρκετά ικανοποιητικά δεν κατάφεραν να πετύχουν το 100% που επέφερε από μόνο του το Μοντέλο Β γεγονός το οποίο σε συνδυασμό με την αύξηση του μέσου χρόνου υπολογισμού δεν επιφέρει οφέλη για πιθανή υιοθέτηση αυτού υβριδικού τρόπου προσέγγισης. Για τα διάφορα k αλλά κυρίως τη περίπτωση του k=2, τα τέλεια ταιριάσματα παρουσίασαν αύξηση αλλά ξανά όχι αρκετά σημαντική για να ληφθεί υπόψη υιοθέτησης του υβριδικού τρόπου τυχαίας επιλογής. Τα αποτελέσματα που παράχθηκαν από τη τυχαία επιλογή μοντέλου παρουσιάζονται στην **Εικόνα 27**.

Πίνακας 2. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά N και θ=3.

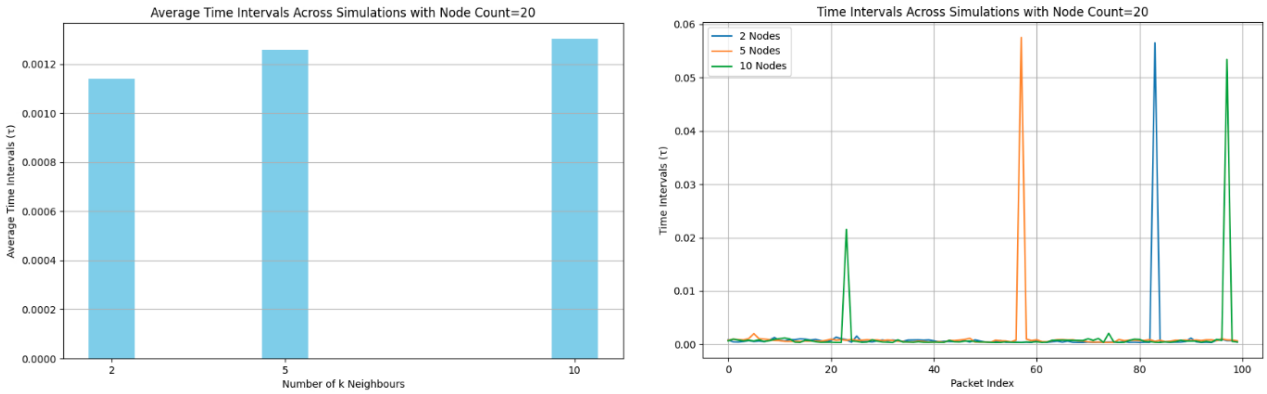
	Τιμές του N				
	5	**10**	**20**	**50**	**100**
Μοντέλο Α	+10%	-48%	-30%	+7%	-24%
Μοντέλο Β	+50%	-30%	-63%	-10%	-45%

Πίνακας 3. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά k και θ=3.

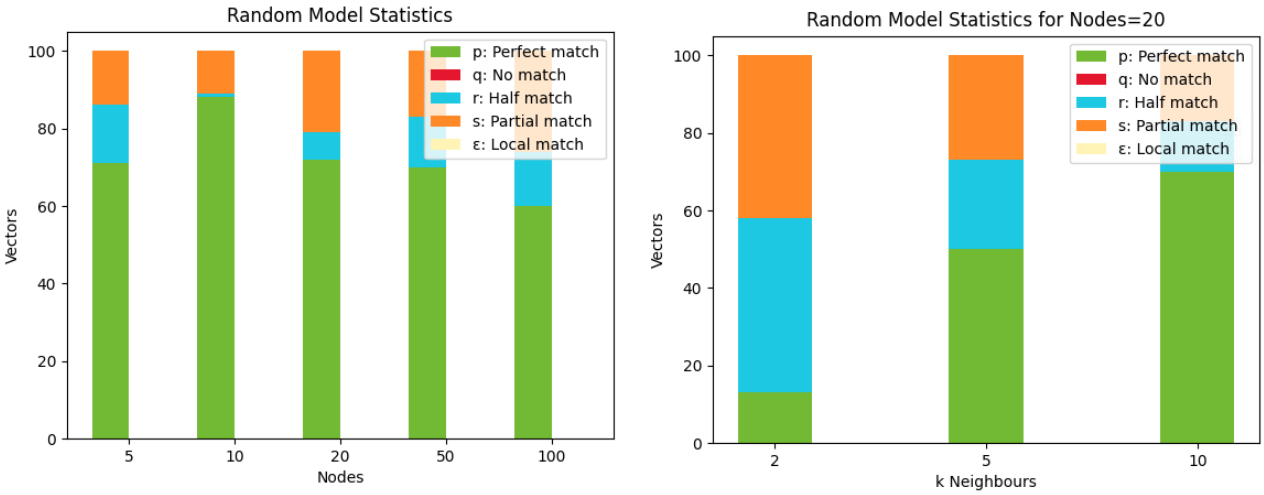
	Τιμές του k		
	2	**5**	**10**
Μοντέλο Α	-39%	+43%	+22%
Μοντέλο Β	-4%	+31%	+14%



Εικόνα 26. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το υβριδικό μοντέλο (θ=3).



Εικόνα 27. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το υβριδικό μοντέλο (θ=3).



Εικόνα 28. Αποτελέσματα μετρικών για το υβριδικό μοντέλο, για διαφορετικά N και διαφορετικά k (θ=3).

