



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**Δυναμική πρόβλεψη βάσει παραμέτρων κλιμάκωσης
που επιτρέπει χρονικά-περιορισμένη παροχή πόρων**

Γιογουρτσόγλου Αβραάμ

Διπλωματική διατριβή για το
Π.Μ.Σ. “Διδακτική της Τεχνολογίας και Ψηφιακά Συστήματα”

Επιβλέπων: Θεμιστοκλέους Μαρίνος

Πειραιάς, Ιούλιος 2014

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μαρίνο Θεμιστοκλέους για το ενδιαφέρον που έδειξε, για τις πολύτιμες συμβουλές του και για την ιδιαίτερη στήριξη που μου παρείχε κατά την διάρκεια αυτής της πορείας μου. Ιδιαίτερες ευχαριστίες θα ήθελα να απευθύνω στον κ. Δημοσθένη Κυριαζή για την πολύτιμη βοήθεια του, στην έρευνα που πραγματοποιήθηκε, για την εκπόνηση της παρούσας διπλωματικής εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω όλους τους συναδέλφους μου, στην ερευνητική ομάδα, με τους οποίους συνεργαστήκαμε άψογα και επιτυχώς όλο αυτό το διάστημα. Τέλος, οφείλω τις πιο θερμές μου ευχαριστίες, στην οικογένεια και τους φίλους μου, που με στηρίζουν όλα αυτά τα χρόνια.

Πανεπιστήμιο Πειραιώς

Ευρετήριο Περιεχομένων

Ευχαριστίες.....	i
Ευρετήριο Περιεχομένων.....	ii
Ευρετήριο Σχημάτων.....	v
Περίληψη Εργασίας.....	vii
Κεφάλαιο 1: Εισαγωγή.....	1
1.1: Περίληψη.....	1
1.2: Ορισμός Υπολογιστικού Νέφους.....	1
1.3: Χρήση του Υπολογιστικού Νέφους.....	3
1.4: Ερευνητικό Πρόβλημα.....	4
1.5: Σκοπός και Αντικειμενικοί Στόχοι.....	6
1.6: Δομή Ερευνητικής Εργασίας.....	6
Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση.....	10
2.1: Εισαγωγή.....	10
2.2: Βασικές Έννοιες Υπολογιστικού Νέφους.....	11
2.2.1 Ορισμός και Χαρακτηριστικά του Υπολογιστικού Νέφους.....	11
2.2.2 Σχέση Υπολογιστικού Νέφους με Υπολογιστικά Πλέγματα.....	12
2.2.3 Μοντέλα Παροχής Υπηρεσιών.....	13
2.2.4 Μοντέλα Εφαρμογής ή Μοντέλα Ανάπτυξης.....	13
2.2.5 Ελαστικότητα στο Υπολογιστικό Νέφος (Elasticity).....	15
2.2.6 Ποιότητα υπηρεσιών (Quality of Service).....	16
2.2.7 Περιγραφή και Ορισμός Σύμβασης Παροχής Υπηρεσιών – Service Level Agreement (SLA).....	17
2.2.8 Τεχνολογίες που χρησιμοποιούνται στο Υπολογιστικό Νέφος.....	18
2.3: Μεθοδολογία Αναζήτησης Σχετικής Βιβλιογραφίας.....	23
2.4: Κριτική Ανάλυση Βιβλιογραφίας.....	25
2.5: Ανοιχτές Θεματικές Περιοχές για Έρευνα.....	30
2.6: Συμπεράσματα.....	31
Κεφάλαιο 3: Δυναμική Πρόβλεψη Απόδοσης Υπηρεσίας σε Νεφοϋπολογιστική.....	33
3.1: Εισαγωγή.....	33
3.2: Προκλήσεις Πρόβλεψης Απόδοσης Υπηρεσίας και Ανάγκης για Πόρους στα Υπολογιστικά Νέφη.....	33
3.3: Μοντέλο Δυναμικής Πρόβλεψης Απόδοσης Υπηρεσίας.....	36

3.4: Συμπεράσματα Προτεινόμενου Μοντέλου	41
Κεφάλαιο 4: Ερευνητική Μεθοδολογία.....	42
4.1: Εισαγωγή.....	42
4.2: Παρουσίαση και Τεκμηρίωση Ερευνητικών Μεθοδολογιών	42
4.2.1 Περιγραφική Προσέγγιση	43
4.2.2 Συσχετιστική Προσέγγιση.....	44
4.2.3 Προ-πειραματική Προσέγγιση	44
4.2.4 Ψευδο-πειραματική Προσέγγιση.....	44
4.2.5 Πειραματική Προσέγγιση.....	44
4.3: Επιλογή Προσέγγισης Μεθοδολογίας Συλλογής Εμπειρικών Δεδομένων	45
4.4: Επεξεργασία Εμπειρικών Δεδομένων.....	46
4.5: Συμπεράσματα Χρήσης της Ερευνητικής Μεθοδολογίας.....	47
Κεφάλαιο 5: Συλλογή Εμπειρικών Δεδομένων	48
5.1: Εισαγωγή.....	48
5.2: Παρουσίαση Εφαρμογής Συλλογής Δεδομένων.....	49
5.2.1 Δομή Εφαρμογής.....	49
5.2.2 Ορίσματα Εκτέλεσης της Εφαρμογής	57
5.3: Διαδικασία Συλλογής Δεδομένων.....	58
5.3.1 Ορισμός Παραμέτρων Εκτέλεσης της Διαδικασίας Συλλογής Δεδομένων.....	58
5.3.2 Παρουσίαση Δείγματος Δεδομένων που Συλλέχθηκαν.....	59
5.4: Συμπεράσματα.....	60
Κεφάλαιο 6: Επεξεργασία Δεδομένων.....	62
6.1: Εισαγωγή.....	62
6.2: Παρουσίαση Εργαλείου “R Project for Statistical Computing”.....	62
6.3: Διαδικασία Επεξεργασίας Δεδομένων.....	65
6.3.1 Παρουσίαση Επεξεργασίας Με Χρήση της Γλώσσας R.....	65
6.3.2 Παρουσίαση Διαγραμμάτων των Δεδομένων, για την Εξαγωγή Συμπερασμάτων.....	66
6.4: Συμπεράσματα.....	72
Κεφάλαιο 7: Προτεινόμενο Μοντέλο Δυναμικής Πρόβλεψης	75
7.1: Εισαγωγή.....	75
7.2: Παρουσίαση Μαθηματικών Εννοιών.....	75
7.2.1 Γραμμικό Υπόδειγμα	76
7.2.2 Εκτίμηση των Συντελεστών του Γραμμικού Υποδείγματος με τη Μέθοδο των Ελαχίστων Τετραγώνων (LS).....	77

7.2.3 Το Πολλαπλό Γραμμικό Υπόδειγμα.....	78
7.2.4 Προβλέψεις και Προβλεπτική Ικανότητα του Γραμμικού Υποδείγματος.....	80
7.2.5 Εκτίμηση Μη Γραμμικών Υποδειγμάτων	85
7.3: Τεκμηρίωση Μορφής Υποδείγματος	87
7.3.1 Καθορισμός Πολλαπλότητας Υποδείγματος.....	88
7.3.2 Καθορισμός Γραμμικότητας του Υποδείγματος.....	88
7.4: Εύρεση Μοντέλων Πρόβλεψης με Χρήση του Eureka	90
7.5: Σύγκριση και Επιλογή Καταλληλότερου Μοντέλου Πρόβλεψης.....	91
7.6: Έλεγχος Μοντέλου Πρόβλεψης.....	95
7.6.1 Επιλογή Κριτηρίων Ελέγχου	95
7.6.2 Διαδικασία Ελέγχου του Μοντέλου Πρόβλεψης	95
7.7: Συμπεράσματα	98
Κεφάλαιο 8: Συμπεράσματα και Μελλοντική Έρευνα.....	100
8.1: Εισαγωγή	100
8.2: Επίτευξη Στόχων Εργασίας.....	102
8.3: Συμπεράσματα Μελέτης.....	104
8.3.1 Συμπεράσματα Αναγκαιότητας Μοντέλων Πρόβλεψης.....	104
8.3.2 Συμπεράσματα Ανάλυσης Μοντέλου Δυναμικής Πρόβλεψης	105
8.4: Μελλοντική Έρευνα.....	105
Βιβλιογραφικές Αναφορές	107

Ευρετήριο Σχημάτων

Σχήμα 1.1: Συνοπτικός πίνακας των δυνατοτήτων για πρόβλεψη πόρων ορισμένων δημόσιων παρόχων υπηρεσιών Υπολογιστικού Νέφους	5
Σχήμα 1.2: Δομή ερευνητικής εργασίας	9
Σχήμα 2.1: Οπτική μοντελοποίηση του ορισμού του υπολογιστικού νέφους με βάση τον ορισμό του NIST	14
Σχήμα 2.2: Παρουσίαση πόρων ηλεκτρονικού υπολογιστή, πριν και μετά την εικονικοποίηση	19
Σχήμα 2.3: Παρουσίαση σχέσης μεταξύ ενορχήστρωσης και χορογραφίας στις υπηρεσίες διαδικτύου.....	20
Σχήμα 2.4: Παρουσίαση σχέσεων μεταξύ υπηρεσιών διαδικτύου (web services), της υπηρεσιοστρεφούς αρχιτεκτονικής (SOA) και του υπολογιστικού νέφους (Cloud Computing)	21
Σχήμα 2.5: Παρουσίαση αρχιτεκτονικής της τεχνολογίας Web 2.0.....	22
Σχήμα 2.6: Παρουσίαση παραδείγματος καταμέτρησης λέξεων ενός κειμένου (word count) με χρήση του μοντέλου προγραμματισμού MapReduce	23
Σχήμα 2.7: Φάσεις βιβλιογραφικής ανασκόπησης (μετάφραση από σελ. 54 του [36])	25
Σχήμα 2.8: Αλληλεπίδραση επιπέδων του Υπολογιστικού Νέφους	30
Σχήμα 3.1: Παρουσίαση ψευδο-κώδικα ροής προγράμματος, όπου ο αριθμός των παραμέτρων επηρεάζει την χρήση πόρων	35
Σχήμα 3.2: Ροή ερευνητικών εργασιών για την δημιουργία δυναμικού μοντέλου πρόβλεψης της απόδοσης μιας υπηρεσίας.....	40
Σχήμα 5.1: Διαδικασία αποστολής παραμέτρων και αιτημάτων	51
Σχήμα 5.2: Διαδικασία απόκρισης των αιτημάτων και συλλογής δεδομένων.....	52
Σχήμα 5.3: Διαδικασία αποστολής παραμέτρων σε γεωγραφική απεικόνιση.....	53
Σχήμα 5.4: Διαδικασία αποστολής παράλληλων αιτημάτων προς την υπηρεσία, σε γεωγραφική απεικόνιση.....	54
Σχήμα 5.5: Διαδικασία απόκρισης στα αιτήματα, σε γεωγραφική απεικόνιση	55
Σχήμα 5.6: Διαδικασία αποστολής των δεδομένων προς την εφαρμογή διαχείρισης, σε γεωγραφική απεικόνιση.....	56
Σχήμα 6.1: Πίνακας μαθηματικών πράξεων στην γλώσσα R.....	64
Σχήμα 6.2: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά δύο.....	67
Σχήμα 6.3: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά δύο	68
Σχήμα 6.4: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τρία	68
Σχήμα 6.5: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τρία.....	68
Σχήμα 6.6: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τέσσερα	69

Σχήμα 6.7: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τέσσερα	69
Σχήμα 6.8: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων (από Άμστερνταμ).....	70
Σχήμα 6.9: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Άμστερνταμ).....	70
Σχήμα 6.10: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Αθήνα)	71
Σχήμα 6.11: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Αθήνα).....	71
Σχήμα 6.12: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Ρώμη).....	71
Σχήμα 6.13: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Ρώμη)	72
Σχήμα 7.1: Σχέση ανάμεσα στις παρατηρούμενες (y_i), τις εκτιμημένες (\hat{y}_i) τιμές της εξαρτημένης μεταβλητής και τα κατάλοιπα $\hat{\epsilon}_i = y_i - \hat{y}_i$	78
Σχήμα 7.2: Ερμηνευμένο και ανερμήνευτο μέρος του δείγματος.....	83
Σχήμα 7.3: Εκτίμηση συντελεστών μη γραμμικού υποδείγματος.....	86
Σχήμα 7.4: Αναπαράσταση της μεταβλητής των αιτημάτων των χρηστών με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα.....	89
Σχήμα 7.5: Αναπαράσταση της μεταβλητής της χιλιομετρικής απόστασης με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα.....	89
Σχήμα 7.6: Αναπαράσταση της μεταβλητής του χρόνου εκτέλεσης με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα.....	89
Σχήμα 7.7: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Αθήνα).....	96
Σχήμα 7.8: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Άμστερνταμ).....	96
Σχήμα 7.9: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Ρώμη)	97
Σχήμα 7.10: Απεικόνιση τιμών χρόνου απόκρισης της υπηρεσίας Wikipedia.org, και τιμών του μοντέλου δυναμικής πρόβλεψης.....	98

Περίληψη Εργασίας

Το βασικό χαρακτηριστικό του υπολογιστικού νέφους (Cloud Computing) είναι η δυνατότητα που προσφέρει για δυναμική προσφορά και κλιμάκωση (βαθμιαία αύξηση ή μείωση) των πόρων. Με τον τρόπο αυτό, ανάλογα με τις απαιτήσεις για πόρους που έχει μια υπηρεσία, ακόμα και αν διαφέρουν ανά χρονικές περιόδους, υπάρχει η δυνατότητα να προσφέρονται χρησιμοποιώντας τα κατάλληλα εργαλεία που βρίσκονται στο επίπεδο της πλατφόρμας του υπολογιστικού νέφους. Ωστόσο η προετοιμασία που χρειάζεται για να προσφερθούν δυναμικά οι πόροι αυτοί, δεν είναι χρονικά αμελητέα. Ο χρόνος, ανάλογα με την κλιμάκωση των πόρων που ζητάει η υπηρεσία, μπορεί να είναι μεγάλος. Η δέσμευση πόρων, έτσι ώστε να είναι διαθέσιμοι σε περίπτωση ανάγκης, δεν είναι λύση, καθώς επιφέρει τη μη βέλτιστη χρήση αυτών, που θα μπορούσαν να προσφερθούν σε άλλες υπηρεσίες. Θα πρέπει λοιπόν, να δημιουργηθεί ένας μηχανισμός που θα προβλέπει την απόδοση της υπηρεσίας, με βάση τις παραμέτρους που την επηρεάζουν (π.χ. πλήθος παράλληλων αιτημάτων προς την υπηρεσία). Συνεπώς, σε περίπτωση πρόβλεψης αύξησης των αναγκών σε πόρους, να ξεκινάει η προετοιμασία τους την κατάλληλη στιγμή, ώστε να είναι έτοιμοι όταν θα χρειαστούν. Αντίστοιχα, σε περίπτωση πρόβλεψης μείωσης των αναγκών σε πόρους, να ξεκινάει η σταδιακή αποδέσμευσή τους.

Η εργασία αυτή ασχολείται με την δημιουργία ενός τέτοιου μηχανισμού δυναμικής πρόβλεψης. Αυτό επιτυγχάνεται με την ανάλυση της συμπεριφοράς μιας υπηρεσίας, η οποία βρίσκεται εγκατεστημένη σε υποδομές υπολογιστικού νέφους, για την δημιουργία ενός μοντέλου πρόβλεψης της απόδοσής της. Σκοπός είναι, με βάση το μοντέλο αυτό, η πρόβλεψη της απόδοσης της υπηρεσίας όταν αλλάζουν οι τιμές των παραμέτρων που την επηρεάζουν. Έτσι, καθορίζοντας ο πάροχος, τη μέγιστη επιθυμητή απόδοση της υπηρεσίας και ανάλογα τις αλλαγές στις παραμέτρους της, θα μπορούν να αυξηθούν είτε να μειωθούν οι πόροι που προσφέρονται σε αυτή. Για την μελέτη και ανάλυση αντιπροσωπευτικών δεδομένων, έγινε επιλογή μιας υπηρεσίας, βάσει συγκεκριμένων κριτηρίων που τέθηκαν. Τα κριτήρια αυτά αφορούν:

- Για την εκτέλεση της, να κάνει χρήση υποδομών υπολογιστικού νέφους.
- Η υπηρεσία, να μη κάνει αυξημένη χρήση πόρων, κατά την εκτέλεση της, αλλά να τους χρησιμοποιεί στην εξυπηρέτηση των αιτημάτων.
- Η υπηρεσία αυτή να χρησιμοποιείται ευρέως από μεγάλο μέρος του πληθυσμού, έτσι ώστε να αποτελεί μια αντιπροσωπευτική υπηρεσία που διαχειρίζεται μεγάλο όγκο δεδομένων.
- Τα αιτήματα προς την υπηρεσία, για την Ευρώπη, να εξυπηρετούνται από τις ίδιες υποδομές υπολογιστικού νέφους, και όχι ανάλογα την γεωγραφική θέση να εξυπηρετούνται τα αιτήματα από διαφορετικές υποδομές.

Ενώ πολλές υπηρεσίες πληρούν τα παραπάνω κριτήρια, θεωρήθηκε σκόπιμο να μελετηθεί η υπηρεσία της εγκυκλοπαιδικής ιστοσελίδας Wikipedia.org. Ο λόγος είναι ότι διαχειρίζεται μεγαλύτερο όγκο δεδομένων από αυτές που αξιολογήθηκαν (Britannica.com, Answers.com, Infoplease.com και Imdb.com). Προφανώς υπάρχουν και άλλες υπηρεσίες που πληρούν τα παραπάνω κριτήρια, η παρούσα όμως εργασία, ασχολήθηκε με την μελέτη συγκεκριμένων.

Κεφάλαιο 1: Εισαγωγή

1.1: Περίληψη

Στο κεφάλαιο αυτό γίνεται αναφορά στο υπολογιστικό νέφος, στα βασικά χαρακτηριστικά του, τα μοντέλα παροχής υπηρεσιών και τα μοντέλα εφαρμογής ή ανάπτυξης. Επίσης παρουσιάζεται η έννοια της ελαστικότητας, της ποιότητας των υπηρεσιών και της διασφάλισης τους μέσω της σύμβασης παροχής υπηρεσιών. Σκοπός είναι η σύνδεση των εννοιών αυτών, με το θέμα της παρούσας εργασίας, που είναι η μελέτη συμπεριφοράς και απόδοσης των υπηρεσιών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, και την δημιουργία αντίστοιχων μοντέλων πρόβλεψης. Στο Κεφάλαιο 2, γίνεται αναλυτικότερη παρουσίαση των εννοιών του υπολογιστικού νέφους, που παρουσιάζονται σε αυτό. Ακόμα, στο Κεφάλαιο 1, παρουσιάζεται η χρήση του υπολογιστικού νέφους από διάφορες γνωστές υπηρεσίες, όπως από μέσα κοινωνικής δικτύωσης και ηλεκτρονικά ταχυδρομεία. Τέλος, παρατίθεται το ερευνητικό πρόβλημα που μελετάται στην παρούσα εργασία, καθώς και ο σκοπός και οι αντικειμενικοί στόχοι αυτής και μια σύντομη περιγραφή της δομής της.

1.2: Ορισμός Υπολογιστικού Νέφους

Υπολογιστικό νέφος ονομάζεται η κατ' αίτηση διαδικτυακή κεντρική διάθεση υπολογιστικών πόρων (όπως δίκτυο, εξυπηρετητές, εφαρμογές και υπηρεσίες) με υψηλή ευελιξία, συνεχή υποστήριξη και υψηλή αυτοματοποίηση. Στο υπολογιστικό νέφος η αποθήκευση, η επεξεργασία και η χρήση δεδομένων, λογισμικού και υπηρεσιών γίνεται διαδικτυακά, μέσω απομακρυσμένων υπολογιστών σε κέντρα υπολογιστικής υποδομής (datacenters). Υπηρεσίες όπως η κατ' αίτηση παροχή εικονικών μηχανών, το διαδικτυακό ηλεκτρονικό ταχυδρομείο, οι μηχανές αναζήτησης και τα κοινωνικά δίκτυα συχνά βασίζονται στην τεχνολογία του υπολογιστικού νέφους. Σύμφωνα με τον ορισμό του Εθνικού Ινστιτούτου Τυποποιήσεων και Τεχνολογίας (NIST - National Institute of Standards and Technology) [1] το υπολογιστικό νέφος ορίζεται ως:

«Το Υπολογιστικό Νέφος (Cloud Computing) είναι ένα μοντέλο που επιτρέπει ευέλικτη, κατόπιν ζήτησης δικτυακή πρόσβαση σε ένα κοινόχρηστο σύνολο διαμορφώσιμων πόρων (π.χ. δίκτυα, κεντρικοί υπολογιστές, αποθηκευτικοί χώροι, εφαρμογές και υπηρεσίες), το οποίο μπορεί να τροφοδοτηθεί γρήγορα και να διατεθεί με ελάχιστη προσπάθεια διαχείρισης ή αλληλεπίδρασης με τον πάροχο της υπηρεσίας. Το μοντέλο του Υπολογιστικού Νέφους προωθεί την διαθεσιμότητα και αποτελείται από πέντε βασικά χαρακτηριστικά, τρία μοντέλα παροχής υπηρεσιών, και τέσσερα μοντέλα ανάπτυξης.» Μετάφραση από [1](σελίδα 2).

Το υπολογιστικό νέφος αποτελείται από πέντε βασικά χαρακτηριστικά. Αυτά είναι η αυτο-εξυπηρέτηση κατ' αίτηση, η ευρεία πρόσβαση στο δίκτυο, η κοινή διάθεση και διαχείριση των πόρων, η ταχεία ελαστικότητα πόρων και τα μετρήσιμα επίπεδα παροχής υπηρεσιών.

Επιπλέον έχουν επινοηθεί μερικά εμπορικά μοντέλα παροχής υπηρεσιών του υπολογιστικού νέφους προς τους καταναλωτές. Τα τρία κυριότερα είναι το λογισμικό νέφους ως υπηρεσία, όπου παρέχεται στον καταναλωτή η δυνατότητα χρήσης εφαρμογών (ή υπηρεσιών) που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, η πλατφόρμα νέφους ως υπηρεσία, όπου προσφέρεται στον καταναλωτή μια πλατφόρμα για την ανάπτυξη ή εγκατάσταση των δικών του εφαρμογών, και το υλικό νέφους ως υπηρεσία, όπου δίνεται η δυνατότητα στον καταναλωτή να δεσμεύσει φυσικούς πόρους από το υπολογιστικό νέφος. Όλα αυτά τα εμπορικά μοντέλα απευθύνονται στην προσφορά των υπηρεσιών του υπολογιστικού νέφους, στους καταναλωτές. Οι υποδομές του όμως, μπορούν να έχουν διαφορετικά μοντέλα εφαρμογής. Δηλαδή, η υποδομή του υπολογιστικού νέφους μπορεί να είναι είτε ιδιωτική, όπου χρησιμοποιείται αποκλειστικά από έναν καταναλωτή (όπως ένας οργανισμός), είτε κοινοτική, όπου μοιράζονται οι υποδομές μεταξύ μιας κοινότητας καταναλωτών με κοινές ανησυχίες (όπως οργανισμοί που ασχολούνται με την ασφάλεια των δεδομένων). Επίσης, η υποδομή μπορεί να είναι δημόσια, όπου διατίθεται στο ευρύ κοινό, είτε υβριδική, που συνδυάζει δύο ή περισσότερα από τα προηγούμενα μοντέλα εφαρμογής.

Ένα από τα πλεονεκτήματα του υπολογιστικού νέφους, είναι η ελαστικότητα που προσφέρει στους πόρους του. Είναι ο βαθμός ευελιξίας, όσο αφορά την ταχύτητα δέσμευσης και αποδέσμευσης των πόρων του, με αυτοματοποιημένο τρόπο, με αποτέλεσμα την δυναμική προσαρμογή των πληροφοριακών συστημάτων στις εκάστοτε ανάγκες του υπολογιστικού νέφους. Έτσι, σε θεωρητικό επίπεδο, γίνεται χρήση των πόρων του με τον βέλτιστο τρόπο, διαμοιράζοντας τους, όπου είναι αναγκαίοι, δίνοντας την εντύπωση στον καταναλωτή, ότι οι δυνατότητες του υπολογιστικού νέφους είναι απεριόριστες. Επιπλέον, με τον τρόπο αυτό, για τον καταναλωτή, οι πόροι είναι πάντοτε διαθέσιμοι και μπορεί να τους προμηθευτεί την στιγμή που τους χρειάζεται. Αυτή η δυνατότητα, προσφέρεται από τα εμπορικά μοντέλα, πληρωμής μόνο των υπηρεσιών που χρησιμοποιούνται κάθε φορά (pay-per-use models).

Όλες αυτές οι δυνατότητες, εξασφαλίζουν την ποιότητα της υπηρεσίας προς τον καταναλωτή. Οι μηχανισμοί ποιότητας των υπηρεσιών παρέχουν εγγυήσεις για την ομαλή εκτέλεση των προσφερόμενων υπηρεσιών. Ορίζονται, λοιπόν, επίπεδα ποιότητας της υπηρεσίας μεταξύ του πελάτη και του παρόχου της. Για τον ορισμό αυτών των επιπέδων, υπογράφεται μια συμβατική συμφωνία, που ονομάζεται σύμβαση παροχής υπηρεσιών, μεταξύ του πελάτη και του παρόχου, η οποία καθορίζει τις εγγυήσεις για την ικανότητα διατήρησης των επιπέδων ποιότητας της υπηρεσίας.

Όπως είναι φανερό, ένα από τα βασικά θέματα που αντιμετωπίζουν οι πάροχοι υπηρεσιών υπολογιστικού νέφους, είναι η εξασφάλιση της ποιότητας των υπηρεσιών που υπόσχονται. Αυτό σημαίνει ότι πρέπει να έχει γίνει μελέτη και ανάλυση των απαιτήσεων για πόρους από τις προσφερόμενες υπηρεσίες. Επιπλέον, θα πρέπει να γίνεται και σωστός διαμοιρασμός των πόρων του υπολογιστικού νέφους, προς τις υπηρεσίες, κάθε χρονική στιγμή, χωρίς να μένουν ανεκμετάλλευτοι πόροι, ενώ υπάρχει ανάγκη χρήσης τους από άλλες υπηρεσίες. Επίσης, η δέσμευση και αποδέσμευσή τους, δεν είναι ακαριαία. Απαιτείται ένα χρονικό διάστημα για τις διαδικασίες αυτές. Συνεπώς, για να είναι διαθέσιμοι οι πόροι από την υπηρεσία, θα πρέπει να έχει ξεκινήσει η διαδικασία προσφοράς τους προς την υπηρεσία,

νωρίτερα (ανάλογα με τον χρόνο προετοιμασίας των πόρων). Μια προσέγγιση του προβλήματος αυτού, είναι η δημιουργία μοντέλων πρόβλεψης των απαιτήσεων για χρήση πόρων, των υπηρεσιών που είναι εγκατεστημένες στις υποδομές. Έτσι, θα μπορεί να γίνει πρόβλεψη της απόδοσης των υπηρεσιών στην διάρκεια προσφοράς τους, ανάλογα την χρήση της ή με βάση άλλες παραμέτρους (π.χ. μεταφορά μεγάλου όγκου δεδομένων), και αντίστοιχα να διαμοιράζονται οι πόροι, κάθε χρονική στιγμή. Αυτή η προσέγγιση θα μελετηθεί και στην παρούσα εργασία. Στην Ενότητα 1.3 θα παρουσιαστούν οι κυριότερες χρήσεις των υπηρεσιών του υπολογιστικού νέφους, παρουσιάζοντας έτσι, την ευρεία χρήση τους από μεγάλο μέρος του πληθυσμού, αλλά και την διαχείριση μεγάλου όγκου δεδομένων που γίνεται.

1.3: Χρήση του Υπολογιστικού Νέφους

Παρόλο που κάποιοι χρήστες μπορεί να μην είναι εξοικειωμένοι με τον όρο, στη πραγματικότητα και στην καθημερινότητά μας ήδη εκμεταλλευόμαστε τα πλεονεκτήματα του υπολογιστικού νέφους. Οι υπηρεσίες που παρέχει μέσω των εφαρμογών του Web 2.0 συμπεριλαμβάνουν μεταξύ άλλων: τα κοινωνικά δίκτυα (π.χ. Facebook και Google+), το ηλεκτρονικό ταχυδρομείο (για παράδειγμα το Gmail, Hotmail και Yahoo!), τη διανομή φωτογραφιών και βίντεο (π.χ. Flickr και το YouTube), διαδικτυακές εφαρμογές (π.χ. Google Docs και το Adobe Express), καθώς και την αποθήκευση των αρχείων - δεδομένων (π.χ. Dropbox, OneDrive και Google Drive) και την επεξεργασία μεγάλου όγκου δεδομένων.

