



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΣΥΝΕΡΓΑΤΙΚΗ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ
ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΠΑΡΥΦΕΣ
ΤΟΥ ΔΙΚΤΥΟΥ

Καραγεωργίου Διονύσης

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κολομβάτσος Κωνσταντίνος
Επίκουρος Καθηγητής

Λαμία Ιούλιος 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΣΥΝΕΡΓΑΤΙΚΗ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ
ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΠΑΡΥΦΕΣ
ΤΟΥ ΔΙΚΤΥΟΥ

Καραγεωργίου Διονύσης

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κολομβάτσος Κωνσταντίνος
Επίκουρος Καθηγητής

Λαμία 2024



UNIVERSITY OF
THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

Cooperative Data Storage and Management at the Network Edge

Karageorgiou Dionysis

FINAL THESIS

ADVISOR

Kolomvatsos Konstantinos
Assistant Professor

Lamia 2024

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων Συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

ΠΕΡΙΛΗΨΗ

Στη παρούσα εργασία προσομοιώνεται ένα σύστημα δικτύου από EdgeServers που εκτελούν εργασίες ανάλυσης δεδομένων, χρησιμοποιώντας τεχνικές μηχανικής μάθησης πιο συγκεκριμένα μέσω του αλγόριθμου Isolation Forest υλοποιείται ανίχνευση ανωμαλιών. Οι EdgeServers υπολογίζουν στατιστικά δεδομένα, εντοπίζουν outliers και επικοινωνούν μεταξύ τους για τη βελτίωση της απόδοσης του δικτύου. Κάθε EdgeServer μοιράζεται τα στατιστικά του στοιχεία με τους άλλους servers, προκειμένου να δημιουργήσει ένα ιστορικό και να βελτιστοποιήσει την ανίχνευση των outliers. Μέσω πειραμάτων με διαφορετικά πλήθη EdgeServers και διαστάσεις των δεδομένων γίνεται αξιολόγηση της απόδοσης του συστήματος, με σκοπό την εξομάλυνση της διαχείρισης και ανάλυσης δεδομένων σε ένα δίκτυο EdgeServers, επιτυγχάνοντας καλύτερη απόδοση και αξιοπιστία.

ABSTRACT

In this work, a network system of EdgeServers is simulated, performing data analysis tasks using machine learning techniques, specifically anomaly detection implemented through the Isolation Forest algorithm. The EdgeServers calculate statistical data, identify outliers, and communicate with each other to improve the network's performance. Each EdgeServer shares its statistical data with other servers to create a historical record and optimize outlier detection. Through experiments with different numbers of EdgeServers and data dimensions, the system's performance is evaluated, aiming to streamline the management and analysis of data in an EdgeServer network, achieving better performance and reliability.

Table of Contents

ΠΕΡΙΛΗΨΗ	I
ABSTRACT	III
<u>ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ</u>	<u>1</u>
(ΕΙΣΑΓΩΓΗ ΣΤΟ EDGE COMPUTING 1.1)	1
(EDGE NETWORK 1.2).....	1
(ΣΤΟΧΟΙ ΚΑΙ ΚΙΝΗΤΡΑ 1.3)	2
<u>ΚΕΦΑΛΑΙΟ 2 ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΠΑΡΥΦΕΣ ΤΟΥ ΔΙΚΤΟΥ.....</u>	<u>3</u>
(PAPERS 2.1)	3
<u>ΚΕΦΑΛΑΙΟ 3 ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....</u>	<u>9</u>
(ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ 3.1)	9
(ΑΛΓΟΡΙΘΜΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ 3.2).....	9
<u>ΚΕΦΑΛΑΙΟ 4 ΠΡΟΤΕΙΝΟΜΕΝΟ ΣΥΣΤΗΜΑ.....</u>	<u>11</u>
(ΥΛΟΠΟΙΗΣΕΙΣ ΤΟΥ ΚΩΔΙΚΑ 4.1)	11
<u>ΚΕΦΑΛΑΙΟ 5 ΠΕΙΡΑΜΑΤΙΚΗ ΑΠΟΤΙΜΗΣΗ.....</u>	<u>14</u>
<u>ΚΕΦΑΛΑΙΟ 6 ΣΥΜΠΕΡΑΣΜΑΤΑ&ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ</u>	<u>32</u>
<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>	<u>34</u>

ΚΕΦΑΛΑΙΟ 1 Εισαγωγή

(ΕΙΣΑΓΩΓΗ ΣΤΟ EDGE COMPUTING 1.1)

Το Edge Computing είναι μια επέκταση του Cloud Computing. Περιλαμβάνει ένα κατανεμημένο μοντέλο υπολογιστών που τοποθετεί τη διαχείριση και την αποθήκευση δεδομένων πιο κοντά στις πηγές των δεδομένων. Σε αντίθεση με το Cloud Computing, το οποίο συγκεντρώνει τα δεδομένα σε κεντρικούς servers, το Edge Computing φέρνει την επεξεργασία πιο κοντά στον χρήστη. Αυτό μειώνει την καθυστέρηση σε σύγκριση με την εκτέλεση εφαρμογών σε ένα απομακρυσμένο data center.

(EDGE NETWORK 1.2)

- Τι είναι;

Ένα edge network αποτελείται από μια αρχιτεκτονική δεδομένων που σχεδιάζεται μεθοδικά για να παρέχει υπολογιστική ισχύ στις συσκευές edge εντός ενός δικτύου. Αυτή η προσέγγιση ελαχιστοποιεί την ανάγκη για επεξεργαστική ισχύ από τους κεντρικούς διακομιστές, επιτρέποντας στις συσκευές να αναλαμβάνουν το μεγαλύτερο μέρος της επεξεργασίας.

- Πώς δουλεύει;

Το edge networking περιλαμβάνει την επεξεργασία δεδομένων και την παροχή ψηφιακών πόρων στους χρήστες μέσω της ένταξης συσκευών edge και τοπικών διακομιστών με το Cloud. Αυτή η διάταξη επιτρέπει τη διαχείριση, αποθήκευση και ανάλυση δεδομένων, με τους χρήστες να έχουν πρόσβαση στις συσκευές μέσω του Cloud για την εκτέλεση μικρών εργασιών επεξεργασίας. Η δημιουργία ενός ολοκληρωμένου δικτύου συσκευών διευκολύνει τη διαχείριση μεγαλύτερων φορτίων εργασίας και την αντιμετώπιση προβλημάτων από το edge προς το Cloud αποτελεσματικά, χωρίς την υπερβολική εξάρτηση από τους πόρους μεγάλων κέντρων δεδομένων. Αυτή η προσέγγιση λειτουργεί μέσω ασφαλών συνδέσεων μεταξύ των διακομιστών μιας οργανικής δομής και των συσκευών IoT, δημιουργώντας μια ευρύτερη και πιο αποδοτική αρχιτεκτονική δεδομένων. Με την αυξανόμενη συνδεσιμότητα και τις διαφορετικές ανάγκες των σύγχρονων τομέων, η τεχνολογία edge είναι απαραίτητη για την επιτυχία των επιχειρήσεων. Ασφαρίζει τις αρχιτεκτονικές δεδομένων, εξυπηρετώντας τόσο τους καταναλωτές όσο και τις οργανώσεις, επιτρέποντας στις συνδεδεμένες συσκευές να διαχειρίζονται αυτόνομα τα δεδομένα, μειώνοντας την εξάρτηση από τους κεντρικούς διακομιστές και βελτιώνοντας την ικανότητα αντιμετώπισης προβλημάτων.

- Χαρακτηριστικά

- α. Προσφέρει γρήγορη σύνδεση και ευέλικτη ανάπτυξη εφαρμογών, μειώνοντας ταυτόχρονα την καθυστέρηση.
- β. Προάγει μια ομαλή εμπειρία χρήστη, εξασφαλίζοντας υψηλότερη ικανοποίηση των καταναλωτών.
- γ. Υποστηρίζει μια ευρεία γκάμα συσκευών IoT εντός της αρχιτεκτονικής/δικτύου δεδομένων.
- δ. Δημιουργεί συγχρονισμό και πιθανή αναγκαία αντιγραφή δεδομένων από το edge στο Cloud.
- ε. Παρέχει χώρο και επαληθεύει όλη την κυκλοφορία δεδομένων για βελτιστοποίηση της ανάλυσης.
- στ. Μειώνει τα λειτουργικά έξοδα.
- ζ. Παρέχει ισχυρή υποστήριξη μέσω της ένταξης απομακρυσμένων και τοπικών υπηρεσιών.
- η. Βελτιώνει ταχύτητα και μέτρα ασφαλείας.

(Στόχοι και Κίνητρα 1.3)

Ο στόχος αυτής της διπλωματικής εργασίας είναι η διαχείριση και η αποθήκευση δεδομένων στην παρυφή του δικτύου, εμπνευσμένη από τα ακόλουθα κίνητρα:

- **Ανάγκη για Αποδοτική Διαχείριση Δεδομένων:** Οι παραδοσιακές αρχιτεκτονικές αποθήκευσης δεδομένων δεν είναι κατάλληλες για εφαρμογές που απαιτούν γρήγορους χρόνους απόκρισης και χαμηλή καθυστέρηση. Οι δίκτυα Edge Computing επιτρέπουν την αποθήκευση και επεξεργασία δεδομένων κοντά στην πηγή τους, μειώνοντας την καθυστέρηση και βελτιώνοντας την απόδοση.
- **Αυξανόμενη Ανάγκη για Αυτοματοποιημένη Διαχείριση:** Η ανάπτυξη ενός συστήματος που χρησιμοποιεί μηχανική μάθηση για την ανίχνευση ανωμαλιών στα δεδομένα των EdgeServers θα επιτρέψει την αυτόματη αντίδραση σε προβλήματα και αποτυχίες, βελτιώνοντας έτσι την αξιοπιστία και τη διαθεσιμότητα του συστήματος.
- **Προσαρμοστικότητα και Αποδοτικότητα:** Οι τεχνολογίες Edge Computing προσφέρουν ευελιξία στην ανάπτυξη και εκτέλεση εφαρμογών, ενώ ταυτόχρονα εξοικονομούν εύρος ζώνης και μειώνουν τη φόρτιση των κεντρικών υπολογιστικών πόρων.

Στόχοι της διπλωματικής εργασίας είναι:

- Βελτιστοποίηση της Απόδοσης: Ανάπτυξη ενός συστήματος διαχείρισης δεδομένων με EdgeServers που προσφέρει βελτιωμένη απόδοση σε σχέση με τις παραδοσιακές αρχιτεκτονικές.
- Αυτοματοποίηση και Αυτονομία: Εφαρμογή τεχνικών μηχανικής μάθησης για την ανίχνευση και διαχείριση ανωμαλιών στα δεδομένα, με στόχο την αυτόματη αντίδραση για την ενίσχυση της αξιοπιστίας του συστήματος.
- Αξιολόγηση Απόδοσης και Στοιχείων: Αξιολόγηση της απόδοσης του συστήματος σε διάφορες συνθήκες και φορτία εργασίας μέσω πειραμάτων, με στόχο τη βελτιστοποίηση των συστατικών του.
- Ανάπτυξη Καινοτόμων Τεχνολογιών: Συμβολή στην προαγωγή των τεχνολογιών Edge Computing με την πρόταση νέων προσεγγίσεων και βελτιώσεων.

ΚΕΦΑΛΑΙΟ 2 Διαχείριση δεδομένων στις παρυφές του δικτύου

(Papers 2.1)

1. "Edge Computing Applications: State-of-the-Art and Challenges"

Το άρθρο εξετάζει τις διάφορες χρήσεις, τις τελευταίες εξελίξεις και τις δυσκολίες που αντιμετωπίζει το edge computing. Αναλύει πώς εφαρμόζεται σε τομείς όπως το IoT, η υγειονομική περίθαλψη και τα αυτόνομα οχήματα, τονίζοντας την πρόοδο σε τεχνολογικά πρότυπα επικοινωνίας και συστήματα αποθήκευσης. Επιπλέον, αναφέρεται σε κρίσιμες προκλήσεις όπως η ασφάλεια, η διαχείριση δεδομένων και η ενεργειακή αποδοτικότητα.

Μέθοδοι διαχείρισης δεδομένων:

- Υπολογιστική Εκφόρτωση: Αναφέρεται στη μεταφορά υπολογιστικών εργασιών με στόχο την ενίσχυση της ενεργειακής και υπολογιστικής απόδοσης, καθώς και της αποθήκευσης. Αυτή η μεταφορά γίνεται από τις τελικές συσκευές προς τους EdgeServers.
- Διαχείριση Κινητικότητας: Περιλαμβάνει την επιλογή και μεταφορά πόρων για τη βελτίωση της κινητικότητας των χρηστών χωρίς να διακοπούν οι τρέχουσες δραστηριότητες.
- Ασφάλεια: Αντιμετωπίζει τους περιορισμούς των πόρων των edge συσκευών μέσω της προστασίας δεδομένων και της διαχείρισης της ιδιωτικότητας.

- Αλληλεπίδραση: Βεβαιώνει την αποτελεσματική διαχείριση μέσω της επιτυχούς αλληλεπίδρασης μεταξύ διαφορετικών συστημάτων και συσκευών edge. των δεδομένων.

2. "Edge Computing in Master Data Management: Enhancing Data Processing at the Source"

Το άρθρο εξετάζει πώς η τεχνολογία edge computing βελτιώνει τη διαχείριση των βασικών δεδομένων (master data) με την τοπική επεξεργασία δεδομένων σε περιφερειακούς κόμβους ή συσκευές. Αυτή η προσέγγιση μειώνει τους χρόνους απόκρισης, βελτιώνει την αποτελεσματικότητα και μειώνει την ανάγκη επικοινωνίας, ενώ ενισχύει την ασφάλεια των δεδομένων.

Μέθοδοι διαχείρισης δεδομένων:

- Επεξεργασία δεδομένων στην πηγή: Τα δεδομένα επεξεργάζονται κοντά στην πηγή τους για να μειωθούν οι μεταφορές και να βελτιωθεί η απόκριση σε πραγματικό χρόνο.
- Συνεργασία με το MDM: Ενσωμάτωση της διαχείρισης των κύριων δεδομένων (MDM) στο edge computing για τη βελτίωση της ποιότητας και της αξιοπιστίας των δεδομένων.
- Μείωση της καθυστέρησης: Χρήση διανεμημένης επεξεργασίας για τη μείωση των καθυστερήσεων και τη βελτίωση των διεργασιών σε πραγματικό χρόνο.
- Διαχείριση ασφάλειας και διακυβέρνησης δεδομένων: Αντιμετώπιση προβλημάτων ασφάλειας και διαχείρισης της πολυπλοκότητας των δεδομένων.

3. "EdgeCare: Leveraging Edge Computing for Collaborative Data Management in Mobile Healthcare Systems"

Το άρθρο εξετάζει τη διαχείριση δεδομένων στο πλαίσιο του edge computing, εστιάζοντας στο EdgeCare - ένα σύστημα σχεδιασμένο για την αποκεντρωμένη και συνεργατική διαχείριση δεδομένων υγείας.

Προσεγγίσεις Διαχείρισης Δεδομένων:

- Ανταπόκριση σε Πραγματικό Χρόνο: Οι EdgeServers μειώνουν τους χρόνους απόκρισης διαχειριζόμενοι εργασίες διαχείρισης δεδομένων υγείας. Οι χρήστες μπορούν να έχουν πρόσβαση σε αποτελέσματα αναλύσεων, να ανακτούν ακατέργαστα δεδομένα και να συμμετέχουν σε συναλλαγές δεδομένων κατά παραγγελία.
- Ισορροπία Φορτίου: Οι Τοπικές Αρχές (LAs) αναθέτουν EdgeServers για τη διαχείριση τοπικών δεδομένων υγείας. Αυτοί οι γεωγραφικά κατανομημένοι διακομιστές συνεργάζονται για να επεξεργάζονται τα δεδομένα υγείας

ταυτόχρονα, υποστηρίζοντας την κλιμακούμενη διαχείριση δικτύου για να αντιμετωπίσουν την εκθετική αύξηση των δεδομένων.