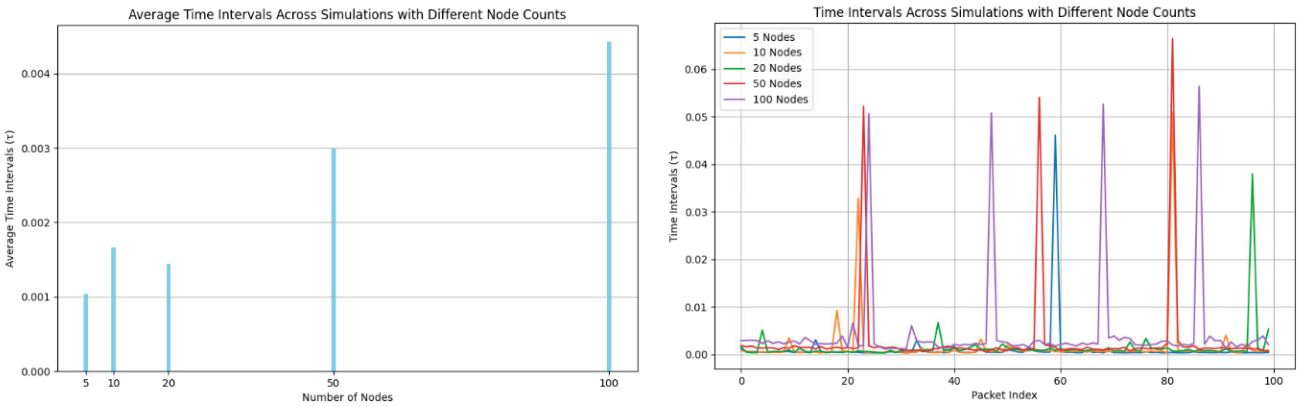
Για θ=4, η τυχαία επιλογή μοντέλου κατάφερε να ρίξει το μέσο χρόνο διαχείρισης πακέτου σε 3εις από τις 5 περιπτώσεις, για τις διάφορες τιμές του N, που εξετάστηκαν όπως φαίνεται στον **Πίνακα 4** και την **Εικόνα 29**. Κυρίως στο μικρό δίκτυο με N=5, η χρόνοι της τυχαίας επιλογής μοντέλου ήταν σχεδόν υποδιπλάσιοι από αυτούς του μοντέλου Β. Τα αποτελέσματα των αντιστοιχίσεων όμως και σε αυτή τη περίπτωση παρόλη την αύξηση του θ, δεν κατάφεραν να ανταγωνιστούν τα αποτελέσματα που έφερε το Μοντέλο Β στις αντίστοιχες περιπτώσεις, **Εικόνα 31**. Επιπλέον οι περιπτώσεις στις οποίες το υβριδικό μοντέλο τυχαίας επιλογής είχε μικρότερους χρόνους από το Μοντέλο Β, είχε συστηματικά λιγότερο από 26% τέλειες αντιστοιχίσεις των εισερχόμενων διανυσμάτων με τους κόμβους του δικτύου. Παρόλα αυτά το υβριδικό μοντέλο παρουσίασε καλύτερα αποτελέσματα από το Μοντέλο Α για όλες τις τιμές του N που δόθηκαν. Όσο αφορά τις προσομοιώσεις με τις διαφορετικές τιμές του k, το υβριδικό μοντέλο παρουσίασε μεγαλύτερους χρόνους για τον υπολογισμό του ομότιμο κόμβου για k=2 και k=5 όπως φαίνεται στην **Εικόνα 30** και **Πίνακα 5**. Συγκεκριμένα στη περίπτωση του k=2 τα τέλεια ταιριάσματα δεν αυξήθηκαν σημαντικά σε σχέση με τα αποτελέσματα του μοντέλου Β ενώ ήταν υποδεέστερα από τα αντίστοιχα αποτελέσματα που επέφερε το Μοντέλο Α. Τέλος για k=10, το υβριδικό μοντέλο εμφάνισε υποδιπλάσιο χρόνο διαχείρισης από το Μοντέλο Α αυξάνοντας παράλληλα την ακρίβεια τέλειου ταιριάσματος κατά 45%. Όμως στη σύγκρισή του με το Μοντέλο Β, η διαφορά στους χρόνους ήταν αμελητέα ενώ τα αποτελέσματα για άλλη μια φορά δεν άγγιξαν το 100% που εμφάνισε το Μοντέλο Β.

Πίνακας 4. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά N και θ=4.

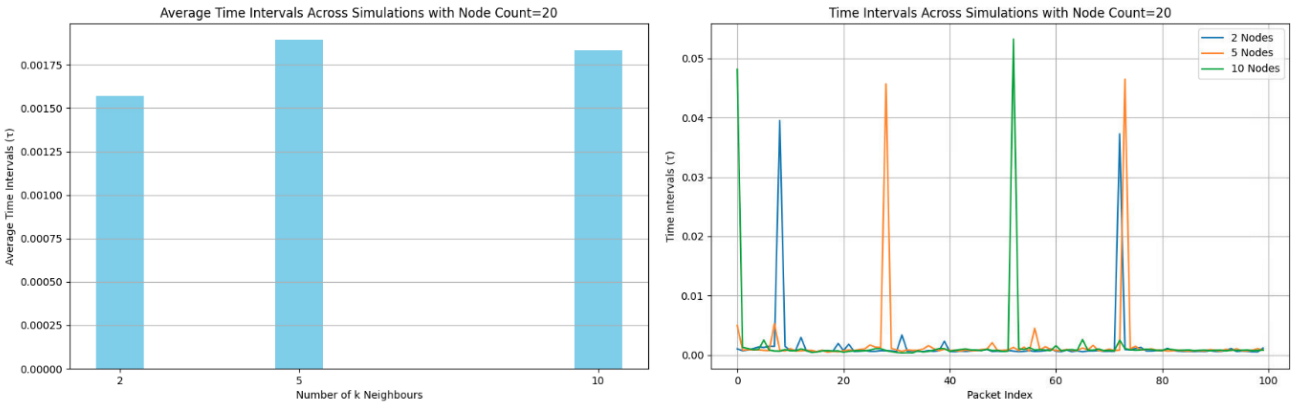
	Τιμές του N				
	5	**10**	**20**	**50**	**100**
Μοντέλο Α	+20%	-30%	+33%	+30%	+25%
Μοντέλο Β	+80%	-30%	+20%	-27%	+9%

Πίνακας 5. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά k και θ=4.