Βάσει της έρευνας της εταιρείας Rackspace, μιας από τις κορυφαίες εταιρείες παροχής υπηρεσιών στο υπολογιστικού νέφους, οι κυριότερες χρήσεις των προσφερόμενων υποδομών της είναι [2]:

- Αποθήκευση και διαμοιρασμός αρχείων
- Διαχείριση βάσεων δεδομένων
- Συστήματα Διαχείρισης Σχέσεων Πελατών (Customer Relationship Management)
- Εφαρμογές ηλεκτρονικού ταχυδρομείου
- Χρήση του μοντέλου πλατφόρμας για εφαρμογές διαδικτύου
- Δημιουργία αντιγράφων ασφαλείας των δεδομένων
- Φιλοξενία ιστοσελίδων
- Ηλεκτρονικά καταστήματα
- Ανάπτυξη και δοκιμή εφαρμογών
- Δημιουργία ιδιωτικών και υβριδικών υπολογιστικών νεφών.

Παρατηρούμε ότι πολλές εταιρείες χρησιμοποιούν για τις υπηρεσίες τους, υποδομές υπολογιστικού νέφους. Ακόμα και για μικρότερες εταιρείες ή εταιρείες που έχουν πρόσφατα ιδρυθεί, μπορούν να ξεκινήσουν χρησιμοποιώντας πόρους του υπολογιστικού νέφους με χαμηλό κόστος και αναλόγως τις ανάγκες, να τους αυξάνουν ή να τους μειώνουν. Για παράδειγμα αν μια εταιρεία δημιουργήσει ένα ηλεκτρικό κατάστημα και το εγκαταστήσει σε υποδομές νέφους, ανάλογα την επισκεψιμότητα του θα μπορεί να προσαρμόζει τους πόρους δυναμικά. Η πρόβλεψη της δυναμικής αυξομείωσης των πόρων

που χρησιμοποιούνται κάθε χρονική στιγμή, θα μελετηθεί στην συνέχεια από την παρούσα εργασία.

1.4: Ερευνητικό Πρόβλημα

Με το υπολογιστικό νέφος (Cloud Computing) μπορούν να προσφερθούν πόροι, που σε θεωρητικό επίπεδο τείνουν στο άπειρο, προσφέροντας κλιμάκωση και ελαστικότητα. Ωστόσο απαιτείται η σωστή διαχείριση των εν λόγω πόρων, για τον διαμοιρασμό τους στις εφαρμογές που τους χρησιμοποιούν. Οι πάροχοι υπηρεσιών υπολογιστικού νέφους, όπως Amazon (Amazon Web Services[3]), Rackspace[4], Google (Google AppEngine[5]), Microsoft (Microsoft Azure[6]), IBM (SoftLayer[7]), SAP[8], προσφέρουν αυτούς τους πόρους βάσει κάποιων οικονομικών μοντέλων. Μερικά παραδείγματα είναι το “Pay-Per-Use Software Model” [9], όπου γίνεται δηλαδή ορθολογισμός του κόστους μόνο με τη χρήση των πόρων και μόνο εφ’ όσον υπάρχει πραγματική ανάγκη γι’ αυτό, και το “Reserved Instance Model” [10], όπου πραγματοποιείται μια εφάπαξ πληρωμή για την δέσμευση συγκεκριμένων πόρων. Στο “Reserved Instance Model”, γίνεται χρήση της τεχνολογίας εικονικοποίησης, έτσι ώστε να κρυφτούν οι υποδομές του υπολογιστικού νέφους, επιτρέποντας σε εφαρμογές να δημιουργήσουν προσαρμόσιμα εικονικά υπολογιστικά και αποθηκευτικά συστήματα.

Οι πελάτες των παρόχων, αν γνωρίζουν τις ανάγκες της υπηρεσίας τους για πόρους, θα μπορούν να επιλέξουν το κατάλληλο μοντέλο για αυτούς. Επίσης όμως και από την πλευρά των παρόχων, σε μοντέλα όπως αυτό του “Pay-Per-Use Software”, θα πρέπει να υπάρχει κάποιος τρόπος έτσι ώστε να προβλέπεται η ανάγκη για πόρους από την υπηρεσία του πελάτη, ώστε να είναι διαθέσιμοι την στιγμή που τους χρειάζεται. Σκοπός είναι λοιπόν, να χρησιμοποιούνται μόνο οι απαραίτητοι πόροι ανά χρονική στιγμή, καλύπτοντας τις ανάγκες της υπηρεσίας. Σε περίπτωση όπου οι ανάγκες της υπηρεσίας είναι μεγαλύτερες από αυτές που προσφέρονται, τότε η υπηρεσία δεν θα είναι πλήρως διαθέσιμη ή ακόμα και καθόλου. Είναι εμφανής λοιπόν, η αναγκαιότητα να υπάρχει μια προσεκτική ανάλυση και πρόβλεψη των αναγκών σε πόρους της υπηρεσίας.

Οι πάροχοι λόγω της ανάγκης αυτής για πρόβλεψη πόρων, έχουν αναπτύξει μηχανισμούς. Οι μηχανισμοί αυτοί προσφέρουν υπηρεσίες για παρακολούθηση, διαχείριση και πρόβλεψη πόρων και υπηρεσιών. Ωστόσο αυτοί οι μηχανισμοί χρησιμοποιούν διαφορετικές τεχνικές μεταξύ τους. Ο πίνακας που ακολουθεί παρουσιάζει συνοπτικά τις δυνατότητες στην εξισορρόπηση φορτίου (Load Balacing), στην πρόβλεψη για τροφοδότηση (Provisioning) και την αυτόματη κλιμάκωση (Auto Scaling) των μηχανισμών που χρησιμοποιούν ορισμένοι κορυφαίοι πάροχοι [11].

Cloud Platforms	Load Balancing	Provisioning	Auto Scaling
Amazon Elastic Compute Cloud	✓	✓	✓
Eucalyptus	✓	✓	×
Microsoft Windows Azure	✓	✓ (fixed templates so far)	✓ (Manual)
Google App Engine	✓	✓	✓
Manjrasoft Aneka	✓	✓	✓
GoGrid Cloud Hosting	✓	✓	✓ (Programmatic way only)

Σχήμα 1.1: Συνοπτικός πίνακας των δυνατοτήτων για πρόβλεψη πόρων ορισμένων δημόσιων παρόχων υπηρεσιών Υπολογιστικού Νέφους

Αναλύοντας τις σύνηθες διαφορετικές υπηρεσίες που θα μπορούσαν να χρησιμοποιούν υπολογιστικούς και αποθηκευτικούς πόρους ενός υπολογιστικού νέφους, παρατηρούμε κοινά σημεία στην επίτευξη της βέλτιστης απόδοσης, βάσει των αναγκών τους. Οι τύποι των εφαρμογών που αναλύθηκαν είναι:

- **Εφαρμογές ιστοσελίδων (websites).** Βασικές παράμετροι που επηρεάζουν την απόδοση: εύρος ζώνης διαδικτύου (internet bandwidth), χρήστες που την επισκέπτονται, γεωγραφική θέση των δεδομένων.
- **Εφαρμογές βάσεων δεδομένων.** Βασικές παράμετροι που επηρεάζουν την απόδοση: εύρος ζώνης διαδικτύου (internet bandwidth), μέγεθος δεδομένων, μέγεθος προς αποθήκευση αρχείων, χρήστες που επικοινωνούν με την βάση, γεωγραφική θέση βάσης, ρυθμός αύξησης των δεδομένων.
- **Εφαρμογές αποθήκευσης δεδομένων (data storage).** Βασικές παράμετροι που επηρεάζουν την απόδοση: εύρος ζώνης διαδικτύου (internet bandwidth), γεωγραφική θέση, ρυθμός αύξησης των δεδομένων, μέγεθος δεδομένων, μέγεθος προς αποθήκευση αρχείων, χρήστες που επεξεργάζονται τα δεδομένα.

Οι κοινές παράμετροι και στους τρεις τύπους των εφαρμογών που αναλύσαμε είναι το εύρος ζώνης διαδικτύου, οι χρήστες που χρησιμοποιούν την υπηρεσία και η γεωγραφική θέση των δεδομένων. Αν θεωρήσουμε ότι οι πάροχοι προσφέρουν το μέγιστο δυνατό εύρος ζώνης διαδικτύου (internet bandwidth) για την πρόσβαση στις υπηρεσίες των πελατών τους, οι παράμετροι που επηρεάζουν τις ανάγκες για πόρους από το υπολογιστικό νέφος, είναι οι χρήστες και η γεωγραφική θέση των δεδομένων. Οπότε αυτοί οι δύο παράμετροι, καθώς και ο χρόνος που εκτελούνται αιτήματα στην υπηρεσία, θα πρέπει να ληφθούν υπόψη στην μελέτη για πρόβλεψη της απόδοσης της υπηρεσία στην διάρκεια του χρόνου εκτέλεσής της.

1.5: Σκοπός και Αντικειμενικοί Στόχοι

Σκοπός της εργασίας αυτής είναι η δημιουργία ενός μοντέλου πρόβλεψης της απόδοσης μιας υπηρεσίας που είναι εγκατεστημένη σε υποδομές υπολογιστικού νέφους. Για την μέτρηση της απόδοσης της, ορίζεται ως μονάδα μέτρησης ο **χρόνος απόκρισης** της υπηρεσίας. Οι παράμετροι που μελετώνται και επηρεάζουν της απόδοση της υπηρεσίας είναι:

Παράμετρος 1: Ο αριθμός των παράλληλων αιτημάτων από τους χρήστες προς την υπηρεσία.

Παράμετρος 2: Η γεωγραφική θέση απ' όπου εκτελούνται τα αιτήματα,

Παράμετρος 3: Ο χρόνος κατά τον οποίο εκτελούνται τα αιτήματα προς την υπηρεσία.

Οι αντικειμενικοί στόχοι για την επίτευξη του σκοπού της εργασίας είναι:

Αντικειμενικός στόχος 1: Ανασκόπηση σχετικής βιβλιογραφίας.

Αντικειμενικός στόχος 2: Δημιουργία μηχανισμού συλλογής των δεδομένων, που σχετίζονται με την απόδοση της υπηρεσίας και των παραμέτρων που την επηρεάζουν.

Αντικειμενικός στόχος 3: Ανάλυση των δεδομένων που προκύπτουν για την εξαγωγή συμπερασμάτων.

Αντικειμενικός στόχος 4: Δημιουργία μοντέλου δυναμικής πρόβλεψης της απόδοσης της υπηρεσίας.

Αντικειμενικός στόχος 5: Έλεγχος και αξιολόγηση του μοντέλου δυναμικής πρόβλεψης.

1.6: Δομή Ερευνητικής Εργασίας

Η δομή της συγκεκριμένης ερευνητικής εργασίας ακολουθεί τη μεθοδολογία όπως περιγράφεται από τους Phillips και Rugh[12], η οποία στηρίζεται σε τέσσερα δομικά στοιχεία ως ακολούθως:

- **Background Theory:** Εστιάζει στο ορισμό του προβλήματος (Κεφάλαιο 1), και στις βιβλιογραφικές έννοιες και αναφορές που σχετίζονται με την ανοιχτή ερευνητική θεματική περιοχή (Κεφάλαιο 2).
- **Focal Theory:** Αναφέρεται στην παραγωγή του μοντέλου, για την επίλυση του ερευνητικού προβλήματος (Κεφάλαιο 3).
- **Data Theory:** Αποτελείται από τον σχεδιασμό της έρευνας μέσω της ερευνητικής μεθοδολογίας (Κεφάλαιο 4), την συλλογή των εμπειρικών δεδομένων (Κεφάλαιο 5), την επεξεργασία τους (Κεφάλαιο 6), την παραγωγή του μοντέλου πρόβλεψης, και την αξιολόγησή του (Κεφάλαιο 7).
- **Novel Contribution:** Παρουσιάζεται μια σύνοψη της έρευνας που πραγματοποιήθηκε στην παρούσα εργασία. Επιπλέον αναφέρονται τα αποτελέσματα της έρευνας που πραγματοποιήθηκε, και πιθανές μελλοντικές επεκτάσεις της.

Στο σημείο αυτό, αναλύεται συνοπτικά η δομή της παρούσας ερευνητικής εργασίας:

Κεφάλαιο 1: Εισαγωγή

Στο πρώτο κεφάλαιο έγινε μια εισαγωγή στο υπολογιστικό νέφος, τα χαρακτηριστικά του, τα μοντέλα παροχής υπηρεσιών και τα μοντέλα εφαρμογής ή ανάπτυξης του. Αυτά αποτελούν έννοιες που συνδέονται με τα ζητήματα που απασχολούν την παρούσα ερευνητική εργασία. Επίσης παρουσιάστηκε το ερευνητικό πρόβλημα που θα μελετηθεί, καθώς και ο σκοπός και οι αντικειμενικοί στόχοι της εργασίας αυτής.

Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση

Στο Κεφάλαιο 2 παρουσιάζονται αναλυτικότερα οι βασικές έννοιες του υπολογιστικού νέφους, που αναφέρθηκαν στο Κεφάλαιο 1, και είναι απαραίτητες για την κατανόηση του ερευνητικού προβλήματος. Ακόμα, γίνεται αναφορά της μεθοδολογίας για την εύρεση της κατάλληλης σχετικής βιβλιογραφίας που μελετήθηκε. Στην Ενότητα 2.4 παρουσιάζεται η κριτική ανάλυση που έγινε, για τις εργασίες που ασχολούνται με παρόμοια ερευνητικά προβλήματα. Επιπλέον, γίνεται περιγραφή των ανοιχτών θεματικών περιοχών για μελλοντική έρευνα και τα συμπεράσματα που προέκυψαν για το συγκεκριμένο κεφάλαιο.

Κεφάλαιο 3: Δυναμική Πρόβλεψη Απόδοσης Υπηρεσίας σε Νεφοϋπολογιστική

Σκοπός του τρίτου κεφαλαίου είναι να εστιάσει στην ανοιχτή ερευνητική θεματική περιοχή που ερευνάται περαιτέρω στην παρούσα εργασία. Αυτή είναι η δυναμική πρόβλεψη της απόδοσης μιας υπηρεσίας, εγκατεστημένης σε υποδομές υπολογιστικού νέφους, βάσει των παραμέτρων κλιμάκωσης της, που επιτρέπει χρονικά περιορισμένη παροχή πόρων. Έτσι, παρουσιάζεται η πρόταση επίλυσης, με την δημιουργία ενός μοντέλου δυναμικής πρόβλεψης, που θα προβλέπει τις τιμές της απόδοσης της υπηρεσίας, βάσει συγκεκριμένων παραμέτρων που την επηρεάζουν.

Κεφάλαιο 4: Ερευνητική Μεθοδολογία

Στο Κεφάλαιο 4, αναφέρονται οι ερευνητική μεθοδολογίες που χρησιμοποιούνται σε παρόμοια ερευνητικά θέματα, καθώς και αυτή που επιλέχθηκε για την παρούσα εργασία. Στην συνέχεια, γίνεται ανάλυση της μεθοδολογίας συλλογής εμπειρικών δεδομένων, και επιλέγονται οι προσεγγίσεις της συσχετιστικής και της πειραματικής. Τέλος, παρουσιάζεται ο τρόπος με τον οποίο επεξεργάζονται τα δεδομένα, και τα συμπεράσματα που προκύπτουν από τη χρήση της ερευνητικής μεθοδολογίας συλλογής εμπειρικών δεδομένων.

Κεφάλαιο 5: Συλλογή Εμπειρικών Δεδομένων

Για την επίτευξη του σκοπού της παρούσας εργασίας, είναι απαραίτητη η συλλογή των απαραίτητων δεδομένων. Στο Κεφάλαιο 5, παρουσιάζεται η ανάπτυξη μηχανισμού για την αυτοματοποίηση της διαδικασίας. Η συγκεκριμένη εφαρμογή, αποτελείται από δύο ξεχωριστά τμήματα. Το πρώτο διαχειρίζεται την όλη διαδικασία, ενώ το δεύτερο βρίσκεται εγκατεστημένο σε υπολογιστές, από διαφορετικές γεωγραφικές θέσεις και πραγματοποιεί τα αιτήματα προς την υπηρεσία. Στο τέλος συγκεντρώνονται όλα τα δεδομένα, στο πρώτο τμήμα της εφαρμογής, που το διαχειρίζεται ο χρήστης. Ένα δείγμα των αποτελεσμάτων της συλλογής των δεδομένων, προβάλλεται στην Ενότητα 5.3.2.

Κεφάλαιο 6: Επεξεργασία Δεδομένων

Σκοπός του έκτου κεφαλαίου, είναι η παρουσίαση της επεξεργασίας των εμπειρικών δεδομένων που συλλέχθηκαν. Η διαδικασία αυτή πραγματοποιήθηκε, με την χρήση του εργαλείου της γλώσσας R. Η συγκεκριμένη γλώσσα χρησιμοποιείται για στατιστική επεξεργασία και ανάλυση δεδομένων. Έτσι, μετά την παρουσίαση του εργαλείου, αναφέρθηκε η επεξεργασία των εμπειρικών δεδομένων και τα συμπεράσματα που προέκυψαν. Τα συμπεράσματα αυτά, σχετίζονται με την ορθότητα εκτέλεσης της διαδικασίας συλλογής των δεδομένων, αλλά και με τη μορφή του υποδείγματος των δεδομένων. Επιπλέον τα δεδομένα μορφοποιήθηκαν, ώστε να αναλυθούν από τον επόμενο στάδιο, για την παραγωγή των μοντέλων πρόβλεψης.

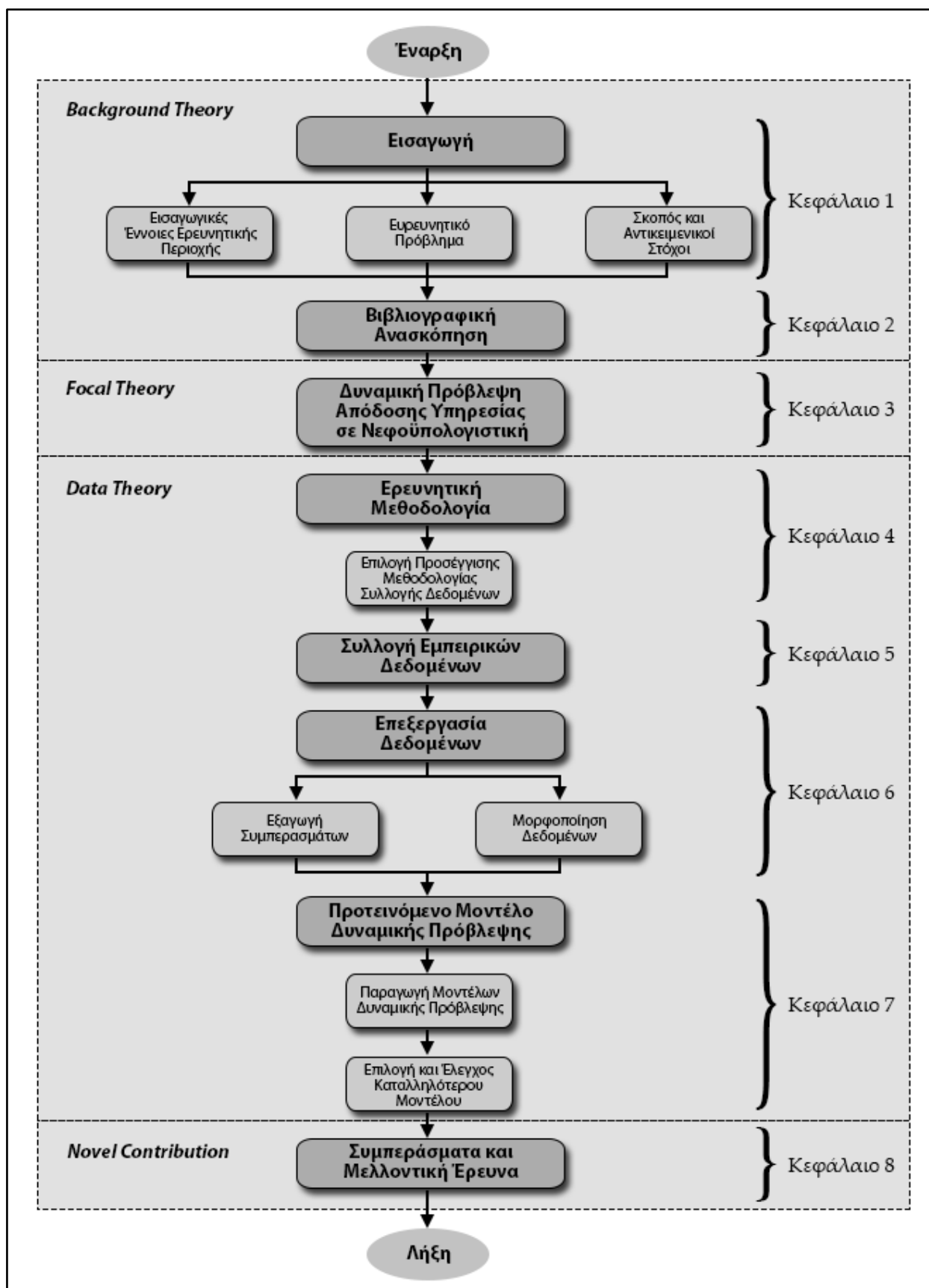
Κεφάλαιο 7: Προτεινόμενο Μοντέλο Δυναμικής Πρόβλεψης

Στο Κεφάλαιο 7, γίνεται παρουσίαση της διαδικασίας παραγωγής των μοντέλων πρόβλεψης και η επιλογή του καταλληλότερου. Αρχικά αναφέρονται οι μαθηματικές έννοιες, όπου χρησιμοποιούνται στην μελέτη των δεδομένων, για την ανάλυση του υποδείγματος τους. Με αυτό τον τρόπο, καθορίζεται η μορφή του υποδείγματος των δεδομένων, που είναι απαραίτητη για την παραγωγή των μοντέλων πρόβλεψης. Στην συνέχεια, μετά την παραγωγή τους, με χρήση του εργαλείου Eureka, επιλέγεται το καταλληλότερο μοντέλο, βάσει συγκεκριμένων κριτηρίων. Τέλος, το συγκεκριμένο μοντέλο, ελέγχεται βάσει της ερευνητικής μεθοδολογίας, που περιγράφει την διαδικασία, με την συλλογή δεδομένων και σύγκρισή τους με αυτά που προβλέπει το μοντέλο.

Κεφάλαιο 8: Συμπεράσματα και Μελλοντική Έρευνα

Στο τελευταίο κεφάλαιο αναφέρονται τα συμπεράσματα που προέκυψαν από την μελέτη του ερευνητικού προβλήματος και αυτά που προέκυψαν από τα αποτελέσματα της προτεινόμενης λύσης. Επίσης παρουσιάζεται η επίτευξη των αντικειμενικών στόχων, καθώς και μελλοντικές επεκτάσεις της παρούσας εργασίας.

Στην συνέχεια ακολουθεί σχήμα που παρουσιάζει την δομή της παρούσας εργασίας, όπως αναφέρθηκε προηγουμένως. Στο σχήμα αυτό, διακρίνονται ευκολότερα τα θέματα που παρουσιάζονται σε κάθε κεφάλαιο και ενότητα.



Σχήμα 1.2: Δομή ερευνητικής εργασίας

Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση

2.1: Εισαγωγή

Τα τελευταία 10 χρόνια, οι τεχνολογίες σχετικά με την εικονικοποίηση (virtualization technologies) πάνω στους υπολογιστικούς πόρους, έχουν γίνει αρκετά δημοφιλείς. Όπως είναι φυσικό, λόγω και της σχετικά πρόσφατης ανάπτυξης των τεχνολογιών αυτών, έχει γίνει εκτεταμένη έρευνα, που σχετίζεται με την πληροφορική. Βέβαια, από την στιγμή που το υπολογιστικό νέφος, έχει γίνει εμπορικό προϊόν, πολλοί πάροχοι, μέσα από τα τμήματα έρευνας και ανάπτυξης τους, κάνουν συνεχείς μελέτες για την βελτιστοποίηση της χρήσης των πόρων του. Η ανάγκη, λοιπόν, για εικονικοποίηση (virtualization) και κοινή χρήση υπολογιστικών υποδομών (shared infrastructures), όπως στα υπολογιστικά νέφη, δημιουργούν μια σειρά από σημαντικές προκλήσεις στους παρόχους που προσφέρουν τις υπηρεσίες αυτές. Εφαρμογές που είναι εγκατεστημένες στις υποδομές αυτές πρέπει να κάνουν χρήση, όσο το δυνατόν, τους ελάχιστους αλλά και απαραίτητους πόρους, κάθε χρονική στιγμή. Αυτό βέβαια πρέπει να γίνεται με τέτοιο τρόπο, έτσι ώστε να πληρούνται οι απαιτήσεις για τα επίπεδα απόδοσης της εφαρμογής.

Το ερευνητικό αυτό πρόβλημα, σχετικά με την επίτευξη της επιθυμητής απόδοσης των υπηρεσιών, και η χρήση των απαραίτητων πόρων ενός υπολογιστικού νέφους, έχει μελετηθεί αρκετές φορές, τα τελευταία 10 χρόνια. Σε ορισμένες περιπτώσεις με την δημιουργία μηχανισμών, διαχείρισης των πόρων του υπολογιστικού νέφους και την δέσμευσή τους για τις εφαρμογές, ανάλογα την δομή τους (π.χ. δέσμευση από ενός φυσικού αποκλειστικού πόρου για την εφαρμογή ή την δέσμευση περισσότερων κοινόχρηστων πόρων). Σε άλλες περιπτώσεις, με την κατηγοριοποίηση των εφαρμογών με βάση τα χαρακτηριστικά τους (π.χ. κύκλος εκτέλεσης της εφαρμογής) και την δημιουργία μοντέλων πρόβλεψης για την απόδοσή τους. Επίσης, όπως αναφέρθηκε και προηγουμένως, και οι εταιρείες, που παρέχουν υπηρεσίες υπολογιστικού νέφους, έχουν αναπτύξει διαφορετικούς μηχανισμούς πρόβλεψης των αναγκών σε πόρους, βάσει των απαιτήσεων απόδοσης των εφαρμογών που φιλοξενούν στις υποδομές τους.

Στο παρόν κεφάλαιο παρατίθεται η σχετική βιβλιογραφία που μελετήθηκε. Παρουσιάζονται αναλυτικότερα οι βασικές έννοιες του υπολογιστικού νέφους, που αναφέρθηκαν στο πρώτο κεφάλαιο, όπου είναι απαραίτητες για την κατανόηση του ερευνητικού προβλήματος. Επιπλέον, γίνεται αναφορά και κριτική ανάλυση άλλων ερευνών για την επίλυση παρόμοιων ερευνητικών θεμάτων, με την πρόβλεψη της απόδοσης των υπηρεσιών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Παρουσιάζονται, λοιπόν, τα κοινά σημεία της εργασίας αυτής με τις άλλες που μελετήθηκαν, καθώς και οι τεχνολογίες ή μέθοδοι επίλυσης που επιλέχθηκαν για να χρησιμοποιηθούν και σε αυτή την εργασία. Επίσης, αναφέρονται οι σχετικές ανοικτές θεματικές περιοχές για μελλοντική έρευνα, με έμφαση σε αυτές που σχετίζονται με την πρόβλεψη απόδοσης και ανάγκης για χρήση επιπλέον πόρων από τις εφαρμογές. Τέλος, προκύπτουν κάποια συμπεράσματα για τα ερευνητικά προβλήματα, που σχετίζονται με την παρούσα εργασία.

2.2: Βασικές Έννοιες Υπολογιστικού Νέφους

2.2.1 Ορισμός και Χαρακτηριστικά του Υπολογιστικού Νέφους

Το Εθνικό Ινστιτούτο Τυποποιήσεων και Τεχνολογίας (NIST - National Institute of Standards and Technology) [1] ορίζει το Υπολογιστικό Νέφος (Cloud Computing) ως εξής:

«Το Υπολογιστικό Νέφος (Cloud Computing) είναι ένα μοντέλο που επιτρέπει ευέλικτη, κατόπιν ζήτησης δικτυακή πρόσβαση σε ένα κοινόχρηστο σύνολο διαμορφώσιμων πόρων (π.χ. δίκτυα, κεντρικοί υπολογιστές, αποθηκευτικοί χώροι, εφαρμογές και υπηρεσίες), το οποίο μπορεί να τροφοδοτηθεί γρήγορα και να διατεθεί με ελάχιστη προσπάθεια διαχείρισης ή αλληλεπίδρασης με τον πάροχο της υπηρεσίας. Το μοντέλο του Υπολογιστικού Νέφους προωθεί την διαθεσιμότητα και αποτελείται από πέντε βασικά χαρακτηριστικά, τρία μοντέλα παροχής υπηρεσιών, και τέσσερα μοντέλα ανάπτυξης.» Μετάφραση από [1](σελίδα 2).