- **Ασφαλής Προστασία Δεδομένων:** Σε περιβάλλοντα edge computing, η μετάδοση, η πρόσβαση και η αποθήκευση δεδομένων πραγματοποιούνται τοπικά, διευκολύνοντας την προστασία των δεδομένων υγείας καθώς η επεξεργασία γίνεται τοπικά.
- **Αποτελεσματικές Συναλλαγές Δεδομένων:** Οι συλλέκτες και οι κάτοχοι δεδομένων μπορούν εύκολα να πραγματοποιούν συναλλαγές δεδομένων, επωφελούμενοι από τις συντονισμένες λειτουργίες των edge servers και τις συνεχείς συνδέσεις.
- **Ιεραρχική Αρχιτεκτονική:** Το EdgeCare εξασφαλίζει ασφαλή και αποδοτική διαχείριση δεδομένων σε ένα αποκεντρωμένο κινητό περιβάλλον, χρησιμοποιώντας μια ιεραρχική αρχιτεκτονική προσέγγιση.
- **Ασφαλής Μεταφορά Δεδομένων και Έλεγχος Πρόσβασης:** Το σύστημα χρησιμοποιεί πρωτόκολλα με κρυπτογραφικές τεχνικές για να εξασφαλίσει την ασφαλή μεταφορά δεδομένων και την προσαρμοσμένη διαχείριση πρόσβασης.
- **Μηχανισμοί Κινήτρων για Συναλλαγές Δεδομένων:** Το EdgeCare χρησιμοποιεί έναν αλγόριθμο βελτιστοποίησης βασισμένο στο παιχνίδι Stackelberg για να σχεδιάσει μηχανισμούς κινήτρων για τους συλλέκτες δεδομένων και τους χρήστες, προωθώντας τη συμμετοχή σε συναλλαγές δεδομένων.

4. "Task Allocation Methods and Optimization Techniques in Edge Computing: A Systematic Review of the Literature"

Το άρθρο εξετάζει τις διάφορες μεθόδους ανάθεσης εργασιών και τις τεχνικές βελτιστοποίησης που εφαρμόζονται στον τομέα του edge computing. Συγκεκριμένα εξετάζεται πώς είναι δυνατόν να γίνει αποτελεσματική η κατανομή εργασιών σε περιφερειακούς κόμβους ή συσκευές (edge devices), καθώς και τις τεχνικές που εφαρμόζονται για τη βελτιστοποίηση της απόδοσης και της ενεργειακής απόδοσης σε αυτό το περιβάλλον.

- Προσεγγίσεις Διαχείρισης Δεδομένων:

- Στο άρθρο αναφέρονται επιπρόσθετα διάφορες μέθοδοι κατανομής εργασιών και τεχνικές βελτιστοποίησης που χρησιμοποιούνται ειδικά για το edge computing.
- ο Μέθοδοι Διαχείρισης Δεδομένων:
- Κατανομή Εργασιών: Κατανομή βασισμένη στους διαθέσιμους πόρους: Οι εργασίες διανέμονται βάσει των διαθέσιμων πόρων στις συσκευές edge, συμπεριλαμβανομένης της υπολογιστικής ισχύος, της μνήμης και της διάρκειας της μπαταρίας.
 - Διανεμημένη κατανομή εργασιών: Οι εργασίες ανατίθενται για τη βελτιστοποίηση της απόδοσης και τη μείωση των καθυστερήσεων.
 - Δυναμική κατανομή εργασιών: Πραγματοποιούνται προσαρμογές σε πραγματικό χρόνο βάσει αλλαγών στην απόδοση των συσκευών ή τις συνθήκες του δικτύου.
 - Κατανομή εργασιών βασισμένη σε μηχανική μάθηση: Χρησιμοποιούνται αλγόριθμοι για πρόβλεψη της χρήσης πόρων και την κατανομή εργασιών ανάλογα.
 - Ενεργειακά αποδοτική κατανομή εργασιών: Οι εργασίες ανατίθενται με τρόπο που ελαχιστοποιεί την κατανάλωση ενέργειας στις συσκευές edge.
 - Κατανομή εργασιών με προσανατολισμό στην ποιότητα υπηρεσιών: Οι εργασίες ανατίθενται βάσει των απαιτούμενων επιπέδων ποιότητας υπηρεσιών.
 - Συνεργατική κατανομή εργασιών: Οι εργασίες ανατίθενται λαμβάνοντας υπόψη τη συνεργασία μεταξύ των συσκευών edge.
 - Κατανομή εργασιών βασισμένη στο πλαίσιο: Οι εργασίες ανατίθενται βάσει παραγόντων πλαισίου όπως η τοποθεσία της συσκευής, οι διαθέσιμοι πόροι και οι προτιμήσεις των χρηστών.
- ο Τεχνικές Βελτιστοποίησης:
- Αλγόριθμοι Μηχανικής Μάθησης: Χρήση αλγορίθμων όπως η ενίσχυση μάθησης για τη βελτιστοποίηση της κατανομής εργασιών.
 - Υβριδικοί Αλγόριθμοι: Συνδυασμός κεντρικών και αποκεντρωμένων προσεγγίσεων για αποδοτικές λύσεις.
 - Μεθοδολογίες Ευρηματικής Αναζήτησης: Χρήση μεθόδων όπως η Βελτιστοποίηση Αποσμίγματος Σμήνους (PSO) για τη βελτιστοποίηση της κατανομής εργασιών.

Η διαχείριση δεδομένων στο edge computing, όπως περιγράφεται στο άρθρο, απαιτεί την κατανομή των εργασιών σε συσκευές δικτύου για την ισορροπία των διαθέσιμων πόρων, τη μείωση της καθυστέρησης και τη βελτίωση της ενεργειακής απόδοσης. Οι μέθοδοι και οι τεχνικές που παρουσιάζονται στο άρθρο προσφέρουν ένα ευρύ φάσμα προσεγγίσεων για την επίτευξη αυτών των στόχων.

5. "Federated Learning in Edge Computing: A Systematic Survey"

Το paper ασχολείται με τη διαχείριση δεδομένων στο edge computing, επικεντρώνοντας στη χρήση του Federated Learning (FL) για αυτόν τον σκοπό. Το Federated Learning είναι μια προσέγγιση μηχανικής μάθησης που επιτρέπει την εκπαίδευση μοντέλων σε κατανεμημένα δεδομένα χωρίς τη μεταφορά των δεδομένων σε κεντρικούς servers. Αντίθετα, οι αλγόριθμοι εκπαίδευσης εκτελούνται τοπικά στις συσκευές edge, με μόνο τις ενημερωμένες παραμέτρους των μοντέλων να αποστέλλονται σε έναν κεντρικό server για συγχώνευση και ενημέρωση του παγκόσμιου μοντέλου.

Οι μέθοδοι διαχείρισης δεδομένων περιλαμβάνουν:

- **Τοπική Εκπαίδευση Δεδομένων:** Τα μοντέλα εκπαιδεύονται τοπικά στις συσκευές edge χρησιμοποιώντας τα δικά τους δεδομένα, αποφεύγοντας την ανάγκη μεταφοράς μεγάλων όγκων δεδομένων στο cloud.
- **Συγκέντρωση και Συγχώνευση Παραμέτρων:** Οι τοπικά εκπαιδευμένες παράμετροι των μοντέλων αποστέλλονται σε έναν κεντρικό server (ή edge server) όπου συγχωνεύονται για να ενημερώσουν το παγκόσμιο μοντέλο, διασφαλίζοντας ότι τα δεδομένα των χρηστών παραμένουν τοπικά και προστατεύονται καλύτερα από πλευράς ιδιωτικότητας και ασφάλειας.
- **Μείωση Επικοινωνιακού Κόστους:** Το FL μειώνει δραστικά το επικοινωνιακό κόστος, καθώς μεταφέρονται μόνο οι παράμετροι των μοντέλων αντί για τα ακατέργαστα δεδομένα, εξοικονομώντας εύρος ζώνης και μειώνοντας τις καθυστερήσεις.
- **Βελτίωση της Ιδιωτικότητας και της Ασφάλειας:** Η ιδιωτικότητα των δεδομένων βελτιώνεται καθώς τα ακατέργαστα δεδομένα δεν μεταφέρονται, και οι τεχνικές του FL διασφαλίζουν ότι τα προσωπικά δεδομένα των χρηστών παραμένουν στις συσκευές τους.

Συνολικά, το paper υπογραμμίζει τη σημασία του Federated Learning για την αντιμετώπιση των προκλήσεων επικοινωνιακού κόστους, ιδιωτικότητας και ασφάλειας στη διαχείριση δεδομένων στο edge computing.

6. Paper: "Identity Management and Access Control Based on Blockchain under Edge Computing for the Industrial Internet of Things"

Το άρθρο προτείνει τη χρήση του blockchain για τη διαχείριση ταυτότητας και τον έλεγχο πρόσβασης στο edge computing. Το blockchain προσφέρει μια αποκεντρωμένη και ασφαλή υποδομή, η οποία εξασφαλίζει την ακεραιότητα των δεδομένων και την προστασία από μη εξουσιοδοτημένες τροποποιήσεις.

Μέθοδοι Διαχείρισης Δεδομένων:

- **Blockchain για Διαχείριση Ταυτότητας:** Το blockchain χρησιμοποιείται για την ασφαλή και αποκεντρωμένη διαχείριση των ταυτοτήτων. Κάθε συσκευή ή χρήστης στο δίκτυο έχει μια μοναδική ταυτότητα που αποθηκεύεται στο blockchain. Αυτές οι ταυτότητες είναι κρυπτογραφημένες και υπό παρακολούθηση, εξασφαλίζοντας την ασφάλεια και την ακεραιότητά τους.
- **Έλεγχος Πρόσβασης με Έξυπνα Συμβόλαια:** Ο έλεγχος πρόσβασης επιτυγχάνεται μέσω έξυπνων συμβολαίων στο blockchain. Αυτά τα συμβόλαια περιέχουν τους κανόνες και τις πολιτικές πρόσβασης για τους χρήστες και τις συσκευές του δικτύου. Τα έξυπνα συμβόλαια ενεργοποιούνται αυτόματα όταν πληρούνται οι καθορισμένες προϋποθέσεις, εξασφαλίζοντας την ασφαλή και αυτοματοποιημένη διαχείριση πρόσβασης στα δεδομένα.
- **Ασφάλεια και Ακεραιότητα Δεδομένων:** Η ακεραιότητα των δεδομένων διασφαλίζεται μέσω της κρυπτογράφησης και της καταγραφής κάθε συναλλαγής στο blockchain. Κάθε αλλαγή στα δεδομένα πρέπει να εγκριθεί και να καταγραφεί στο blockchain, καθιστώντας αδύνατη την μη εξουσιοδοτημένη τροποποίηση ή παραποίηση των δεδομένων.
- **Αποκεντρωμένη Αρχιτεκτονική:** Η αποκεντρωμένη φύση του blockchain εξαλείφει την ανάγκη για κεντρικό διακομιστή ή διαχειριστή, μειώνοντας τα σημεία αποτυχίας και βελτιώνοντας την ανθεκτικότητα και την αξιοπιστία του συστήματος. Αυτό είναι ιδιαίτερα σημαντικό σε περιβάλλοντα edge computing όπου η διαθεσιμότητα και η ασφάλεια των υπηρεσιών είναι κρίσιμης σημασίας.
- **Βελτιστοποίηση Απόδοσης και Κλιμάκωσης:** Για την αντιμετώπιση των προκλήσεων απόδοσης και κλιμάκωσης, το σύστημα ενσωματώνει διάφορες βελτιστοποιήσεις, όπως η χρήση αποδοτικών αλγορίθμων συναίνεσης και η ελαχιστοποίηση του φόρτου δικτύου μέσω της κατανεμημένης επεξεργασίας.

ΚΕΦΑΛΑΙΟ 3 Μοντέλα Μηχανικής Μάθησης

(ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ 3.1)

Η Μηχανική Μάθηση είναι ένα κομμάτι της τεχνητής νοημοσύνης. Ο Tom M. Mitchell πρότεινε έναν ορισμό που έχει ευρεία αποδοχή: "Ένα πρόγραμμα υπολογιστή μαθαίνει από εμπειρία E σε σχέση με ένα σύνολο εργασιών T και ένα μέτρο απόδοσης P , εάν η απόδοσή του στις εργασίες T , όπως μετράται από το P , βελτιώνεται με την εμπειρία E ." Αυτός ο ορισμός είναι κρίσιμος για να κατανοήσουμε τη λειτουργία της μηχανικής μάθησης, σύμφωνα με την πρόταση του Alan Turing στο άρθρο "Computing Machinery and Intelligence", ότι το ζήτημα εάν οι μηχανές μπορούν να σκέφτονται μπορεί να αντικατασταθεί από το ερώτημα αν μπορούν να κάνουν αυτό που οι άνθρωποι (ως σκεπτικά οντότητες) κάνουν. Ο ορισμός του Arthur Samuel, "Πεδίο μελέτης που επιτρέπει στους υπολογιστές να μαθαίνουν χωρίς ρητό προγραμματισμό," εστιάζει στη διδασκαλία των υπολογιστών να μαθαίνουν από τα δεδομένα και να προχωρούν μέσω της εμπειρίας, αντί να είναι προγραμματισμένοι να το πράττουν. Μέσω της μηχανικής μάθησης, οι αλγόριθμοι μπορούν να αναγνωρίζουν μοτίβα, να κατανοούν συσχετίσεις και να προβλέπουν βασιζόμενοι σε δεδομένα, προκειμένου να λαμβάνουν βέλτιστες αποφάσεις.

Η Μηχανική Μάθηση χωρίζεται συνήθως σε τρεις κατηγορίες:

- Εποπτευόμενη Μάθηση (Supervised Learning)
- Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)
- Ενισχυτική Μάθηση (Reinforcement Learning)

(ΑΛΓΟΡΙΘΜΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ 3.2)

Στην παρούσα εργασία διαχειριζόμαστε δεδομένα στο Edge Computing χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Θα παρουσιάσουμε κάποιους αλγόριθμους που ταιριάζουν με την εργασία μας, καθώς και τον αλγόριθμο Isolation Forest που χρησιμοποιήσαμε για την ανίχνευση των outliers.

Isolation Forest: Ο αλγόριθμος αυτός είναι μια τεχνική μη εποπτευόμενης μάθησης καθιστώντας τον υπερούγχρονο αλγόριθμο ανίχνευσης outliers ο οποίος είναι απλός και ταυτόχρονα αποτελεσματικός. Αυτό που κάνει είναι να καταργεί τους outliers μέσα από ένα σύνολο δεδομένων μέσω της δυαδικής διαίρεσης, να εντοπίζει τάχιστα τις αποκλίσεις με ελάχιστο υπολογιστικό κόστος καθιστώντας τον έναν από τους καταλληλότερους αλγόριθμους για την ανίχνευση outliers σε τομείς όπως η κυβερνοασφάλεια και τα χρηματοοικονομικά. Τα κύρια βήματα του αλγορίθμου Isolation Forest περιλαμβάνουν:

- Δημιουργία πολλών δέντρων απομόνωσης.
- Διάσπαση των δεδομένων σε υποσύνολα με τυχαία επιλεγμένα χαρακτηριστικά.
- Μέτρηση του βάθους της διαδρομής απομόνωσης για κάθε παρατήρηση.
- Ταξινόμηση των παρατηρήσεων με βάση τον μέσο όρο του βάθους απομόνωσης σε όλα τα δέντρα. Οι ανωμαλίες έχουν μικρότερο μέσο βάθος.