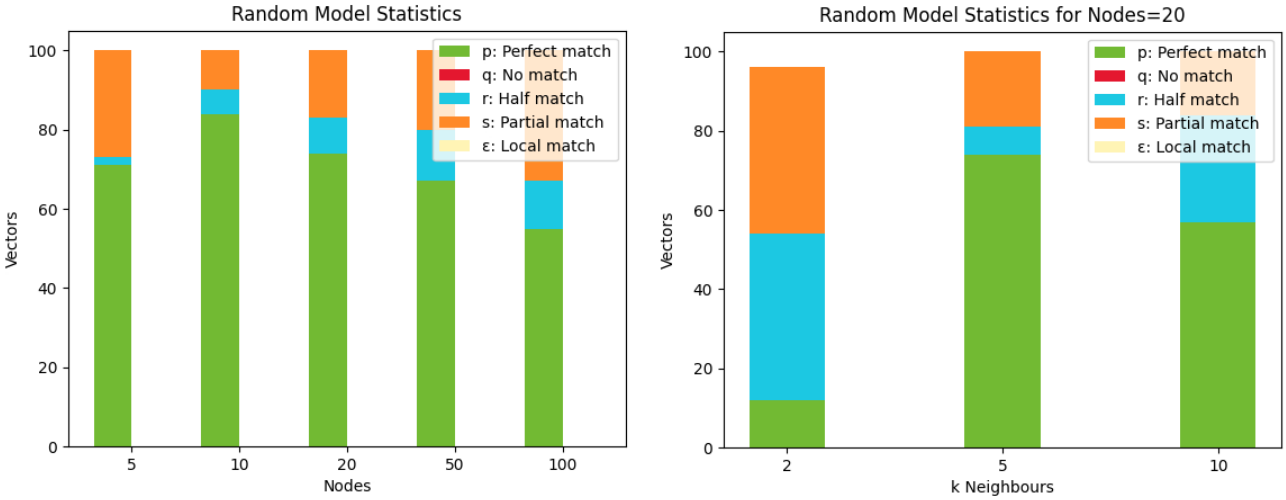
	Τιμές του k		
	2	**5**	**10**
Μοντέλο Α	-14%	-5.5%	+102%
Μοντέλο Β	-0.6%	-22%	+1%



Εικόνα 29. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το υβριδικό μοντέλο (θ=4).



Εικόνα 30. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το υβριδικό μοντέλο (θ=4).



Εικόνα 31. Αποτελέσματα μετρικών για το υβριδικό μοντέλο, για διαφορετικά N και διαφορετικά k (θ=4).

Τέλος για θ=5, το υβριδικό μοντέλο επέφερε μεγαλύτερους χρόνους διαχείρισης σε όλες τις προσομοιώσεις διαφορετικού μεγέθους δικτύου, **Εικόνα 32** και **Πίνακας 6**. Η ακρίβεια του μοντέλου ήταν αρκετά ικανοποιητική παρόλο που ξανά δεν κατάφερε να φτάσει το 100% που το Μοντέλο Β κατάφερε στις αντίστοιχες

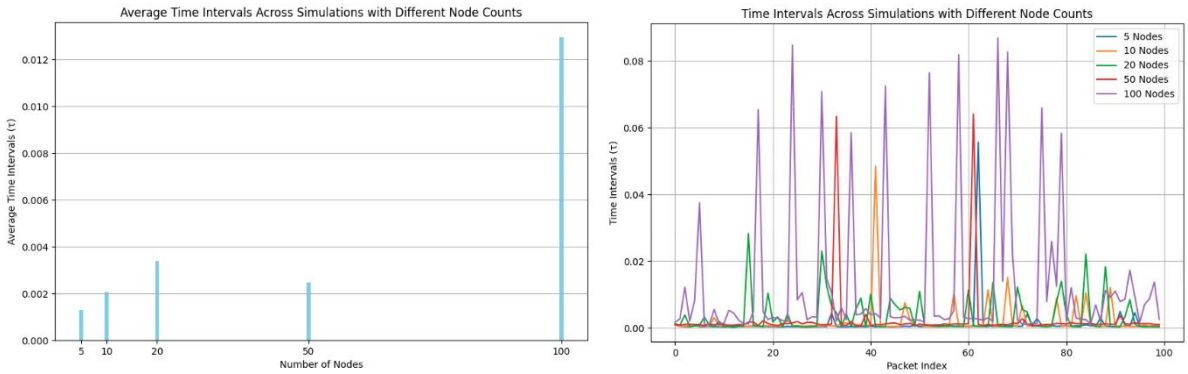
μετρήσεις, **Εικόνα 34**. Οι προσομοιώσεις για τα διαφορετικά k, **Εικόνα 33** και **Πίνακας 7**, έδειξαν αξιοσημείωτες μειώσεις στο χρόνο διαχείρισης πακέτων για k=5 όπου το υβριδικό μοντέλο ήταν περισσότερο από δύο φορές ταχύτερο στην κατηγοριοποίηση του μοντέλου σημειώνοντας παράλληλα ποσοστό τέλειων ταιριασμάτων άνω του 70%. Στις περιπτώσεις όπου k=2 και k=10, το υβριδικό μοντέλο ήταν χειρότερο χρονικά και από τα δύο προτεινόμενα μοντέλα. Στο κομμάτι της ακρίβειας για k=2 παρόλο που τα αποτελέσματα ήταν καλύτερα από αυτά του μοντέλου B, δεν κατάφεραν να ξεπεράσουν αυτά του μοντέλου A το οποίο είχε κατά 85% περισσότερα τέλεια ταιριάσματα. Για k=10 το Μοντέλο B είχε 100% ακρίβεια αντιστοίχισης του εισερχόμενου πακέτου με τον πιο ταιριαστό κόμβο σε αντίθεση με το υβριδικό μοντέλο το οποίο σε κατάφερε να κατηγοριοποιήσει σωστά το 79% των πακέτων σε περισσότερο χρόνο.