Όπως αναφέρεται στον ορισμό του NIST[1] τα ουσιώδη χαρακτηριστικά του υπολογιστικού νέφους είναι τα ακόλουθα:

- **Αυτο-εξυπηρέτηση κατ' αίτηση (On-demand self-service):** Ένας καταναλωτής μπορεί να δεσμεύσει από μόνος του τους υπολογιστικούς πόρους που χρειάζεται, όπως χρόνο στον server και αποθηκευτικό χώρο στο δίκτυο, ανάλογα με τις ανάγκες του αυτόματα, χωρίς να απαιτείται ανθρώπινη αλληλεπίδραση με το φορέα παροχής κάθε υπηρεσίας.
- **Ευρεία πρόσβαση στο δίκτυο:** Οι δυνατότητες είναι διαθέσιμες μέσω του δικτύου και προσβάσιμες μέσω τυποποιημένων μηχανισμών που προωθούν την χρήση από ετερογενείς thin ή thick client πλατφόρμες (π.χ. κινητά τηλέφωνα, φορητούς υπολογιστές και PDAs).
- **Κοινή διάθεση και διαχείριση των πόρων:** Οι υπολογιστικοί πόροι του παρόχου χρησιμοποιούνται για να εξυπηρετήσουν πολλαπλούς καταναλωτές με τη χρήση του μοντέλου πολλαπλών μισθωτών (multi-tenant), με τους διάφορους φυσικούς και εικονικούς πόρους να ανατίθενται δυναμικά και εκ νέου ανάλογα με τη ζήτηση των καταναλωτών. Υπάρχει μια αίσθηση ανεξαρτησίας από τον τόπο στο γεγονός ότι ο πελάτης δεν έχει γενικά κανέναν έλεγχο ή γνώση σχετικά με την ακριβή τοποθεσία των παρεχόμενων πόρων, αλλά μπορεί να είναι σε θέση να προσδιορίζει την τοποθεσία σε ένα υψηλότερο επίπεδο αφαίρεσης (π.χ. χώρα, κράτος, ή datacenter). Παραδείγματα πόρων αποτελούν οι αποθηκευτικοί χώροι, η επεξεργασία, η μνήμη, το bandwidth του δικτύου, καθώς και οι εικονικές μηχανές.
- **Ταχεία ελαστικότητα πόρων:** Οι πόροι μπορούν να δεσμευτούν προς χρήση γρήγορα και ελαστικά, σε ορισμένες περιπτώσεις αυτόματα, έτσι ώστε να εμφανιστούν άμεσα ως μη διαθέσιμοι (scale out) και επίσης να αποδεσμευτούν γρήγορα για να εμφανιστούν ξανά ως διαθέσιμοι (scale in). Για τον καταναλωτή, οι διαθέσιμες δυνατότητες για δέσμευση και χρήση συχνά φαίνεται να είναι απεριόριστες και μπορούν να αγοραστούν ανά πάσα στιγμή και σε οποιαδήποτε ποσότητα.

- **Μετρήσιμα επίπεδα παροχής υπηρεσιών:** Τα συστήματα υπολογιστικών νεφών ελέγχουν και βελτιστοποιούν αυτόματα τη χρήση των πόρων, αξιοποιώντας μια δυνατότητα μέτρησης σε κάποιο επίπεδο αφαίρεσης που είναι κατάλληλο για το είδος της υπηρεσίας (π.χ. αποθήκευση, επεξεργασία, bandwidth, ενεργοί λογαριασμοί χρηστών). Η χρήση των πόρων μπορεί να παρακολουθείται, να ελέγχεται, και να παρουσιάζεται με τη μορφή reports, παρέχοντας διαφάνεια τόσο για τον πάροχο όσο και για τον καταναλωτή της χρησιμοποιούμενης υπηρεσίας.

Η επίτευξη των παραπάνω στόχων γίνεται μέσω των Υπηρεσιοστρεφών Υποδομών [13]. Σε αυτές, οι βασικές λειτουργίες παρέχονται με τη μορφή υπηρεσίας, δηλαδή με την έκθεση μιας κατάλληλης διεπαφής μέσω δικτύου, μέσω της οποίας ένας καταναλωτής ή ένα λογισμικό διαχείρισης μπορεί αυτόματα να καλέσει για να πραγματοποιηθεί μια πράξη (π.χ. εκκίνηση μιας εικονικής μηχανής, εκκίνηση μιας υπηρεσίας λογισμικού κλπ.)

2.2.2 Σχέση Υπολογιστικού Νέφους με Υπολογιστικά Πλέγματα

Τα υπολογιστικά πλέγματα [14] ξεκίνησαν σαν μία πολλά υποσχόμενη τεχνολογία η οποία ενσωμάτωνε ετερογενείς υπολογιστικούς πόρους μεγάλης κλίμακας σε ένα κοινό σύστημα διαχείρισης. Οι πόροι αυτοί πιθανόν να άνηκαν στον ίδιο ή διαφορετικούς παρόχους (και αντίστοιχες γεωγραφικές τοποθεσίες) και η διαχείρισή τους επιτυγχανόταν μέσω των υπηρεσιοστρεφών αρχιτεκτονικών και υλοποιήσεων. Το υπηρεσιοστρεφές πλαίσιο φρόντιζε να παρέχει την απαραίτητη διασύνδεση και κοινή διαχείριση.

Παρόλο που η τεχνολογία αυτή ξεκίνησε πολύ δυναμικά, περιορίστηκε τελικά σε ακαδημαϊκούς και ερευνητικούς ρόλους ή σε πολύ συγκεκριμένες εφαρμογές που μπορούσαν να εκμεταλλευτούν τις δυνατότητές τους αλλά και να περιορίσουν τα μειονεκτήματά τους. Το βασικό μειονέκτημα ήταν η έλλειψη απομόνωσης μεταξύ των χρηστών, που δημιουργούσε προβλήματα σχετικά με την εγγύηση της παρεχόμενης υπηρεσίας, την ασφάλεια αλλά κυρίως με τις διαφορετικές απαιτήσεις περιβάλλοντος κάθε εφαρμογής. Έτσι, δύο εφαρμογές με διαφορετικές ανάγκες π.χ. σε λειτουργικό σύστημα, βιβλιοθήκες ή εκδόσεις λογισμικού δεν μπορούσαν να εκτελούνται στον ίδιο φυσικό πόρο.

Η βασική διαφορά των Υπολογιστικών Νεφών έγκειται στο γεγονός ότι χρησιμοποιούν την τεχνολογία της εικονικοποίησης (virtualization). Μέσω αυτής της τεχνικής, κάθε χρήστης είναι εντελώς απομονωμένος από τους υπόλοιπους και μπορεί να χειρίζεται κατά το δοκούν την διαμόρφωση του περιβάλλοντος εκτέλεσης των εφαρμογών του. Το γεγονός αυτό επέτρεψε στα Υπολογιστικά Νέφη να γίνουν σχεδόν αμέσως εμπορικά εκμεταλλεύσιμα και να μπορούν να χρησιμοποιηθούν για υπολογιστική ισχύ γενικού σκοπού. Μέσω αυτής της τεχνολογίας άτομα ή επιχειρήσεις έχουν πρόσβαση σε ελαστικούς υπολογιστικούς πόρους ή λογισμικό με πληρωμή ανάλογα με τη χρήση, χωρίς την ανάγκη κεφαλαιακής επένδυσης εξαρχής ή μετέπειτα συντήρησης. Έτσι μπορούν να τα προσαρμόζουν ανάλογα με τις ανάγκες τους ή τη ζήτηση, μειώνοντας το κόστος.

2.2.3 Μοντέλα Παροχής Υπηρεσιών

Υπάρχουν τρία μοντέλα παροχής υπηρεσιών νέφους με βάση τις τρεις θεμελιώδεις κατηγορίες οι οποίες συχνά αναφέρονται και ως μοντέλο SPI (Software, Platform or Infrastructure as a service) [15].

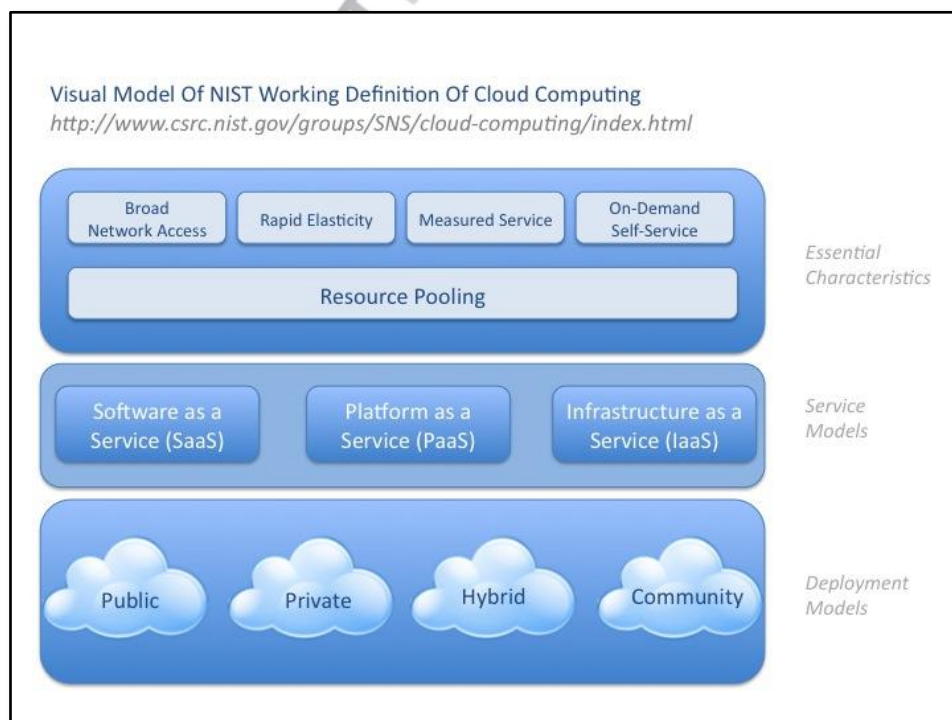
- **Λογισμικό νέφους ως υπηρεσία (Cloud Software as a Service - SaaS):** Η δυνατότητα που παρέχεται στον καταναλωτή είναι να χρησιμοποιεί τις εφαρμογές του παρόχου που τρέχουν σε μια υποδομή υπολογιστικού νέφους. Οι εφαρμογές είναι προσβάσιμες από διάφορες συσκευές μέσω μιας διεπαφής, όπως ένα πρόγραμμα περιήγησης στο ιντερνέτ (π.χ. web-based email). Ο καταναλωτής δεν έχει τη διαχείριση ή τον έλεγχο της χρησιμοποιούμενης υποδομής του νέφους, συμπεριλαμβανομένων των δικτύων, των λειτουργικών συστημάτων, των αποθηκευτικών μονάδων, ή ακόμα και μεμονωμένων δυνατοτήτων της εφαρμογής, με την πιθανή εξαίρεση κάποιων περιορισμένων, ειδικά για τον χρήστη ρυθμίσεων παραμετροποίησης των εφαρμογών. Παραδείγματα αυτής της κατηγορίας είναι το GoogleApps και το Office 365 της Microsoft.
- **Πλατφόρμα νέφους ως υπηρεσία (Cloud Platform as a Service - PaaS):** Η δυνατότητα που παρέχεται στον καταναλωτή είναι να αναπτύσσει πάνω στην υποδομή του υπολογιστικού νέφους, εφαρμογές που έχει δημιουργήσει ή εφαρμογές που έχει αποκτήσει, οι οποίες έχουν δημιουργηθεί με χρήση γλωσσών προγραμματισμού και εργαλείων που υποστηρίζονται από τον πάροχο. Ο καταναλωτής δεν διαχειρίζεται ούτε ελέγχει τη σχετική υποδομή του νέφους, που συμπεριλαμβάνει τα δίκτυα, τα λειτουργικά συστήματα ή τα αποθηκευτικά μέσα, αλλά έχει τον έλεγχο των εφαρμογών που έχουν αναπτυχθεί, και ενδεχομένως, των παραμετροποιήσεων του περιβάλλοντος φιλοξενίας των εφαρμογών. Παραδείγματα αυτής της κατηγορίας είναι το Azure της Microsoft και το Google App Engine.
- **Δομή του νέφους ως υπηρεσία (Cloud Infrastructure as a Service - IaaS):** Η δυνατότητα που παρέχεται στον καταναλωτή είναι να μπορεί να δεσμεύσει προς χρήση επεξεργαστική ισχύ, αποθηκευτικά μέσα, δίκτυα, και άλλους θεμελιώδεις υπολογιστικούς πόρους, όπου ο καταναλωτής είναι σε θέση να αναπτύξει και να εκτελέσει αυθαίρετο λογισμικό, το οποίο μπορεί να περιλαμβάνει λειτουργικά συστήματα και εφαρμογές. Ο καταναλωτής δεν έχει τη διαχείριση ή τον έλεγχο της χρησιμοποιούμενης υποδομής του υπολογιστικού νέφους, αλλά έχει τον έλεγχο των λειτουργικών συστημάτων, των αποθηκευτικών μέσων, των εφαρμογών που έχουν αναπτυχθεί και πιθανόν κάποιον περιορισμένο έλεγχο επιλεγμένου εξοπλισμού δικτύωσης (π.χ. τείχος προστασίας στα δίκτυα). Πάροχοι τέτοιων υπηρεσιών είναι η Amazon(EC2), Microsoft, Rackspace και Orange Business Service.

2.2.4 Μοντέλα Εφαρμογής ή Μοντέλα Ανάπτυξης

Το υπολογιστικό νέφος διατίθεται/προσφέρεται σε διαφορετικές μορφές: δημόσια νέφη, νέφη κοινότητας, ιδιωτικά Νέφη και Υβριδικά νέφη που συνδυάζουν και την δημόσια και την ιδιωτική μορφή λειτουργίας.

- **Ιδιωτικά νέφη (Private cloud):** Η υποδομή του υπολογιστικού νέφους λειτουργεί αποκλειστικά και μόνο για έναν. Η διαχείρισή της μπορεί να γίνεται από τον ίδιο τον οργανισμό ή από τρίτους και μπορεί να βρίσκεται εντός ή εκτός των εγκαταστάσεων του οργανισμού.
- **Νέφη κοινότητας (Community cloud):** Η υποδομή του υπολογιστικού νέφους μοιράζεται μεταξύ πολλών οργανισμών και υποστηρίζει μια συγκεκριμένη κοινότητα που έχει κοινές ανησυχίες (π.χ. αποστολή, απαιτήσεις ασφαλείας, πολιτική και θέματα συμμόρφωσης). Η διαχείρισή της μπορεί να γίνεται από τον ίδιο τον οργανισμό ή από τρίτους και μπορεί να βρίσκεται εντός ή εκτός των εγκαταστάσεων του οργανισμού.
- **Δημόσια νέφη (Public cloud):** Η υποδομή του υπολογιστικού νέφους διατίθεται στο ευρύ κοινό ή σε μια μεγάλη ομάδα εταιρειών και ανήκει σε έναν οργανισμό που πουλά υπηρεσίες υπολογιστικού νέφους. Προσφέρεται, λοιπόν, ένα σύνολο τυποποιημένων εφαρμογών και υπηρεσιών κατ' αίτηση χωρίς ο χρήστης να γνωρίζει ούτε που ακριβώς φιλοξενούνται τα δεδομένα του ούτε που ακριβώς πραγματοποιούνται οι επεξεργασίες των δεδομένων του. Πρόκειται για ένα μοντέλο ανάπτυξης γενικά δεκτικό για χρήση από ευρύ κοινό.
- **Υβριδικό Νέφος (Hybrid cloud):** Η cloud υποδομή είναι μια σύνθεση από δύο ή περισσότερα clouds (private, community or public) τα οποία παραμένουν μοναδικές οντότητες, αλλά συνδέονται μεταξύ τους με τυποποιημένη ή αποκλειστική τεχνολογία που επιτρέπει τη φορητότητα δεδομένων και εφαρμογών (π.χ. εξισορρόπηση φόρτου εργασίας μεταξύ των clouds).

Στο Σχήμα 1.1 παρουσιάζεται οπτικά το μοντέλο του υπολογιστικού νέφους βάσει του ορισμού που προτάθηκε από το NIST, και παρουσιάστηκε προηγουμένως.



Σχήμα 2.1: Οπτική μοντελοποίηση του ορισμού του υπολογιστικού νέφους με βάση τον ορισμό του NIST

2.2.5 Ελαστικότητα στο Υπολογιστικό Νέφος (Elasticity)

Η ελαστικότητα έχει αρχικά οριστεί στη φυσική ως μια ιδιότητα του υλικού να έχει την ικανότητα να επιστρέψει στην αρχική του κατάσταση μετά από μια παραμόρφωση. Στην οικονομικά θεωρία, ανεπίσημα, η ελαστικότητα υποδηλώνει την ευαισθησία μιας εξαρτημένης μεταβλητής, στις αλλαγές μίας ή περισσότερων άλλων μεταβλητών [16].

Και στις δύο περιπτώσεις, η ελαστικότητα είναι μια διαισθητική έννοια και μπορεί να περιγραφεί με ακρίβεια χρησιμοποιώντας μαθηματικούς τύπους. Η έννοια της ελαστικότητας έχει μεταφερθεί στο πλαίσιο του υπολογιστικού νέφους (Cloud Computing) και θεωρείται ως ένα από τα κεντρικά χαρακτηριστικά του μοντέλου του [17].

Για εμπορικούς σκοπούς, ο όρος ελαστικότητα χρησιμοποιείται σε μεγάλο βαθμό στις διαφημίσεις των παρόχων υποδομών υπολογιστικού νέφους, ακόμη και στην ονομασία κάποιων συγκεκριμένων-εξειδικευμένων προϊόντων ή υπηρεσιών τους. Ακόμα κι αν κατέβαλαν μεγάλες προσπάθειες για να μπορέσουν τα συστήματα αυτά του υπολογιστικού νέφους να συμπεριφέρονται με έναν ελαστικό τρόπο, δεν υπάρχει κοινή και ακριβή κατανόηση του όρου μέχρι στιγμής. Έτσι, δεν έχουν προταθεί τρόποι για να ποσοτικοποιηθεί, ώστε να συγκρίνετε η ελαστική συμπεριφορά. Για να υπογραμμιστεί η παρατήρηση αυτή, παραθέτουμε πέντε ορισμούς της ελαστικότητας που αποδεικνύουν την ασυνεπή χρήση και τη μη σαφή κατανόηση του όρου:

1. ODCA, Compute Υποδομών-as-a-Service [18]
"[...] ορίζει την ελαστικότητα ως παραμετροποίηση και επεκτασιμότητα της λύσης [...] Συγκεκριμένα, είναι η δυνατότητα να αναβαθμίζει (scale up) και να μειώνει (scale down) τους πόρους με βάση το φόρτο εργασίας."
2. NIST Ορισμός του Cloud Computing [1]
"Ταχεία ελαστικότητα πόρων: Οι πόροι μπορούν να δεσμευτούν προς χρήση γρήγορα και ελαστικά, σε ορισμένες περιπτώσεις αυτόματα, έτσι ώστε να εμφανιστούν άμεσα ως μη διαθέσιμοι (scale out) και επίσης να αποδεσμευτούν γρήγορα για να εμφανιστούν ξανά ως διαθέσιμοι (scale in). Για τον καταναλωτή, οι διαθέσιμες δυνατότητες για δέσμευση και χρήση συχνά φαίνεται να είναι απεριόριστες και μπορούν να αγοραστούν ανά πάσα στιγμή και σε οποιαδήποτε ποσότητα."
3. IBM, Thoughts on Cloud, Edwin Schouten, 2012 [19]
"Η ελαστικότητα είναι ουσιαστικά μια «μετονομασία» της κλιμάκωσης [...]" και «αφαιρεί» οποιαδήποτε χειροκίνητη εργασία που απαιτείται για να αυξήσει ή να μειώσει την απόδοση."
4. Rich Wolski, CTO, Eucalyptus, 2011 [20]
"Η ελαστικότητα μετρά την ικανότητα του υπολογιστικού νέφους να αντιστοιχίσει ένα απλό αίτημα του χρήστη σε διαφορετικούς πόρους."
5. Reuven Cohen, 2009 [21]
"Η ελαστικότητα είναι «η ποσοτική δυνατότητα διαχείρισης, μέτρησης, πρόβλεψης και προσαρμογής μιας εφαρμογής βάσει των αναγκών της σε πραγματικό χρόνο, που τοποθετείται σε υποδομή που χρησιμοποιεί ένα συνδυασμό τοπικών και απομακρυσμένων υπολογιστικών πόρων."

Οι ορισμοί (1), (2) και (3) από κοινού περιγράφουν την ελαστικότητα ως κλιμάκωση των πόρων του συστήματος για να αυξήσει ή να μειώσει τις δυνατότητες του, σύμφωνα με τις οποίες οι ορισμοί (1), (2) και (5) αναφέρουν ρητά ότι το μέγεθος των πόρων που προβλέπονται συνδέονται κατά κάποιον τρόπο με την πρόσφατη ζήτηση ή τον φόρτο εργασίας. Σε αυτά τα δύο σημεία φαίνεται να υπάρχει κάποια συναίνεση. Οι ορισμοί (4) και (5) προσπαθούν να συλλάβουν την ελαστικότητα με κάποιον γενικό όρο ως «ποσοτικά» την ικανότητα του συστήματος να χειριστεί τις αιτήσεις των εφαρμογών χρησιμοποιώντας διαφορετικούς πόρους. Και τα δύο αυτά τα ορίσματα, δεν δίνουν συγκεκριμένες λεπτομέρειες σχετικά με τα βασικά χαρακτηριστικά της ελαστικότητας, ούτε παρέχουν καμία συμβουλή για το πως η ελαστικότητα μπορεί να μετρηθεί. Ο ορισμός (3) προβλέπει ότι δεν χρειάζεται κάποια χειροκίνητη εργασία, ενώ στον ορισμό του NIST (2), οι διαδικασίες που επιτρέπουν την ελαστικότητα, δεν χρειάζεται να είναι πλήρως αυτοματοποιημένες. Επιπλέον, ο ορισμός του NIST προσθέτει το επίθετο «ταχεία» στην ελαστικότητα και καταλήγει στην ιδεαλιστική εικόνα της «τέλειας» ελαστικότητας, όπου ατέλειωτοι πόροι είναι διαθέσιμοι με την κατάλληλη πρόβλεψη σε οποιοδήποτε χρονικό σημείο, με έναν τρόπο που ο τελικός χρήστης δεν αντιλαμβάνεται καμία μεταβλητότητα των επιδόσεων.

Έτσι λαμβάνοντας υπόψη τους προηγούμενους ορισμούς αλλά και την μελέτη του [22], προκύπτει το συμπέρασμα ότι η ελαστικότητα παρουσιάζεται ως:

Ο βαθμός ευελιξίας, όσο αφορά την ταχύτητα δέσμευσης και αποδέσμευσης πόρων, με αυτοματοποιημένο τρόπο, με αποτέλεσμα την δυναμική προσαρμογή των πληροφοριακών συστημάτων στις εκάστοτε ανάγκες ενός υπολογιστικού νέφους. Στο μυαλό του πελάτη καταναλωτή οι δυνατότητες του υπολογιστικού νέφους είναι απεριόριστες, ενώ οι πόροι που διαθέτει τείνουν στο άπειρο. Επιπλέον είναι πάντοτε διαθέσιμες και ο πελάτης μπορεί να προμηθευτεί κάθε φορά όσες ακριβώς χρειάζεται.

Η ταχύτητα, που κάθε φορά ο πάροχος του υπολογιστικού νέφους προσαρμόζεται στις απαιτήσεις του πελάτη, εξαρτάται από τις συμβάσεις (SLA – Service Level Agreements) που έχει συνάψει μαζί του. Αναλυτικότερα για την συμβατική συμφωνία μεταξύ του παρόχου υπηρεσιών υπολογιστικού νέφους και των πελατών του, αναφέρονται στην Ενότητα «2.2.7 Περιγραφή και Ορισμός Σύμβασης Παροχής Υπηρεσιών».

2.2.6 Ποιότητα υπηρεσιών (Quality of Service)

Στο πεδίο των τηλεπικοινωνιών και των δικτύων υπολογιστών ο όρος ποιότητα υπηρεσιών (Quality of Service, QoS) αναφέρεται σε μηχανισμούς διασφάλισης της στατικής ανάθεσης δικτυακών πόρων σε συνδέσεις οι οποίες το απαιτούν. Η ποιότητα υπηρεσιών υλοποιείται με απόδοση προτεραιοτήτων στις διαφορετικές συνδέσεις ενός δικτύου, έτσι ώστε όσες χρειάζονται σταθερούς πόρους (π.χ. εφαρμογές πραγματικού χρόνου, όπως βιντεοδιάσκεψη ή άλλες υπηρεσίες πολυμέσων) να είναι βέβαιο ότι τους διαθέτουν. Οι εν λόγω πόροι διασφαλίζουν χαρακτηριστικά της σύνδεσης όπως τον απαιτούμενο ρυθμό μετάδοσης δεδομένων, την απαιτούμενη καθυστέρηση, μεταβολή της καθυστέρησης, πιθανότητα απώλειας πακέτων κλπ. Οι μηχανισμοί ποιότητας υπηρεσιών παρέχουν εγγυήσεις για τη σταθερότητα ενός ή περισσότερων από τα χαρακτηριστικά αυτά της σύνδεσης υπό συνθήκες συμφόρησης και περιορισμένης χωρητικότητας του

τηλεπικοινωνιακού καναλιού. Επίσης η ποιότητα υπηρεσιών είναι απαραίτητη μόνο σε δίκτυα μεταγωγής πακέτων, αφού σε δίκτυα μεταγωγής κυκλώματος ο τύπος και τα χαρακτηριστικά κάθε σύνδεσης γίνονται αντικείμενο διαπραγμάτευσης κατά την εγκαθίδρυση της τελευταίας και παραμένουν σταθερά μέχρι τον τερματισμό της.

Τα υπολογιστικά περιβάλλοντα που προσφέρονται από το υπολογιστικό νέφος, μπορούν να εγγυηθούν στους χρήστες την ποιότητα της υπηρεσίας που προσφέρουν (π.χ. αποδόσεις του υλικού, όπως επεξεργαστική ισχύ, ταχύτητα αποθήκευσης δεδομένων και μέγεθος χρήσης της κύριας μνήμης) [23].

Καθώς, λοιπόν, ο αριθμός των χρηστών του διαδικτύου συνεχίζει να αυξάνεται, αυξάνονται και οι απαιτήσεις απόδοσης των υπηρεσιών που προσφέρονται. Επιπλέον πολλές από τις τελευταίες διαδικτυακές υπηρεσίες (π.χ. βιντεοδιάσκεψη, επεξεργασία βίντεο) απαιτούν υψηλές ταχύτητες δικτύου. Η απόδοση του δικτύου είναι ένα στοιχείο ενδιαφέροντος τόσο για τον χρήστη, όσο και για τον πάροχο υπηρεσιών. Έτσι οι πάροχοι υπηρεσιών διαδικτύου και υπολογιστικού νέφους, πρέπει να εφαρμόζουν τεχνικές και τεχνολογίες για να παρέχουν την καλύτερη δυνατή εξυπηρέτηση προς τους πελάτες τους.

Όταν ορισθούν τα επίπεδα της ποιότητας της υπηρεσίας μεταξύ του πελάτη και του παρόχου, μπορούν να υπογράψουν μια συμβατική συμφωνία που ονομάζεται σύμβαση παροχής υπηρεσιών (Service Level Agreement), η οποία καθορίζει τις εγγυήσεις για την ικανότητα ενός δικτύου ή πρωτοκόλλου να δώσει την απόδοση που βασίζεται σε αμοιβαία συμφωνηθέντα μέτρα. Σε άλλες προσεγγίσεις οι πόροι διατηρούνται σε κάθε βήμα της υπηρεσίας, βάσει των προκαθορισμένων πόρων που έχουν ζητηθεί από τον πελάτη.

2.2.7 Περιγραφή και Ορισμός Σύμβασης Παροχής Υπηρεσιών – Service Level Agreement (SLA)

Ως «Σύμβαση Παροχής Υπηρεσιών» - ή όπως αναφέρεται σε συντομία στη βιβλιογραφία SLA - Service Level Agreement, ορίζεται η σχέση ανάμεσα σε ένα πάροχο υπηρεσιών και τον πελάτη της υπηρεσίας. Ένα SLA περιγράφει την παρεχόμενη υπηρεσία, περιγράφει τους στόχους της υπηρεσίας αυτής καθορίζοντας τις υποχρεώσεις τόσο από τον πάροχο της υπηρεσίας όσο και του πελάτη. Μια τέτοια συμφωνία μπορεί να καλύπτει πολλαπλές υπηρεσίες ή να αφορά διαφορετικούς πελάτες, ταυτόχρονα[24, 25].

Ένα SLA διαδραματίζει κεντρικό ρόλο στον κύκλο ζωής μιας υπηρεσίας, καθώς προβάλλει ορισμένες απαιτήσεις τόσο από τους «Παρόχους Υπηρεσιών Υπολογιστικού Νέφους» όσο και από τους «Πελάτες Υπηρεσιών Υπολογιστικού Νέφους». Θα μπορούσαμε να πούμε ότι ένα επιτυχημένο SLA θα πρέπει να εκπληρώνει τους εξής στόχους:

- Να περιγράφει με ακρίβεια τις υπηρεσίες τις οποίες εμποτεύει.
- Να καθορίσει την ποιότητα της υπηρεσίας με λεπτομέρεια.
- Να περιγράφει λεπτομερώς τους «Βασικούς Δείκτες Απόδοσης» - Key Performance Indicators - (KPI) [26].