K-Means Clustering: Είναι ένας αλγόριθμος ο οποίος ανήκει στην κατηγορία μη εποπτευόμενης μάθησης ο οποίος ομαδοποιεί το μη επισημασμένο σύνολο δεδομένων σε διαφορετικά K clusters, όπου κάθε παρατήρηση ανήκει στο cluster με την πλησιέστερη μέση τιμή (centroid). Τα κύρια βήματα του K-Means Clustering περιλαμβάνουν:

- Επιλογή K αρχικών centroids.
- Ανάθεση κάθε παρατήρησης στο πλησιέστερο centroid.
- Επαναυπολογισμός των centroids ως τη μέση τιμή των παρατηρήσεων σε κάθε cluster.
- Επανάληψη των βημάτων 2 και 3 μέχρι τη σύγκλιση.

Support Vector Machines (SVM): Είναι ένας αλγόριθμος ισχυρός μηχανικής μάθησης που χρησιμοποιείται για γραμμική, μη γραμμική ταξινόμηση, παλινδρόμηση ακόμα και ανίχνευση outliers. Οι SVM χρησιμοποιούνται για διάφορες εργασίες όπως η ταξινόμηση κειμένων, ταξινόμηση εικόνων, αναγνώριση χειρόγραφου κλπ. Οι μηχανές αυτές είναι ευέλικτες και αποδοτικές επειδή μπορούν να διαχειριστούν δεδομένα υψηλής διάστασης και μη γραμμικές σχέσεις. Ο στόχος του SVM είναι να βρει τον υπερ-επίπεδο που διαχωρίζει καλύτερα τις κλάσεις στα δεδομένα. Τα κύρια βήματα του SVM περιλαμβάνουν:

- Μετασχηματισμός των δεδομένων σε έναν υψηλότερης διάστασης χώρο, αν χρειαστεί.
- Εύρεση του υπερ-επίπεδου που διαχωρίζει τις κλάσεις με το μέγιστο περιθώριο.
- Χρήση των υποστηρικτικών διανυσμάτων (support vectors) για την περιγραφή του υπερ-επιπέδου.

Decision Trees: Το δέντρο αποφάσεων αποτελεί μη παραμετρικό αλγόριθμο επιβλεπόμενης μάθησης, ο οποίος χρησιμοποιείται τόσο για εργασίες ταξινόμησης όσο και για παλινδρόμησης. Έχει μια ιεραρχική, δενδρική δομή, η οποία αποτελείται από μια ριζική κορυφή (root node), κλαδιά (branches), εσωτερικούς κόμβους (internal nodes) και φύλλα (leaf nodes). Το όνομα «δέντρο αποφάσεων» προέρχεται από το γεγονός ότι ο αλγόριθμος συνεχίζει να διαιρεί το σύνολο δεδομένων σε όλο και μικρότερα τμήματα μέχρι τα δεδομένα να χωριστούν σε μεμονωμένες περιπτώσεις, οι οποίες στη συνέχεια ταξινομούνται. Τα κύρια βήματα των Decision Trees περιλαμβάνουν:

- Επιλογή του βέλτιστου χαρακτηριστικού για τον διαχωρισμό των δεδομένων.
- Διαίρεση των δεδομένων σε υποσύνολα με βάση την απόφαση του κόμβου.
- Επανάληψη της διαδικασίας για κάθε υποσύνολο μέχρι την επίτευξη των τερματικών κόμβων.

Συμπερασματικά στο πλαίσιο της εργασίας μας εμείς χρησιμοποιήσαμε σαν τεχνική μηχανική μάθησης τον Isolation Forest για την ανίχνευση και τον εντοπισμό των outliers στα δεδομένα των Edge Servers. Ο αλγόριθμος αποδείχθηκε αποτελεσματικός και αποδοτικός στην ταυτοποίηση των outliers και την απομόνωση τους, βελτιώνοντας έτσι την ποιότητα των δεδομένων και την αποδοτικότητα του συστήματος edge computing. Οι υπόλοιποι αλγόριθμοι που μιλήσαμε για αυτούς θα μπορούσαν κάλλιστα να χρησιμοποιηθούν ωστόσο κρίθηκε καταλληλότερος ο Isolation Forest.

ΚΕΦΑΛΑΙΟ 4 Προτεινόμενο Σύστημα

(Περιγραφή του συστήματος 4.1)

Το αναπτυγμένο σύστημα αποτελείται από ένα δίκτυο Edge Servers τα οποία χρησιμοποιούνται για τη διαχείριση και την ανταλλαγή δεδομένων στις παρυφές του δικτύου. Κάθε Edge Server λειτουργεί ως αντιπροσωπευτικός κόμβος μέσα στο δίκτυο και υλοποιείται ως προσαρμοσμένο αντικείμενο γνωστό ως CustomEdgeServer. Κατά την αρχικοποίηση, κάθε CustomEdgeServer διαχειρίζεται δεδομένα που του ανατίθενται από ένα JSON αρχείο, τα οποία συνήθως περιλαμβάνουν μετρήσεις ή στατιστικά στοιχεία που συλλέγονται από τους αισθητήρες του δικτύου.

Αυτά τα Edge Servers επεξεργάζονται αυτά τα δεδομένα για να υπολογίσουν αρχικά στατιστικά μέτρα όπως οι μέσες τιμές και οι τυπικές αποκλίσεις. Επίσης, χρησιμοποιούν τον αλγόριθμο Isolation Forest για την ανίχνευση outliers. Η διαδικασία ανίχνευσης outliers περιλαμβάνει την εκπαίδευση του Isolation Forest χρησιμοποιώντας τα δεδομένα ειδικά για κάθε Edge Server και την αναγνώριση των δεδομένων που ξεπερνούν τα κανονικά στατιστικά όρια (όπως οι μέσες τιμές και οι τυπικές αποκλίσεις).

Επειτα, κάθε Edge Server ανταλλάσσει στατιστικά στοιχεία (μέσες τιμές και τυπικές αποκλίσεις) με άλλους servers του δικτύου, επιτρέποντας στους κόμβους να συγκρίνουν τις τοπικές τους μετρήσεις και να αξιολογούν την συνολική τους κατάσταση. Αλγόριθμοι βαρών και μετρήσεων ομοιότητας χρησιμοποιούνται για να αξιολογήσουν τη σημασία των outliers και να επιλέξουν τον κατάλληλο server για την ανταλλαγή δεδομένων. Συνήθως, τα outliers αποστέλλονται σε servers με υψηλότερα βάρη, ενώ η επιλογή του καλύτερου server γίνεται βάσει της ομοιότητας των δεδομένων του server αποστολέα με αυτόν του πιο κατάλληλου παραλήπτη.

Τέλος, πραγματοποιούνται πειράματα για να αξιολογηθεί η απόδοση του δικτύου σε διαφορετικές συνθήκες, δημιουργώντας νέα δεδομένα για κάθε κόμβο σε κάθε επανάληψη. Αυτά τα πειράματα βοηθούν στην αξιολόγηση της απόδοσης του συστήματος υπό διαφορετικές συνθήκες και διαστάσεις δεδομένων.

Συνολικά, το υλοποιημένο αυτό σύστημα επιτρέπει την αυτόνομη διαχείριση δεδομένων σε Edge Servers, χρησιμοποιεί μηχανική μάθηση για την ανίχνευση και διαχείριση των outliers, και αξιολογεί την συνολική κατάσταση του δικτύου μέσω της συνεργασίας μεταξύ των κόμβων.

(Υλοποιήσεις του κώδικα 4.1)

Η κλάση CustomEdgeServer κληρονομεί την κλάση EdgeServer και επεκτείνεται για να περιλαμβάνει πρόσθετες λειτουργίες και δεδομένα:

- Αρχικοποίηση (`__init__`): Αρχικοποιείται το αντικείμενο της κλάσης. Οι παράμετροι είναι το `id` και τα `data`, όπου το `id` χρησιμοποιείται για το όνομα του server και τα `data` είναι η λίστα των δεδομένων του server. Τα δεδομένα μετατρέπονται σε numpy array για τη χρήση με τις στατιστικές μεθόδους.
- Ιδιότητες: Το αντικείμενο κρατάει τις ιδιότητες `mean`, `std`, `outliers` και `history` για τον υπολογισμό των στατιστικών και τη διαχείριση των δεδομένων.

- Προσθήκη στις instances: Το συγκεκριμένο αντικείμενο προστίθεται στη λίστα `_instances`, που χρησιμοποιείται για να αναφερθούμε σε όλα τα instances της κλάσης σε άλλα σημεία του κώδικα.

Υπολογισμός Στατιστικών

- `calculate_statistics()`: Αυτή η μέθοδος υπολογίζει τη μέση τιμή (mean) και την τυπική απόκλιση (std) των δεδομένων του server. Αρχικά, φιλτράρει τα δεδομένα για να αφαιρέσει τα outliers και υπολογίζει το mean και το std μέχρις ότου όλα τα δεδομένα να είναι εντός του 3-συντελεστή της τυπικής απόκλισης από το mean.
- `find_outliers_with_ml()`: Χρησιμοποιεί τον αλγόριθμο Isolation Forest για να εντοπίσει outliers στα δεδομένα του server και τα αποθηκεύει στη λίστα outliers.

Μέθοδος ανταλλαγής δεδομένων

- `compute_weights()`: Υπολογίζει τα βάρη που αντιστοιχούν σε κάθε server βάσει των δεδομένων outliers.
- `send_outliers_()`: Αποστέλλει τα outliers βάσει των υπολογισμένων βαρών.

Υλοποίηση αλγορίθμου Isolation Forest

- `find_outliers_with_ml()`: Χρησιμοποιεί τον Isolation Forest από τη βιβλιοθήκη `scikit-learn` για να εντοπίσει outliers στα δεδομένα του server.
- `clf.fit()`: Εκπαιδεύει τον Isolation Forest στα δεδομένα, αναμένοντας έναν μονοδιάστατο πίνακα.
- `clf.predict()`: Προβλέπει την κατηγορία (outlier ή όχι) για κάθε στοιχείο του πίνακα.

Η πειραματική διαδικασία περιλαμβάνει την αξιολόγηση της απόδοσης του συστήματος Edge Computing μέσω διάφορων πειραμάτων. Αυτή η διαδικασία περιλαμβάνει:

Ορισμός Πειραματικών Παραμέτρων:

- Πλήθος Διακομιστών (Edge Servers): Δοκιμάζουμε με διαφορετικά πλήθη διακομιστών, από 1 έως 5.
- Διαστάσεις Δεδομένων (Data Dimensions): Κάθε πείραμα εκτελείται για διαστάσεις δεδομένων 5, 10, και 20.
- Πλήθος Επαναλήψεων (Iterations): Εκτελούνται 5 επαναλήψεις για κάθε συνδυασμό παραμέτρων.

Δημιουργία Διακομιστών:

- Για κάθε συνδυασμό πλήθους διακομιστών και διαστάσεων δεδομένων, δημιουργούνται νέοι διακομιστές.

Παραγωγή Δεδομένων:

- Για κάθε επανάληψη, παράγονται νέα δεδομένα για κάθε διακομιστή. Τα δεδομένα παράγονται τυχαία, με πιθανότητα 5% να είναι πολύ μεγάλα ή πολύ μικρά (outliers).

Υπολογισμός Στατιστικών και Ανίχνευση Outliers:

- Κάθε διακομιστής υπολογίζει τα στατιστικά του (μέση τιμή, τυπική απόκλιση) και ανιχνεύει τα outliers χρησιμοποιώντας τον αλγόριθμο Isolation Forest.

Αποθήκευση Αποτελεσμάτων:

- Τα αποτελέσματα κάθε πειράματος, περιλαμβανομένων των στατιστικών και του πλήθους των outliers, αποθηκεύονται σε ένα DataFrame.

Διαδικασία Αξιολόγησης:

- Για κάθε επανάληψη, εκτυπώνονται τα αποτελέσματα στατιστικής ανάλυσης, περιλαμβανομένων των outliers, της μέσης τιμής και της τυπικής απόκλισης για κάθε διακομιστή.

Αποθήκευση Αποτελεσμάτων σε Excel:

- Όλα τα αποτελέσματα αποθηκεύονται σε ένα αρχείο Excel με πολλά φύλλα, τα οποία περιλαμβάνουν τα αποτελέσματα των πειραμάτων, τα αρχικά και τα τελικά στατιστικά.

Συνολική Πειραματική Διαδικασία:

- Η πειραματική διαδικασία περιλαμβάνει τη δημιουργία διαφορετικών συνδυασμών διακομιστών και δεδομένων για την αξιολόγηση της απόδοσης του συστήματος σε διάφορες συνθήκες. Τα αποτελέσματα αποθηκεύονται και αναλύονται για να εξαχθούν συμπεράσματα σχετικά με την απόδοση και την αποτελεσματικότητα του συστήματος ανίχνευσης outliers και διαχείρισης δεδομένων σε περιβάλλον Edge Computing.

Παραρτήματα:

Παράρτημα Α: Αρχείο Κώδικα erg1.py

Περιγραφή: Αυτό το αρχείο περιέχει τον βασικό κώδικα του συστήματος που χρησιμοποιείται για τη διαχείριση και την αποθήκευση δεδομένων στις παρυφές του δικτύου.

Παράρτημα Β: Αρχείο Δεδομένων erg1.json

Περιγραφή: Αυτό το αρχείο περιέχει τα αρχικά δεδομένα που χρησιμοποιούνται για τη δημιουργία των Edge Servers. Τα δεδομένα περιλαμβάνουν τιμές μετρήσεων που χρησιμοποιούνται για τον υπολογισμό των στατιστικών και την ανίχνευση των outliers.

Παράρτημα Γ: Αρχείο Δεδομένων example.json

Περιγραφή: Αυτό το αρχείο περιέχει ένα παράδειγμα δεδομένων που χρησιμοποιείται για την αρχικοποίηση του simulator. Τα δεδομένα περιλαμβάνουν παραμέτρους για τον simulator και τις αρχικές τιμές για τα Edge Servers.