Πίνακας 6. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά N και θ=5.

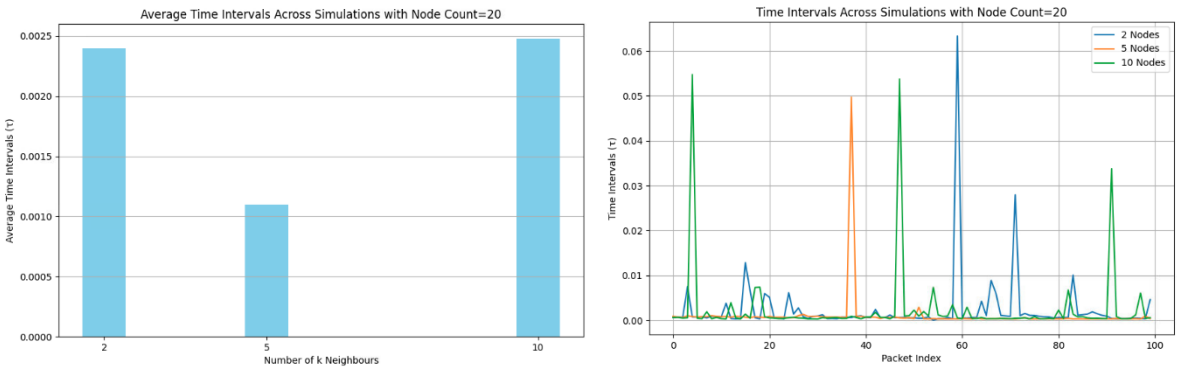
	Τιμές του N				
	5	**10**	**20**	**50**	**100**
Μοντέλο A	-67%	-10%	-47%	-8%	-55%
Μοντέλο B	-33%	-23%	-54%	0%	-74%

Πίνακας 7. Στατιστικά σύγκρισης χρόνων της τυχαίας επιλογής μοντέλου σε σχέση με τα προτεινόμενα μοντέλα, για διαφορετικά k και θ=5.

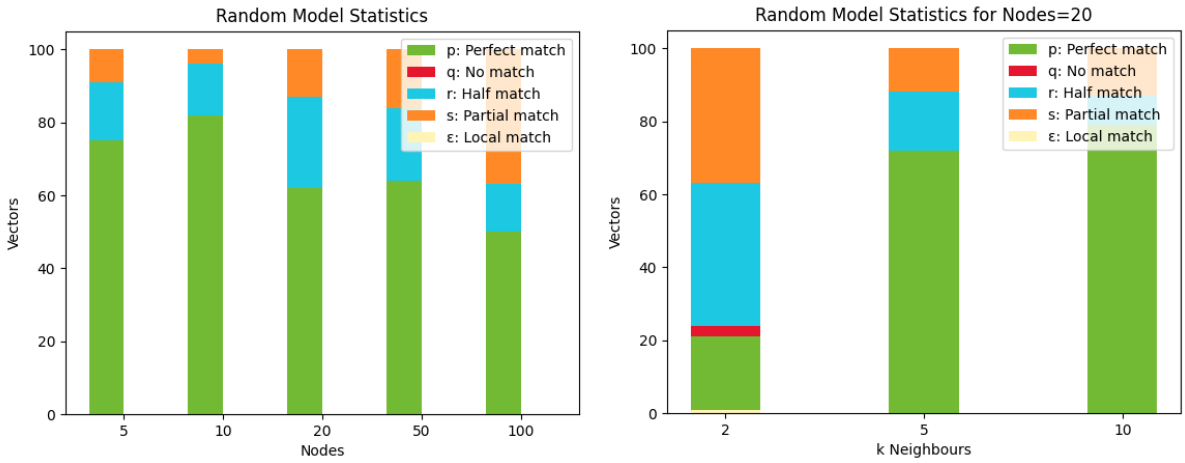
	Τιμές του k		
	2	**5**	**10**
Μοντέλο A	-17%	+123%	-24%
Μοντέλο B	-27%	+64%	-22%



Εικόνα 32. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με N=20 και μεταβλητό k για το υβριδικό μοντέλο (θ=5).



Εικόνα 33. Μέσος χρόνος και χρονικές διακυμάνσεις επεξεργασίας εισερχόμενων πακέτων, με k=3 και μεταβλητό N για το υβριδικό μοντέλο (θ=5).



Εικόνα 34. Αποτελέσματα μετρικών για το υβριδικό μοντέλο, για διαφορετικά N και διαφορετικά k (θ=5).

7.3 ΠΕΡΙΟΡΙΣΜΟΙ

Ένας σημαντικός περιορισμός αυτής της έρευνας αποτελεί η απουσία δοκιμαστικών σεναρίων στον πραγματικό κόσμο. Λόγω των δυσκολιών στην εξαγωγή πραγματικών συνόλων δεδομένων από περιβάλλοντα edge computing, η μελέτη βασίζεται σε προσομοιωμένα δεδομένα που δημιουργήθηκαν με χρήση της γλώσσας προγραμματισμού Python. Ενώ τα προσομοιωμένα δεδομένα παρέχουν έναν ελεγχόμενο χώρο για την πειραματική διερεύνηση, τα ευρήματα ενδέχεται να μην αποτυπώνουν πλήρως τις περίπλοκες και μεταβαλλόμενες συνθήκες που χαρακτηρίζουν πραγματικά περιβάλλοντα edge computing. Τα δεδομένα που χρησιμοποιούνται σε αυτήν τη μελέτη είναι προσομοιωμένα για να αντιστοιχούν σε συνθήκες edge computing. Παρότι έχουν γίνει προσπάθειες για να διασφαλιστεί η ρεαλιστικότατα των προσομοιωμένων δεδομένων, είναι σημαντικό να αναγνωριστεί ότι ορισμένες πτυχές της πραγματικής πολυπλοκότητας, των δυναμικών περιβαλλοντικών παραγόντων και των απρόβλεπτων προκλήσεων ενδέχεται να μην

αντανακλώνται πλήρως στα προσομοιωμένα σενάρια. Οι περιορισμοί αυτοί θα αντιμετωπιστούν σε μελλοντική έρευνα.