Για να καθοριστεί με επιτυχία ένα SLA θα πρέπει οι «Πάροχοι Υπηρεσιών Υπολογιστικού Νέφους» να έχουν κατανοήσει κάποια βασικά στοιχεία όπως:

- Τους ρόλους και τις υποχρεώσεις οι οποίες καθορίζονται μέσα από το SLA [26].
- Να είναι σε θέση να περιγράψουν με ακρίβεια και συγκεκριμένα την υπηρεσία, για την οποία θα δημιουργηθεί το SLA.
- Να γνωρίζει τις απαιτήσεις των υπηρεσιών που θα εποπτεύεται από το συγκεκριμένο SLA, ώστε να μπορεί να καθορίσει ανάλογα κλειδιά που ταιριάζουν με την περίπτωση.
- Να μπορεί να ορίσει επίπεδα στην υπηρεσία με βάση κρίσιμα χαρακτηριστικά απόδοσης της υπηρεσίας.

2.2.8 Τεχνολογίες που χρησιμοποιούνται στο Υπολογιστικό Νέφος

Ορισμένες από τις τεχνολογίες που χρησιμοποιεί το Υπολογιστικό νέφος, συνοψίζονται παρακάτω:

- **Τεχνολογία εικονικοποίησης (Virtualization technology):**

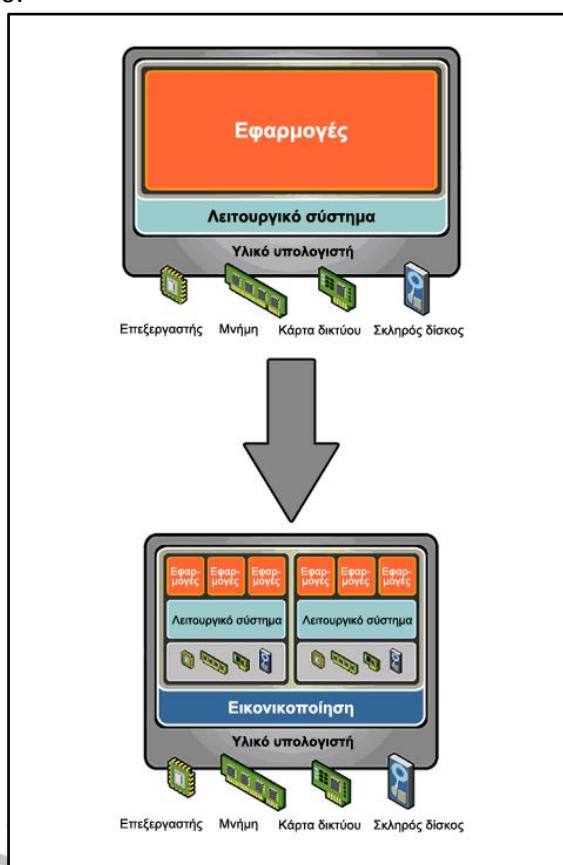
Η τεχνολογία της εικονικοποίησης χωρίζει και απομονώνει το υλικό (hardware) των υπολογιστών και έτσι παρέχει ευέλικτες και κλιμακωτές υπολογιστικές πλατφόρμες. Το υλικό αποτελεί το χαμηλότερο επίπεδο στις υποδομές του υπολογιστικού νέφους και περιλαμβάνει την προσωρινή μνήμη αποθήκευσης (RAM), αποθηκευτικούς χώρους δεδομένων (σκληροί δίσκοι), επεξεργαστική ισχύ (κεντρικές μονάδες επεξεργασίας – CPU), κάρτες δικτύου και ότι άλλες συσκευές που μπορούν να χρησιμοποιηθούν από τα υπολογιστικά συστήματα (π.χ. αισθητήρες συλλογής δεδομένων). Στο σχήμα 2.2, παρουσιάζονται στο πάνω μέρος, οι υπολογιστικοί πόροι όταν χρησιμοποιούνται από ένα σύνολο εφαρμογών, που είναι εγκατεστημένες σε έναν υπολογιστή, και στο κάτω μέρος, μετά την διαδικασία της εικονικοποίησης, όταν οι ίδιοι πόροι διαμοιράζονται σε περισσότερους εικονικούς υπολογιστές με δικά τους σύνολα εφαρμογών ο καθένας. Οι τεχνικές εικονικών υπολογιστικών πόρων (όπως VMware[27], Citrix XenServer[28] και προϊόντα της Microsoft[29]), προσφέρουν στους καταναλωτές, υπολογιστικές υποδομές κατά ζήτηση (on demand).

Η τεχνολογία αυτή όμως, δεν απευθύνεται μόνο στο υλικό των υπολογιστών. Χρησιμοποιείται επίσης και σε επίπεδο λογισμικού (π.χ. όταν οι προσφερόμενες υπηρεσίες χρησιμοποιούν το υλικό των υπολογιστών). Έτσι γίνεται ένας συνδυασμός του υλικού και του λογισμικού, ώστε να καλυφθούν οι απαιτήσεις προς το υπολογιστικό νέφος [30]. Η εικονικοποίηση εφαρμογών χρησιμοποιεί εικονικά πακέτα λογισμικού για να τοποθετήσει εφαρμογές, και δεδομένα αντί για συμβατικές διαδικασίες εγκατάστασης. Τα πακέτα εφαρμογών μπορούν να ενεργοποιηθούν ή και να απενεργοποιηθούν άμεσα, να επανέλθουν στην αρχική τους παραμετροποίηση, επομένως και να μετριαστεί αρκετά ο κίνδυνος της ανάμιξης τους με άλλες εφαρμογές ενώ τρέχουν στον δικό τους υπολογιστικό χώρο. Οι τεχνικές εικονικοποίησης δικτύων υπολογιστών, όπως η VPN[31], προσφέρει στους χρήστες προσαρμοσμένα δικτυακά περιβάλλοντα, ώστε να έχουν πρόσβαση στους πόρους του υπολογιστικού νέφους.

Οι τεχνικές εικονικοποίησης είναι απαραίτητες, δεδομένου ότι επιτρέπουν ευέλικτες και κλιμακωτές υπηρεσίες υλικού. Έτσι μια κοινή ερμηνεία της

τεχνολογίας της εικονικοποίησης είναι η αντιστοιχία των φυσικών πόρων ενός υπολογιστικού συστήματος, σε περισσότερες λογικές αναπαραστάσεις ή λογικά τμήματα. Η τεχνολογία αυτή θεωρείται απαραίτητη σε μια δυναμικά προσφερόμενη υποδομή υλικού ενός υπολογιστικού νέφους, καθώς παρέχει σημαντικά πλεονεκτήματα στην απομόνωση, διαμοιρασμό και διαχείριση των πόρων.

Με βάση αυτά που αναφέρθηκαν, για τον διαμοιρασμό των φυσικών πόρων, στις εφαρμογές, στο παρακάτω σχήμα παρουσιάζεται η τεχνολογία της εικονικοποίησης, όπου προσθέτει ένα επιπλέον επίπεδο μεταξύ του υλικού των υπολογιστών και του λογισμικού. Έτσι, οι ίδιοι πόροι μπορούν να διαμοιράζονται στο λογισμικό.

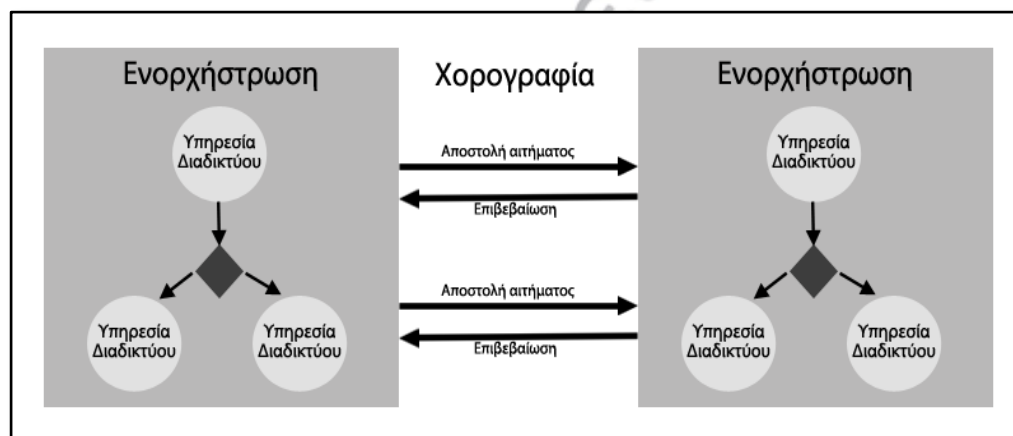


Σχήμα 2.2: Παρουσίαση πόρων ηλεκτρονικού υπολογιστή, πριν και μετά την εικονικοποίηση

- **Ενορχήστρωση της ροής εκτέλεσης της υπηρεσίας και της εργασίας (Orchestration of service flow and workflow - choreography):** Το υπολογιστικό νέφος προσφέρει μια σειρά από πρότυπα υπηρεσιών κατά ζήτηση, τα οποία μπορούν να συνθέσουν υπηρεσίες. Έτσι, τα υπολογιστικά νέφη θα πρέπει να προσφέρουν στους χρήστες, αυτόματα την ενορχήστρωση υπηρεσιών από διαφορετικές πηγές και από διαφορετικούς τύπους ροών υπηρεσιών ή ροών εργασιών. Οι όροι ενορχήστρωση και χορογραφία περιγράφουν δύο πτυχές της

δημιουργίας επιχειρηματικών διαδικασιών από σύνθετες υπηρεσίες ενός υπολογιστικού νέφους. Οι δύο αυτοί όροι επικαλύπτονται κάπως, αλλά στην σχήμα 2.3 απεικονίζεται ξεκάθαρα η σχέση μεταξύ τους, σε ένα υψηλό επίπεδο. Η ενορχήστρωση αναφέρετε σε μια εκτελέσιμη διαδικασία – υπηρεσία που μπορεί να αλληλεπιδράσει με άλλες εσωτερικές ή εξωτερικές διαδικασίες – υπηρεσίες. Οι αλληλεπιδράσεις γίνονται με την ανταλλαγή μηνυμάτων. Περιλαμβάνουν επιχειρηματική λογική και μια τάξη στην σειρά εκτέλεσης κάθε εργασίας, δημιουργώντας ένα πολυσταδιακό μοντέλο. Η ενορχήστρωση αντιπροσωπεύει τον έλεγχο πάνω σε ένα διαδικαστικό κομμάτι υπηρεσίας. Η χορογραφία της χορογραφίας είναι περισσότερο συλλογική και επιτρέπει σε κάθε εμπλεκόμενη διαδικασία να περιγράψει τον ρόλο της στην αλληλεπίδραση. Η χορογραφία «παρακολουθεί» τα μηνύματα μεταξύ των πηγών και των τμημάτων των διαδικασιών. Συνήθως τα μηνύματα αυτά προέρχονται από την αλληλεπίδραση μεταξύ διαδικτυακών υπηρεσιών και όχι μιας συγκεκριμένης επιχειρηματικής διαδικασίας που ένας και μόνο φορέας εκτελεί [32].

Στο παρακάτω σχήμα παρουσιάζεται η διαφορά μεταξύ της διαδικασίας της ενορχήστρωσης και της χορογραφίας, όπως αναφέρθηκε προηγουμένως.

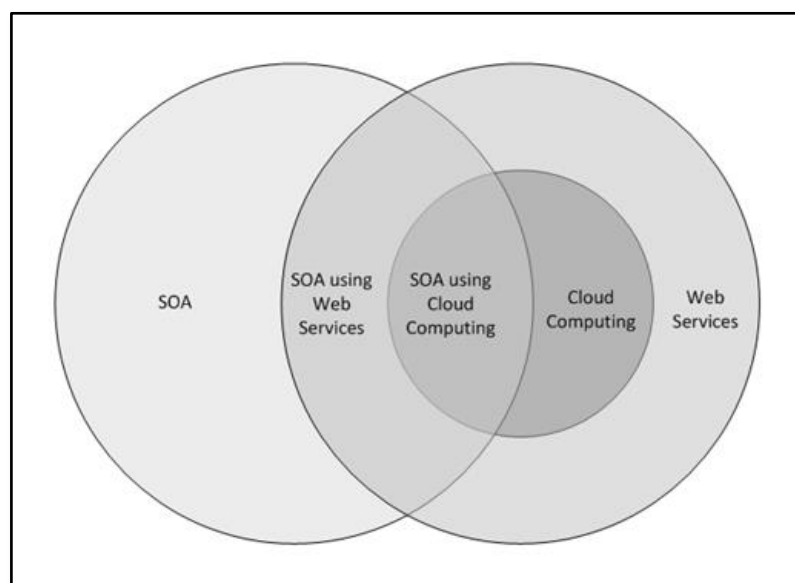


Σχήμα 2.3: Παρουσίαση σχέσης μεταξύ ενορχήστρωσης και χορογραφίας στις υπηρεσίες διαδικτύου

- Υπηρεσίες διαδικτύου (web services) και υπηρεσιοστρεφής αρχιτεκτονική (SOA - Service Oriented Architecture):**
 Οι υπηρεσίες του υπολογιστικού νέφους συνήθως προσφέρονται ως υπηρεσίες ιστού (web services), και ακολουθούν τα ίδια πρότυπα όπως WSDL[33, 34], SOAP[34, 35] και UDDI[34]. Η οργάνωση και ενορχήστρωση των υπηρεσιών στο υπολογιστικό νέφος διαχειρίζονται με βάση την υπηρεσιοστρεφή αρχιτεκτονική. Οι υπηρεσίες αυτές, θα μπορούσαν να χρησιμοποιηθούν από ένα περιβάλλον εφαρμογών υπηρεσιοστρεφούς αρχιτεκτονικής, έτσι ώστε να είναι διαθέσιμες σε διάφορες καταμεμημένες πλατφόρμες και να είναι προσβάσιμες μέσω διαδικτύου. Στο παρακάτω διάγραμμα Βεν, απεικονίζονται οι σχέσεις μεταξύ υπηρεσιών διαδικτύου, της υπηρεσιοστρεφούς αρχιτεκτονικής και του υπολογιστικού νέφους. Οι υπηρεσίες διαδικτύου ενσωματώνουν το υπολογιστικό νέφος στο διάγραμμα αυτό, επειδή το υπολογιστικό νέφος χρησιμοποιεί της υπηρεσίες σε όλα τα επίπεδα

του (μπορεί να υπάρχουν εξαιρέσεις αλλά είναι σπάνιες). Είναι δυνατόν, ωστόσο, να χρησιμοποιούνται οι υπηρεσίες διαδικτύου και σε άλλες περιπτώσεις και εκτός του υπολογιστικού νέφους. Η χρήση των υπηρεσιών διαδικτύου μπορεί να είναι μέρος μιας υπηρεσιοστρεφούς αρχιτεκτονικής, αλλά μπορεί και όχι. Οι υπηρεσίες θα μπορούσαν να είναι απλώς μια σύνδεση. Τέλος, είναι δυνατόν να έχουμε υπηρεσιοστρεφή αρχιτεκτονική χωρίς την χρήση υπηρεσιών διαδικτύου για τις συνδέσεις [36].

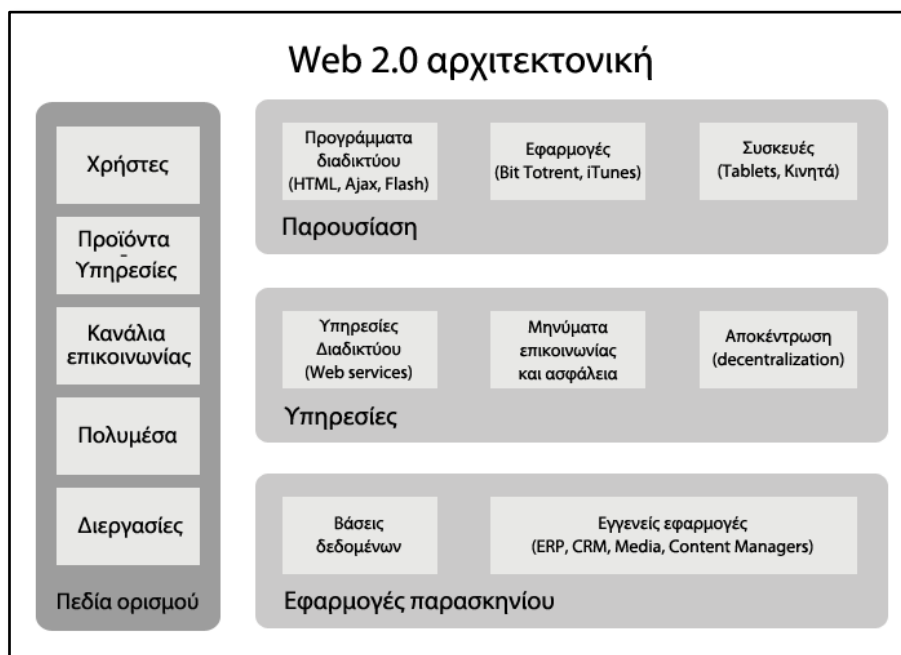
Στο ακόλουθο σχήμα, παρουσιάζεται η σχέση μεταξύ των υπηρεσιών διαδικτύου, της υπηρεσιοστρεφούς αρχιτεκτονικής και του υπολογιστικού νέφους.



Σχήμα 2.4: Παρουσίαση σχέσεων μεταξύ υπηρεσιών διαδικτύου (web services), της υπηρεσιοστρεφούς αρχιτεκτονικής (SOA) και του υπολογιστικού νέφους (Cloud Computing)

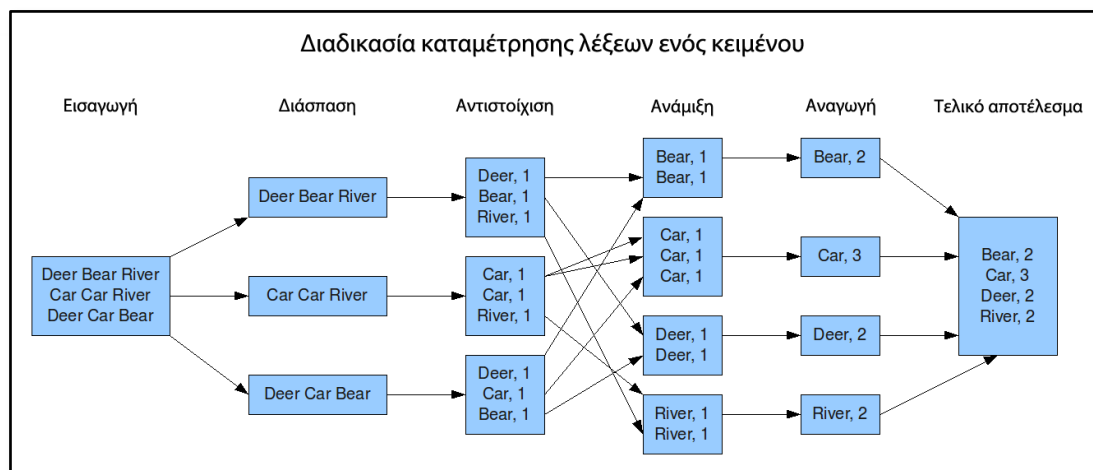
- **Web 2.0 τεχνολογία:**
Το Web 2.0 είναι μια σχετικά πρόσφατη τεχνολογία που περιγράφει τις καινοτόμες τάσεις της χρήσης της World Wide Web τεχνολογίας και του σχεδιασμού του διαδικτύου που έχει ως στόχο να ενισχύσει τη δημιουργικότητα, την ανταλλαγή πληροφοριών, την συνεργασία και την λειτουργικότητα του διαδικτύου. Η βασική ιδέα πίσω από το Web 2.0 είναι η βελτίωση της διασύνδεσης και της διαδραστικότητας των εφαρμογών διαδικτύου. Το νέο πρότυπο για την ανάπτυξη και την πρόσβαση στις διαδικτυακές εφαρμογές, επιτρέπει στους χρήστες να έχουν πρόσβαση στο διαδίκτυο πιο εύκολα και αποτελεσματικά. Ο χρήστης θα μπορεί να δρα στον Παγκόσμιο Ιστό όπως δρούσε μέχρι τώρα στον υπολογιστή του. Οι πιο ειδικοί μιλούν για έναν νέο τρόπο σχεδίασης των ιστοσελίδων ο οποίος θα βασίζεται στην διάδραση του χρήστη. Θα επιτρέπει στον χρήστη να αλλάξει τόσο το περιβάλλον της σελίδας όσο και να παρέμβει στο περιεχόμενό της. Χαρακτηριστικές εφαρμογές του Web 2.0 είναι τα κοινωνικά μέσα (social media), οι εγκυκλοπαιδικές σελίδες (wiki) και οι σελίδες ευρείας ενημέρωσης (blog). Πολλές από τις εντολές διάδρασης που χαρακτηρίζουν την λειτουργία του Web 2.0 μας είναι ήδη γνωστές από διάφορες ιστοσελίδες κοινωνικής δικτύωσης όπως το facebook ή το youtube

για παράδειγμα. Τέτοιες εκφράσεις είναι η αναζήτηση (search), η λέξη κλειδί (tag), η παράθεση συνδέσμων (links) ή η συγγραφή (authoring) όπως λειτουργεί σε πολλές σελίδες όπου οι χρήστες μπορούν να δημιουργήσουν άρθρα αλλά και να ανανεώσουν ή να διαγράψουν ήδη υπάρχοντα. Οι υπηρεσίες του υπολογιστικού νέφους είναι και αυτές εφαρμογές διαδικτύου, οι οποίες είναι διαθέσιμες κατά ζήτηση. Έτσι και το υπολογιστικό νέφος υιοθετεί την Web 2.0 τεχνολογία. Στο Σχήμα 1.5 γίνεται εικονική απεικόνιση της Web 2.0 αρχιτεκτονικής.



Σχήμα 2.5: Παρουσίαση αρχιτεκτονικής της τεχνολογίας Web 2.0

- **Προγραμματιστικό μοντέλο (Programming model):**
Οι χρήστες οδηγούνται στο υπολογιστικό νέφος με δεδομένα και εφαρμογές. Ορισμένα μοντέλα προγραμματισμού θα πρέπει να προταθούν ώστε οι χρήστες να προσαρμοστούν στις υποδομές του νέφους. Για λόγους απλότητας και εύκολης πρόσβασης στις υπηρεσίες του νέφους, το μοντέλο προγραμματισμού, δεν θα πρέπει να είναι υπερβολικά πολύπλοκο ή πολύ καινοτόμο για τους τελικούς χρήστες. Το MapReduce είναι ένα μοντέλο προγραμματισμού και μια σχετική υλοποίηση για επεξεργασία και παραγωγή μεγάλων δεδομένων, που χρησιμοποιήθηκε από την Google στις υποδομές της. Το μοντέλο MapReduce [37] περιλαμβάνει κατ' αρχάς την εφαρμογή μιας λειτουργίας αντιστοίχισης (map) σε ορισμένα αρχεία δεδομένων, ένα σύνολο από ζεύγη κλειδιά / τιμές (key/value pairs), και στη συνέχεια επεξεργάζεται μια λειτουργία "αναγωγής" (reduce) σε όλα τα δεδομένα που μοιράζονται το ίδιο κλειδί. Το Map-Reduce-Merge υπόδειγμα προσθέτει μια λειτουργία "συγχώνευσης". Η μέθοδος Hadoop εξελίσσει το μοντέλο MapReduce, με μια πλατφόρμα εκτέλεσης εφαρμογών σε μεγάλες υπολογιστικές μονάδες (clusters). Έτσι υλοποιεί το μοντέλο MapReduce και παρέχει ένα κατακεμημένο σύστημα αρχείων, το Hadoop Distributed File System. Το MapReduce και το Hadoop χρησιμοποιείται πλέον από μεγάλες εταιρείες στις υποδομές του υπολογιστικού νέφους τους, όπως Yahoo!, Intel και HP.



Σχήμα 2.6: Παρουσίαση παραδείγματος καταμέτρησης λέξεων ενός κειμένου (word count) με χρήση του μοντέλου προγραμματισμού MapReduce

2.3: Μεθοδολογία Αναζήτησης Σχετικής Βιβλιογραφίας

Εκτός από τις βασικές έννοιες του υπολογιστικού νέφους που αναλύθηκαν στην Ενότητα 2.2, θα πρέπει να μελετηθούν εις βάθος και οι έρευνες που έχουν γίνει για θέματα σχετικά με την απόδοση των εφαρμογών, που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Έτσι, με την μελέτη των σχετικών ερευνητικών εργασιών, θα επιλεγθούν οι κατάλληλες μέθοδοι και τεχνικές για την επίλυση παρόμοιων θεμάτων, ώστε να χρησιμοποιηθούν και στην παρούσα εργασία. Επίσης, θα αποσαφηνιστεί και η συνεισφορά της παρούσας εργασίας στο χώρο της μελέτης της ποιότητας των υπηρεσιών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους.

Για την εύρεση της σχετικής βιβλιογραφίας, επιλέχθηκε η κατάλληλη μεθοδολογία αναζήτησης, με συνδυασμό των μεθοδολογιών αναζήτησης που προτάθηκαν από τον Cooper[38] και τον Bregeton[39]. Η μεθοδολογία αυτή, χρησιμοποιήθηκε στην εργασία [40] για παρόμοια έρευνα. Σκοπός της Ενότητας 2.3 δεν είναι να γίνει εις βάθος ανάλυση της μεθοδολογίας, αλλά βάσει αυτής να περιγραφούν οι τρεις φάσεις της μεθοδολογίας έρευνας.

Φάση 1: Σχεδίαση

- **Σκοπός:** Ο σκοπός της βιβλιογραφικής ανασκόπησης είναι η εύρεση εργασιών που σχετίζονται με την μελέτη της ποιότητας των εφαρμογών, που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους.
- **Ερευνητικό πρωτόκολλο:** Το ερευνητικό πρωτόκολλο που ακολουθήθηκε, συνοψίζεται στις φάσεις της ερευνητικής μεθοδολογίας, που αναλύονται στην παρούσα ενότητα. Δεν κατασκευάστηκε κάποιο πρωτόκολλο για παρουσίαση, καθώς η έρευνα πραγματοποιήθηκε από ένα άτομο.

Φάση 2: Διεξαγωγή έρευνας

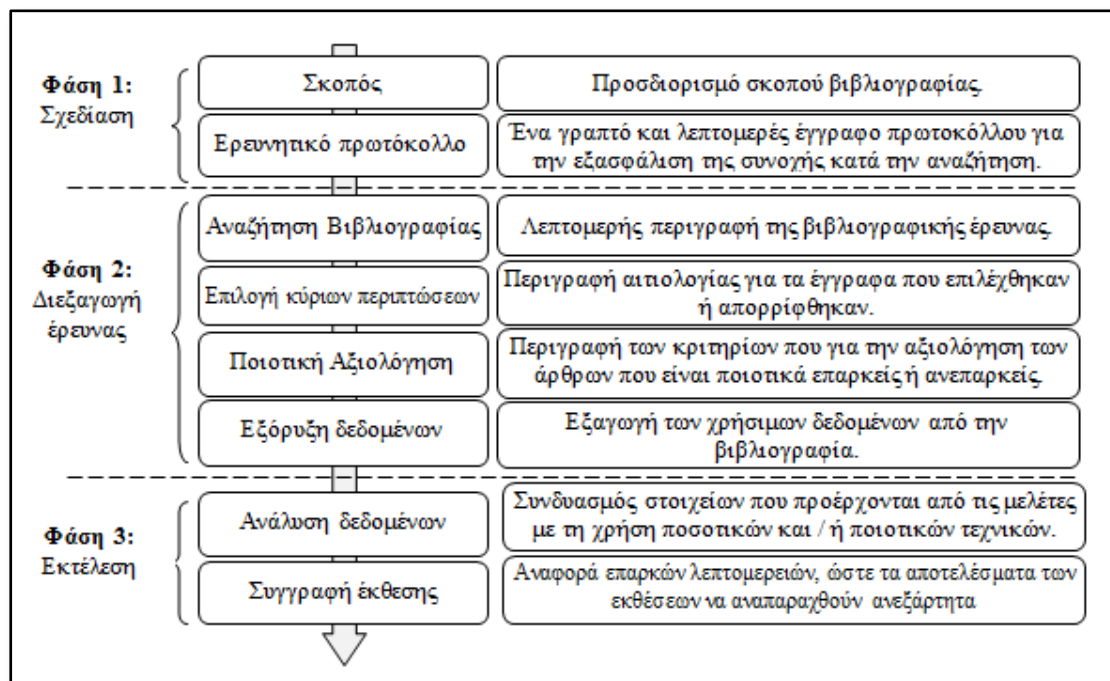
- **Αναζήτηση βιβλιογραφίας:** Για την εύρεση της σχετικής βιβλιογραφίας, μελετήθηκαν εργασίες από διάφορες πηγές, όπως ερευνητικά περιοδικά, ερευνητικά συνέδρια, βιβλία και ιστοσελίδες. Οι λέξεις ή φράσεις κλειδιά που χρησιμοποιήθηκαν στις αναζητήσεις, σχετίζονταν με την μελέτη συμπεριφοράς των εφαρμογών, που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, για την δημιουργία μοντέλων πρόβλεψης της απόδοσης τους. Οι φράσεις που χρησιμοποιήθηκαν είναι: α) “Μοντέλα πρόβλεψης απόδοσης υπηρεσιών στο νέφος”, β) “Μελέτη απόδοσης εφαρμογών”, γ) “Prediction models for adaptive resource provisioning in the cloud”, δ) “Just in time scalability for cloud applications” ε) “Predict behavior of cloud services”, στ) “Prediction Framework for Cloud Computing Services”. Οι βάσεις δεδομένων που χρησιμοποιήθηκαν για την εύρεση της βιβλιογραφίας, ήταν ιστοσελίδες όπως IEEEExplore και Google Scholar.
- **Επιλογή κύριων περιπτώσεων:** Η επιλογή εδώ, έγινε με βάση αυτά που αναφέρονται και στην πρώτη φάση της μεθοδολογίας. Δηλαδή, μελέτες σχετικές την ποιότητα των εφαρμογών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Εξαιρέθηκαν εργασίες που έγιναν με σκοπό μόνο την προσαρμογή των πόρων του υπολογιστικού νέφους στις υπηρεσίες, χωρίς την μελέτη της συμπεριφοράς τους, επειδή στην παρούσα εργασία δεν θα μελετηθεί η συμπεριφορά και οι προσαρμογή των πόρων του υπολογιστικού νέφους, στις ανάγκες των εφαρμογών. Επίσης εξαιρέθηκαν εργασίες που δεν ήταν γραμμές στα ελληνικά ή στα αγγλικά, λόγω έλλειψης γνώσης κατανόησης άλλων γλωσσών.
- **Ποιοτική αξιολόγηση:** Όπως αναφέρθηκε, οι εργασίες που μελετήθηκαν προέρχονται από ερευνητικά περιοδικά, ερευνητικά συνέδρια και ακαδημαϊκά βιβλία. Οι πηγές αυτές θεωρούνται αρκετά αξιόπιστες, καθώς είτε έχουν αξιολογηθεί από ειδικούς του χώρου (peer-review), είτε έχουν συγγραφεί από καταξιωμένους ειδικούς που σχετίζονται με το συγκεκριμένο. Οι ιστοσελίδες που μελετήθηκαν, αποτελούν της επίσημες ιστοσελίδες των μεγαλύτερων παρόχων υπηρεσιών υπολογιστικού νέφους, όπου μελετούν συχνά θέματα που σχετίζονται με την απόδοση των προσφερόμενων υπηρεσιών τους.
- **Εξόρυξη δεδομένων:** Όλα τα παραπάνω που αναφέρθηκαν (άρθρα, ακαδημαϊκά βιβλία), συνθέτουν το κύριο σώμα της βιβλιογραφίας που μελετήθηκε και για το οποίο εξήχθησαν τα δεδομένα.