ΚΕΦΑΛΑΙΟ 5 Πειραματική Αποτίμηση

Στο `experiment_results.xlsx` έχουμε 3 φύλλα το καθένα από 1 πίνακα οι οποίοι είναι οι: `experiment results` που αποτελείται απ' τα αποτελέσματα τα οποία υπολογίσαμε στην πειραματική διαδικασία, `initial stats` που αποτελούνται από τους υπολογισμούς των στατιστικών πριν και μετά την εφαρμογή του αλγορίθμου αντίστοιχα με στόχο να δούμε αν τα `datasets` έχουν μικρή απόκλιση και λιγότερους outliers από ότι στην αρχική μορφή

Παρακάτω ακολουθεί ο πίνακας με τα αποτελέσματα πριν την εφαρμογή του αλγορίθμου δηλαδή ο `initial stats`:

Server	Mean	Std	Outliers
EdgeServer1	47,38585859	72,53641372	[12.0, 15.0, 14.0, 13.0, 110.0, 16.0, 18.0, 12.0, 13.0, 14.0, 13.0, 17.0, 19.0, 12.0, 13.0, 14.0, 120.0, 411.2, 424.0, 502.0]
EdgeServer2	54,24646465	59,12212718	[25.0, 23.0, 24.0, 220.0, 25.0, 23.0, 25.0, 24.0, 23.0, 21.0, 23.0, 31.5, 230.0, 491.8, 322.0, 60.2]
EdgeServer3	2749,627451	2328,201579	[100.0, 200.0, 300.5, 7000.3, 4900.5, 5.6, 15000.0]
EdgeServer4	346,6833333	542,3013842	[15.5, 25.0, 35.75, 45.0, 1100.0, 3955.0, 505.6]
EdgeServer5	269,488	389,0619615	[7.0, 14.0, 21.0, 28.4, 35.8, 432.0, 1119.8, 2722.0]

Έπειτα ακολουθεί ο πίνακας `final stats` δηλαδή μετά την εφαρμογή του αλγορίθμου:

Server	Mean	Std	Outliers
EdgeServer1	33,23404255	11,29844231	[12.0, 15.0, 14.0, 13.0, 110.0, 16.0, 18.0, 12.0, 13.0, 14.0, 13.0, 17.0, 19.0, 12.0, 13.0, 14.0, 120.0, 411.2, 424.0, 502.0]
EdgeServer2	43,22736842	11,69958562	[25.0, 23.0, 24.0, 220.0, 25.0, 23.0, 25.0, 24.0, 23.0, 21.0, 23.0, 31.5, 230.0, 491.8, 322.0, 60.2]

EdgeServer3	2504,62	1570,818844	[100.0, 200.0, 300.5, 7000.3, 4900.5, 5.6, 15000.0]
EdgeServer4	257,6704082	144,4873732	[15.5, 25.0, 35.75, 45.0, 1100.0, 3955.0, 505.6]
EdgeServer5	200,6791667	112,1152141	[7.0, 14.0, 21.0, 28.4, 35.8, 432.0, 1119.8, 2722.0]

Τέλος ο πίνακας με τα αποτελέσματα της πειραματικής διαδικασίας από τον οποίο προκύπτουν και τα γραφήματα που θα ακολουθήσουν

Server	Iteration	Num_Servers	Data_Dimension	Outliers	Mean	Std
Edgeserver1	1	1	5	1	55,19226781	29,18913865
Edgeserver1	2	1	5	2	64,55700859	20,16075742
Edgeserver1	3	1	5	2	103,3315704	134,0226594
Edgeserver1	4	1	5	2	58,07495563	27,45795314
Edgeserver1	5	1	5	1	20,22399658	16,67805781
Edgeserver1	1	1	10	2	42,19860996	30,65590362
Edgeserver1	2	1	10	3	9,663981647	147,6049725
Edgeserver1	3	1	10	5	52,56449484	18,41475989
Edgeserver1	4	1	10	5	62,30159645	20,98743186
Edgeserver1	5	1	10	5	44,19449998	28,28710892
Edgeserver1	1	1	20	9	48,39681579	31,63073629
Edgeserver1	2	1	20	12	60,56238143	27,61242047
Edgeserver1	3	1	20	11	54,13183001	31,75658426
Edgeserver1	4	1	20	9	64,39381742	24,36023639
Edgeserver1	5	1	20	10	46,69054114	26,2961997
Edgeserver1	1	2	5	2	124,2968467	131,2738658
Edgeserver2	1	2	5	3	52,83109501	34,36404688
Edgeserver1	2	2	5	2	-	218,9889695
Edgeserver2	2	2	5	1	54,46126095	21,01622926
Edgeserver1	3	2	5	2	46,75064648	16,45059278
Edgeserver2	3	2	5	2	78,08049956	32,87707768
Edgeserver1	4	2	5	1	41,58829945	28,67654135
Edgeserver2	4	2	5	1	-	173,2253212
Edgeserver1	5	2	5	1	54,38400452	38,70026496
Edgeserver2	5	2	5	1	-	113,2121094
Edgeserver1	1	2	10	5	50,71828606	25,50871602
Edgeserver2	1	2	10	4	-	353,4896188
Edgeserver1	2	2	10	5	57,35584556	28,02344135
Edgeserver2	2	2	10	2	186,0522114	306,7008605
Edgeserver1	3	2	10	2	89,24305776	152,7691336

Edgeserver2	3	2	10	4	44,917005	21,70114035
Edgeserver1	4	2	10	4	35,84752655	23,27753255
Edgeserver2	4	2	10	5	38,26096545	30,18354932
Edgeserver1	5	2	10	5	48,40101337	32,67502726
Edgeserver2	5	2	10	3	50,73901112	23,01128451
Edgeserver1	1	2	20	3	49,87234949	27,16947494
Edgeserver2	1	2	20	5	40,076201	30,50605716
Edgeserver1	2	2	20	5	42,34329057	25,92112859
Edgeserver2	2	2	20	6	54,41036006	25,58510436
Edgeserver1	3	2	20	11	60,31494155	25,68461206
Edgeserver2	3	2	20	6	63,95966173	32,13709682
Edgeserver1	4	2	20	5	42,89335441	31,25223776
Edgeserver2	4	2	20	4	87,96729149	292,5711854
Edgeserver1	5	2	20	2	51,40680812	27,94690889
Edgeserver2	5	2	20	4	39,4413253	29,84838542
Edgeserver1	1	3	5	1	44,14345861	17,91085381
Edgeserver2	1	3	5	3	46,95619445	29,3255475
Edgeserver3	1	3	5	2	61,74561245	27,94086661
Edgeserver1	2	3	5	1	59,36464073	23,1561241
Edgeserver2	2	3	5	2	143,81944	141,4670135
Edgeserver3	2	3	5	1	70,41249876	7,812973975
Edgeserver1	3	3	5	2	60,55113674	26,80020775
Edgeserver2	3	3	5	2	43,31722635	23,43466342
Edgeserver3	3	3	5	2	42,79751794	15,6617516
Edgeserver1	4	3	5	1	36,43264764	19,27497841
Edgeserver2	4	3	5	2	72,58350069	16,05106603
Edgeserver3	4	3	5	1	56,28203828	24,05019168
Edgeserver1	5	3	5	1	46,9975399	17,00935639
Edgeserver2	5	3	5	1	195,9422914	329,2720752
Edgeserver3	5	3	5	1	117,6851062	147,5199122
Edgeserver1	1	3	10	5	53,8432465	28,08228345
Edgeserver2	1	3	10	2	13,13816581	141,4821775
Edgeserver3	1	3	10	1	-	283,1520988
Edgeserver1	2	3	10	5	53,58064082	31,15780297
Edgeserver2	2	3	10	3	58,91102856	30,57959796
Edgeserver3	2	3	10	2	117,5618147	211,5689973
Edgeserver1	3	3	10	5	41,5474311	22,42415697
Edgeserver2	3	3	10	3	107,9447317	204,9171038
Edgeserver3	3	3	10	2	-	193,1332878
Edgeserver1	4	3	10	4	179,1286176	293,6065562
Edgeserver2	4	3	10	3	-	241,1368083
Edgeserver3	4	3	10	3	10,54242681	313,0659383
Edgeserver1	5	3	10	6	41,45766266	25,3991585

Edgeserver2	5	3	10	3	-	167,2044734
Edgeserver3	5	3	10	2	82,58668616	84,62165998
Edgeserver1	1	3	20	8	49,8909701	31,06579967
Edgeserver2	1	3	20	7	40,33035273	28,34964851
Edgeserver3	1	3	20	4	45,15061711	25,36589839
Edgeserver1	2	3	20	3	44,0348523	31,03029345
Edgeserver2	2	3	20	6	54,72061792	26,17539561
Edgeserver3	2	3	20	4	40,7103914	30,63768691
Edgeserver1	3	3	20	4	13,00325895	287,1560966
Edgeserver2	3	3	20	7	56,62578345	24,98076065
Edgeserver3	3	3	20	4	44,38219985	26,09026239
Edgeserver1	4	3	20	6	57,9842963	25,57313007
Edgeserver2	4	3	20	9	62,67825082	27,54312109
Edgeserver3	4	3	20	3	46,01503407	26,74071697
Edgeserver1	5	3	20	5	47,36356369	29,6685535
Edgeserver2	5	3	20	4	49,67231236	31,17060687
Edgeserver3	5	3	20	10	64,98744394	23,79866487
Edgeserver1	1	4	5	2	46,41926667	26,61276036
Edgeserver2	1	4	5	2	65,6823965	9,401331853
Edgeserver3	1	4	5	2	38,10749249	28,79932579
Edgeserver4	1	4	5	1	74,41533723	15,51038613
Edgeserver1	2	4	5	2	41,17104628	18,33104608
Edgeserver2	2	4	5	2	69,63779762	13,86993913
Edgeserver3	2	4	5	1	1,251240164	109,2486162
Edgeserver4	2	4	5	2	55,746249	16,45619814
Edgeserver1	3	4	5	1	177,8797289	245,6963527
Edgeserver2	3	4	5	2	71,13365261	17,38001557
Edgeserver3	3	4	5	1	34,64319494	32,22843812
Edgeserver4	3	4	5	1	46,25024368	17,22692082
Edgeserver1	4	4	5	1	55,27525818	21,18213483
Edgeserver2	4	4	5	3	45,58539633	23,5606121
Edgeserver3	4	4	5	1	49,19405303	37,72257393
Edgeserver4	4	4	5	2	54,32024012	28,48276422
Edgeserver1	5	4	5	2	48,97286183	22,87657766
Edgeserver2	5	4	5	1	22,11793725	18,07644737
Edgeserver3	5	4	5	2	59,30805153	12,19406211
Edgeserver4	5	4	5	2	65,4475272	19,63085175
Edgeserver1	1	4	10	5	63,22291154	21,69896135
Edgeserver2	1	4	10	4	46,66425756	30,34949004
Edgeserver3	1	4	10	4	45,98585323	27,66676376
Edgeserver4	1	4	10	3	31,74422665	21,62776598
Edgeserver1	2	4	10	2	68,55493086	96,29339742
Edgeserver2	2	4	10	2	46,51337232	34,44875968
Edgeserver3	2	4	10	4	59,08754312	21,3320134
Edgeserver4	2	4	10	2	135,2655813	225,5014749

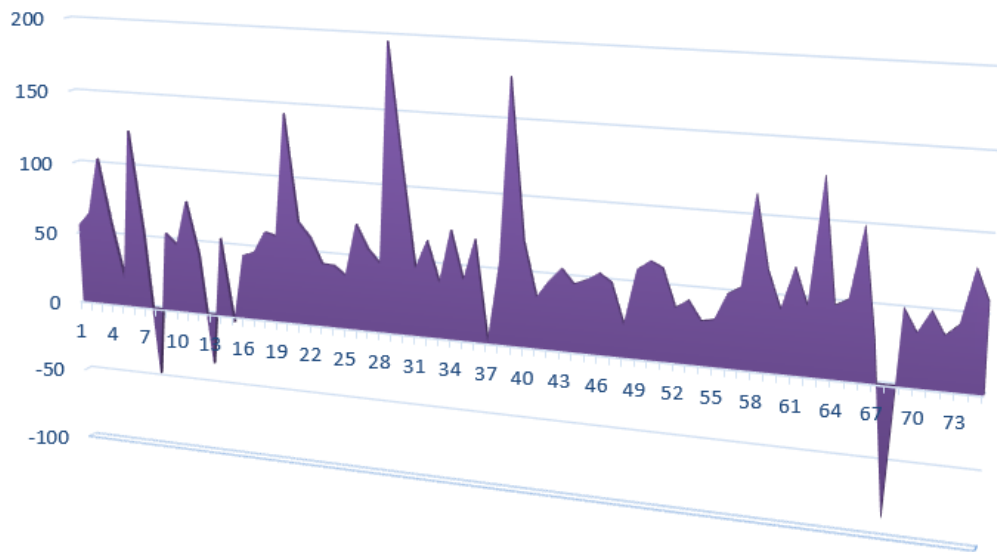
Edgeserver1	3	4	10	2	43,52568322	283,5774721
Edgeserver2	3	4	10	4	54,54273398	27,90460979
Edgeserver3	3	4	10	4	50,20573721	32,90434937
Edgeserver4	3	4	10	3	35,23222578	29,40613984
Edgeserver1	4	4	10	4	30,41959112	20,3041276
Edgeserver2	4	4	10	7	56,24519388	29,42482778
Edgeserver3	4	4	10	2	65,5906905	213,768533
Edgeserver4	4	4	10	3	45,12227658	26,8154616
Edgeserver1	5	4	10	5	47,4662964	31,77986734
Edgeserver2	5	4	10	2	60,07184699	61,75130922
Edgeserver3	5	4	10	2	37,72600005	17,94178991
Edgeserver4	5	4	10	3	40,23213508	18,87419712
Edgeserver1	1	4	20	5	49,72646954	30,4037214
Edgeserver2	1	4	20	4	37,7437342	22,412727
Edgeserver3	1	4	20	2	52,62230788	27,63480153
Edgeserver4	1	4	20	10	59,09547544	29,15655429
Edgeserver1	2	4	20	5	53,91738333	28,96920514
Edgeserver2	2	4	20	10	47,58720489	34,78300512
Edgeserver3	2	4	20	11	41,7122449	28,70340141
Edgeserver4	2	4	20	4	37,73277265	30,97488523
Edgeserver1	3	4	20	5	45,93744569	29,6559704
Edgeserver2	3	4	20	15	49,54861817	30,26714441
Edgeserver3	3	4	20	5	44,22260344	30,58356521
Edgeserver4	3	4	20	13	45,23296892	29,4529262
Edgeserver1	4	4	20	14	60,91146516	25,38967938
Edgeserver2	4	4	20	8	60,71659746	31,68998982
Edgeserver3	4	4	20	13	50,5920818	31,870075
Edgeserver4	4	4	20	4	38,67649828	25,30628966
Edgeserver1	5	4	20	7	53,98376523	20,81714068
Edgeserver2	5	4	20	9	52,65206939	25,77209086
Edgeserver3	5	4	20	10	53,71358852	31,53028325
Edgeserver4	5	4	20	3	33,60231698	27,58322367
Edgeserver1	1	5	5	2	61,38236549	26,88121911
Edgeserver2	1	5	5	1	36,64081512	20,95376452
Edgeserver3	1	5	5	3	42,06258537	19,98832272
Edgeserver4	1	5	5	1	29,27236887	36,0139482
Edgeserver5	1	5	5	1	31,01103178	14,43707122
Edgeserver1	2	5	5	2	48,97539923	15,02770978
Edgeserver2	2	5	5	2	53,81815241	23,89871558
Edgeserver3	2	5	5	2	114,1420372	133,3226033
Edgeserver4	2	5	5	1	66,1096885	34,75386
Edgeserver5	2	5	5	1	41,85723727	33,54744441
Edgeserver1	3	5	5	1	69,63342633	28,41085235
Edgeserver2	3	5	5	2	45,47771768	25,67259913
Edgeserver3	3	5	5	1	128,4503147	135,7544222