ΚΕΦΑΛΑΙΟ 8 Συμπεράσματα

Σε αυτήν τη μελέτη, εξετάσαμε τον κρίσιμο τομέα της αντιστοίχισης εισερχόμενων πακέτων σε κόμβους στην ακμή του δικτύου, χρησιμοποιώντας δύο διακριτικά μοντέλα, Μοντέλο Α και Μοντέλο Β, για να αντιμετωπίσουμε τις προκλήσεις που σχετίζονται με αυτό το έργο. Επιπλέον, προσομοιώθηκε ένα μοντέλο τυχαίας επιλογής, το οποίο επέλεγε δυναμικά είτε το Μοντέλο Α είτε το Μοντέλο Β για την κατηγοριοποίηση των εισερχομένων πακέτων. Η σφαιρική αξιολόγηση και ανάλυση αυτών των μοντέλων παρήγαγαν αξιοσημείωτα ευρήματα, στον τομέα της δυναμικής διαχείρισης δεδομένων. Το Μοντέλο Α, έδειξε να αποδίδει καλύτερα σε μικρομεσαία δίκτυα κόμβων με μεσαίο προς μεγάλο βαθμό ευαισθησίας (θ). Η απόδοσή του παρόλα αυτά επισκιάστηκε από το βελτιωμένο Μοντέλο Α, το οποίο πέτυχε το απόλυτο 100% σε ακρίβεια για τιμές του k > 3, ενώ παράλληλα εμφάνισε και μικρότερους χρόνους σε πολλές από τις περιπτώσεις έναντι του Β, κερδίζοντάς το τόσο σε ακρίβεια όσο και σε μέσο χρόνο εκτέλεσης. Αυτό πιθανόν να οφείλεται στο ότι το Μοντέλο Α χρησιμοποιεί τη παράθεση ότι τα εισερχόμενα πακέτα διαμορφώνουν τα φίλτρα δεδομένων με τις τιμές που είναι κεντραρισμένες γύρω από μια τιμή c_j . Αυτή η προσέγγιση, ενώ παρέχει μια λεπτομερή εξέταση των διαστημάτων, εισάγει έναν προσανατολισμό προς τιμές που είναι συγκεντρωμένες κοντά στο κέντρο του διαστήματος, γεγονός που μπορεί να δημιουργήσει ασυμφωνίες στις τελικές μετρήσεις ομοιότητας. Αντίθετα, το Μοντέλο Β υιοθετεί μια πιο αποσπασμένη και γενική προσέγγιση για τον υπολογισμό αποστάσεων, επιτρέποντας μια πιο ευρεία και λιγότερο προκατειλημμένη θεώρηση που του δίνει πλεονέκτημα στη καλύτερη αντιστοιχία. Τέλος το υβριδικό μοντέλο, ενώ κατάφερε να επιφέρει καλύτερα αποτελέσματα ακρίβειας σε σχέση με το Μοντέλο Α, παρουσίασε αστάθεια στις χρονικές του μετρήσεις σε σύγκριση με τη χρήση των μοντέλων μεμονωμένα.

Οι αποδόσεις και των 2 μοντέλων αφήνουν περιθώριο για βελτιστοποίηση μέσω της περαιτέρω δοκιμής των παραμέτρων με σκοπό τη βελτίωση του χρόνου εκτέλεσης. Ως επέκταση της παρούσας μελέτης, ένας σημαντικός τομέας που θα εξεταστεί σε μελλοντική έρευνα είναι η καλύτερη αξιοποίηση των φίλτρων για την εξαγωγή γνώσης. Αυτή η προσέγγιση αποσκοπεί στην πολύπλευρή αντιμετώπιση της ετερογένειας σε δίκτυα IoT, χρησιμοποιώντας πιο αποδοτικά τα υπάρχοντα

δεδομένα των υπάρχοντών κόμβων για τη δημιουργία προβλέψεων. Η δυναμική αξιοποίηση των φίλτρων θα επιτρέψει την πιο αποτελεσματική προσαρμογή των κόμβων στις μεταβαλλόμενες απαιτήσεις του περιβάλλοντος, αυξάνοντας την ευελιξία και την αποδοτικότητα του συστήματος.

.ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] H. Hua, Y. Li, T. Wang, N. Dong, W. Li, and J. Cao, "Edge Computing with Artificial Intelligence: A Machine Learning Perspective," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–35, Sep. 2023, doi: 10.1145/3555802.

[2] T. Ouyang, X. Chen, L. Zeng, and Z. Zhou, "Cost-Aware Edge Resource Probing for Infrastructure-Free Edge Computing: From Optimal Stopping to Layered Learning," in *2019 IEEE Real-Time Systems Symposium (RTSS)*, Hong Kong, Hong Kong: IEEE, Dec. 2019, pp. 380–391. doi: 10.1109/RTSS46320.2019.00041.

[3] "Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper," Cisco. Accessed: Nov. 06, 2023. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[4] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, and M. Imran, "The Role of Edge Computing in Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 110–115, Nov. 2018, doi: 10.1109/MCOM.2018.1700906.

[5] M. H. A. Awadalla, "Task Mapping and Scheduling in Wireless Sensor Networks," 2013.

[6] M. Bal, W. Shen, and H. Ghenniwa, "Collaborative signal and information processing in wireless sensor networks: A review," in *2009 IEEE International Conference on Systems, Man and Cybernetics*, San Antonio, TX, USA: IEEE, Oct. 2009, pp. 3151–3156. doi: 10.1109/ICSMC.2009.5346152.

[7] K. Muthukrishnan, M. Lijding, και P. Havinga, 'Towards Smart Surroundings: Enabling Techniques and Technologies for Localization', στο *Location- and Context-Awareness*, T. Strang και C. Linnhoff-Popien, Επιμ., στο Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, σσ. 350–362. doi: 10.1007/11426646_32.