Φάση 3: Εκτέλεση

- **Ανάλυση δεδομένων:** Για τη σχετική βιβλιογραφία, έγινε ποιοτική ανάλυση. Μελετήθηκαν οι τεχνικές και οι μεθοδολογίες που χρησιμοποιήθηκαν για την επιλογή των καταλληλότερων για την επίλυση και του θέματος της συγκεκριμένης εργασίας.
- **Συγγραφή έκθεσης:** Στην Ενότητα 2.4 παρουσιάζεται η κριτική ανάλυση της βιβλιογραφίας που έγινε για την παρούσα εργασία. Όπως αναφέρθηκε, η ανάλυση αυτή είναι ποιοτική και έχει ως σκοπό την επιλογή των καταλληλότερων τεχνικών

και μεθοδολογιών για την επίλυση του προβλήματος που παρουσιάζεται στο Κεφάλαιο 3.

Στο παρακάτω σχήμα παρουσιάζονται οι τρεις φάσεις της μεθοδολογίας αναζήτησης της σχετικής βιβλιογραφίας που παρουσιάστηκε στην Ενότητα 2.3.



Σχήμα 2.7: Φάσεις βιβλιογραφικής ανασκόπησης (μετάφραση από σελ. 54 του [36])

2.4: Κριτική Ανάλυση Βιβλιογραφίας

Τα τελευταία 10 χρόνια, έχουν γίνει έρευνες σχετικά με την πρόβλεψη για πόρους, από τις εφαρμογές που είναι εγκατεστημένες σε εικονικούς υπολογιστές (Virtual Machines), ενός υπολογιστικού νέφους. Οι μελέτες συνήθως γίνονται πάνω σε συγκεκριμένες υποδομές, οι οποίες διαφέρουν ανά έρευνα, ορίζοντας παραμέτρους απόδοσης και παραμέτρους που την επηρεάζουν. Μερικές μελέτες όμως, όπως και της παρούσας εργασίας, γίνονται πάνω σε άγνωστους φυσικούς πόρους του υπολογιστικού νέφους.

Μελέτες σχετικά με την πρόβλεψη χρήσης πόρων από τις εφαρμογές που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, συχνά γίνονται με την δημιουργία και εγκατάσταση αντιπροσωπευτικής υποδομής. Ένα παράδειγμα παρουσιάζεται στο [11] όπου δημιουργήθηκε ένα περιβάλλον υπολογιστικού νέφους (InterCloud), που διευκολύνει την έγκαιρη και κλιμακωτή ανάγκη για πόρους από τις εφαρμογές που είναι εγκατεστημένες, βάσει των κριτηρίων της ποιότητας της υπηρεσίας (QoS) για τις συνθήκες του φόρτου εργασίας, των πόρων και των δικτυακών επικοινωνιών. Το περιβάλλον αυτό υποστηρίζει δυναμική αυξομείωση των πόρων (εικονικοί υπολογιστές, υπηρεσίες, αποθηκευτικοί χώροι και βάσεις δεδομένων), βάσει των ξαφνικών απαιτήσεων των εφαρμογών κατά την διάρκεια εκτέλεσης τους. Μια χρήσιμη μελέτη, καθώς εκτός από τους υπολογιστικούς πόρους που συνέθεσαν το περιβάλλον που αναπτύχθηκε, δημιουργήθηκαν και μηχανισμοί

για την άμεση αυξομείωση των πόρων σε ομόσπονδα υπολογιστικά συστήματα. Επίσης μελετήθηκε και το πρόβλημα με την φυσική γεωγραφική θέση των υπολογιστικών και αποθηκευτικών πόρων που συνθέτουν το υπολογιστικό νέφος, σε σχέση με την γεωγραφική θέση των αιτημάτων των χρηστών προς τις υπηρεσίες του νέφους αυτού. Έτσι, τα ομόσπονδα υπολογιστικά συστήματα έπρεπε να τοποθετηθούν σε διαφορετικές γεωγραφικές περιοχές και να διαχειρίζονται σωστά τα αιτήματα, για την επίτευξη καλύτερων αποδόσεων των υπηρεσιών. Η μελέτη της γεωγραφικής απόστασης των αιτημάτων για χρήση, από τις υποδομές των υπηρεσιών, είναι αρκετά χρήσιμη, καθώς επηρεάζει την απόδοση της υπηρεσίας. Η προσέγγιση, λοιπόν, της εργασίας [11] για την μελέτη των αιτημάτων που προέρχονται από διαφορετικές γεωγραφικές θέσεις, είναι ενδιαφέρουσα και χρήσιμη για την μελέτη παρόμοιων θεμάτων.

Δυστυχώς, λόγω της συνεχής και αλματώδης ανάπτυξης και σχεδιασμού νέων αρχιτεκτονικών στην κατασκευή του υλικού (hardware) των υπολογιστών, έχει ως αποτέλεσμα κάποια συμπεράσματα που έχουν προκύψει από διάφορες μελέτες, να μην έχουν ισχύ. Μερικά χαρακτηριστικά παραδείγματα είναι ότι οι νέες κεντρικές μονάδες επεξεργασίας των υπολογιστών, έχουν παραπάνω από έναν πυρήνα (έως και 8, αυτή την στιγμή – 7/2014), προσφέροντας έτσι μεγαλύτερη υπολογιστική ισχύ. Επίσης και στο κύριο αποθηκευτικό μέσο που είναι ο σκληρός δίσκος, χρησιμοποιούνται τεχνολογίες χωρίς μηχανικά (κινούμενα) μέρη και μαγνητικούς δίσκους με κεφαλές ανάγνωσης, που ίσχυε μέχρι πρόσφατα, αλλά κυκλώματα συνδεδεμένων λογικών πυλών (chips) για την αποθήκευση των δεδομένων (SSD technology[41]), προσφέροντας μεγαλύτερες ταχύτητες διαβάσματος και εγγραφής των δεδομένων, αυξάνοντας τις προσφερόμενες δυνατότητες του υπολογιστικού νέφους.

Μια ενδιαφέρουσα μελέτη έχει γίνει στο [42] σχετικά με την πρόβλεψη της επίδρασης συγκεκριμένων σημαντικών παραμέτρων, στην επίδοση εικονικών υπολογιστών (Virtual Machines) ενός υπολογιστικού νέφους. Οι παράμετροι αυτοί είναι τα ποσοστά κατανομής των φυσικών πόρων, προσαρμογή πόρων σε πραγματικό χρόνο και η δημιουργία εικονικών υπολογιστών την ίδια χρονική στιγμή σε κοινούς φυσικούς πόρους. Οι προμηθευτές υποδομής (IaaS Providers) μοιράζουν τους πόρους του υπολογιστικού νέφους, στους προμηθευτές πλατφόρμας (PaaS Providers), όπου πάνω εκεί εγκαθίστανται οι εφαρμογές των πελατών. Αν όμως δεν έχει γίνει πρόβλεψη για τον σωστό διαμοιρασμό των πόρων αυτών, υπάρχει η πιθανότητα κάποια στιγμή να μην υπάρχουν διαθέσιμοι πόροι για τις εφαρμογές. Έτσι θα πρέπει να γίνεται σωστός συνδυασμός των εφαρμογών που βρίσκονται πάνω στους ίδιους φυσικούς πόρους. Για την μελέτη αυτή, έγινε καταμέτρηση της απόδοσης των εικονικών υπολογιστών, με την χρήση προγραμμάτων συγκριτικής αξιολόγησης (benchmark), όταν σε αυτούς εκτελούνται κάποιες εφαρμογές. Οι μετρήσεις έγιναν, σε εικονικούς υπολογιστές που ήταν εγκατεστημένοι σε υποδομές υπολογιστικού νέφους, έχοντας διαφορετικά ποσοστά στον χρόνο χρήσης της κεντρικής μονάδας επεξεργασίας (CPU) με έναν πυρήνα, σε κεντρική μονάδα επεξεργασίας με παραπάνω από έναν πυρήνα και σε περισσότερες από μια κεντρικές μονάδες επεξεργασίας. Για την δημιουργία συνάρτησης πρόβλεψης, της απόδοσης των εικονικών υπολογιστών, έγινε χρήση νευρωνικών δικτύων υπολογιστών, καθώς και σύγκριση με πολύ-παραμετρική μαθηματική παλινδρόμηση. Η συνάρτηση αυτή είχε ως αποτέλεσμα, μια πολύ καλή προσέγγιση στην πραγματική συμπεριφορά στην απόδοση των εικονικών μηχανών, με

μικρά ποσοστά λαθών (μέση τιμή σφάλματος 8,78%). Μια αρκετά χρήσιμη εργασία για ανοιχτά ερευνητικά θέματα, που σχετίζονται με την μελέτη απόδοσης υπηρεσιών σενεφροϋπολογιστικής, καθώς κύριος σκοπός της εργασίας είναι η δημιουργία ενός μοντέλου πρόβλεψης της απόδοσης των υπηρεσιών. Έτσι, με την κατάλληλη χρήση των τεχνητών νευρωνικών δικτύων υπολογιστών, πάνω στα δεδομένα που έχουν συλλεχθεί, δημιουργούνται μοντέλα πρόβλεψης της απόδοσης των υπηρεσιών. Ένα θέμα που δεν αναλύεται αρκετά είναι τα κριτήρια επιλογής του καταλληλότερου μοντέλου πρόβλεψης. Βέβαια, παρουσιάζονται τα μοντέλα πρόβλεψης με τις ποσοστιαίες μέσες τιμές απόκλισης, αλλά θα μπορούσε να γίνει μεγαλύτερη ανάλυση σε περισσότερες παραμέτρους για την επιλογή του καταλληλότερου μοντέλου. Επίσης για την αξιολόγηση των αποδόσεων των υπηρεσιών, χρησιμοποιείται η τιμή απόδοσης της εφαρμογής βάσει συγκριτικής αξιολόγησης (benchmark). Σε παρόμοιες μελέτες, πάνω σε υπηρεσίες ιστοσελίδων, θα μπορούσε να χρησιμοποιηθεί άλλη μονάδα αξιολόγησης, όπως ο χρόνος απόκρισης της υπηρεσίας στα αιτήματα χρήσης της. Ο λόγος που δεν έγινε στην εργασία [42], είναι επειδή ασχολείται με υπηρεσίες που χρησιμοποιούν περισσότερο υπολογιστική ισχύ, έχοντας περισσότερες ανάγκες για πόρους κατά την εκτέλεση τους, και όχι τόσο αυξημένη ζήτηση για παράλληλη χρήση.

Μια ακόμα εργασία πάνω στην δημιουργία μοντέλων πρόβλεψης επίδοσης εφαρμογών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, μελετήθηκε στο [43]. Στην εργασία αυτή παρουσιάζονται μελέτες του κύκλου ζωής υπηρεσιοστρεφών εφαρμογών, με την βοήθεια του εργαλείου GNU Octave. Η IRMOS πλατφόρμα που χρησιμοποιείται στην μελέτη, χρησιμοποιεί υποδομές υπολογιστικού νέφους για την εκτέλεση των εγκατεστημένων εφαρμογών. Με την χρήση XML αρχείων γίνεται η περιγραφή των απαιτήσεων των εφαρμογών αυτών στην πλατφόρμα. Σκοπός, λοιπόν, είναι η καταγραφή και μελέτη του κύκλου ζωής των εφαρμογών για την παρουσίαση των ελαχίστων απαιτήσεων τους. Με την βοήθεια του εργαλείου GNU Octave έγινε η μελέτη των μοντέλων πρόβλεψης επίδοσης των εφαρμογών. Στο τέλος παρουσιάζονται οι μέθοδοι πρόβλεψης με την πιθανότητα για λάθη και δοκιμή των μοντέλων αυτών πάνω σε μια εφαρμογή που κωδικοποιεί ροές δεδομένων σε MPEG4 κωδικοποίηση. Ιδιαίτερα χρήσιμη είναι η ανάλυση των παραμέτρων του κύκλου ζωής των υπηρεσιοστρεφών εφαρμογών που μελετώνται. Ο τρόπος ανάλυσης των παραμέτρων αυτών, με εργαλείο GNU Octave, μπορούν να χρησιμοποιηθεί και σε παρόμοιες αναλύσεις κύκλου ζωής άλλων εφαρμογών, ακόμα και με διαφορετικά εργαλεία (όπως το εργαλείο της γλώσσας R). Τέλος, και σε αυτή την εργασία τα μοντέλα πρόβλεψης προκύπτουν ύστερα από επεξεργασία των δεδομένων με την χρήση τεχνητών νευρωνικών δικτύων, που σαν πρακτική είναι χρήσιμη σε τέτοια ερευνητικά θέματα.

Στην επόμενη εργασία [44], που μελετήθηκε, παρουσιάζεται μια πλατφόρμα πρόβλεψης επίδοσης μιας υπηρεσιοστρεφούς εφαρμογής κατά την διάρκεια εκτέλεσης της. Στην πλατφόρμα αυτή, με την χρήση του εργαλείου GNU Octave, εφαρμόζονται μοντέλα πρόβλεψης. Επίσης παρουσιάζεται μια μεθοδολογία για την τροποποίηση, απλών-τυπικών υπηρεσιών, σε παραμετρικές υπηρεσίες με πολλαπλές καταστάσεις λειτουργίας. Έτσι, βάσει των συνθηκών εκτέλεσης που εφαρμόζονται, τροποποιείται η προαναφερθείσα πλατφόρμα πρόβλεψης, για να επιτευχθούν καλύτερες αποδόσεις στις υπηρεσίες αυτές. Αυτό πραγματοποιείται από μια διαδικασία 5 βημάτων, έτσι ώστε να γίνει μια ανάλυση

στον σχεδιασμό και στην εκτέλεση της υπηρεσίας, για να κατηγοριοποιηθεί σε σχέση με άλλες. Τέλος, ένας αλγόριθμος εκτίμησης δημιουργεί κανόνες απόδοσης για τις εφαρμογές, με την χρήση τεχνητών νευρωνικών δικτύων. Ο αλγόριθμος αυτός αντιμετωπίζει τις εφαρμογές ως «μαύρα κουτιά», χωρίς δηλαδή να γνωρίζει τους κανόνες που ορίζουν τα επίπεδα απόδοσης της κάθε εφαρμογής. Η προαναφερθείσα πλατφόρμα πρόβλεψης επιδόσεων υπηρεσιών, είναι εγκατεστημένη πάνω στην πλατφόρμα IRMOS, όπου χρησιμοποιεί δομές υπολογιστικού νέφους. Η εργασία αυτή αποτελεί μια ακόμα χρήσιμη μελέτη πάνω στην ανάλυση των εφαρμογών, και τον προσδιορισμό των επιπέδων απόδοσης για παρόμοιες εφαρμογές. Και σε αυτή την εργασία, γίνεται χρήση των τεχνητών νευρωνικών δικτύων για την δημιουργία των μοντέλων πρόβλεψης των αποδόσεων των εφαρμογών.

Στην εργασία [45] παρουσιάζεται μια δύο επιπέδων, γενική προσέγγιση, στην διαχείριση των επιπέδων του υπολογιστικού νέφους (SaaS, PaaS, IaaS), βασισμένη στην συμπεριφορά των χρηστών και των υπηρεσιών. Η προσέγγιση αυτή παρέχει εκτιμήσεις για τα χαρακτηριστικά των πόρων σε χαμηλό επίπεδο, με την ανάλυση των πληροφοριών στο υψηλό επίπεδο που σχετίζεται με τις απαιτήσεις της υπηρεσίας (π.χ. αριθμός χρηστών, συχνότητα αιτημάτων, γεωγραφική θέση που προέρχονται τα αιτήματα). Έτσι γίνεται πρόβλεψη της αναμενόμενης συμπεριφοράς του χρήστη της υπηρεσίας και προσαρμογή των πόρων του υπολογιστικού νέφους που βρίσκονται στο χαμηλό επίπεδο. Επίσης προσδιορίζει τα πρότυπα σε υψηλό επίπεδο, μέσω της ανάλυσης χρονοσειρών (time series analysis), και στην συνέχεια μετατρέπονται σε χαρακτηριστικά του χαμηλού επιπέδου με την χρήση τεχνητών νευρωνικών δικτύων. Παρουσιάζονται λοιπόν, δυο προσεγγίσεις για τα επίπεδα (υψηλό και χαμηλό) του υπολογιστικού νέφους. Η μια αφορά την δημιουργία μηχανισμού ανάλυσης χρονοσειρών, για τον προσδιορισμό και εκτίμηση των παραμέτρων του υψηλού επιπέδου, μαζί με τις απαιτήσεις εκτέλεσης της υπηρεσίας. Στην δεύτερη προσέγγιση, δημιουργήθηκε ένας μηχανισμός για την μετάφραση και μετατροπή των παραμέτρων του υψηλού επιπέδου, σε απαιτήσεις πόρων του χαμηλού επιπέδου. Και οι δύο μηχανισμοί βασίστηκαν στα τεχνητά νευρωνικά δίκτυα υπολογιστών, τα οποία χρησιμοποιήθηκαν ξεχωριστά για κάθε υπηρεσία και κάθε υπολογιστική υποδομή για την δημιουργία μηχανισμών πρόβλεψης συμπεριφοράς. Η μελέτη του υψηλού επιπέδου της υποδομής του υπολογιστικού νέφους, ως προς τις απαιτήσεις των υπηρεσιών σε πόρους, είναι αρκετά χρήσιμη για αντίστοιχες μελέτες σε υπηρεσίες νεφοϋπολογιστικής, που δεν είναι γνωστές οι υποδομές του χαμηλότερου επιπέδου. Έτσι, με παρόμοια βήματα ανάλυσης, μπορούν να προσδιοριστούν τα επίπεδα απόδοσης των υπηρεσιών στην διάρκεια εκτέλεσης τους.

Η εργασία [46] επικεντρώνεται σε μια αυτόματη και δυναμική κατανομή πόρων ενός υπολογιστικού νέφους, που συνδέεται με την εκτέλεση εφαρμογών που είναι εγκατεστημένες σε αυτό. Προσεγγίζει το πρόβλημα αυτό της κατανομής, ως μια χρονικά περιορισμένη χρήση πόρων για κάθε εργασία. Μέσα σε αυτούς τους περιορισμούς, η υποδομή τους υπολογιστικού νέφους, πρέπει να εξυπηρετήσει τις απαιτήσεις της εφαρμογής. Αυτό γίνεται με την χρήση μηχανισμού για την καταγραφή των πιο σημαντικών λειτουργιών τους, που ζητάνε πόρους, για να εκτελεστούν στο περιορισμένο χρονικό διάστημα. Έτσι παρουσιάζεται ο σχεδιασμός, η υλοποίηση και η αξιολόγηση του μηχανισμού για την υποστήριξη των εφαρμογών σε περιβάλλοντα υπολογιστικού νέφους.

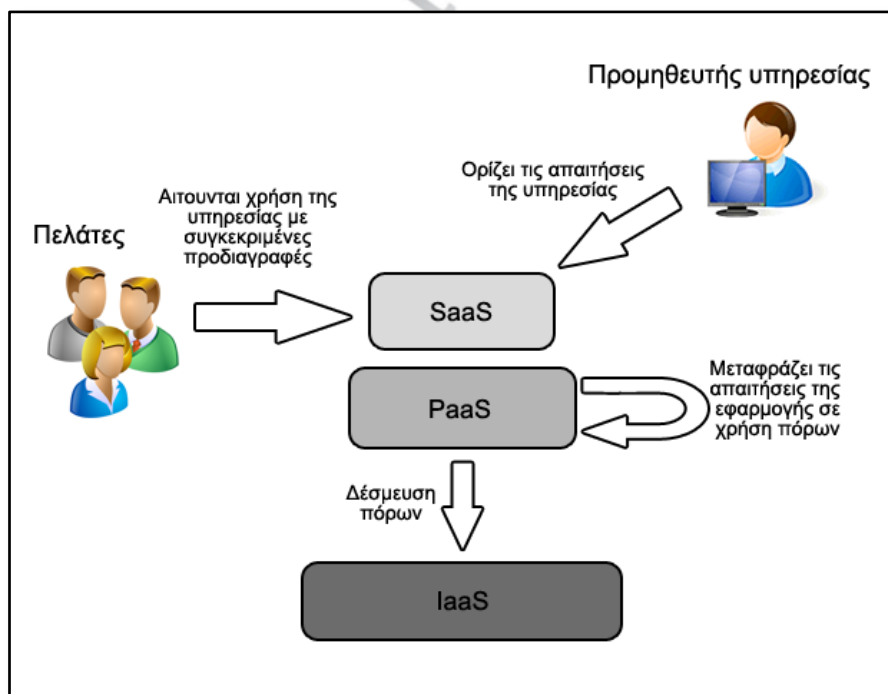
Το βασικό στοιχείο αυτής της έρευνας, είναι η δημιουργία ενός αλγορίθμου για την πρόβλεψη της ανάγκης για πόρους. Οι παράμετροι που χρησιμοποιούνται στο μοντέλο αυτό, που επηρεάζουν την απόδοση των εφαρμογών, είναι η κατανομή της προσωρινής μνήμης και η κατανομή της επεξεργαστικής ισχύς. Η ανάλυση των εφαρμογών που έγινε σε αυτή την εργασία, είναι χρήσιμη στην εξαγωγή συμπερασμάτων σχετικά με τις παραμέτρους που επηρεάζουν την απόδοση τους. Επίσης παρουσιάζονται τρόποι για τον προσδιορισμό των απαιτήσεων για πόρους, από τον υπολογιστικό νέφος, κατά την διάρκεια εκτέλεσης των εφαρμογών. Για παράδειγμα σε εφαρμογές ιστού, οι απαιτήσεις για πόρους αυξάνονται όταν επεξεργάζονται τα αιτήματα για χρήση, και όχι τόσο στην προσφορά των δεδομένων που επιστρέφονται. Επιπλέον, ένα χαρακτηριστικό παράδειγμα παραμέτρου που επηρεάζει τις αποδόσεις σε τέτοιες εφαρμογές είναι ο αριθμός των παράλληλων αιτημάτων για χρήση.

Μια άλλη προσέγγιση ακολουθεί η εργασία [47]. Σε αυτή χρησιμοποιούνται τα προφίλ, που αποτυπώνουν την γνώση για κλιμάκωση σε διάφορους τύπους εφαρμογών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Η προσέγγιση αυτή, αυτοματοποιεί την ανάπτυξη και την κλιμάκωση των εφαρμογών στις υποδομές αυτές. Η άμεση επεκτασιμότητα επιτυγχάνεται χωρίς να αντιστοιχίζεται σε συγκεκριμένες υποδομές υπολογιστικού νέφους. Έτσι, η αντιστοιχία των εφαρμογών με συγκεκριμένα προφίλ, επιτρέπει την πρόβλεψη ανάγκης πόρων κατά την διάρκεια εκτέλεσης τους. Τα προφίλ αυτά, χρησιμοποιούνται επίσης από τις εφαρμογές και για να ορίσουν τις απαιτήσεις τους στο υπολογιστικό νέφος, χωρίς να χρειάζεται διαφοροποίηση στο τρόπο παρουσίασης των απαιτήσεων αυτών, ανάλογα τις υποδομές. Οι πληροφορίες των προφίλ που χρησιμοποιούνται σε αυτή την εργασία, μπορούν να χρησιμοποιηθούν στην μελέτη εφαρμογών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Για παράδειγμα, σε μελέτες απόδοσης των εφαρμογών ιστού, μπορεί να αναλυθεί το επιλεγμένο προφίλ που παρουσιάστηκε στην εργασία για παρόμοιες εφαρμογές. Από το προφίλ αυτό μπορούν να εξαχθούν συμπεράσματα, σχετικά με την χρήση ποιών χαρακτηριστικών των πόρων επηρεάζουν την απόδοση των εφαρμογών. Σε εφαρμογές ιστού, κύριο χαρακτηριστικό που επηρεάζει την απόδοση είναι η προσωρινή μνήμη της κεντρικής μονάδας επεξεργασίας.

Στην εργασία [48] αναπτύχθηκαν στρατηγικές πρόβλεψης με βάση τις απαιτήσεις για πόρους, υπηρεσιών που είναι εγκατεστημένες σε υπολογιστικό νέφος. Η ανάλυση και η δημιουργία μοντέλου πρόβλεψης γίνεται με την χρήση τεχνητών νευρωνικών δικτύων υπολογιστών και γραμμικής παλινδρόμησης. Για την καταμέτρηση της απόδοσης των υπηρεσιών χρησιμοποιήθηκε πρόγραμμα συγκριτικής αξιολόγησης (benchmark). Για την συνάρτηση πρόβλεψης χρησιμοποιήθηκαν αλγόριθμοι μηχανικής εκμάθησης. Η υπηρεσία που μελετήθηκε η συμπεριφορά της ήταν ένα ηλεκτρονικό κατάστημα. Η χρήση αλγορίθμων μηχανικής εκμάθησης, σε συνδυασμό με τα τεχνητά νευρωνικά δίκτυα υπολογιστών, για την δημιουργία των μοντέλων πρόβλεψης, είναι ιδιαίτερα χρήσιμη. Δημιουργούνται μοντέλα που έχουν μικρό ποσοστά αποκλίσεων από την πραγματικότητα. Μπορούν λοιπόν να χρησιμοποιηθεί η τακτική αυτή και σε άλλα παρόμοια ερευνητικά θέματα, που σχετίζονται με την απόδοση των υπηρεσιών.

2.5: Ανοιχτές Θεματικές Περιοχές για Έρευνα

Σε ένα σύνθετο πολυστρωματικό οικοσύστημα, όπως είναι το Υπολογιστικό Νέφος, είναι ιδιαίτερα κρίσιμη η διαχείριση του επιπέδου της πλατφόρμας (PaaS). Στο επίπεδο αυτό πρέπει να γίνεται σωστά η μετάφραση των καθορισμένων από τον καταναλωτή, παραμέτρων εφαρμογής (φόρτος εργασίας και ποιότητα υπηρεσίας), στις ιδιότητες του επιπέδου υλικού (IaaS), που προσφέρει τους φυσικούς πόρους. Έτσι ο πάροχος του επιπέδου πλατφόρμας, θα πρέπει να ανταλλάξει πληροφορίες με τον πάροχο επιπέδου υλικού. Αυτό όμως είναι αρκετά δύσκολο για τεχνικούς αλλά και επιχειρηματικούς λόγους. Επίσης όμως είναι εξαιρετικά δύσκολο, ειδικά για τους παρόχους εφαρμογών (SaaS) και υποδομής (IaaS), να εκθέσουν τις εσωτερικές λειτουργικές παραμέτρους και τις διαδικασίες των επιπέδων τους σε άλλα επίπεδα, λόγω τους κινδύνου να έχουν πρόσβαση σε αυτές τις πληροφορίες, ανταγωνιστές τους. Από την άλλη, η ανταλλαγή αυτών των πληροφοριών δεν είναι πάντα τεχνικά εφικτή. Κάθε επίπεδο έχει ένα σύνολο παραμέτρων που, που μπορεί να διαφέρει από τα υπόλοιπα, και με αυτές τις παραμέτρους έχει αναπτύξει και αντίστοιχους μηχανισμούς για την παροχή υπηρεσιών. Επιπλέον, συχνά γίνεται χρήση εξωτερικών βιβλιοθηκών ή κώδικα τρίτων (third party software), για τα οποία δεν υπάρχει γνώση της δομής τους, δυσχεραίνοντας ακόμα περισσότερο της ροή της πληροφορίας προς τα άλλα επίπεδα. Τέλος, η έλλειψη προτύπων για τα Υπολογιστικά Νέφη είναι ένα άλλο μεγάλο ζήτημα, που περιορίζει τον αποτελεσματικό διαμοιρασμό των πληροφοριών, μεταξύ των επιπέδων του. Μερικές προσπάθειες πάνω στο ζήτημα αυτό έχουν γίνει στα OCCI[49] και ONV[50].