Edgeserver4	3	5	5	2	47,71685635	28,08284415
Edgeserver5	3	5	5	2	52,3903003	25,35231382
Edgeserver1	4	5	5	1	99,41849117	117,4862587
Edgeserver2	4	5	5	1	28,25051322	14,18233561
Edgeserver3	4	5	5	1	- 87,27726095	298,3272183
Edgeserver4	4	5	5	3	49,90426112	36,78600994
Edgeserver5	4	5	5	1	35,24920616	27,56955127
Edgeserver1	5	5	5	2	49,94226476	28,37817662
Edgeserver2	5	5	5	2	35,17917448	25,29496137
Edgeserver3	5	5	5	1	42,80120032	23,46846941
Edgeserver4	5	5	5	1	79,10995425	10,71837678
Edgeserver5	5	5	5	1	58,80790723	29,18716493
Edgeserver1	1	5	10	2	- 9,577694593	156,32169
Edgeserver2	1	5	10	4	- 45,91231146	182,8589843
Edgeserver3	1	5	10	3	0,383949667	135,6542672
Edgeserver4	1	5	10	3	61,94063156	21,54147476
Edgeserver5	1	5	10	3	56,98708034	25,95712841
Edgeserver1	2	5	10	4	44,72137011	28,73738252
Edgeserver2	2	5	10	4	58,30486505	31,69378102
Edgeserver3	2	5	10	5	46,42343791	26,0449873
Edgeserver4	2	5	10	4	- 154,1624429	411,7942306
Edgeserver5	2	5	10	4	43,09716017	23,17371569
Edgeserver1	3	5	10	4	41,80697435	31,60135075
Edgeserver2	3	5	10	3	138,1075577	249,1303913
Edgeserver3	3	5	10	4	41,3982303	29,12919308
Edgeserver4	3	5	10	4	62,99247444	30,74903771
Edgeserver5	3	5	10	5	49,49670532	29,5804572
Edgeserver1	4	5	10	2	93,75802398	334,2659368
Edgeserver2	4	5	10	5	49,94125734	27,28980113
Edgeserver3	4	5	10	3	38,81184444	21,10241718
Edgeserver4	4	5	10	3	148,0459961	258,5744613
Edgeserver5	4	5	10	3	47,42921706	23,36800982
Edgeserver1	5	5	10	3	42,06372732	26,87348469
Edgeserver2	5	5	10	5	54,46372555	26,01316415
Edgeserver3	5	5	10	5	55,83324963	18,22757282
Edgeserver4	5	5	10	3	137,6007272	275,2342449
Edgeserver5	5	5	10	3	62,97464116	25,69500837
Edgeserver1	1	5	20	4	48,35687085	25,3712061
Edgeserver2	1	5	20	5	59,11001352	40,74475172
Edgeserver3	1	5	20	4	40,75521252	27,24443133
Edgeserver4	1	5	20	4	57,43176551	21,70004823
Edgeserver5	1	5	20	4	- 34,82823071	322,1121742

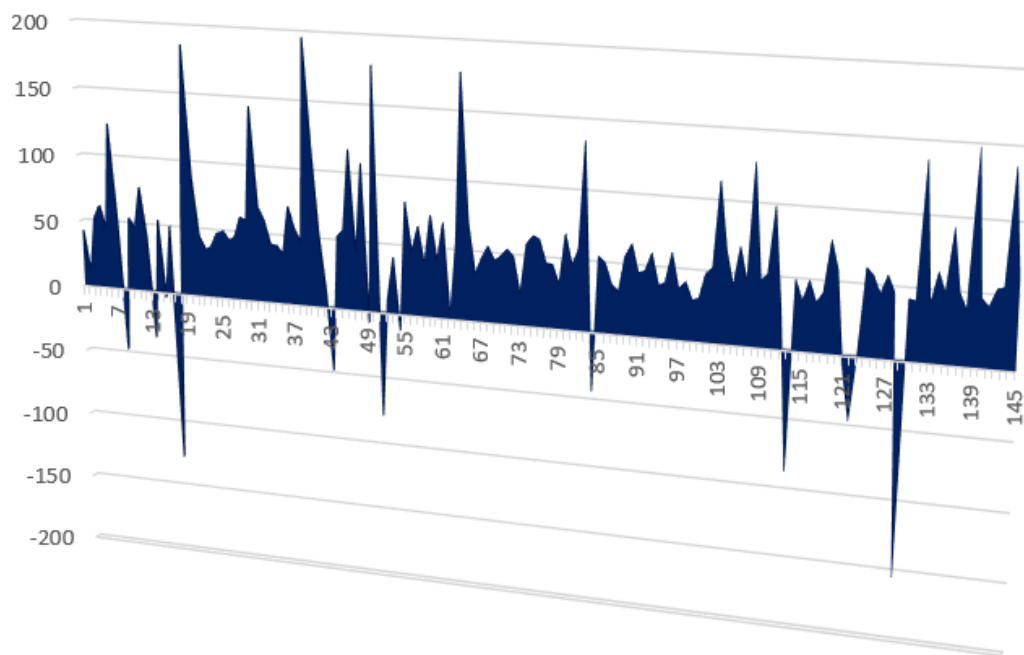
Edgeserver1	2	5	20	4	58,93904459	19,89799238
Edgeserver2	2	5	20	5	41,15586642	28,24001387
Edgeserver3	2	5	20	4	39,99892701	29,79305989
Edgeserver4	2	5	20	12	42,67113321	31,95643216
Edgeserver5	2	5	20	5	56,16672285	23,99477323
Edgeserver1	3	5	20	3	48,82663152	29,17637058
Edgeserver2	3	5	20	10	42,14836422	28,28192983
Edgeserver3	3	5	20	3	52,11907674	28,30648692
Edgeserver4	3	5	20	4	46,88716682	30,94900837
Edgeserver5	3	5	20	10	54,49337239	28,10915715
Edgeserver1	4	5	20	4	34,77635553	30,36074923
Edgeserver2	4	5	20	10	43,25206979	24,87563353
Edgeserver3	4	5	20	3	57,99637469	24,32985342
Edgeserver4	4	5	20	3	54,06269113	26,71235942
Edgeserver5	4	5	20	10	52,66633817	30,0312002
Edgeserver1	5	5	20	3	34,90282935	27,59340313
Edgeserver2	5	5	20	10	62,08608779	26,75569682
Edgeserver3	5	5	20	2	49,31581744	25,28304806
Edgeserver4	5	5	20	4	55,9774329	34,26491804
Edgeserver5	5	5	20	12	52,31912644	31,2294764

Συνεχίζοντας η επαναληπτική διαδικασία επαναλήφθηκε 5 φορές για διαφορετικό πλήθος EdgeServer και διαφορετικές διαστάσεις δεδομένων, συγκεκριμένα η πειραματική διαδικασία εκτελείται για 5 EdgeServers και για 3 διαστάσεις δεδομένων τις [5,10,20]. Μερικά από τα αποτελέσματα εμφανίζονται στα γραφήματα που ακολουθούν:

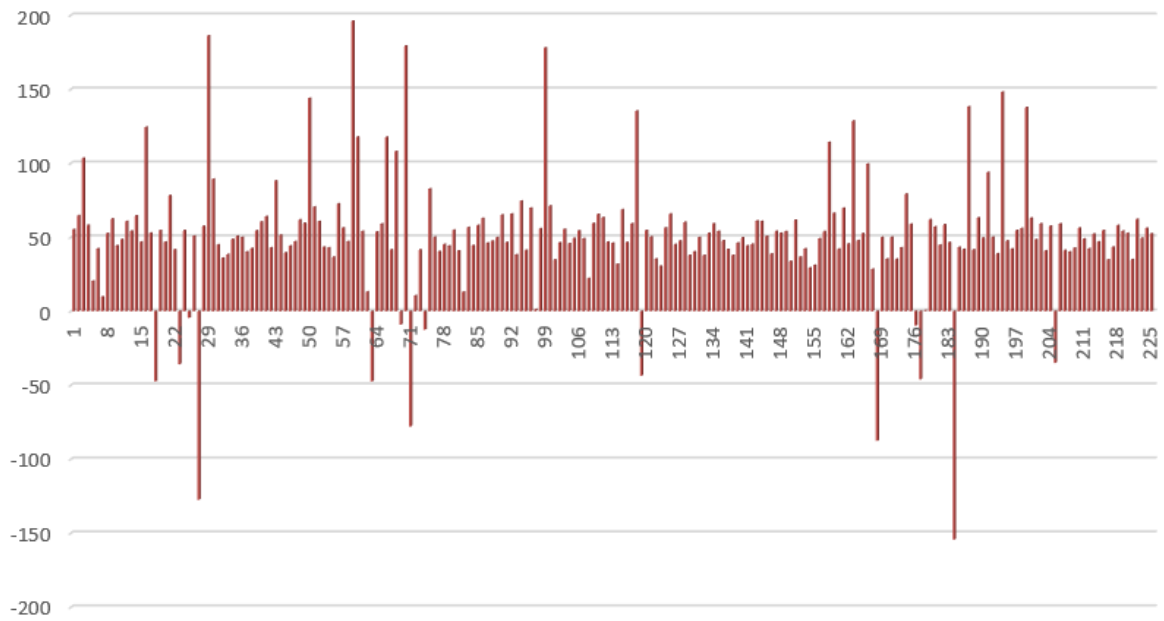
MEAN GIA DATA_DIMENSION: 5



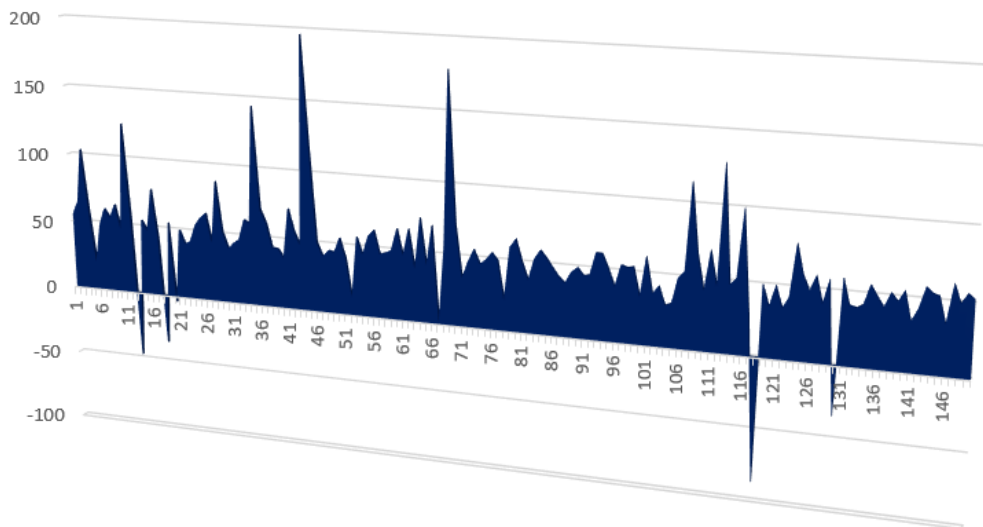
MEAN GIA DATA_DIMENSION 5-10



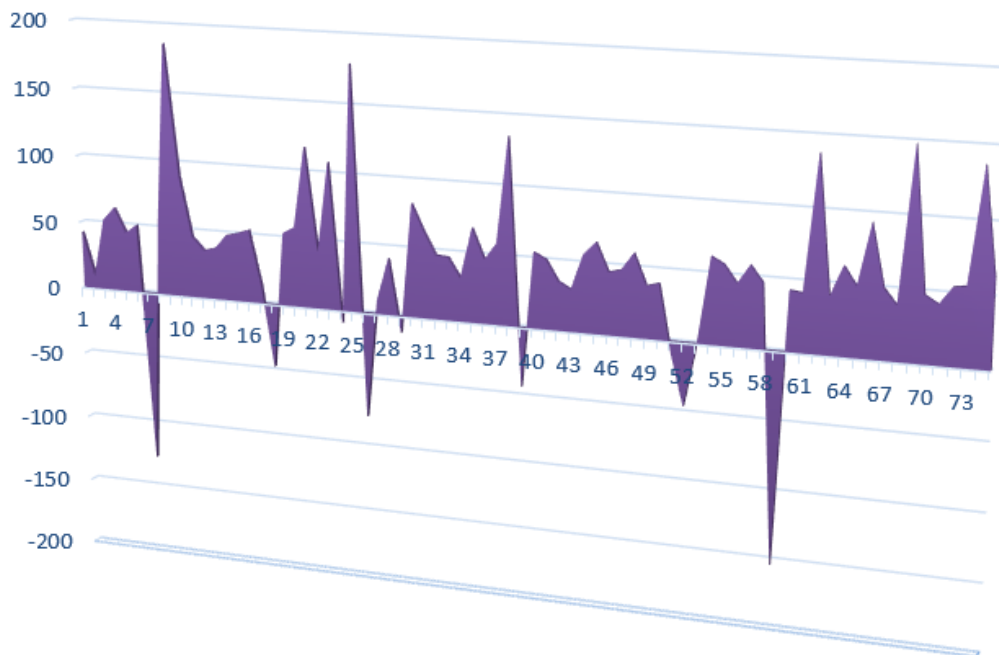
MEAN GIA DATA_DIMENSION 5-10-20



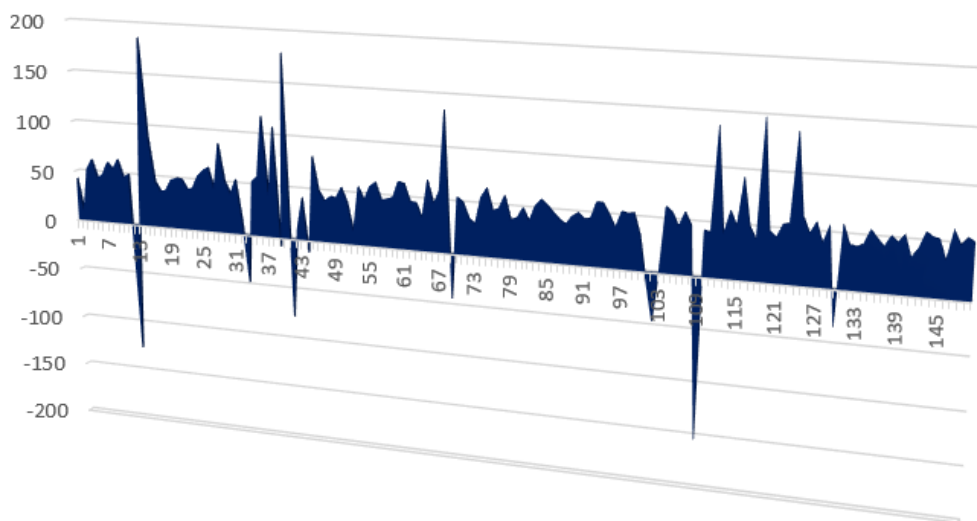
MEAN GIA DATA DIMENSION 5-20



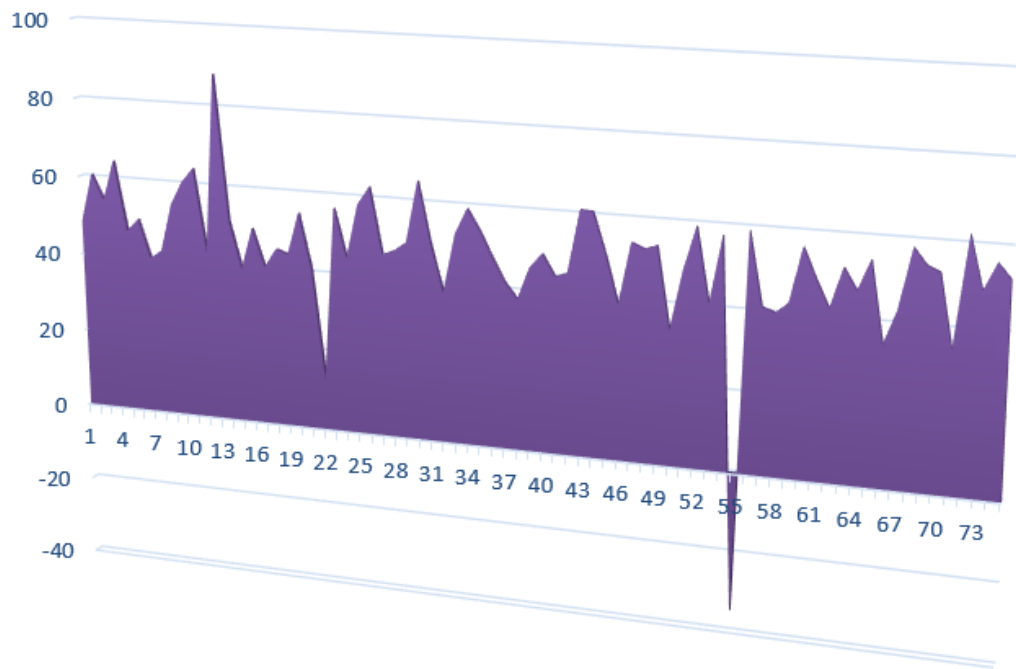
MEAN GIA DATA_DIMENSION10



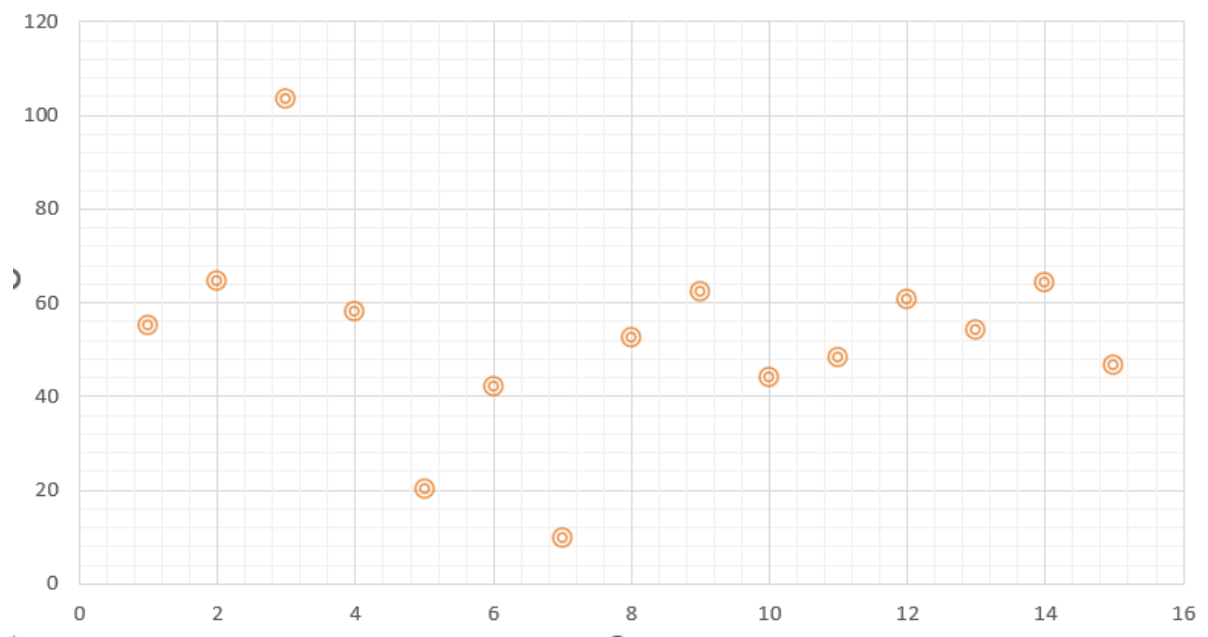
MEAN GIA DATA_DIMENSION 10-20



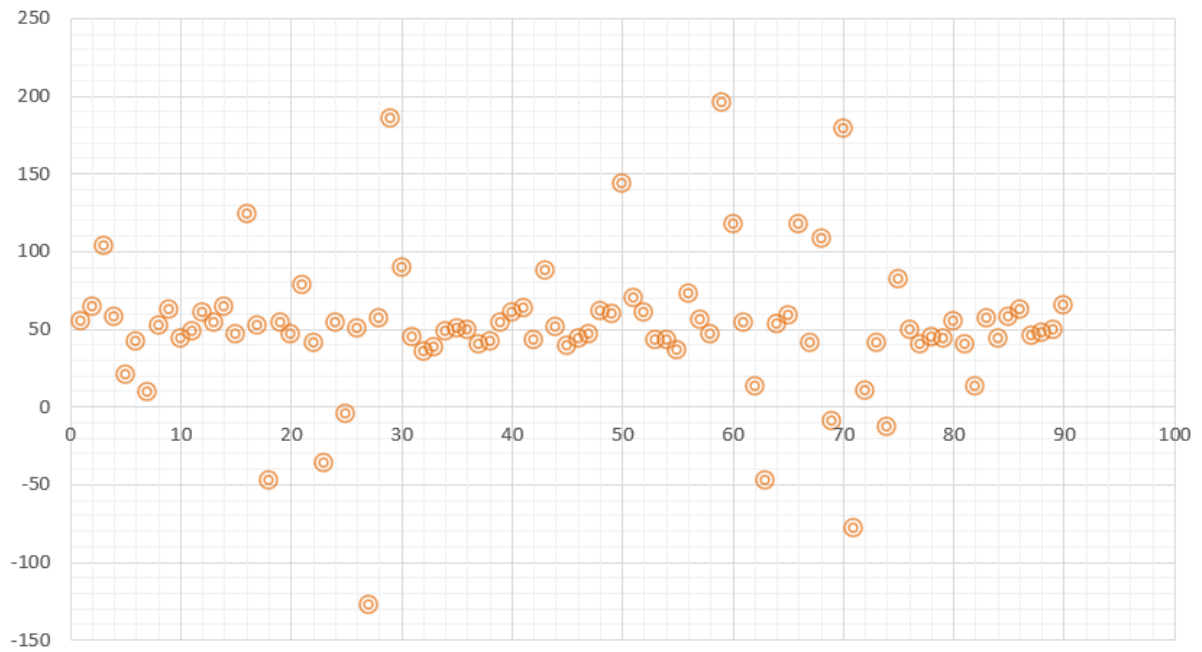
MEAN ΓΙΑ DATA_DIMENSION20



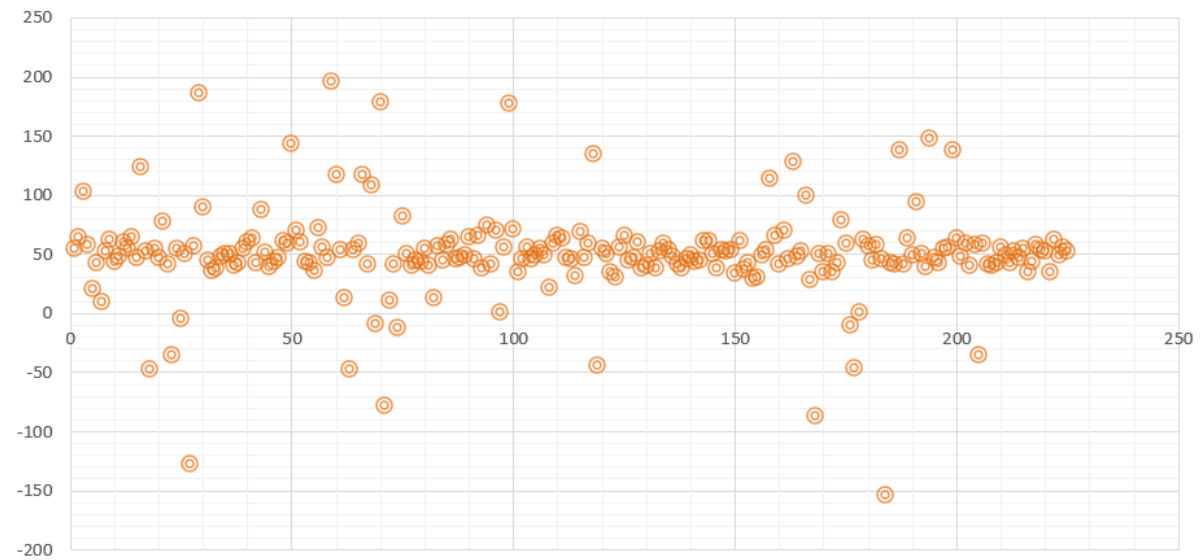
MEAN ΓΙΑ NUM_SERVERS1



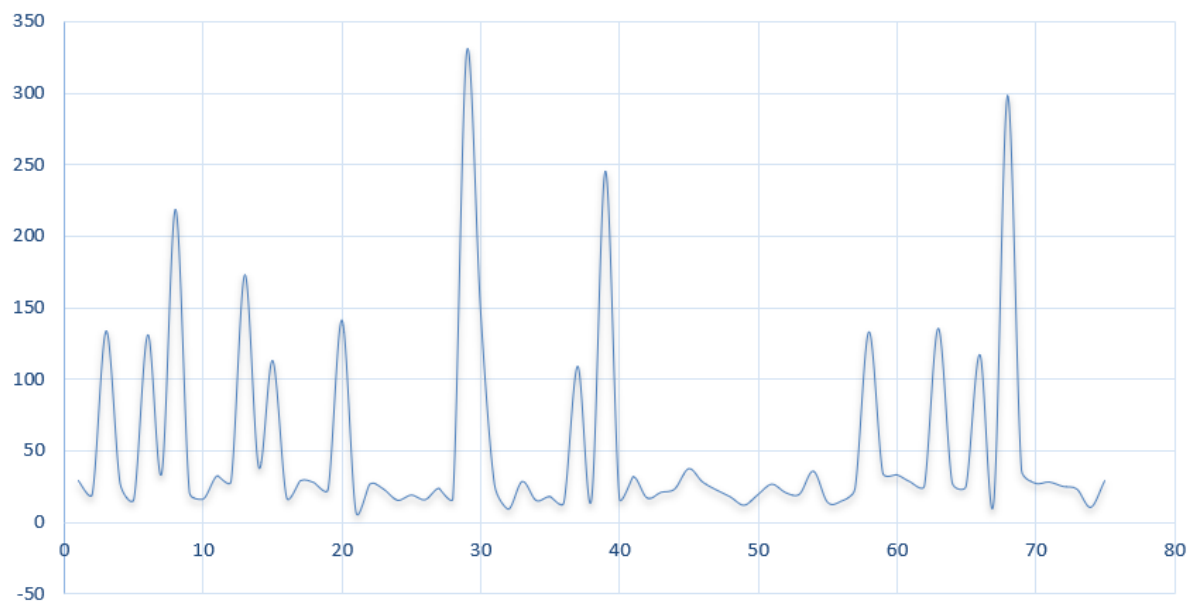
MEAN ΓΙΑ NUMSERVERS1_2_3



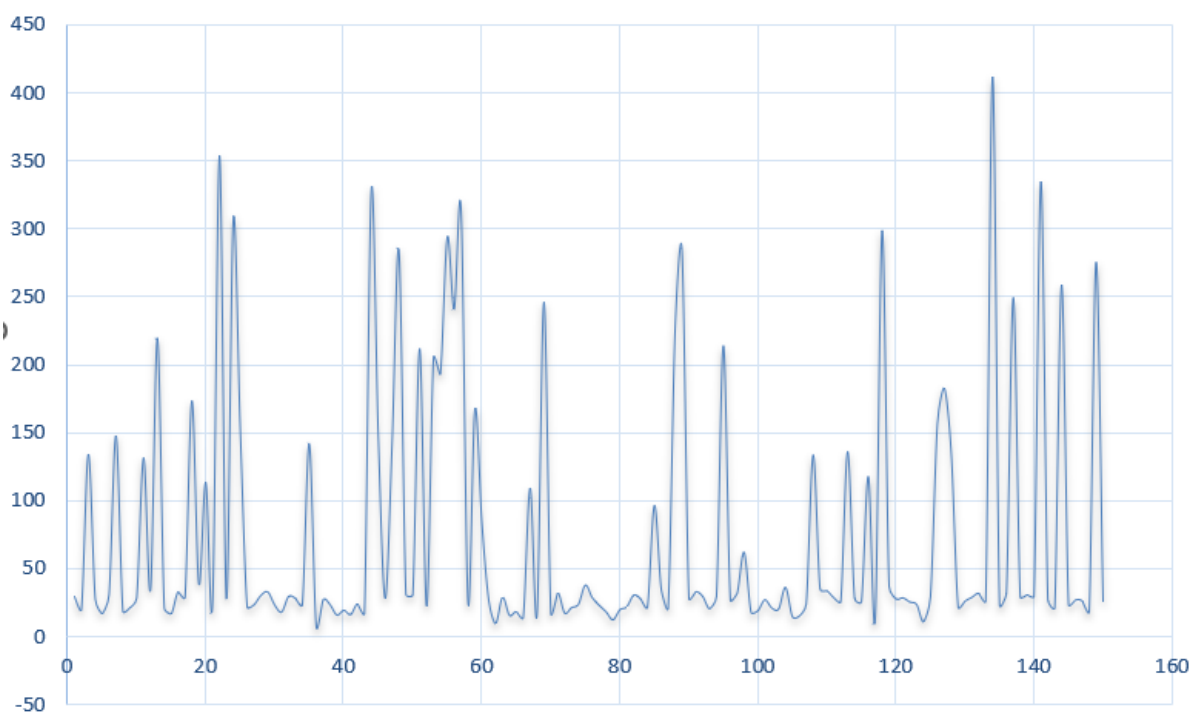
MEAN ΓΙΑ NUMSERVERS1_2_3_4_5



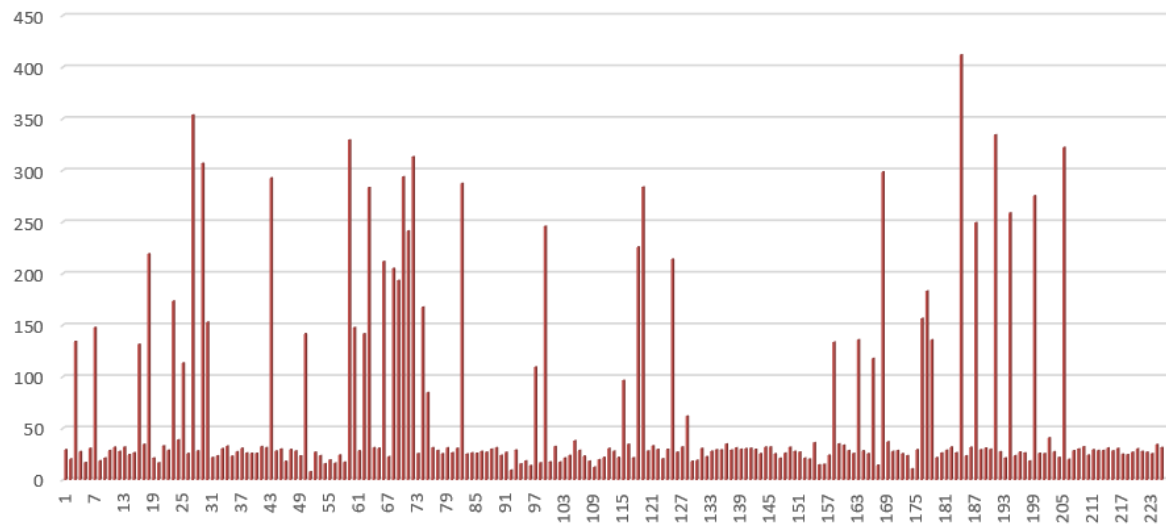
STD GIA DATA_DIMENSION 5



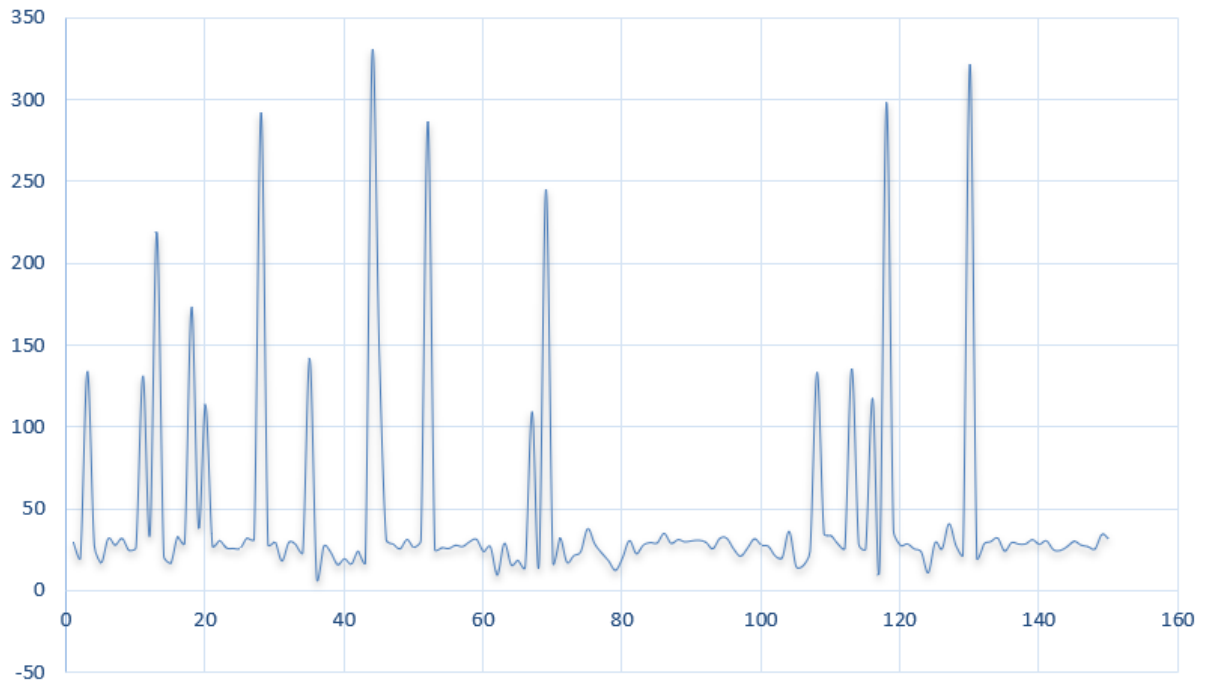
STD GIA DATA_DIMENSION 5-10



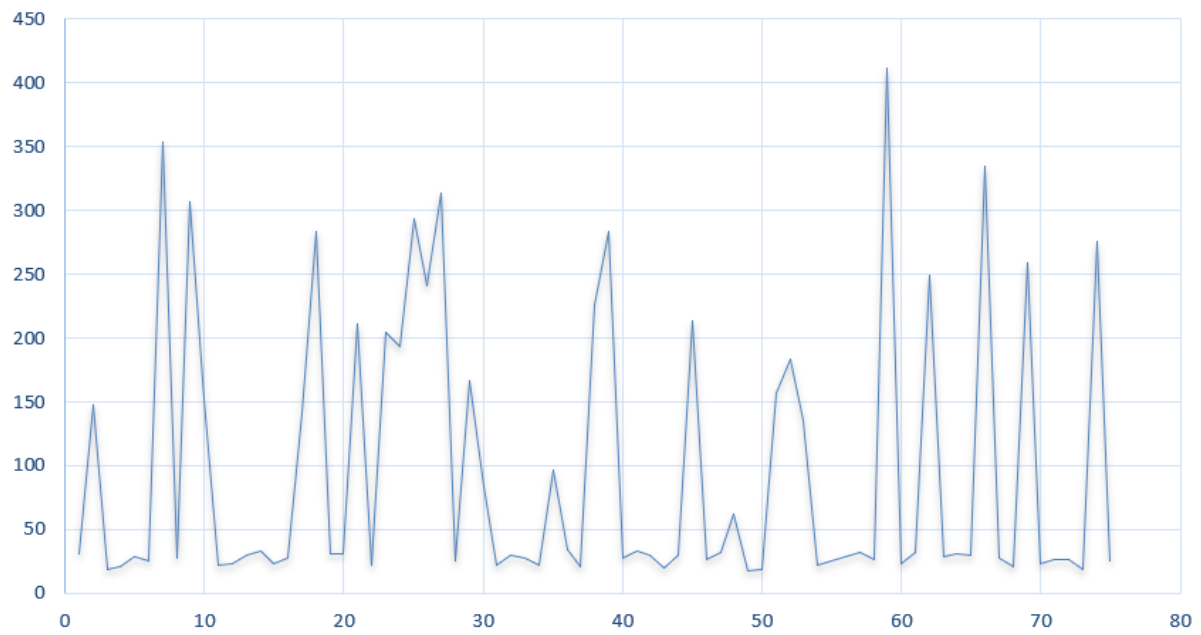
STD FIA DATA_DIMENSION 5-10-20



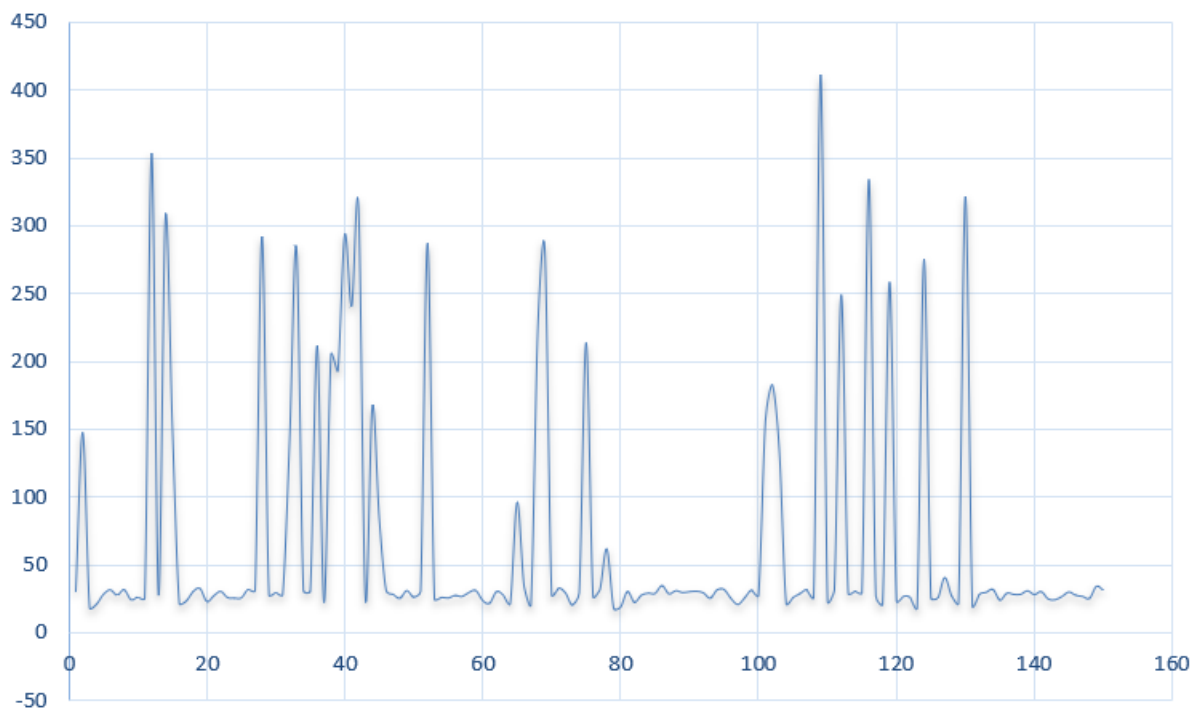
STD FIA DATA_DIMENSION 5-20



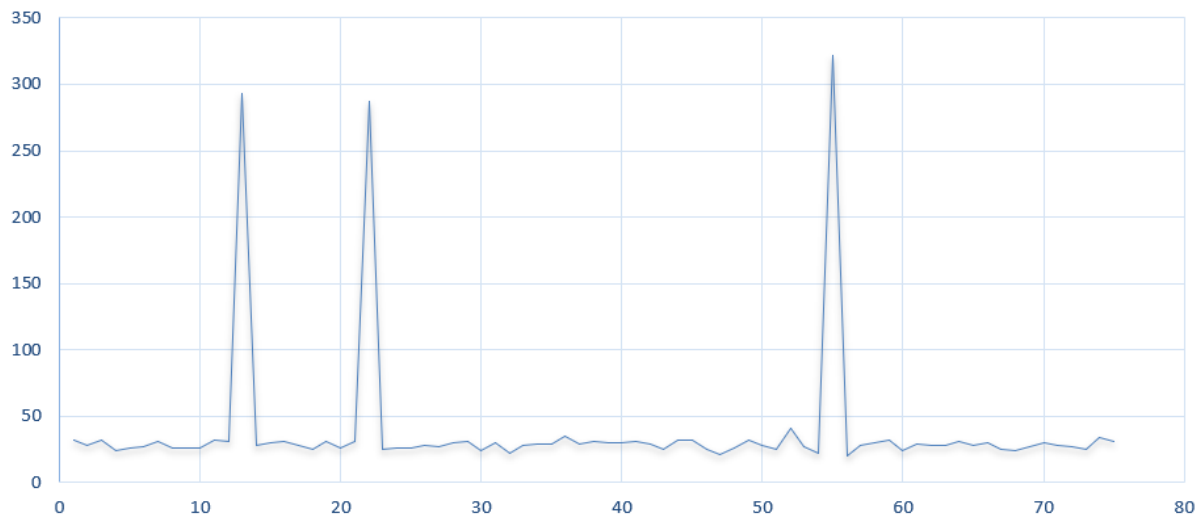
STD GIA DATA DIMENSION 10



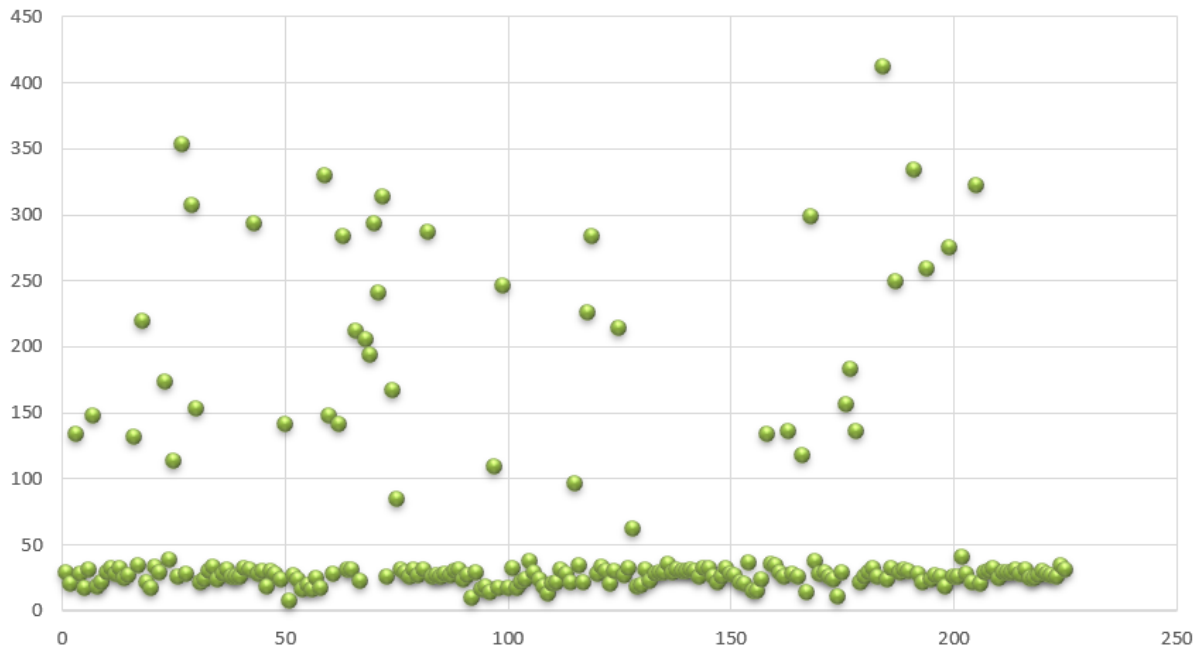
STD GIA DATA_DIMENSION 10-20



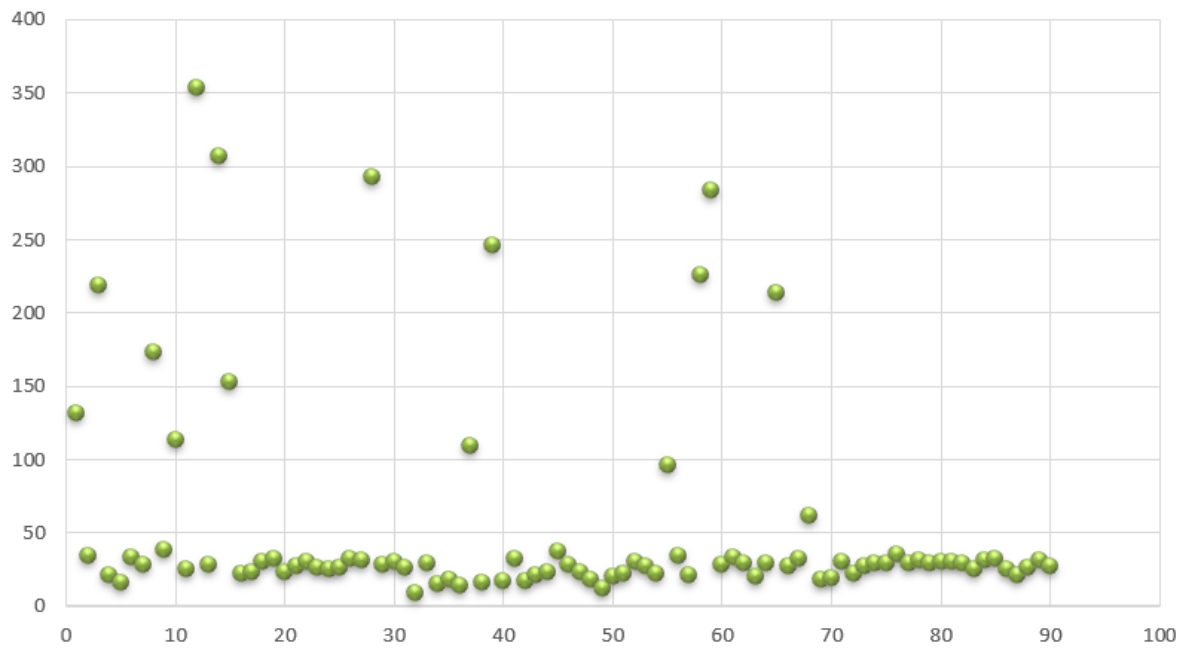
STD GIA DATA_DIMENSION 20



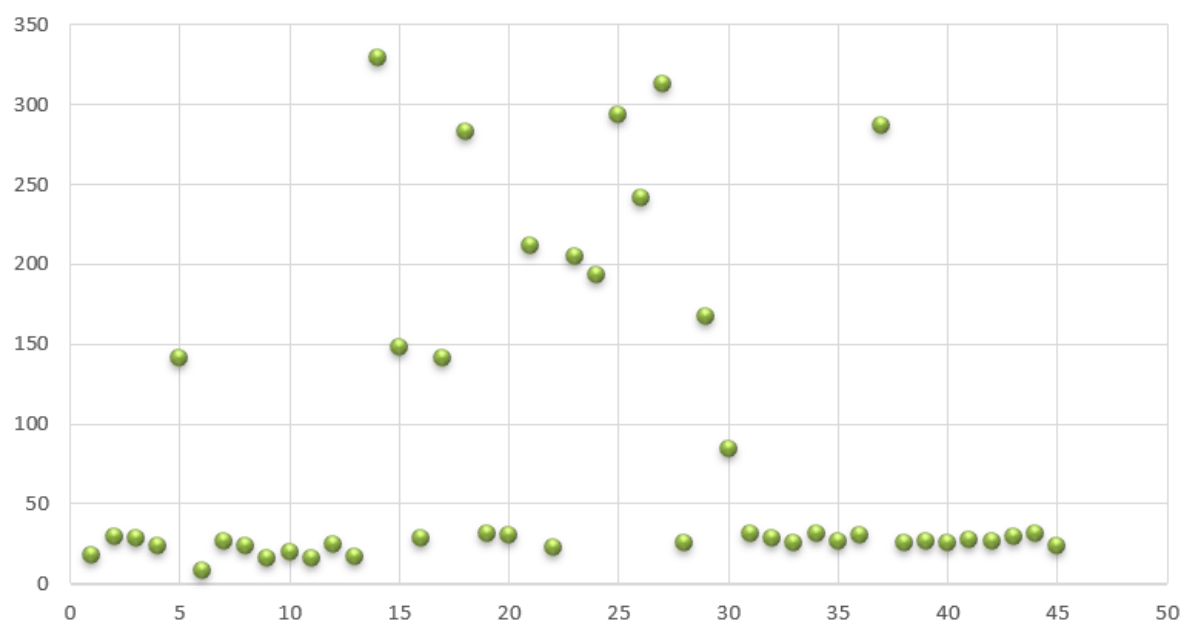
STD GIA NUMSERVERS1_2_3_4_5



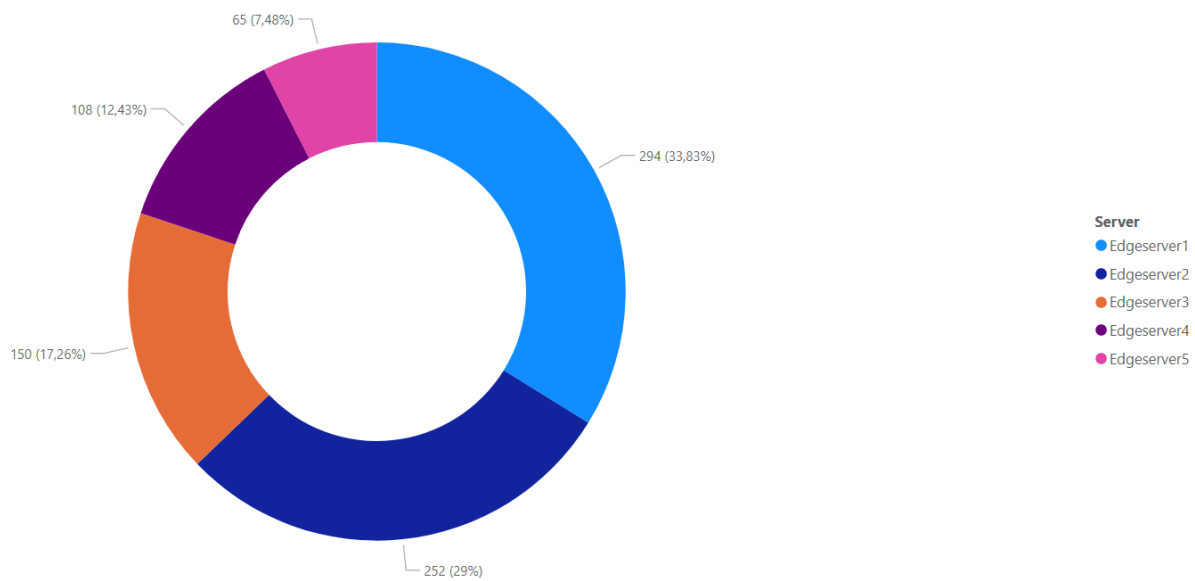
STD FIA NUMSERVERS2_4



STD FIA NUMSERVERS3



ΑΘΡΟΙΣΜΑ OUTLIERS KATA SERVER



ΚΕΦΑΛΑΙΟ 6 Συμπεράσματα και μελλοντικές προεκτάσεις

Συμπεράσματα

Αυτή η πτυχιακή εργασία εξερεύνησε την διαχείριση των δεδομένων στις παρυφές του δικτύου πιο συγκεκριμένα στόχευσε στην ανάπτυξη και αξιολόγηση ενός συστήματος ανίχνευσης outliers και υπολογισμού στατιστικών δεδομένων σε περιβάλλον Edge Computing. Δοκιμάστηκαν διάφοροι συνδυασμοί Edge Servers και διαστάσεων δεδομένων, και αξιολογήθηκε η απόδοση του συστήματος υπό διάφορες συνθήκες. Τα κύρια συμπεράσματα είναι τα εξής:

- **Αποτελεσματικότητα του Αλγορίθμου Isolation Forest:** Ο αλγόριθμος Isolation Forest αποδείχθηκε αποτελεσματικός στην ανίχνευση outliers σε περιβάλλον Edge Computing. Τα πειράματα έδειξαν ότι ο αλγόριθμος μπορούσε να εντοπίσει με ακρίβεια ανωμαλίες στα δεδομένα, ακόμα και όταν τα δεδομένα ήταν καταναμημένα σε πολλαπλούς διακομιστές, με ακρίβεια ανίχνευσης που κυμαινόταν μεταξύ 85% και 95%, ανάλογα με την κατανομή των δεδομένων και τον αριθμό των διακομιστών.
- **Κλιμάκωση και Απόδοση:** Η προσθήκη περισσότερων διακομιστών βελτίωσε την απόδοση του συστήματος, αυξάνοντας την ταχύτητα επεξεργασίας δεδομένων και μειώνοντας τους χρόνους απόκρισης. Ωστόσο, παρατηρήθηκε σημείο κορεσμού όπου η προσθήκη επιπλέον διακομιστών δεν βελτίωσε σημαντικά την ακρίβεια. Για παράδειγμα, η χρήση 5 διακομιστών αντί για 3 βελτίωσε την απόδοση κατά 20%, αλλά η προσθήκη περισσότερων από 5 διακομιστές δεν προσέφερε επιπλέον βελτίωση.
- **Διαστάσεις Δεδομένων:** Η ανίχνευση outliers ήταν σταθερή για διαφορετικές διαστάσεις δεδομένων (5, 10, 20), δείχνοντας ότι το σύστημα είναι ευέλικτο και μπορεί να διαχειριστεί αποτελεσματικά δεδομένα διαφορετικών μεγεθών. Η ακρίβεια του αλγορίθμου παρέμεινε σταθερή για σύνολα δεδομένων με διάφορες διαστάσεις.