[8] L. Hu και D. Evans, 'Localization for mobile sensor networks', στο *Proceedings of the 10th annual international conference on Mobile computing and networking*, Philadelphia PA USA: ACM, Σεπτεμβρίου 2004, σσ. 45–57. doi: 10.1145/1023720.1023726.

{9] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*. Springer Science & Business Media, 2012.

[10] P. Zhou, W. Chen, S. Ji, H. Jiang, L. Yu, and D. Wu, "Privacy-Preserving Online Task Allocation in Edge-Computing-Enabled Massive Crowdsensing," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7773–7787, Oct. 2019, doi: 10.1109/JIOT.2019.2903515.

[11] E. I. Gaura, J. Brusey, M. Allen, R. Wilkins, D. Goldsmith, and R. Rednic, "Edge Mining the Internet of Things," *IEEE Sensors J.*, vol. 13, no. 10, pp. 3816–3825, Oct. 2013, doi: 10.1109/JSEN.2013.2266895.

[12] L. Lei, H. Xu, X. Xiong, K. Zheng, W. Xiang, and X. Wang, "Multi-user Resource Control with Deep Reinforcement Learning in IoT Edge Computing." arXiv, Jun. 18, 2019. Accessed: Nov. 07, 2023. [Online]. Available: http://arxiv.org/abs/1906.07860

[13] F. Khan, A. Ur Rehman, J. Zheng, M. A. Jan, and M. Alam, "Mobile crowdsensing: A survey on privacy-preservation, task management, assignment models, and incentives mechanisms," *Future Generation Computer Systems*, vol. 100, pp. 456–472, Nov. 2019, doi: 10.1016/j.future.2019.02.014.

[14] M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu, "Multi-task assignment for crowdsensing in mobile social networks," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, Kowloon, Hong Kong: IEEE, Apr. 2015, pp. 2227–2235. doi: 10.1109/INFOCOM.2015.7218609.

[15] M. Xiao, J. Wu, L. Huang, R. Cheng, and Y. Wang, "Online Task Assignment for Crowdsensing in Predictable Mobile Social Networks," *IEEE Trans. on Mobile Comput.*, vol. 16, no. 8, pp. 2306–2320, Aug. 2017, doi: 10.1109/TMC.2016.2616473.

[16] G. Bajaj and P. Singh, "Mew: A Plug-n-Play Framework for Task Allocation in Mobile Crowdsensing," in *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications*, Delft Netherlands: ACM, Nov. 2017, pp. 19–24. doi: 10.1145/3139243.3139246.

[17] D. Verma και G. Sharma, ‘Massive Data Management in Healthcare Systems using Edge Computing and Internet of Things’, στο *2022 International Conference on Futuristic Technologies (INCOFT)*, Belgaum, India: IEEE, Νοεμβρίου 2022, σσ. 1–6. doi: 10.1109/INCOFT55651.2022.10094535.
[18] X. Li, X. Huang, C. Li, R. Yu, και L. Shu, ‘EdgeCare: Leveraging Edge Computing for Collaborative Data Management in Mobile Healthcare Systems’, *IEEE Access*, τ. 7, σσ. 22011–22025, 2019, doi: 10.1109/ACCESS.2019.2898265.
[19] M. N. K. Boulos, J. T. Wilson, and K. A. Clauson, ‘‘Geospatial blockchain: Promises, challenges, and scenarios in health and healthcare,’’ Int. J. Health Geogr., vol. 17, p. 25, Jul. 2018.
[20] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin, και W. Zhen, ‘A Blockchain-Based Trusted Data Management Scheme in Edge Computing’, *IEEE Trans. Ind. Inf.*, τ. 16, τχ. 3, σσ. 2013–2021, Μαρτίου 2020, doi: 10.1109/TII.2019.2933482.
[21] K. Kolomvatsos και C. Anagnostopoulos, ‘A Proactive Statistical Model Supporting Services and Tasks Management in Pervasive Applications’, *IEEE Trans. Netw. Serv. Manage.*, τ. 19, τχ. 3, σσ. 3020–3031, Σεπτεμβρίου 2022, doi: 10.1109/TNSM.2022.3161663.
[22] P. Bellavista, A. Corradi, L. Foschini, και D. Scotece, ‘Differentiated Service/Data Migration for Edge Services Leveraging Container Characteristics’, *IEEE Access*, τ. 7, σσ. 139746–139758, 2019, doi: 10.1109/ACCESS.2019.2943848.
[23] C. Anagnostopoulos, S. Hadjiefthymiades, και K. Kolomvatsos, ‘Accurate, Dynamic, and Distributed Localization of Phenomena for Mobile Sensor Networks’, *ACM Trans. Sen. Netw.*, τ. 12, τχ. 2, σσ. 1–59, Μαΐου 2016, doi: 10.1145/2882966.
[24] K. Kolomvatsos, C. Anagnostopoulos, και S. Hadjiefthymiades, ‘Data Fusion and Type-2 Fuzzy Inference in Contextual Data Stream Monitoring’, *IEEE Trans. Syst. Man Cybern, Syst.*, τ. 47, τχ. 8, σσ. 1839–1853, Αυγούστου 2017, doi: 10.1109/TSMC.2016.2560533.
[25] K. Kolomvatsos, C. Anagnostopoulos, και S. Hadjiefthymiades, ‘An efficient environmental monitoring system adopting data fusion, prediction, & fuzzy logic’, στο *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Corfu, Greece: IEEE, Ιουλίου 2015, σσ. 1–6. doi: 10.1109/IISA.2015.7388070.
[26] O. M. Elzeki, M. Z. Rashad, και M. A. Elsoud, ‘Overview of Scheduling Tasks in Distributed Computing Systems’, τ. 2, τχ. 3, 2012.
[27] L.-T. Hsieh, H. Liu, Y. Guo, και R. Gazda, ‘Task Management for Cooperative Mobile Edge Computing’, στο *2020 IEEE/ACM Symposium on Edge Computing (SEC)*, San Jose, CA, USA: IEEE, Νοεμβρίου 2020, σσ. 352–357. doi: 10.1109/SEC50012.2020.00051.
[28] A. Karanika, P. Oikonomou, K. Kolomvatsos, και T. Loukopoulos, ‘A Demand-driven, Proactive Tasks Management Model at the Edge’, στο *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Glasgow, United Kingdom: IEEE, Ιουλίου 2020, σσ. 1–8. doi: 10.1109/FUZZ48607.2020.9177653.
[29] K. Kolomvatsos και C. Anagnostopoulos, ‘Multi-criteria optimal task allocation at the edge’, *Future Generation Computer Systems*, τ. 93, σσ. 358–372, Απριλίου 2019, doi: 10.1016/j.future.2018.10.051.
[30] K. Bhardwaj, P. Agrawal, A. Gavrilovska, και K. Schwan, ‘AppSachet: Distributed App Delivery from the Edge Cloud’, στο *Mobile Computing, Applications, and Services*, τ. 162, S. Sigg, P. Nurmi, και F. Salim, Επιμ., στο Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 162., Cham: Springer International Publishing, 2015, σσ. 89–106. doi: 10.1007/978-3-319-29003-4_6.
[31] P. Simoens, Y. Xiao, P. Pillai, Z. Chen, K. Ha, και M. Satyanarayanan, ‘Scalable crowd-sourcing of video from mobile devices’, στο *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, Taipei Taiwan: ACM, Ιουνίου 2013, σσ. 139–152. doi: 10.1145/2462456.2464440.