Σχήμα 2.8: Αλληλεπίδραση επιπέδων του Υπολογιστικού Νέφους

Ένα σημαντικό κενό στα Υπολογιστικά Νέφη, όπως προσφέρονται από τους παρόχους, είναι η έλλειψη εγγυήσεων απόδοσης πραγματικού χρόνου και ποιότητας της υπηρεσίας. Στις

υπηρεσίες αυτές, ο πάροχος πρέπει να λαμβάνει υπόψη του το βασικό χαρακτηριστικό αυτών, που είναι η πραγματοποίηση μιας υπολογιστικής πράξης από την υπηρεσία μέσα σε κάποιο χρονικό διάστημα. Επιπλέον, δεν μπορούν οι πάροχοι να εγγυηθούν απόδοση της υπηρεσίας, με βάση συγκεκριμένες παραμέτρους που θα ορίσει ο πελάτης. Όπως το παράδειγμα μιας ιστοσελίδας, όπου κύριο χαρακτηριστικό απόδοσης είναι ο χρόνος απόκρισης. Συνήθως οι εγγυήσεις που προσφέρουν και αναφέρονται στις συμβάσεις παροχής υπηρεσιών (SLAs), σχετίζονται με την διαθεσιμότητα των εικονικών υπολογιστών (π.χ. το 99.95% του χρόνου).

Ένα ακόμα σημαντικό θέμα που υπεισέρχεται σε διαμοιραζόμενες υποδομές (όπως αυτή τους Υπολογιστικού Νέφους), είναι η αλληλεπίδραση μεταξύ των εικονικών πόρων που εκτελούνται σε ένα φυσικό πόρο και ανήκουν σε διαφορετικούς χρήστες. Αυτή η αλληλεπίδραση μπορεί να ποικίλει ανάλογα με τον την αρχιτεκτονική του φυσικού πόρου και το είδος των εφαρμογών που εκτελούνται μέσα στους εικονικούς υπολογιστές.

Τέλος, ένα σημαντικό θέμα που σχετίζεται άμεσα με την απόδοση των εφαρμογών που είναι εγκατεστημένες σε Υπολογιστικά Νέφη, είναι η πρόβλεψη ανάγκης για πόρους από αυτά, για την ικανοποίηση των απαιτήσεων των αποδόσεών τους. Κατά τον χρόνο εκτέλεσης μιας υπηρεσίας, υπάρχουν αυξομειώσεις στις απαιτήσεις για πόρους από το υπολογιστικό νέφος. Θα πρέπει να υπάρχει, λοιπόν, κάποιος μηχανισμός που θα προβλέπει τις ανάγκες για πόρους, πριν αυτές ζητηθούν, έτσι ώστε να έχουν προετοιμαστεί για να μπορούν να προσφερθούν όταν θα χρειάζονται. Αυτό μπορεί να γίνει με την πρόβλεψη της απόδοσης της υπηρεσίας, με βάση τις παραμέτρους που την επηρεάζουν. Έτσι, το μοντέλο πρόβλεψης θα προβλέπει την απόδοση της υπηρεσίας για τις τιμές των παραμέτρων που την επηρεάζουν.

2.6: Συμπεράσματα

Η ευρεία χρήση σύγχρονων αλλά και πρόσφατων τεχνολογιών, όπως αυτή του υπολογιστικού νέφους, επιφέρει μια σειρά θεμάτων που πρέπει να μελετηθούν και να επιλυθούν. Επιπλέον, όπως αναφέρθηκε, η δομή του υπολογιστικού νέφους, δημιουργεί πρόσθετα προβλήματα. Όλα αυτά τα θέματα έχουν ως επίκεντρο την εγγυημένη υποστήριξη των εφαρμογών που είναι εγκατεστημένες σε αυτό. Η υποστήριξη αυτή, των εφαρμογών που προσφέρονται σαν υπηρεσίες, σχετίζεται με διάφορα θέματα, όπως η διαθεσιμότητα της στο χρόνο, η διαθεσιμότητα της στον αριθμό των αιτημάτων για χρήση που δέχεται, καθώς και στην διατήρηση των επιθυμητών αποδόσεών της. Όλα αυτά τα θέματα θα πρέπει να ορίζονται στα συμβόλαια παροχής υπηρεσιών.

Ανοιχτές θεματικές περιοχές, σχετίζονται με την δημιουργία μηχανισμών για την μετάφραση των απαιτήσεων των εφαρμογών, σε ιδιότητες του επιπέδου του υλικού. Επίσης, υπάρχουν αρκετές θεματικές περιοχές που σχετίζονται με την εγγύηση των αποδόσεων των εφαρμογών, είτε ορίζοντας συγκεκριμένες παραμέτρους για κάθε εφαρμογή, είτε ορίζοντας συγκεκριμένους φυσικούς πόρους που θα της προσφέρονται και θα έχουν τις αντίστοιχες δυνατότητες, είτε την αποφυγή αλληλεπίδρασης με άλλες εφαρμογές που είναι εγκατεστημένες στους ίδιους φυσικούς πόρους.

Το υπολογιστικό νέφος παρέχει μηχανισμούς για να υποστηρίξει την διατήρηση των επιθυμητών αποδόσεων της κάθε υπηρεσίας. Το πρόβλημα όμως που δημιουργείται είναι η έγκαιρη ενεργοποίηση αυτών, καθώς ο χρόνος εκτέλεσης τους δεν είναι αμελητέος. Οι προκλήσεις πρόβλεψης της απόδοσης μιας εφαρμογής και η πρόβλεψη ανάγκης για πόρους είναι αρκετές. Όπως αναφέρθηκε στις προηγούμενες ενότητες, με το πρόβλημα πρόβλεψης της συμπεριφοράς των εφαρμογών, έχουν ασχοληθεί αρκετές μελέτες. Σε αυτές τις μελέτες σκοπός ήταν η δημιουργία μοντέλων πρόβλεψης της απόδοσης των εφαρμογών. Για την δημιουργία των μοντέλων αυτών, άλλοτε γινόταν καταγραφή των συμπεριφορών τους κατά της διάρκεια εκτέλεσης τους, άλλοτε τμηματοποίηση της εκτέλεσης τους και μοντελοποίηση της κάθε φάσης, και άλλοτε αντιστοιχία των εφαρμογών με έναν αριθμό από προφίλ που αποτύπωναν την γνώση για κλιμάκωση.

Στο Κεφάλαιο 3, θα παρουσιαστεί η επιλογή της ανοιχτής ερευνητικής θεματικής ενότητας, που σχετίζεται με την δημιουργία μοντέλων πρόβλεψης απόδοσης υπηρεσιών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Με βάση τη μελέτη τις σχετικής βιβλιογραφίας που έγινε, χρησιμοποιούνται οι τεχνικές και οι μεθοδολογίες για την ανάλυση της συμπεριφοράς της υπηρεσίας που διερευνάται, που έχει ως σκοπό την δημιουργία μοντέλων πρόβλεψης της απόδοσής της. Έτσι, στο Κεφάλαιο 3, περιγράφεται ακόμα, η διαδικασία που ακολουθήθηκε για την μελέτη και ανάλυση της υπηρεσίας.

Κεφάλαιο 3: Δυναμική Πρόβλεψη Απόδοσης Υπηρεσίας σε Νεφοϋπολογιστική

3.1: Εισαγωγή

Όπως αναφέρθηκε στο Κεφάλαιο 2, η νεφοϋπολογιστική έχει γίνει αντικείμενο μελέτης και ανάλυσης, τόσο σε ακαδημαϊκό όσο και σε επιχειρηματικό επίπεδο. Η υιοθέτηση λύσεων νεφοϋπολογιστικής θεωρείται βέλτιστη πρακτική [51], σε επίπεδο παροχής υπηρεσιών πληροφορικής και τηλεπικοινωνιών. Μεγάλο κομμάτι της έρευνας στρέφεται προς τη κατεύθυνση της μεγιστοποίησης της απόδοσης των εφαρμογών που χρησιμοποιούν υποδομές υπολογιστικού νέφους, λαμβάνοντας υπόψη περιορισμούς κόστους. Συγκεκριμένα οι πάροχοι των εφαρμογών, αποσκοπούν στη μεγιστοποίηση της πελατειακής τους βάσης, αυξάνοντας παράλληλα το επίπεδο ποιότητας των παρεχόμενων υπηρεσιών.

Καθώς ο φόρτος χρήσης των εφαρμογών δεν είναι σταθερός, οι ανάγκες τους σε πόρους μεταβάλλονται δυναμικά, και εξαρτώνται από παράγοντες όπως: (i) είδος υπηρεσίας (π.χ. υπηρεσίες υποβολής φορολογικών δηλώσεων που χρησιμοποιούνται συγκεκριμένες χρονικές περιόδους ανά έτος), (ii) ώρες αιχμής δικτύου, (iii) μερίδα πελατών που απευθύνεται η εφαρμογή (π.χ. φοιτητές, εργαζόμενους, ανέργους, κλπ). Προκειμένου τα επίπεδα της ποιότητας της παρεχόμενης υπηρεσίας να βρίσκονται σε υψηλό σημείο, ο πάροχος της εφαρμογής πρέπει να μπορεί να προβλέπει δυναμικά και με ακρίβεια κάθε χρονική στιγμή τις προκύπτουσες ανάγκες για πόρους, ώστε να ανταπεξέλθει στον προκείμενο φόρτο εργασίας. Για να επιτευχθεί αυτό, χρειάζονται μοντέλα τα οποία μπορούν να αναγνωρίσουν πρότυπα (patterns) χρήσης πόρων, λαμβάνοντας υπόψη παραμέτρους που ορίζονται από τον πάροχο της εφαρμογής.

Η ανάγκη για δυναμική πρόβλεψη των αναγκών της εφαρμογής με σκοπό την μεγιστοποίηση της «επένδυσης» σε πόρους, έχει στρέψει μεγάλο κομμάτι της ερευνητικής προσπάθειας προς τη δημιουργία δυναμικών μοντέλων πρόβλεψης, που επιλύουν το πρόβλημα που περιγράφηκε παραπάνω. Στο παρόν κεφάλαιο, γίνεται μια προσπάθεια καταγραφής των προκλήσεων στον τομέα της πρόβλεψης της απόδοσης, ώστε να φανεί η πληθώρα των διαφορετικών παραμέτρων που επηρεάζουν την απόδοση μιας υπηρεσίας. Από τις διάφορες παραμέτρους που θα παρουσιαστούν και θα αναλυθούν, θα γίνει επιλογή και συνδυασμός αυτών που θα μοντελοποιηθούν στην παρούσα ερευνητική εργασία.

3.2: Προκλήσεις Πρόβλεψης Απόδοσης Υπηρεσίας και Ανάγκης για Πόρους στα Υπολογιστικά Νέφη

Το υπολογιστικό νέφος προσφέρει μηχανισμούς για αύξηση ή μείωση των πόρων του, που προσφέρονται για την εκτέλεση μιας υπηρεσίας. Το πρόβλημα όμως είναι, αυτοί οι πόροι να είναι διαθέσιμοι την στιγμή που τους χρειάζεται η κάθε υπηρεσία, που είναι εγκατεστημένη σε αυτό. Οπότε η ανάγκη για παραγωγή μοντέλων πρόβλεψης, έτσι ώστε η

διαδικασία προετοιμασίας των πόρων να ξεκινάει την κατάλληλη στιγμή, για να είναι έτοιμοι όταν χρειάζονται, είναι προφανής.

Όπως αναφέρθηκε στο Κεφάλαιο 2, το επίπεδο της πλατφόρμας (PaaS), έχει την ευθύνη της μετάφρασης/μετατροπής των παραμέτρων που χρησιμοποιούνται στο επίπεδο της εφαρμογής/υπηρεσίας (SaaS). Έτσι, οι παράμετροι του φόρτου εργασίας (Load Balancing) και οι παράμετροι της ποιότητας της υπηρεσίας (Quality of Service), θα πρέπει να μετατραπούν σε παραμέτρους του επιπέδου του υλικού (IaaS), δηλαδή υπολογιστικούς, αποθηκευτικούς ή άλλους πόρους (π.χ. ταχύτητα κεντρικής μονάδας επεξεργασίας, πλήθος επεξεργαστών, μέγεθος προσωρινής μνήμης, μέγεθος μνήμης αποθήκευσης δεδομένων, κλπ). Επιπλέον το γεγονός ότι υπάρχουν πληροφορίες που δεν μπορούν να είναι διαθέσιμες μεταξύ των επιπέδων του υπολογιστικού νέφους, όπως η δομή του κώδικα των εφαρμογών στο επίπεδο της εφαρμογής, και οι προδιαγραφές των φυσικών πόρων που προσφέρονται από το επίπεδο του υλικού, κάνουν ακόμα πιο δύσκολο το έργο της μετάφρασης/μετατροπής του επιπέδου της πλατφόρμας.

Επίσης υπάρχουν και άλλα προβλήματα και περιορισμοί που υπεισέρχονται λόγω της μορφής και της δομής του υπολογιστικού νέφους. Ένα από αυτά είναι ο περιορισμός των συμβολαίων παροχής υπηρεσιών (SLAs) για κάθε ξεχωριστό επίπεδο, προκειμένου να επιτευχθούν με ακρίβεια οι απαιτήσεις του καθενός επιπέδου. Ένα ακόμα πρόβλημα είναι η ανίχνευση του συσχετισμού μεταξύ των εισόδων και των εξόδων του μοντέλου απόδοσης. Το πρόβλημα που δημιουργείται είναι ότι ανάλογα των αριθμό των παραμέτρων που εισάγονται στην εφαρμογή, μπορεί να ενεργοποιούνται διαφορετικά τμήματα του κώδικα, που θα χρειάζονται περισσότερους ή λιγότερους πόρους για την επεξεργασία τους. Μερικά παραδείγματα, ψευδο-κώδικα εφαρμογών, παρουσιάζονται στο Σχήμα 3.1. Στο πρώτο παράδειγμα του σχήματος, για την ίδια εφαρμογή, ανάλογα με τις παραμέτρους που εισάγονται, εκτελούνται διαφορετικά κομμάτια του κώδικα της, που ενεργοποιούν διαφορετικές διαδικασίες, οι οποίες διαφέρουν στις απαιτήσεις τους για πόρους. Στο δεύτερο παράδειγμα του Σχήματος 3.1, παρουσιάζεται ο κώδικας μιας εφαρμογής, όπου ανάλογα το πλήθος των δεδομένων που εισάγονται (π.χ. αρχεία μέσα σε φάκελο προς επεξεργασία), αυξάνεται και ο χρόνος χρήσης των πόρων ή ακόμα και οι απαιτήσεις για δέσμευση πόρων. Τα παραδείγματα αυτά έχουν ως σκοπό την κατανόηση του προβλήματος που υπάρχει, στην παραγωγή βέλτιστων μοντέλων πρόβλεψης για κάθε εφαρμογή. Τέλος, η γεωγραφική τοποθεσία των φυσικών πόρων του επιπέδου του υλικού, αλλά και η γεωγραφική τοποθεσία των αιτήσεων χρήσης της υπηρεσίας στο επίπεδο της εφαρμογής, δημιουργούν επιπλέον προβλήματα στην παραγωγή μοντέλων πρόβλεψης.

```
// Παράδειγμα 1
1. Αν έχει εισαχθεί τιμή στην παράμετρο A μόνο, τότε:
2. {
3.   Εκτέλεση διαδικασίας που χρησιμοποιεί μικρό αριθμό πόρων
4. }
5. Αλλιώς αν έχουν εισαχθεί τιμές για τις παραμέτρους A και B, τότε:
6. {
7.   Εκτέλεση διαδικασίας που χρησιμοποιεί μεγάλο αριθμό πόρων
8. }

// Παράδειγμα 2
1. Για κάθε αρχείο στον φάκελο που δόθηκε:
2. {
3.   Εκτέλεση διαδικασίας
4. }
```

Σχήμα 3.1: Παρουσίαση ψευδο-κώδικα ροής προγράμματος, όπου ο αριθμός των παραμέτρων επηρεάζει την χρήση πόρων

Όπως αναφέρθηκε στην Ενότητα 2.4, έχουν γίνει αρκετές μελέτες σχετικά με την πρόβλεψη ανάγκης για πόρους μιας υπηρεσίας που είναι εγκατεστημένη σε υποδομές υπολογιστικού νέφους. Η μελέτη λοιπόν, για την δημιουργία μοντέλων πρόβλεψης, μπορεί να περιλαμβάνει από τις υποδομές του υπολογιστικού νέφους, μέχρι τις διαφορετικές παραμέτρους απόδοσης της κάθε εφαρμογής. Έτσι μπορούν να δημιουργηθούν γενικά μοντέλα πρόβλεψης συμπεριφοράς των εφαρμογών, κατηγοριοποιώντας τις εφαρμογές αυτές που εκτελούνται πάνω στις ίδιες ή διαφορετικές υποδομές ενός υπολογιστικού νέφους, αλλά και συγκεκριμένα μοντέλα για κάθε υπηρεσία. Στην δεύτερη περίπτωση, μπορεί να έχει γίνει εγκατάσταση μια υποδομής ενός υπολογιστικού νέφους, όπου έχει σκοπό την χρήση του μόνο από μια υπηρεσία. Τέτοια παραδείγματα θα μπορούσαν να ήταν τα κοινωνικά δίκτυα που χρησιμοποιούνται από μεγάλο αριθμό χρηστών, όπως το Facebook ή το Twitter, όπου έχουν δημιουργηθεί για αυτά αποκλειστικές υποδομές υπολογιστικού νέφους. Ένα ακόμα παράδειγμα θα μπορούσε να ήταν μια ηλεκτρονική εγκυκλοπαίδεια, όπως το Wikipedia.org, όπου οι υποδομές του υπολογιστικού νέφους της, βρίσκονται σε διαφορετικές γεωγραφικές θέσεις, για την αποδοτικότερη εκτέλεση της υπηρεσίας.

Τέλος, η μελέτη για την δημιουργία μοντέλων πρόβλεψης, μπορεί να γίνει με χρήση πραγματικών δεδομένων μιας ευρείας χρήσης υπηρεσίας, γνωρίζοντας όλες τις παραμέτρους όλων των επιπέδων, είτε από προσομοιώσεις σε μικρότερη κλίμακα (τόσο σε επίπεδο παραμέτρων εκτέλεσης ή χρήσης της υπηρεσίας, όσο και σε επίπεδο του υλικού της υποδομής του υπολογιστικού νέφους). Επίσης, χρήσιμη είναι η μελέτη της συμπεριφοράς υπηρεσιών, χωρίς να υπάρχει πλήρης γνώση της υποδομής του υλικού, βγάζοντας όμως κάποια χρήσιμα συμπεράσματα (π.χ. πότε υπάρχει ανάγκη για αύξηση των προσφερόμενων πόρων), δημιουργώντας δυναμικά μοντέλα πρόβλεψης της απόδοσης της υπηρεσίας. Τέτοιες μελέτες, περιλαμβάνουν παραμέτρους που δεν σχετίζονται με το υλικό του υπολογιστικού νέφους (π.χ. χρήση προσωρινής μνήμης, ταχύτητα επεξεργασίας δεδομένων), αλλά με την χρήση της υπηρεσίας. Γι' αυτό το λόγο, γίνεται μελέτη και ανάλυση παραμέτρων όπως, είναι τα παράλληλα αιτήματα για χρήση που δέχεται η υπηρεσία, αλλά και την γεωγραφική απόσταση απ' όπου προέρχονται.

3.3: Μοντέλο Δυναμικής Πρόβλεψης Απόδοσης Υπηρεσίας

Η υιοθέτηση λύσεων νεφροϋπολογιστικής, για την παροχή υπηρεσιών, θεωρείται βέλτιστη πρακτική, λόγω των πλεονεκτημάτων που προσφέρει και έχουν αναλυθεί στο Κεφάλαιο 2 (π.χ. διαμοιρασμός φυσικών πόρων ανάλογα τις ανάγκες των εφαρμογών). Για τον λόγο αυτό, έχει γίνει αντικείμενο μελέτης από ερευνητές, τόσο σε ακαδημαϊκό όσο και σε επιχειρηματικό επίπεδο, έχοντας ως στόχο τη μεγιστοποίηση της απόδοσης των εφαρμογών που χρησιμοποιούν υποδομές υπολογιστικού νέφους. Με αυτό τον τρόπο, επιτυγχάνεται αύξηση των επιπέδων ποιότητας των υπηρεσιών που προσφέρουν οι πάροχοι.

Καθώς ο φόρτος χρήσης των παρεχόμενων υπηρεσιών δεν είναι σταθερός, μεταβάλλεται παράλληλα και οι ανάγκες της εφαρμογής σε πόρους, από το υπολογιστικό νέφος. Παράγοντες που επηρεάζουν αυτές τις μεταβολές είναι: (i) είδος υπηρεσίας (π.χ. υπηρεσίες υποβολής φορολογικών δηλώσεων που χρησιμοποιούνται συγκεκριμένες χρονικές περιόδους ανά έτος). (ii) ώρες αιχμής δικτύου, (iii) μερίδα πελατών που απευθύνεται η εφαρμογή (π.χ. φοιτητές, εργαζόμενους, ανέργους, κλπ). Αυτός είναι ένας βασικός λόγος όπου η υιοθέτηση λύσεων νεφροϋπολογιστικής, θεωρείται βέλτιστη πρακτική. Γίνεται έτσι χρήση ενός βασικού χαρακτηριστικού του υπολογιστικού νέφους, που είναι η δυνατότητα που παρέχει για δυναμική προσφορά και κλιμάκωση (υπολογιστικών, αποθηκευτικών και άλλων) πόρων. Με αυτόν τον τρόπο, ανάλογα με τις απαιτήσεις για πόρους που έχει μια υπηρεσία, ακόμα και αν διαφέρουν ανά χρονικές περιόδους, υπάρχει η δυνατότητα να προσφέρονται, χρησιμοποιώντας τα κατάλληλα εργαλεία, που βρίσκονται στο επίπεδο της πλατφόρμας του υπολογιστικού νέφους. Καθώς όμως, η προετοιμασία για την προσφορά των πόρων στις υπηρεσίες - εφαρμογές, χρειάζεται αρκετό χρόνο, ανάλογα την κλιμάκωση των πόρων που ζητάει η υπηρεσία, θα πρέπει να υπάρχουν μηχανισμοί που θα προβλέπουν την δυναμική ανάγκη των υπηρεσιών για πόρους. Έτσι, θα επιτυγχάνεται η απαιτούμενη απόδοση των υπηρεσιών που είναι εγκατεστημένες στο υπολογιστικό νέφος, κάθε χρονική στιγμή.

Στην Ενότητα 2.4 αναφέρθηκαν οι σχετικές μελέτες που έχουν γίνει για την ανάλυση των εφαρμογών που είναι εγκατεστημένες στο υπολογιστικό νέφος, και την δημιουργία μηχανισμών που προβλέπουν την δυναμική ανάγκη για πόρους από τις υποδομές. Κυρίως ασχολούνται με την επίτευξη των απαιτούμενων επιδόσεων τους, σε ένα μοντέλο διαμοιραζόμενων πόρων, όπως είναι αυτό του υπολογιστικού νέφους. Δημιουργήθηκαν, έτσι, κατάλληλα μοντέλα πρόβλεψης, της συμπεριφοράς των εφαρμογών, και με την κατασκευή αντίστοιχων μηχανισμών, διαμοιράζονται δυναμικά οι πόροι του υπολογιστικού νέφους στις εφαρμογές, ανάλογα τις ανάγκες τους κάθε χρονική στιγμή. Οι μελέτες αυτές όμως, είχαν ως σκοπό την δημιουργία μηχανισμών που δουλεύουν με μοντέλα πρόβλεψης, για γενικούς τύπους εφαρμογών, πάνω σε συγκεκριμένες υποδομές υπολογιστικού νέφους. Αυτό αναλύθηκε στην βιβλιογραφία και συγκεκριμένα στην εργασία [44], που αντιμετωπίζει τις εφαρμογές σαν «μαύρα κουτιά» και τις κατηγοριοποιεί μέσα από συγκεκριμένα βήματα, και στην εργασία [47], όπου χρησιμοποιούνται συγκεκριμένα προφίλ, που αποτυπώνουν την γνώση για κλιμάκωση σε διάφορους τύπους εφαρμογών. Στις μελέτες αυτές, υποστηρίζεται ότι οι μηχανισμοί αυτοί δουλεύουν για τους

περισσότερους τύπους εφαρμογών. Σε άλλες εργασίες όμως, που παρουσιάστηκαν στην Ενότητα 2.4, μελετάται για το πρόβλημα αυτό, συγκεκριμένοι τύποι εφαρμογών, έχοντας έτσι μεγαλύτερη επιτυχία στην δημιουργία καταλληλότερων μηχανισμών πρόβλεψης (όπως για παράδειγμα στην εργασία [48], όπου μελετάται μια εφαρμογή ηλεκτρονικού καταστήματος). Σε κάθε περίπτωση όμως, μπορεί να γίνει χρήση των τεχνικών και των μεθοδολογιών, που χρησιμοποιούν οι εργασίες αυτές για την μελέτη παρόμοιων ερευνητικών θεματικών περιοχών.

Στην Ενότητα 2.5 παρουσιάστηκαν οι ανοιχτές θεματικές περιοχές για έρευνα, που σχετίζονται με την ανάλυση της συμπεριφοράς των εφαρμογών, για την επίτευξη των απαιτούμενων επιδόσεών τους, με χρήση του υπολογιστικού νέφους. Οι περισσότερες ερευνητικές περιοχές, σχετίζονται με τους πόρους του υπολογιστικού νέφους, και πως θα διαμοιράζονται σωστά στις εφαρμογές – υπηρεσίες. Πως δηλαδή θα μεταφράζονται οι απαιτήσεις των εφαρμογών που βρίσκονται στο υψηλό επίπεδο του υπολογιστικού νέφους (επίπεδο εφαρμογών), σε συγκεκριμένους πόρους που βρίσκονται στο επίπεδο του υλικού. Επίσης αναφέρθηκαν θέματα, που έχουν ως σκοπό την δημιουργία δυναμικών μοντέλων πρόβλεψης, χωρίς την γνώση των υποδομών που χρησιμοποιεί το υπολογιστικό νέφος, μελετώντας όμως την συμπεριφορά των εφαρμογών με βάση συγκεκριμένες παραμέτρους, που σχετίζονται με την χρήση της υπηρεσίας.

Η παρούσα εργασία ασχολείται με την δημιουργία ενός μοντέλου δυναμικής πρόβλεψης της απόδοσης μιας υπηρεσίας – εφαρμογής, που είναι εγκατεστημένη σε υποδομές υπολογιστικού νέφους, βάσει των παραμέτρων κλιμάκωσης της, που επιτρέπει χρονικά περιορισμένη παροχή πόρων. Αυτό επιτυγχάνεται με την ανάλυση της συμπεριφοράς της υπηρεσίας, ορίζοντας συγκεκριμένες παραμέτρους που μελετώνται και επηρεάζουν την απόδοσή της. Σκοπός είναι, με βάση το μοντέλο αυτό, η πρόβλεψη της απόδοσης της υπηρεσίας, όταν αλλάζουν οι τιμές των παραμέτρων που την επηρεάζουν. Έτσι, καθορίζοντας, ο πάροχος της υπηρεσίας, τη μέγιστη επιθυμητή απόδοση της και ανάλογα τις αλλαγές στις παραμέτρους της, θα μπορούν να αυξηθούν είτε να μειωθούν οι πόροι που προσφέρονται σε αυτή.