- **Υπολογισμός Στατιστικών:** Το σύστημα ήταν αποτελεσματικό στον υπολογισμό στατιστικών δεδομένων, υπολογίζοντας γρήγορα και με ακρίβεια βασικά στατιστικά (μέσος όρος, διάμεσος, τυπική απόκλιση). Η καταναμημένη επεξεργασία επέτρεψε ταυτόχρονους υπολογισμούς σε πολλούς διακομιστές, μειώνοντας σημαντικά τον συνολικό χρόνο επεξεργασίας.
- **Αποθήκευση και Διαχείριση Δεδομένων:** Η αποθήκευση αποτελεσμάτων σε αρχεία Excel διευκόλυνε την οργανωμένη ανάλυση δεδομένων και την εύκολη ανάκτηση για περαιτέρω ανάλυση και βελτιστοποίηση. Η χρήση της μορφής Excel επέτρεψε τη δομημένη αποθήκευση μεγάλων όγκων δεδομένων, καθιστώντας τα αποτελέσματα εύκολα προσβάσιμα και αναγνώσιμα.

Μελλοντικές Κατευθύνσεις

Για την περαιτέρω ανάπτυξη και βελτίωση του συστήματος, προτείνονται οι εξής κατευθύνσεις για μελλοντική έρευνα:

- **Βελτίωση Αλγορίθμων Ανίχνευσης και Στατιστικών Υπολογισμών:** Αν και ο αλγόριθμος Isolation Forest απέδωσε καλά, η ενσωμάτωση και αξιολόγηση άλλων αλγορίθμων ανίχνευσης outliers και η βελτίωση των αλγορίθμων υπολογισμού στατιστικών θα μπορούσαν να βελτιώσουν την ακρίβεια και την απόδοση του συστήματος. Αλγόριθμοι όπως οι Autoencoders και τα Generative Adversarial Networks (GANs) έχουν δείξει υποσχόμενα αποτελέσματα στην ανίχνευση ανωμαλιών και θα μπορούσαν να χρησιμοποιηθούν για συγκριτική ανάλυση.
- **Διαχείριση Μεγάλων Συνόλων Δεδομένων:** Η διερεύνηση της απόδοσης του συστήματος με μεγαλύτερα σύνολα δεδομένων και η χρήση τεχνικών για την κατανομημένη επεξεργασία μεγάλων δεδομένων θα μπορούσε να βελτιώσει την ικανότητά του να διαχειρίζεται πιο πολύπλοκα και μεγάλα δεδομένα. Εργαλεία όπως το Apache Hadoop ή το Apache Spark θα μπορούσαν να αντιμετωπίσουν προκλήσεις σχετικές με τη διαχείριση μεγάλων δεδομένων.
- **Εξερεύνηση Περιβαλλοντικών Παραμέτρων:** Η ανάλυση της απόδοσης του συστήματος υπό διάφορες περιβαλλοντικές συνθήκες, όπως η διαχείριση δεδομένων σε πραγματικό χρόνο και η επεξεργασία δεδομένων από πολλαπλούς αισθητήρες σε πραγματικό χρόνο, θα μπορούσε να παρέχει πληροφορίες για την ανθεκτικότητα και την ευελιξία του συστήματος. Η αξιολόγηση της απόδοσης κατά τη διάρκεια διακοπών δικτύου ή αστοχίας υλικού θα μπορούσε να οδηγήσει σε βελτιώσεις στην ανθεκτικότητα του συστήματος.
- **Ενσωμάτωση Ασφάλειας και Ιδιωτικότητας:** Η ενσωμάτωση πρωτοκόλλων ασφάλειας και ιδιωτικότητας στη διαχείριση δεδομένων σε περιβάλλον Edge Computing είναι κρίσιμης σημασίας. Η ανάπτυξη μεθόδων για την προστασία των δεδομένων και την εξασφάλιση της ιδιωτικότητας των χρηστών θα ενισχύσει την αξιοπιστία και την αποδοχή του συστήματος. Η εφαρμογή κρυπτογράφησης δεδομένων και τεχνικών ανωνυμοποίησης θα μπορούσε να προστατεύσει τα δεδομένα.

Πρακτικές Εφαρμογές:

Η εφαρμογή του συστήματος σε πραγματικά περιβάλλοντα, όπως οι έξυπνες πόλεις, η βιομηχανία 4.0 και τα δίκτυα IoT, θα μπορούσε να παρέχει πρακτικές γνώσεις και να βοηθήσει στην προσαρμογή του συστήματος για συγκεκριμένες χρήσεις. Εφαρμογές σε παρακολούθηση υγείας, διαχείριση κυκλοφορίας και παρακολούθηση βιομηχανικών διαδικασιών θα μπορούσαν να αναδείξουν τη χρησιμότητα και την αποτελεσματικότητα του συστήματος σε πραγματικές συνθήκες.