[32] T. Wang, J. Zhou, A. Liu, M. Z. A. Bhuiyan, G. Wang, και W. Jia, 'Fog-Based Computing and Storage Offloading for Data Synchronization in IoT', *IEEE Internet Things J.*, τ. 6, τχ. 3, σσ. 4272–4282, Ιουνίου 2019, doi: 10.1109/JIOT.2018.2875915.
[33] Y. Yao, Q. Cao, και A. V. Vasilakos, 'EDAL: An Energy-Efficient, Delay-Aware, and Lifetime-Balancing Data Collection Protocol for Heterogeneous Wireless Sensor Networks', *IEEE/ACM Trans. Networking*, τ. 23, τχ. 3, σσ. 810–823, Ιουνίου 2015, doi: 10.1109/TNET.2014.2306592.
[34] A. Zhou, S. Wang, J. Li, Q. Sun, και F. Yang, 'Optimal mobile device selection for mobile cloud service providing', *J Supercomput*, τ. 72, τχ. 8, σσ. 3222–3235, Αυγούστου 2016, doi: 10.1007/s11227-016-1704-0.
[35] C. Anagnostopoulos, S. Hadjiefthymiades, και K. Kolomvatsos, 'Time-optimized user grouping in Location Based Services', *Computer Networks*, τ. 81, σσ. 220–244, Απριλίου 2015, doi: 10.1016/j.comnet.2015.02.017.
[36] S. Sardellitti, G. Scutari, και S. Barbarossa, 'Joint Optimization of Radio and Computational Resources for Multicell Mobile-Edge Computing', *IEEE Trans. on Signal and Inf. Process. over Networks*, τ. 1, τχ. 2, σσ. 89–103, Ιουνίου 2015, doi: 10.1109/TSIPN.2015.2448520.
[37] M. Breitbach, D. Schafer, J. Edinger, και C. Becker, 'Context-Aware Data and Task Placement in Edge Computing Environments', στο *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, Kyoto, Japan: IEEE, Μαρτίου 2019, σσ. 1–10. doi: 10.1109/PERCOM.2019.8767386.
[38] K. Kolomvatsos, C. Anagnostopoulos, M. Koziri, και T. Loukopoulos, 'Proactive & Time-Optimized Data Synopsis Management at the Edge', *IEEE Trans. Knowl. Data Eng.*, σσ. 1–1, 2020, doi: 10.1109/TKDE.2020.3021377.
[39] K. Kolomvatsos, 'A proactive inference scheme for data-aware decision making in support of pervasive applications', *Future Generation Computer Systems*, τ. 136, σσ. 193–204, Νοεμβρίου 2022, doi: 10.1016/j.future.2022.05.018.
[40] G. Boulougaris και K. Kolomvatsos, 'A QoS-aware, Proactive Tasks Offloading Model for Pervasive Applications', στο *2022 9th International Conference on Future Internet of Things and Cloud (FiCloud)*, Rome, Italy: IEEE, Αυγούστου 2022, σσ. 24–31. doi: 10.1109/FiCloud57274.2022.00011.
[41] Y. Yu και V. K. Prasanna, 'Energy-balanced task allocation for collaborative processing in networked embedded systems', στο *Proceedings of the 2003 ACM SIGPLAN conference on Language, compiler, and tool for embedded systems*, San Diego California USA: ACM, Ιουνίου 2003, σσ. 265–274. doi: 10.1145/780732.780769.
[42] A. Razavinegad, 'Task Allocation In Robot Mobile Wireless Sensor Networks', τ. 3, τχ. 6, 2014.
[43] J. Yang, H. Zhang, Y. Ling, C. Pan, και W. Sun, 'Task Allocation for Wireless Sensor Network Using Modified Binary Particle Swarm Optimization', *IEEE Sensors J.*, τ. 14, τχ. 3, σσ. 882–892, Μαρτίου 2014, doi: 10.1109/JSEN.2013.2290433.
[44] X. Hu και B. Xu, 'Task Allocation Mechanism Based on Genetic Algorithm in Wireless Sensor Networks', στο *Applied Informatics and Communication*, D. Zeng, Επιμ., στο Communications in Computer and Information Science. Berlin, Heidelberg: Springer, 2011, σσ. 46–58. doi: 10.1007/978-3-642-23214-5_7.
[45] Coltin, B., and N. Veloso. "'Mobile Robot Task Allocation in Hybrid Wireless Sensors Networks', in 2010 Int." *Conf. on Intelligent Robots and Systems*. 2010.
[46] K. Kolomvatsos και C. Anagnostopoulos, 'A probabilistic model for assigning queries at the edge', *Computing*, τ. 102, τχ. 4, σσ. 865–892, Απριλίου 2020, doi: 10.1007/s00607-019-00767-8.
[47] M. Soula, A. Karanika, K. Kolomvatsos, C. Anagnostopoulos, και G. Stamoulis, 'Intelligent tasks allocation at the edge based on machine learning and bio-inspired algorithms', *Evolving Systems*, τ. 13, τχ. 2, σσ. 221–242, Απριλίου 2022, doi: 10.1007/s12530-021-09379-0.
[48] A. Rath, A. Hristoskova, και S. Klein, 'PFilter: Privacy-Aware and Secure Data Filtering at the Edge for Distributed Edge Analytics', στο *Artificial Intelligence Applications and Innovations. AIAI 2021 IFIP WG 12.5 International Workshops*, I. Maglogiannis, J. Macintyre, και L. Iliadis, Επιμ., στο IFIP