Η χρήση του μοντέλου αυτού, βοηθάει επίσης, στην εξαγωγή κάποιων πολύ χρήσιμων συμπερασμάτων, σχετικά με το πόσο επηρεάζεται η απόδοσή της υπηρεσίας από την κάθε παράμετρο. Έτσι, αν κάποια παράμετρος επηρεάζει περισσότερο την απόδοση της υπηρεσίας, σε σχέση με τις υπόλοιπες που μελετώνται, μπορούν να γίνουν κινήσεις στο επίπεδο του υλικού, που θα βελτιώνουν την συμπεριφορά της. Αν για παράδειγμα, από το μοντέλο πρόβλεψης, προκύψουν τα συμπεράσματα, ότι η γεωγραφική απόσταση απ' όπου προέρχονται τα αιτήματα για χρήση της υπηρεσίας, επηρεάζει αρκετά την απόδοσή της, τότε θα μπορούσε να επεκταθούν οι υποδομές του υπολογιστικού νέφους, σε περισσότερες γεωγραφικές θέσεις. Με αυτό τον τρόπο, θα μειωθεί η απόσταση των υποδομών του υπολογιστικού νέφους, από τις γεωγραφικές θέσεις απ' όπου προέρχονται τα αιτήματα για χρήση της υπηρεσίας, και έτσι, θα επιτευχθεί η επιθυμητή απόδοση, με χρήση ίσως λιγότερων πόρων. Η ανάλυση αυτή των συμπερασμάτων, σε σχέση με το κόστος αλλαγής στις υποδομές του υπολογιστικού νέφους, δεν αποτελεί σκοπό αυτής της εργασίας. Στην εργασία αυτή, παρουσιάζεται ο βαθμός επιρροής της κάθε παραμέτρου που μελετάται, στην απόδοση της υπηρεσίας.

Αρχικά, για να την δημιουργία του μοντέλου πρόβλεψης, θα πρέπει να μελετηθεί η απόδοση μιας υπηρεσίας, που χρησιμοποιεί υποδομές υπολογιστικού νέφους, σε σχέση με τις παραμέτρους που την επηρεάζουν. Οπότε, για την μελέτη και ανάλυση αντιπροσωπευτικών δεδομένων, έγινε επιλογή μιας υπηρεσίας, βάσει συγκεκριμένων κριτηρίων που τέθηκαν. Τα κριτήρια αυτά αφορούν:

- Για την εκτέλεση της, να κάνει χρήση υποδομών υπολογιστικού νέφους.
- Η υπηρεσία, να μη κάνει αυξημένη χρήση πόρων κατά την εκτέλεση της, αλλά να τους χρησιμοποιεί στην εξυπηρέτηση των αιτημάτων.
- Η υπηρεσία αυτή να χρησιμοποιείται ευρέως από μεγάλο μέρος του πληθυσμού, έτσι ώστε να αποτελεί μια αντιπροσωπευτική υπηρεσία που διαχειρίζεται μεγάλο όγκο δεδομένων.
- Τα αιτήματα προς την υπηρεσία, για την Ευρώπη, να εξυπηρετούνται από τις ίδιες υποδομές υπολογιστικού νέφους, και όχι ανάλογα την γεωγραφική θέση να εξυπηρετούνται τα αιτήματα από διαφορετικές υποδομές.

Ενώ πολλές υπηρεσίες πληρούν τα παραπάνω κριτήρια, κρίθηκε σκόπιμο να μελετηθεί η υπηρεσία της εγκυκλοπαιδικής ιστοσελίδας www.wikipedia.org. Ο λόγος είναι ότι διαχειρίζεται μεγαλύτερο όγκο δεδομένων από αυτές που αξιολογήθηκαν (www.britannica.com, www.answers.com, www.infoplease.com και www.imdb.com). Προφανώς υπάρχουν και άλλες υπηρεσίες που πληρούν τα παραπάνω κριτήρια, η παρούσα όμως εργασία, ασχολήθηκε με την μελέτη συγκεκριμένων.

Στην συνέχεια, ορίστηκαν οι παράμετροι που επηρεάζουν την απόδοσή τους. Για την επιλογή των καταλληλότερων παραμέτρων η έρευνα πραγματοποιήθηκε με βάση την βιβλιογραφία που παρουσιάστηκε στα δύο πρώτα κεφάλαια. Έτσι, οι παράμετροι που θα μελετηθούν και επηρεάζουν την απόδοση της υπηρεσίας είναι:

- τα παράλληλα αιτήματα από τους χρήστες προς την υπηρεσία,
- η γεωγραφική θέση απ' όπου εκτελούνται τα αιτήματα,
- ο χρόνος κατά τον οποίο εκτελούνται τα αιτήματα προς την υπηρεσία.

Για την συλλογή και επεξεργασία των παραμέτρων αυτών της υπηρεσίας, θα πρέπει να γίνει χρήση των κατάλληλων τεχνικών και μεθοδολογιών. Από την βιβλιογραφία που μελετήθηκε και παρουσιάστηκε στην Ενότητα 2.4, για την επίλυση του ερευνητικού αυτού προβλήματος, επιλέχθηκαν οι καταλληλότερες, με βάση την δυνατότητα χρήσης τους σε παρόμοια θέματα, όπως είναι η παρούσα εργασία.

Ένα κοινό σημείο της εργασίας [45] με την παρούσα, είναι η μελέτη της παραμέτρου των παράλληλων αιτημάτων προς την υπηρεσία. Για τον λόγο αυτό, κατασκευάστηκε εφαρμογή που προσομοιώνει τα παράλληλα αιτήματα προς την υπηρεσία, έτσι ώστε να καταγραφεί η συμπεριφορά της υπηρεσίας. Επίσης η εφαρμογή αυτή, πραγματοποιούσε συλλογή όλων των δεδομένων για τις παραμέτρους που μελετούσε η εργασία. Και για την παρούσα εργασία κρίθηκε απαραίτητο η ανάπτυξη αντίστοιχης εφαρμογής, που προσομοιώνει τα παράλληλα αιτήματα προς την υπηρεσία, αλλά και συλλέγει τα δεδομένα για τις παραμέτρους που μελετά.

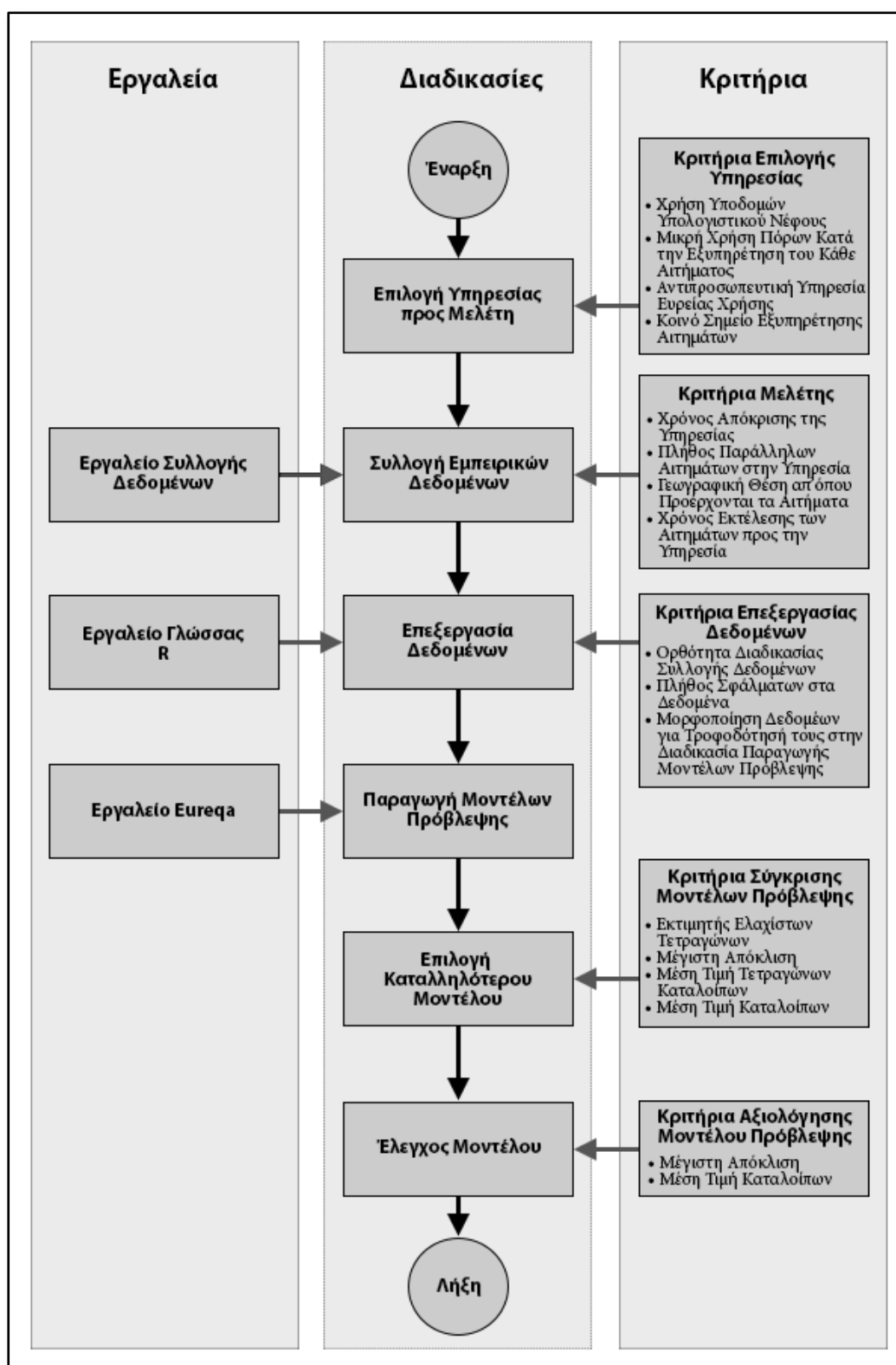
Για την επεξεργασία των δεδομένων που συλλέγονται για την υπηρεσία που μελετάται, όπως αναφέρθηκε στις εργασίες [43] και [44], έγινε χρήση του εργαλείου GNU Octave. Ένα παρόμοιο εργαλείο χρησιμοποιείται και στην παρούσα εργασία. Το εργαλείο αυτό είναι το “R Project”[52]. Η επιλογή του συγκεκριμένου έγινε, καθώς προϋπήρχε γνώση πάνω στο εργαλείο της γλώσσας R, και ο τρόπος επεξεργασίας των δεδομένων από τις εργασίες [43] και [44], μπορούν να αναπαραχθεί και στο εργαλείο της γλώσσας R.

Επίσης, στις εργασίες [42], [43], [44] και [48] έγινε χρήση αλγορίθμων τεχνητής μάθησης και των τεχνητών νευρωνικών δικτύων υπολογιστών, για την δημιουργία του μοντέλου πρόβλεψης. Έτσι, με την ανάλυση των δεδομένων που έχουν συλλεχθεί και επεξεργαστεί, και με τα κατάλληλα εργαλεία που χρησιμοποιούν αλγορίθμους μηχανικής μάθησης και νευρωνικά δίκτυα υπολογιστών, παράγονται μοντέλα που προβλέπουν τις τιμές της απόδοσης της υπηρεσίας που μελετάται. Οπότε και στην παρούσα εργασία, για την δημιουργία των μοντέλων πρόβλεψης, γίνεται χρήση παρόμοιου εργαλείου, πάνω στο οποίο προϋπάρχει γνώση, και είναι το Eureka [53]. Το εργαλείο αυτό, όπως παρουσιάζεται αναλυτικότερα στο Κεφάλαιο 7, κάνει χρήση μηχανικής εκμάθησης και των τεχνητών νευρωνικών δικτύων υπολογιστών, για την παραγωγή μοντέλων πρόβλεψης, με βάση τα εισαχθέντα δεδομένα.

Στις εργασίες [42] και [48], για την αξιολόγηση της απόδοσης της υπηρεσίας που μελετάται, σαν μονάδας μέτρησης χρησιμοποιήθηκε ο βαθμός ενός προγράμματος συγκριτικής αξιολόγησης. Στην παρούσα εργασία, κρίθηκε καταλληλότερη μονάδα μέτρησης, ο χρόνος απόκρισης της υπηρεσίας. Αυτό οφείλετε στο γεγονός, ότι η υπηρεσία που μελετάται είναι υπηρεσία ιστοσελίδας, που έχει ως αντικειμενικό σκοπό την σχεδόν άμεση εξυπηρέτηση των αιτημάτων προς αυτή.

Στο Σχήμα 3.2, παρουσιάζεται η ροή των ερευνητικών εργασιών, που πραγματοποιούνται για την επίλυση της θεματικής περιοχής, που σχετίζεται με την δημιουργία δυναμικού μοντέλου πρόβλεψης της απόδοσης της υπηρεσίας, βάσει των παραμέτρων κλιμάκωσής της. Παρουσιάζονται έτσι, οι διαδικασίες που ακολουθήθηκαν σε κάθε βήμα, βάσει συγκεκριμένων κριτηρίων, όπου ήταν απαραίτητο. Επιπλέον, για την πραγματοποίηση των διαδικασιών, ήταν απαραίτητη η χρήση συγκεκριμένων εργαλείων, τα οποία αναφέρονται στο Σχήμα 3.2. Συνεπώς, βάσει του σχήματος, αρχικά γίνεται επιλογή της υπηρεσίας που μελετάται, βάσει συγκεκριμένων κριτηρίων που αναφέρονται. Στην συνέχεια, με βάση τα κριτήρια μελέτης, όπου προσδιορίζουν τις παραμέτρους της υπηρεσίας που μελετώνται, πραγματοποιείται συλλογή των εμπειρικών δεδομένων της υπηρεσίας. Τα δεδομένα αυτά, αφού αποθηκευτούν, επόμενο στάδιο είναι η επεξεργασία τους. Η επεξεργασία των δεδομένων, γίνεται με την χρήση του εργαλείου της γλώσσας R, που χρησιμοποιείται για την επεξεργασία δεδομένων και την εκτέλεση στατιστικών υπολογισμών. Η διαδικασία αυτή πραγματοποιείται για την εξαγωγή συμπερασμάτων, σχετικά με την ορθότητα των δεδομένων που συλλέχθηκαν, αλλά και για την μορφοποίηση τους ώστε να χρησιμοποιηθούν στην συνέχεια. Στο επόμενο στάδιο, τα δεδομένα αυτά, εισάγονται στην εφαρμογή Eureka, όπου κάνοντας χρήση μοντέλων μηχανικής μάθησης και νευρωνικών δικτύων υπολογιστών, παράγει μοντέλα πρόβλεψης. Τα μοντέλα που προκύπτουν, στην συνέχεια, συγκρίνονται με βάση συγκεκριμένα κριτήρια, που αναφέρονται συνοπτικά στο Σχήμα 3.2, για την επιλογή του καταλληλότερου μοντέλου. Το επιλεγμένο μοντέλο, βάσει

συγκεκριμένων κριτηρίων, θα αξιολογηθεί για την εγκυρότητά του, σε πραγματικά δεδομένα. Έτσι, θα προκύψουν κάποια συμπεράσματα για την ικανότητα που προσφέρει το μοντέλο αυτό, για την πρόβλεψη της απόδοσης της υπηρεσίας που μελετάται.



Σχήμα 3.2: Ροή ερευνητικών εργασιών για την δημιουργία δυναμικού μοντέλου πρόβλεψης της απόδοσης μιας υπηρεσίας

3.4: Συμπεράσματα Προτεινόμενου Μοντέλου

Στο παρόν κεφάλαιο, παρουσιάστηκαν οι προκλήσεις που υπάρχουν για την δημιουργία μοντέλων πρόβλεψης της απόδοσης υπηρεσιών, που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους. Οι προκλήσεις αυτές, παρουσιάζονται είτε λόγω της δομής του υπολογιστικού νέφους, με τα ξεχωριστά επίπεδα των εφαρμογών και του υλικού του, όπου δεν είναι εύκολη η μεταφορά της πληροφορίας μεταξύ των επιπέδων, είτε λόγω άλλων περιορισμών του (π.χ. περιορισμοί των SLAs, συσχέτιση μεταξύ εισόδων και εξόδων των δεδομένων των εφαρμογών). Προκλήσεις, επίσης υπάρχουν και στην πρόβλεψη της απόδοσης εφαρμογών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, χωρίς την μελέτη παραμέτρων του υλικού, αλλά παραμέτρων που σχετίζονται με την χρήση της υπηρεσίας (π.χ. το πλήθος των παράλληλων αιτημάτων για χρήση της υπηρεσίας και η γεωγραφική θέση, απ' όπου προέρχονται).

Η παρούσα εργασία ασχολείται με την δημιουργία ενός μοντέλου δυναμικής πρόβλεψης της απόδοσης μιας υπηρεσίας – εφαρμογής, που είναι εγκατεστημένη σε υποδομές υπολογιστικού νέφους, βάσει των παραμέτρων κλιμάκωσης της, που επιτρέπει χρονικά περιορισμένη παροχή πόρων. Έτσι, στην Ενότητα 3.3 αναφέρθηκε το προτεινόμενο μοντέλο για την επίλυση της ερευνητικής θεματικής περιοχής. Για την κατασκευή αυτού του μοντέλου, επιλέχθηκε η διερεύνηση υπηρεσίας που πληροί συγκεκριμένα κριτήρια. Ο λόγος είναι η μελέτη συγκεκριμένου τύπου υπηρεσιών (που επηρεάζεται η απόδοσή τους από τις επιλεγμένες παραμέτρους), για την δημιουργία μοντέλων πρόβλεψης. Έτσι, το μοντέλο δυναμικής πρόβλεψης της απόδοσης, θα περιλαμβάνει τις τιμές των παραμέτρων που επιλέχθηκαν, καθώς και τις βαρύτητες τους, και θα προβλέπει για τις τιμές αυτές, την απόδοση της υπηρεσίας.

Για την επίλυση της θεματικής περιοχής που επιλέχθηκε, χρησιμοποιήθηκαν τεχνικές και μεθοδολογίες από την σχετική βιβλιογραφία, που παρουσιάστηκε στο Κεφάλαιο 2. Αναφέρθηκε έτσι, στην Ενότητα 3.3, με ποιόν τρόπο χρησιμοποιούνται οι τεχνικές και οι μεθοδολογίες αυτές από την παρούσα εργασία, και ποιες είναι οι διαφοροποιήσεις που έγιναν. Οι ερευνητική μεθοδολογία που ακολουθήθηκε, παρουσιάζεται αναλυτικότερα στο Κεφάλαιο 4. Μπορεί στην παρούσα εργασία να εξετάζεται μια συγκεκριμένη υπηρεσία, αλλά η ίδια διαδικασία μπορεί να ακολουθηθεί και για παρόμοιες υπηρεσίες που πληρούν τα κριτήρια που τέθηκαν. Με αυτόν τον τρόπο, είναι δυνατή η δημιουργία δυναμικών μοντέλων πρόβλεψης για την κάθε μια ξεχωριστά.

Κεφάλαιο 4: Ερευνητική Μεθοδολογία

4.1: Εισαγωγή

Στο Κεφάλαιο 2 παρουσιάστηκαν θεματικές ερευνητικές περιοχές που είτε μελετήθηκαν από άλλες εργασίες, είτε παραμένουν ανοιχτές προς μελέτη. Μια από αυτές τις ανοιχτές θεματικές περιοχές επιλέχθηκε για περεταίρω ανάλυση από την παρούσα εργασία, και παρουσιάστηκε αναλυτικά στην Ενότητα 3.3. Η θεματική περιοχή που επιλέχθηκε σχετίζεται με την δυναμική πρόβλεψη της απόδοσης υπηρεσιών που είναι εγκατεστημένες σε υποδομές υπολογιστικού νέφους, βάσει των παραμέτρων κλιμάκωσής τους. Για την εύρεση της καταλληλότερης προσέγγισης της ερευνητικής μεθοδολογίας, αρχικά μελετήθηκαν οι μεθοδολογίες που χρησιμοποιήθηκαν στην σχετική βιβλιογραφία (Ενότητα 2.2). Στην συνέχεια, ερευνήθηκαν περεταίρω οι επιλεγμένες μεθοδολογίες της βιβλιογραφίας, βάσει των διαφορετικών προσεγγίσεων που μπορεί να ακολουθηθούν.

Στο κεφάλαιο αυτό γίνεται παρουσίαση των ερευνητικών μεθόδων που ακολουθούνται σε παρόμοιες έρευνες, που σχετίζονται με την ανάλυση της συμπεριφοράς εφαρμογών για την δημιουργία μοντέλων πρόβλεψης της απόδοσης τους. Επιπλέον πραγματοποιείται σε βάθος ανάλυση, της συγκεκριμένης που ακολουθήθηκε στην παρούσα εργασία. Τέλος, θα αναφερθεί η χρησιμότητα της επιλεγμένης ερευνητικής μεθοδολογίας και σε άλλες εργασίες, που μελετάνε την απόδοση υπηρεσιών, λαμβάνοντας υπόψη παραμέτρους που αφορούν την χρήση τους και όχι παραμέτρους του υλικού των υποδομών τους.

4.2: Παρουσίαση και Τεκμηρίωση Ερευνητικών Μεθοδολογιών

Για την επιλογή της κατάλληλης ερευνητικής μεθοδολογίας, που θα χρησιμοποιηθεί στην παρούσα εργασία, μελετήθηκε αρχικά η συλλογή εμπειρικών δεδομένων που χρησιμοποιήθηκε από τις εργασίες της σχετικής βιβλιογραφίας. Θεωρείται από τους ερευνητές της σχετικής βιβλιογραφίας, ως μια κατάλληλη μεθοδολογία, καθώς η δημιουργία μοντέλων πρόβλεψης της απόδοσης των υπηρεσιών βασίζεται σε πραγματικά δεδομένα, που συλλέγονται κατά την εκτέλεση των υπηρεσιών. Έτσι, βάσει της προσέγγισης που ακολουθείται (αναφέρονται στην επόμενη παράγραφο), για την μεθοδολογία συλλογής εμπειρικών δεδομένων, γίνεται καταγραφή της απόδοσης των υπηρεσιών, και των παραμέτρων κλιμάκωσής τους, όπου στην συνέχεια θα επεξεργαστούν για την δημιουργία των μοντέλων πρόβλεψης της απόδοσης.

Με βάση την ανάλυση των μεθοδολογιών που παρουσιάζονται στο [54], οι βασικές ερευνητικές προσεγγίσεις συλλογής των εμπειρικών δεδομένων (με σκοπό είτε την περιγραφή νέων δεδομένων, είτε τον έλεγχο υποθέσεων) είναι:

- Περιγραφική προσέγγιση (observational approach)
- Συσχετιστική προσέγγιση (correlational approach)
- Προ-πειραματική προσέγγιση (pre-experimental approach)

- Ψευδο-πειραματική σχεδιασμοί (quasi-experimental approach)
- Πειραματικοί σχεδιασμοί (experimental approach)

Στις επόμενες ενότητες παρουσιάζονται οι παραπάνω προσεγγίσεις της μεθοδολογίας, για την επιλογή της καταλληλότερης για τη μελέτη της παρούσας εργασίας. Επιπλέον, όπως αναφέρεται στο [55], οι εμπειρικές μέθοδοι συλλογής δεδομένων περιλαμβάνουν υποθετικά δείγματα και ελέγχους των μεταβλητών. Περιλαμβάνει λοιπόν την έννοια της υπόθεσης πάνω στα εμπειρικά δεδομένα. Αυτό σημαίνει ότι πραγματοποιούνται κάποιες ερευνητικές υποθέσεις, όπως για παράδειγμα η καταλληλότητα των δεδομένων ή των παραμέτρων για την εξαγωγή συμπερασμάτων. Όπως είναι λογικό, η έννοια της υπόθεσης περιλαμβάνει την περίπτωση σφάλματος, σχετικά με τη γενίκευση των εξαγόμενων συμπερασμάτων.

4.2.1 Περιγραφική Προσέγγιση

Η προσέγγιση αυτή αποσκοπεί στην καταγραφή και περιγραφή ψυχολογικών φαινομένων. Μέθοδοι συλλογής δεδομένων στις οποίες μπορεί να στηρίζεται είναι η συστηματική παρατήρηση (όταν το υπό μελέτη φαινόμενο είναι ορατό και απτό), ψυχομετρικά τεστ (για καταγραφή του τι αισθάνεται κάποιος), καταγραφή φυσιολογικών δεδομένων με ειδικά όργανα (πχ. καρδιακή λειτουργία), μέθοδοι ανασκόπησης (ερωτηματολόγιο, συνέντευξη) για την καταγραφή σκέψεων, στάσεων, πεποιθήσεων.

Μέθοδος επιλογής όταν:

- Ο χειρισμός των μεταβλητών είναι πρακτικά ή δεοντολογικά αδύνατον να πραγματοποιηθεί.
- Δεν είναι δυνατή η ταξινόμηση των συμμετεχόντων σε ομάδες ή συνθήκες της έρευνας με τυχαίο τρόπο.
- Στόχος της έρευνας είναι η απλή καταγραφή του φαινομένου (και όχι η πρόβλεψη ή ερμηνεία του).

Πλεονέκτημα:

- Η συλλογή δεδομένων γίνεται κάτω από φυσικές συνθήκες (φυσικό περιβάλλον, λεκτική επικοινωνία, κλπ.). Η παρέμβαση του ερευνητή σε αυτή τη μέθοδο είναι όσο το δυνατόν μικρότερη.

Μειονεκτήματα:

- Δεν συλλέγονται στοιχεία σχετικά με τον προσδιορισμό των αιτιωδών σχέσεων μεταξύ των μεταβλητών, καθώς υπάρχουν πολλές μεταβλητές που μεταβάλλονται συγχρόνως. Η περιγραφική μέθοδος επιτρέπει τον προσδιορισμό της χρονικής αλληλουχίας των μεταβολών μεταξύ τους (πχ προβολή σκηνών βίας στην τηλεόραση-επιθετική συμπεριφορά, αύξηση φορολογίας-δυσανεμία προς την κυβέρνηση). Η διαπίστωση χρονικών σχέσεων ωστόσο δεν αποτελεί απόδειξη αιτιωδών σχέσεων (μόνο περιστασιακή ένδειξη).
- Ο ερευνητής παρεμβαίνει λιγότερο στη διαδικασία, ωστόσο η επιστημονική αξία και η ακρίβεια των αποτελεσμάτων εξαρτάται από την υποκειμενική κρίση του.

Η αξία της περιγραφικής μεθόδου έγκειται στο ότι μπορεί να συλλεχθούν στοιχεία που θα διαμορφώσουν συγκεκριμένες υποθέσεις (πχ σχετικά με τη σπουδαιότητα διαφορετικών μεταβλητών και τις μεταξύ τους σχέσεις) και θα περιγράψουν τις βασικές συνιστώσες του προβλήματος. Ιδιαίτερα χρήσιμη στα αρχικά στάδια της διερεύνησης ενός φαινομένου.

4.2.2 Συσχετιστική Προσέγγιση

Χρησιμοποιείται για υψηλότερο επίπεδο κατανόησης ενός φαινομένου. Η μέθοδος αυτή συνίσταται στη συλλογή δεδομένων για την καταγραφή της συνδιακύμανσης μεταξύ μεταβλητών, προκειμένου να οδηγηθεί ο ερευνητής σε συμπεράσματα για τη σχέση που υπάρχει μεταξύ τους. Εξετάζεται η σχέση μεταξύ των διαφορετικών τιμών μιας μεταβλητής με τη συστηματική διακύμανση των τιμών μιας άλλης μεταβλητής.

Ο χειρισμός της ανεξάρτητης μεταβλητής δεν είναι άμεσος, αλλά γίνεται μέσω επιλογής, κατά τη συλλογή δεδομένων, ενός εύρους τιμών της μεταβλητής (η συλλογή δεδομένων να καλύπτει ικανοποιητικά όλο το εύρος τιμών της μεταβλητής, ή όλων των κατηγοριών της μεταβλητής). Με τη χρήση στατιστικών μεθόδων επιτρέπεται η μελέτη της συσχέτισης μεταξύ πολλών μεταβλητών. Ένα μειονέκτημα της συσχετιστικής προσέγγισης είναι ότι, η διαπίστωση σχέσεων συνάφειας δεν μεταφράζεται απαραίτητα στην ύπαρξη αιτιωδών σχέσεων (δεν είναι γνωστό αν υπάρχει αιτιότητα και προς ποια κατεύθυνση).

4.2.3 Προ-πειραματική Προσέγγιση

Η προσέγγιση αυτή αναφέρεται σε μελέτες περίπτωσης που ολοκληρώνονται σε μία προσπάθεια, και στοχεύουν στην περιγραφή των μεταβλητών και όχι στο χειρισμό τους. Οι προ-πειραματικοί σχεδιασμοί δεν είναι ιδιαίτερα δημοφιλείς, ακριβώς επειδή αφορούν μελέτες περίπτωσης και δεν προχωρούν σε συγκρίσεις. Ωστόσο, πολλές μελέτες περίπτωσης μπορούν να οδηγήσουν στην περιγραφή θεμελιωδών σχέσεων.