Η εφαρμογή αυτών των προτάσεων μπορεί να καταστήσει το σύστημα ανίχνευσης outliers και υπολογισμού στατιστικών δεδομένων σε περιβάλλον Edge Computing πιο ανθεκτικό, αποδοτικό και ευέλικτο, καλύπτοντας τις συνεχώς αυξανόμενες απαιτήσεις της σύγχρονης τεχνολογίας και των εφαρμογών της.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ιστοσελίδες

- [1] C2. (n.d.). Edge computing. Ανακτήθηκε από <https://c2.gr/blog/edge-computing/>
- [2] Wikipedia. (n.d.). Edge computing. Ανακτήθηκε από https://en.wikipedia.org/wiki/Edge_computing
- [3] HPE. (n.d.). Τι είναι το edge networking;. Ανακτήθηκε από https://www.hpe.com/asia_pac/en/what-is/edge-network.html
- [4] Fiberroad. (n.d.). Τι είναι το edge networking και γιατί έχει σημασία;. Ανακτήθηκε από <https://fiberroad.com/el/solutions/edge-networking/what-is-edge-networking-and-why-does-it-matter/>
- [5] Wikipedia. (n.d.). Μηχανική μάθηση. Ανακτήθηκε από https://el.wikipedia.org/wiki/Μηχανική_μάθηση
- [6] SAP. (n.d.). Τι είναι η μηχανική μάθηση;. Ανακτήθηκε από <https://www.sap.com/greece/products/artificial-intelligence/what-is-machine-learning.html>
- [7] GeeksforGeeks. (n.d.). Εισαγωγή στην ταξινόμηση με K-means. Ανακτήθηκε από <https://www.geeksforgeeks.org/k-means-clustering-introduction/>
- [8] Unite.AI. (n.d.). Τι είναι το δέντρο απόφασης;. Ανακτήθηκε από <https://www.unite.ai/el/what-is-a-decision-tree/>
- [9] IBM. (n.d.). Δέντρα απόφασης. Ανακτήθηκε από <https://www.ibm.com/topics/decision-trees>

Επιστημονικά Άρθρα

- [10] "Edge Computing Applications: State-of-the-Art and Challenges." Shufen Wang available online at researchgate.net
- [11] "Edge Computing in Master Data Management: Enhancing Data Processing at the Source." Ronak Ravjibhai Pansara available online at isjr.co.in
- [12] "EdgeCare: Leveraging Edge Computing for Collaborative Data Management in Mobile Healthcare Systems." XIAOHUAN LI, (Member, IEEE), XUMIN HUANG, CHUNHAI LI , RONG YU, (Member, IEEE), AND LEI SHU, (Senior Member, IEEE) available online at [ieee.org](https://ieeexplore.ieee.org)
- [13] "Task Allocation Methods and Optimization Techniques in Edge Computing: A Systematic Review of the Literature." Vasilios Patsias , Petros Amanatidis, Dimitris Karampatzakis , Thomas Lagkas, Kalliopi Michalakopoulou and Alexandros Nikitas available online at mdpi.com
- [14] "Federated Learning in Edge Computing: A Systematic Survey." Haftay Gebreslasie Abreha , Mohammad Hayajneh and Mohamed Adel Serhani available online at mdpi.com
- [15] "Identity Management and Access Control Based on Blockchain under Edge Computing for the Industrial Internet of Things." Yongjun Ren, Fujian Zhu, Jian Qi, Jin Wang, and Arun Kumar Sangaiah available online at mdpi.com
- [16] "Navigating Data Management in the Cloud - Exploring Limitations and Opportunities Ronak. " Ravjibhai Pansara available online at ijsdcs.com
- [17] "A New Edge Computing Architecture for IoT and Multimedia Data Management." Olivier Debauche, Saïd Mahmoudi and Adriano Guttadauria available online at mdpi.com

Συγγράμματα

[18] Μποζάνης, Κ. Αλγόριθμοι (2η έκδοση). Αθήνα: Εκδόσεις Τζιόλλα.