Advances in Information and Communication Technology. Cham: Springer International Publishing, 2021, σσ. 305–310. doi: 10.1007/978-3-030-79157-5_25.
[49] S. Zoller, A. Reinhardt, και R. Steinmetz, 'Distributed data filtering in logistics wireless sensor networks based on transmission relevance', στο *37th Annual IEEE Conference on Local Computer Networks*, Clearwater Beach, FL, USA: IEEE, Οκτωβρίου 2012, σσ. 256–259. doi: 10.1109/LCN.2012.6423622.
[50] T. Xia, C. Jin, X. Zhou, και A. Zhou, 'Filtering Duplicate Items over Distributed Data Streams', στο *Advances in Web-Age Information Management*, W. Fan, Z. Wu, και J. Yang, Επιμ., στο Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, σσ. 779–784. doi: 10.1007/11563952_80.
[51] S. Deng, D. Yue, A. Zhou, X. Fu, L. Yang, και Y. Xue, 'Distributed content filtering algorithm based on data label and policy expression in active distribution networks', *Neurocomputing*, τ. 270, σσ. 159–169, Δεκεμβρίου 2017, doi: 10.1016/j.neucom.2017.03.087.
[52] G. Koloniari, N. Ntarmos, E. Pitoura, και D. Souravlias, 'One is enough: distributed filtering for duplicate elimination', στο *Proceedings of the 20th ACM international conference on Information and knowledge management*, Glasgow Scotland, UK: ACM, Οκτωβρίου 2011, σσ. 433–442. doi: 10.1145/2063576.2063643.
[53] H. Nikpey Somehsaraei, S. Ghosh, S. Maity, P. Pramanik, S. De, και M. Assadi, 'Automated Data Filtering Approach for ANN Modeling of Distributed Energy Systems: Exploring the Application of Machine Learning', *Energies (19961073)*, τ. 13, τχ. 14, σ. 3750, Ιουλίου 2020, doi: 10.3390/en13143750.
[54] X. Xu και W. Dai, 'Data Acquisition, Filtering and Buffering Protocol Design for Edge Computing Nodes', στο *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, Palma de Mallorca, Spain: IEEE, Ιουλίου 2021, σσ. 1–6. doi: 10.1109/INDIN45523.2021.9557533.
[55] Y. J. Kim, C.-H. Park, και M. Yoon, 'FILM: Filtering and Machine Learning for Malware Detection in Edge Computing', *Sensors*, τ. 22, τχ. 6, Art. τχ. 6, Ιανουαρίου 2022, doi: 10.3390/s22062150.
[56] F. Li και D. Wang, '5G Network Data Migration Service Based on Edge Computing', *Symmetry*, τ. 13, τχ. 11, Art. τχ. 11, Νοεμβρίου 2021, doi: 10.3390/sym13112134.
[57] X. Peng, R. Chen, K. Yu, F. Ye, και W. Xue, 'An Improved Weighted K-Nearest Neighbor Algorithm for Indoor Localization', *Electronics*, τ. 9, τχ. 12, Art. τχ. 12, Δεκεμβρίου 2020, doi: 10.3390/electronics9122117.
[58] K. Kolomvatsos, C. Anagnostopoulos, and S. Hadjiefthymiades, "A time optimized scheme for top-k list maintenance over incomplete data streams," *Information Sciences*, vol. 311, pp. 59–73, Aug. 2015, doi: 10.1016/j.ins.2015.03.035.
[59] Q. Li *et al.*, "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3347–3366, Apr. 2023, doi: 10.1109/TKDE.2021.3124599.
[60] N. Harth, C. Anagnostopoulos, H.-J. Voegel, and K. Kolomvatsos, "Local & Federated Learning at the network edge for efficient predictive analytics," *Future Generation Computer Systems*, vol. 134, pp. 107–122, Sep. 2022, doi: 10.1016/j.future.2022.03.030.
[61] N. A. Sulieman, L. Ricciardi Celsi, W. Li, A. Zomaya, and M. Villari, "Edge-Oriented Computing: A Survey on Research and Use Cases," *Energies (19961073)*, vol. 15, no. 2, p. 452, Jan. 2022, doi: 10.3390/en15020452.
[62] D. Curran-Everett, 'Explorations in statistics: the assumption of normality', *Advances in Physiology Education*, τ. 41, τχ. 3, σσ. 449–453, Σεπτεμβρίου 2017, doi: 10.1152/advan.00064.2017.
[63] T. K. Kim και J. H. Park, 'More about the basic assumptions of t-test: normality and sample size', *Korean J Anesthesiol*, τ. 72, τχ. 4, σσ. 331–335, Απριλίου 2019, doi: 10.4097/kja.d.18.00292.
[64] Y.-A. L. Borgne, 'Bias-Variance trade-off characterization in a classification problem What differences with regression ?'.