4.2.4 Ψευδο-πειραματική Προσέγγιση

Οι σχεδιασμοί αυτού του τύπου αφορούν έρευνες που στοχεύουν στη διερεύνηση αιτιωδών σχέσεων και πραγματοποιούνται εκτός εργαστηρίου. Βασικό χαρακτηριστικό τους είναι ότι για πρακτικούς ή δεοντολογικούς λόγους δεν μπορεί να γίνει τυχαία δειγματοληψία. Έτσι, το δείγμα δεν θεωρείται αντιπροσωπευτικό του πληθυσμού και μπαίνει σε κίνδυνο η δυνατότητα γενίκευσης των αποτελεσμάτων. Ωστόσο, αυτού του τύπου οι ερευνητικοί σχεδιασμοί μπορεί να είναι πολύ χρήσιμοι, καθώς διερευνούν τα φαινόμενα στις πραγματικές τους συνθήκες και όχι στις τεχνητές συνθήκες που διαμορφώνονται εντός ενός πειραματικού εργαστηρίου. Αυτού του τύπου οι σχεδιασμοί προσφέρουν αυξημένη δυνατότητα γενίκευσης των αποτελεσμάτων (εξωτερική εγκυρότητα), με το κόστος του μειωμένου βαθμού ελέγχου της πειραματικής διαδικασίας (εσωτερική εγκυρότητα).

4.2.5 Πειραματική Προσέγγιση

Ο ερευνητής ελέγχει τις μεταβλητές που τον ενδιαφέρουν, και παρεμβαίνει άμεσα στο χειρισμό τους, με σκοπό τη διερεύνηση αιτιωδών σχέσεων. Οι συνθήκες διεξαγωγής ενός

πειράματος πρέπει να είναι όσο το δυνατόν περισσότερο ελεγχόμενες, προκειμένου να αποφύγουμε την επίδραση άσχετων ή ανεπιθύμητων μεταβλητών. Ελέγχει επίσης τις μεταβλητές που είναι άσχετες ή ανεπιθύμητες, απαλείφοντας τις ή κρατώντας τις σταθερές.

Στην πειραματική διαδικασία έχουμε δύο συνθήκες για την καταγραφή των δεδομένων:

- Πειραματική/ές συνθήκη/ες (experimental condition-group)
- Συνθήκη ελέγχου (control condition-group)

Στην πειραματική συνθήκη ο ερευνητής χειρίζεται τη μεταβλητή, ενώ στην συνθήκη ελέγχου όχι. Αυτή θα πρέπει να είναι η μόνη διαφορά μεταξύ των δύο. Αν υπάρξει διαφορά στα σκορ μεταξύ των δύο συνθηκών, τότε θεωρούμε ότι αυτή είναι απόρροια του πειραματικού χειρισμού (experimental treatment), μόνο όταν αυτή είναι η μοναδική διαφορά μεταξύ των δύο συνθηκών.

Πλεονεκτήματα:

- Εξέταση αιτιωδών σχέσεων,
- Διεξαγωγή σε αυστηρά ελεγχόμενο περιβάλλον-ουσιώδες χαρακτηριστικό της επιστημονικής μεθόδου. Θεωρείται ότι αποτελεί το πρότυπο της επιστημονικής έρευνας.

Μειονέκτημα:

- Ο υπερβολικός έλεγχος καθιστά τα αποτελέσματα τεχνητά, ίσως μη γενικεύσιμα σε πραγματικές καταστάσεις.

4.3: Επιλογή Προσέγγισης Μεθοδολογίας Συλλογής Εμπειρικών Δεδομένων

Όπως αναφέρθηκε στην Ενότητα 4.2, στις εργασίες της σχετικής βιβλιογραφίας η ερευνητική μέθοδος που ακολουθήθηκε ήταν η μεθοδολογία συλλογής δεδομένων με πειραματική προσέγγιση. Στην παρούσα εργασία, βάσει και της περεταίρω ανάλυσης της μεθοδολογίας συλλογής εμπειρικών δεδομένων, φαίνεται σκόπιμη η χρήση της, καθώς η πρόβλεψη της απόδοσης της υπηρεσίας βασίζεται στις τιμές των παραμέτρων που την επηρεάζουν. Για τον λόγο αυτό, πραγματοποιείται καταγραφή της συμπεριφοράς της απόδοσης για τις συγκεκριμένες τιμές των παραμέτρων της. Οι προσεγγίσεις της μεθοδολογίας συλλογής δεδομένων, που επιλέχθηκαν στην παρούσα εργασία, με βάση αυτά που παρουσιάστηκαν στις προηγούμενες ενότητες, είναι αυτές της συσχετιστικής και της πειραματικής προσέγγισης. Γίνεται έτσι ένας συνδυασμός και των δύο προσεγγίσεων.

Η επιλογή της συσχετιστικής προσέγγισης οφείλεται στην συνδιακύμανση μεταξύ των μεταβλητών. Στην παρούσα εργασία εξετάζεται η σχέση μεταξύ των διαφορετικών τιμών ορισμένων προκαθορισμένων ανεξάρτητων μεταβλητών με τη συστηματική διακύμανση των τιμών της εξαρτημένης μεταβλητής. Ο χειρισμός των ανεξάρτητων μεταβλητών δεν είναι άμεσος, αλλά γίνεται μέσω επιλογής, κατά τη συλλογή δεδομένων, ενός εύρους τιμών της μεταβλητής. Έτσι οι τιμές των ανεξάρτητων μεταβλητών θα πρέπει να συλλέγονται για το μεγαλύτερο δυνατό εύρος τιμών τους.

Η επιλογή της πειραματικής προσέγγισης οφείλεται στον έλεγχο των μεταβλητών που έχουν προκαθοριστεί. Ο συνδυασμός με την προηγούμενη προσέγγιση, δεν επιτρέπει τον πλήρη έλεγχο, αλλά μπορεί να υπάρχει μερικώς έλεγχος. Όπως για παράδειγμα, στην συλλογή των δεδομένων, να μπορούν να οριστούν όρια στις τιμές κάποιων μεταβλητών, χωρίς όμως να επηρεάζουν τις υπόλοιπες ανεξάρτητες μεταβλητές. Έτσι, ορίστηκαν όρια στην μέγιστη τιμή της παραμέτρου των παράλληλων αιτημάτων προς την υπηρεσία. Η εφαρμογή που εκτελεί τις αιτήσεις, δέχεται ως παράμετρο, την μέγιστη τιμή των παράλληλων αιτημάτων και κάθε χρονική στιγμή εκτελεί έναν τυχαίο αριθμό για αυτές (π.χ. την χρονική στιγμή 1 εκτελούνται 163 παράλληλα αιτήματα προς την υπηρεσία, και την χρονική στιγμή 2 εκτελούνται 76 παράλληλα αιτήματα). Για κάθε εκτέλεση των αιτημάτων, τα δεδομένα που συλλέγονται, αποθηκεύονται σε αρχεία.

4.4: Επεξεργασία Εμπειρικών Δεδομένων

Επόμενο στάδιο της εργασίας αυτής, μετά την συλλογή των εμπειρικών δεδομένων, είναι η μελέτη και η επεξεργασία τους. Μέσα από αυτή την διαδικασία, προκύπτουν κάποια χρήσιμα συμπεράσματα για την μελέτη της απόδοσης της υπηρεσίας, που χρησιμοποιούνται από τα επόμενα στάδια της μελέτης (π.χ. αν η απόδοση της υπηρεσίας επηρεάζεται από άλλες παραμέτρους που δεν μελετώνται από την συγκεκριμένη εργασία). Στην συνέχεια, τα επεξεργασμένα δεδομένα, από το προηγούμενο στάδιο μελέτης, εισάγονται στο εργαλείο Eureka [53], για την έναρξη της διαδικασίας παραγωγής των μοντέλων πρόβλεψης. Το Eureka είναι ένα εργαλείο όπου χρησιμοποιεί τεχνικές μηχανικής μάθησης και τεχνητά νευρωνικά δίκτυα υπολογιστών. Όπως αναφέρθηκε στην Ενότητα 3.3, οι τεχνικές της μηχανικής μάθησης και των τεχνητών νευρωνικών δικτύων υπολογιστών, χρησιμοποιήθηκαν και από τις εργασίες της βιβλιογραφίας, για την δημιουργία μηχανισμών πρόβλεψης. Αναλυτικότερη παρουσίαση του εργαλείου Eureka, γίνεται στο έβδομο κεφάλαιο. Στο τέλος της διαδικασίας αυτής, θα προκύψουν ορισμένα μοντέλα πρόβλεψης της απόδοσης της υπηρεσίας που μελετάται.

Τα μοντέλα πρόβλεψης, που έχουν παραχθεί στο προηγούμενο στάδιο της διαδικασίας, αξιολογούνται με βάση συγκεκριμένα αντικειμενικά κριτήρια, για την επιλογή του καταλληλότερου. Τα κριτήρια αυτά προέκυψαν από την μαθηματική ανάλυση μοντέλων πρόβλεψης και αναφέρονται αναλυτικά στο Κεφάλαιο 7 (π.χ. μέση τιμή απόκλισης προβλεπόμενων τιμών από τις πραγματικές). Στην συνέχεια, το επιλεγμένο μοντέλο, αφού παρουσιαστεί αναλυτικά, γίνονται κάποιοι έλεγχοι και δοκιμές, για να αποδειχθεί η ορθότητα του. Η διαδικασία για τον έλεγχο του δυναμικού μοντέλου πρόβλεψης, γίνεται με την συλλογή δεδομένων, με τον ίδιο μηχανισμό που έχει αναπτυχθεί για την συλλογή των εμπειρικών δεδομένων. Τα δεδομένα αυτά ύστερα από επεξεργασία, πραγματοποιείται σύγκριση τους με τα δεδομένα που προκύπτουν από το μοντέλο πρόβλεψης που έχει επιλεγεί. Όσο πιο μικρή απόκλιση υπάρχει μεταξύ των πραγματικών και των προβλεπόμενων δεδομένων, τόσο πιο αξιόπιστο θεωρείται το μοντέλο.

4.5: Συμπεράσματα Χρήσης της Ερευνητικής Μεθοδολογίας

Για την επίλυση του ερευνητικού προβλήματος, με την δημιουργία ενός μοντέλου πρόβλεψης απόδοσης της υπηρεσίας που μελετάται, παρουσιάστηκε στο παρόν κεφάλαιο, η ερευνητική μεθοδολογία. Γίνεται, λοιπόν, χρήση της μεθοδολογίας συλλογής εμπειρικών δεδομένων, βασιζόμενη στην συσχετιστική και την πειραματική προσέγγιση. Ο λόγος που επιλέχθηκε η συσχετιστική προσέγγιση της μεθοδολογίας, είναι η συνδιακύμανση που υπάρχει, μεταξύ των μεταβλητών. Δηλαδή, η σχέση μεταξύ των διαφορετικών τιμών ορισμένων προκαθορισμένων ανεξάρτητων μεταβλητών (παράμετροι που επηρεάζουν την απόδοση), με τη συστηματική διακύμανση των τιμών της εξαρτημένης μεταβλητής (απόδοση υπηρεσίας). Ο συνδυασμός της προηγούμενης προσέγγισης με την πειραματική, έχει ως αποτέλεσμα τον μερικό έλεγχο των μεταβλητών. Έτσι, μπορούν να οριστούν όρια στις τιμές κάποιων μεταβλητών, χωρίς όμως να επηρεάζουν τις υπόλοιπες ανεξάρτητες μεταβλητές (π.χ. ορισμός μέγιστης τιμής των παράλληλων αιτημάτων προς την υπηρεσία, κατά την συλλογή των δεδομένων).

Στη συνέχεια, παρουσιάστηκε η επεξεργασία των δεδομένων που έχουν συλλεχθεί, από την οποία θα προκύψουν κάποια χρήσιμα συμπεράσματα για την επίλυση του προβλήματος (π.χ. η επιρροή της κάθε μεταβλητής στην απόδοση της υπηρεσίας). Τα επεξεργασμένα δεδομένα, από το προηγούμενο στάδιο μελέτης, εισάγονται στο εργαλείο Eureka, για την έναρξη της διαδικασίας παραγωγής των μοντέλων πρόβλεψης. Από τα μοντέλα που θα προκύψουν, θα επιλεγεί με βάση συγκεκριμένα αντικειμενικά κριτήρια, το καταλληλότερο. Τα κριτήρια αυτά βασίζονται σε μαθηματικές έννοιες που παρουσιάζονται στο Κεφάλαιο 7. Για το επιλεγμένο μοντέλο, στην συνέχεια θα ελεγχθεί η μέση τιμή απόκλισης των προβλεπόμενων τιμών του, από την πραγματικότητα. Θα γίνει δηλαδή, συλλογή νέων δεδομένων όπου θα συγκριθούν με τα αποτελέσματα του μοντέλου πρόβλεψης. Με αυτόν τον τρόπο θα ελεγχθεί η εγκυρότητα του συγκεκριμένου μοντέλου.

Η παρούσα εργασία, για την δημιουργία του δυναμικού μοντέλου πρόβλεψης απόδοσης, ασχολείται με μια συγκεκριμένη υπηρεσία (ηλεκτρονική εγκυκλοπαίδεια Wikipedia.org). Η διαδικασία για την συλλογή και την επεξεργασία των δεδομένων, καθώς και η ερευνητική μεθοδολογία, που παρουσιάστηκαν στο παρόν κεφάλαιο, μπορούν να χρησιμοποιηθούν για την δημιουργία δυναμικών μοντέλων πρόβλεψης, σε εφαρμογές που πληρούν τα κριτήρια που παρουσιάστηκαν στην Ενότητα 3.3. Με αυτό τον τρόπο μπορούν να κατασκευαστούν, δυναμικά μοντέλα πρόβλεψης της απόδοσης για κάθε υπηρεσία και με τον ορισμό επιθυμητών επιπέδων απόδοσης, να σχεδιαστούν κατάλληλοι μηχανισμοί για προσφορά πόρων στην υπηρεσία, από τις υποδομές του υπολογιστικού νέφους. Η κατασκευή τέτοιων μηχανισμών δεν αποτελεί ερευνητικό τομέα της παρούσας εργασίας, αλλά με την χρήση των δυναμικών μοντέλων που θα προκύψουν, θα μπορούσε να αποτελέσει ερευνητική θεματική ενότητα μιας επόμενης μελέτης.

Κεφάλαιο 5: Συλλογή Εμπειρικών Δεδομένων

5.1: Εισαγωγή

Στο Κεφάλαιο 3 παρουσιάστηκε η μελέτη της ερευνητικής θεματικής περιοχής της παρούσας εργασίας, που σχετίζεται με την δημιουργία δυναμικού μοντέλου πρόβλεψης της απόδοσης μιας υπηρεσίας, βάσει των παραμέτρων κλιμάκωσής της. Για την διερεύνηση της ερευνητικής περιοχής, στο Κεφάλαιο 4, επιλέχθηκε η ερευνητική μεθοδολογία συλλογής εμπειρικών δεδομένων με χρήση της συσχετιστικής και της πειραματικής προσέγγισης. Η επιλογή της συνδυαστικής και της πειραματικής προσέγγισης, προέκυψε λόγω της συνδιακύμανσης που υπάρχει μεταξύ των μεταβλητών (σχέση των παραμέτρων που μελετώνται, με την απόδοση της υπηρεσίας), και του μερικού ελέγχου της παραμέτρου των παράλληλων αιτημάτων προς την υπηρεσία, χωρίς να επηρεάζεται η συμπεριφορά των υπόλοιπων παραμέτρων.

Έχοντας έτσι, ορίσει την ερευνητική μεθοδολογία, της οποίας γίνεται χρήση, στο παρόν κεφάλαιο θα παρουσιαστεί ο τρόπος συλλογής των εμπειρικών δεδομένων. Θα παρουσιαστεί λοιπόν, η ανάπτυξη της εφαρμογής που αναλαμβάνει την διαδικασία συλλογής των δεδομένων. Η εφαρμογή αυτή, αποτελείται από δύο τμήματα. Το πρώτο τμήμα, είναι υπεύθυνο για τον ορισμό των παραμέτρων εκτέλεσης της διαδικασίας (π.χ. υπηρεσία στην οποία θα εκτελεστούν τα αιτήματα, ορισμός μέγιστου αριθμού παράλληλων αιτημάτων, κλπ), και χειρίζεται από έναν κεντρικό υπολογιστή. Το δεύτερο τμήμα, είναι υπεύθυνο για την εκτέλεση των αιτημάτων προς την υπηρεσία και την αποστολή τους, όταν ολοκληρωθεί η διαδικασία, στο πρώτο τμήμα της εφαρμογής.

Επιπλέον, με τη χρήση της εφαρμογής που αναπτύχθηκε, πραγματοποιήθηκε και η διαδικασία συλλογής των εμπειρικών δεδομένων, για την παρούσα εργασία. Οι παράμετροι που λόγω της επιρροής τους στην απόδοσή της υπηρεσίας, συλλέχθηκαν τα δεδομένα, είναι: (i) ο αριθμός των παράλληλων αιτημάτων προς την υπηρεσία, (ii) η γεωγραφική απόσταση των υποδομών του υπολογιστικού νέφους, από τους υπολογιστές που πραγματοποιούν τα αιτήματα χρήσης στην υπηρεσία και, (iii) ο χρόνος εκτέλεσης των αιτημάτων προς την υπηρεσία.

Σκοπός του κεφαλαίου, είναι η παρουσίαση της διαδικασίας συλλογής εμπειρικών δεδομένων. Αρχικά, στην Ενότητα 5.2, αναλύεται η εφαρμογή που αναπτύχθηκε, σχετικά με τον τρόπο που συλλέγει τα δεδομένα, ώστε οι τιμές των παραμέτρων που καταγράφονται να είναι αντιπροσωπευτικές της συμπεριφοράς της υπηρεσίας. Στην Ενότητα 5.3 με την χρήση της εφαρμογής αυτής, παρουσιάζεται η διαδικασία που ακολουθήθηκε για την συλλογή των εμπειρικών δεδομένων για την υπηρεσία Wikipedia.org. Τέλος, στην Ενότητα 5.4, αναφέρονται τα συμπεράσματα της διαδικασίας, με την χρήση της εφαρμογής.

5.2: Παρουσίαση Εφαρμογής Συλλογής Δεδομένων

5.2.1 Δομή Εφαρμογής

Για την συλλογή των δεδομένων, αναπτύχθηκε μια εφαρμογή που θα εκτελεί την διαδικασία. Για την ανάπτυξη της έγινε χρήση της γλώσσας προγραμματισμού Java, και του εργαλείου Net Beans IDE 7.4 [56]. Όπως έχει αναφερθεί στο Κεφάλαιο 3, οι παράμετροι που επηρεάζουν την απόδοση της υπηρεσίας που μελετάται, είναι ο αριθμός των παράλληλων αιτημάτων προς την υπηρεσία, η χιλιομετρική απόσταση των υπολογιστών που πραγματοποιούν τα αιτήματα, από τις υποδομές του υπολογιστικού νέφους και ο χρόνος κατά τον οποίο εκτελούνται τα αιτήματα.

Παράμετρος Εκτέλεσης Παράλληλων Αιτημάτων προς την Υπηρεσία

Η εφαρμογή, έχει ως σκοπό πέρα από την συλλογή των δεδομένων για τις παραμέτρους που ορίστηκαν, και των δεδομένα για την απόδοση της υπηρεσίας, τον μερικό έλεγχο των παράλληλων αιτημάτων προς την υπηρεσία, για την πραγματοποίηση της διαδικασίας. Κατασκευάστηκε λοιπόν, μια εφαρμογή που δημιουργεί μεγάλο αριθμό ταυτόχρονων αιτημάτων, προσομοιώνοντας τους χρήστες που αιτούνται την εμφάνιση σελίδων, από την υπηρεσία. Για να μην επηρεάζει την δειγματοληψία, η ταχύτητα λήψης των δεδομένων της σελίδας, που μπορεί να οφείλεται στο εύρος ζώνης του δικτύου του κάθε υπολογιστή με το διαδίκτυο, ή την επεξεργαστική ισχύ του υπολογιστή για την εμφάνιση των δεδομένων, η εφαρμογή κάνει αίτηση για εμφάνιση της σελίδας και μόλις δεχθεί τον κωδικό απόκρισης τότε τερματίζει την αίτηση. Ο λόγος που έγινε αυτή η προσέγγιση είναι επειδή δεν εξετάζονται παράμετροι του υλικού του υπολογιστή που πραγματοποιεί τα αιτήματα. Έτσι, η εφαρμογή δεν απαιτεί μεγάλη υπολογιστική ισχύ, από τον υπολογιστή που εκτελείται η διαδικασία συλλογής των δεδομένων, για την δημιουργία και εκτέλεση της κάθε αίτησης.

Με τον όρο «κωδικός απόκρισης», εννοείται ο κωδικός κατάστασης (status code) του πρωτοκόλλου HTTP που απαντάει ο εξυπηρετητής της ιστοσελίδας όταν του γίνεται αίτηση για την εμφάνιση κάποιας σελίδας του ιστοτόπου [57]. Για παράδειγμα όταν ο κωδικός απόκρισης είναι 200, τότε το αίτημα έχει πετύχει και οι πληροφορίες της αίτησης μπορούν να επιστραφούν. Αν ο κωδικός είναι 400 τότε το αίτημα δεν μπόρεσε να γίνει αντιληπτό από τον εξυπηρετητή και δεν επιστρέφεται κάποια επιπλέον πληροφορία. Η εφαρμογή λοιπόν που αναπτύχθηκε εξετάζει και τον κωδικό απόκρισης του πρωτοκόλλου HTTP και αν είναι το 200 τότε αποθηκεύει το χρόνο απόκρισης, αλλιώς σε κάθε άλλη περίπτωση αποθηκεύει την αίτηση ως εσφαλμένη για να εξετασθεί.

Παράμετρος Απόστασης Προέλευσης Αιτημάτων από τις Υποδομές της Υπηρεσίας

Για την μελέτη της παραμέτρου της χιλιομετρικής απόστασης του υπολογιστή που κάνει την αίτηση για εμφάνιση της σελίδας, μέχρι τις υποδομές του υπολογιστικού νέφους που είναι εγκατεστημένη η υπηρεσία, θα πρέπει να βρεθεί τρόπος να συλλέγονται τα δεδομένα από διαφορετικές αποστάσεις. Για να γίνει σωστά η καταγραφή με βάση το προηγούμενο, θα πρέπει η εφαρμογή να εκτελείται παράλληλα σε υπολογιστές που είναι σε διαφορετικές αποστάσεις από τις υποδομές. Για τον λόγο αυτό η εφαρμογή αναπτύχθηκε σε δύο τμήματα. Για τα δύο αυτά τμήματα, δημιουργήθηκαν ξεχωριστά εκτελέσιμα αρχεία, όπου

μπορούν να θεωρηθούν και δύο ξεχωριστές εφαρμογές. Η πρώτη εφαρμογή δέχεται από τον χρήστη τις παραμέτρους εκτέλεσης της διαδικασίας, στέλνοντάς 'τες με το κατάλληλο ρυθμό στην άλλη που εκτελεί τις αιτήσεις προς την υπηρεσία. Έτσι, η πρώτη ορίζεται ως εφαρμογή εισαγωγής δεδομένων και διαχείρισης της όλης διαδικασίας συλλογής εμπειρικών δεδομένων, και η δεύτερη ως εφαρμογή εκτέλεσης των αιτημάτων προς την υπηρεσία.

Για την συλλογή των εμπειρικών δεδομένων, η πρώτη είναι εγκατεστημένη σε έναν υπολογιστή όπου γίνεται η διαχείριση της όλης διαδικασίας. Η δεύτερη είναι εγκατεστημένη στους υπολογιστές που είναι σε διαφορετικές αποστάσεις από τις υποδομές της υπηρεσίας, και δέχεται τις πληροφορίες (από την παραπάνω εφαρμογή) για την εκτέλεση των αιτημάτων προς την υπηρεσία. Μετά το τέλος της κάθε διαδικασίας και αφού έχει γίνει η συλλογή των δεδομένων από την δεύτερη εφαρμογή, όλα τα δεδομένα στέλνονται στην πρώτη, όπου τα συλλέγει και τα αποθηκεύει ξεχωριστά για τον κάθε υπολογιστή που εκτέλεσε τα αιτήματα, στην επιλεγμένη υπηρεσία.

Στην δεύτερη εφαρμογή το μόνο που χρειάζεται να ορισθεί είναι η πόρτα του TCP/IP πρωτοκόλλου [58], μέσω της οποίας θα περιμένει τον αριθμό των αιτήσεων που θα πρέπει να εκτελέσει. Επιπλέον έχει κουμπιά για την έναρξη και λήξη της διαδικασίας, αναμένοντας τις παραμέτρους εκτέλεσης. Αυτό σημαίνει ότι, κατά την έναρξη της δεύτερης εφαρμογής, παραμένει σε κατάσταση αναμονής, μέχρις ότου μέσω του πρωτοκόλλου TCP/IP, λάβει τα απαραίτητα δεδομένα για ξεκινήσει η διαδικασία εκτέλεσης των αιτημάτων προς την υπηρεσία.

Παράμετρος του Χρόνου Εκτέλεσης των Αιτημάτων προς την Υπηρεσία

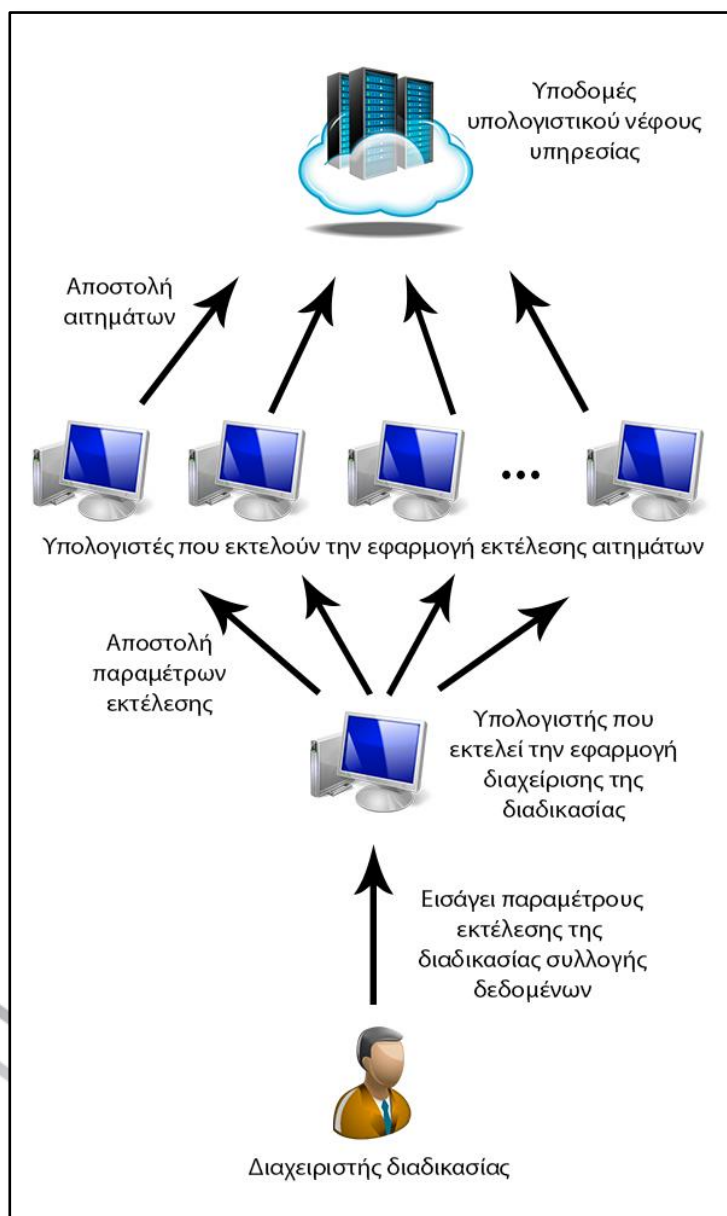
Λαμβάνοντας υπόψη, την παράμετρο του χρόνου εκτέλεσης των αιτημάτων προς την υπηρεσία, θα πρέπει να καταγράφονται οι χρονικές στιγμές κατά την οποίες εκτελούνται οι αιτήσεις. Για τον λόγο αυτό, στην καταγραφή των δεδομένων από την εφαρμογή που εκτελεί τις αιτήσεις, αποθηκεύονται και οι χρονικές στιγμές όπου εκτελούνται τα αιτήματα. Με τον τρόπο αυτό μπορεί να συγκριθεί η συμπεριφορά της υπηρεσίας, στις αιτήσεις για χρήση της, σε διαφορετικές χρονικές στιγμές.

Στις προηγούμενες παραγράφους ορίστηκε ο τρόπος με τον οποίο θα προσομοιώνονται, από την εφαρμογή, τα αιτήματα των χρηστών για χρήση της υπηρεσίας, από διαφορετικές γεωγραφικές θέσεις. Έτσι θα καταγράφεται η συμπεριφορά της υπηρεσίας, για την εξυπηρέτηση μεγάλου πλήθους αιτημάτων χρήσης της. Επιπλέον, αναφέρθηκε ο τρόπος με τον οποίο θα αποθηκεύονται τα δεδομένα για τις παραμέτρους της υπηρεσίας, που επηρεάζουν την απόδοσή της, έχοντας αντιπροσωπευτική μορφή αναπαράστασης της κατάστασής τους. Τα δεδομένα αρχικά συλλέγονται στην εφαρμογή που εκτελεί τα αιτήματα προς την υπηρεσία, και στην συνέχεια αποστέλλονται, μέσω του πρωτοκόλλου TCP/IP, στην εφαρμογή διαχείρισης της διαδικασίας.

Σχηματική Παρουσίαση Εκτέλεσης της Διαδικασίας Συλλογής των Δεδομένων

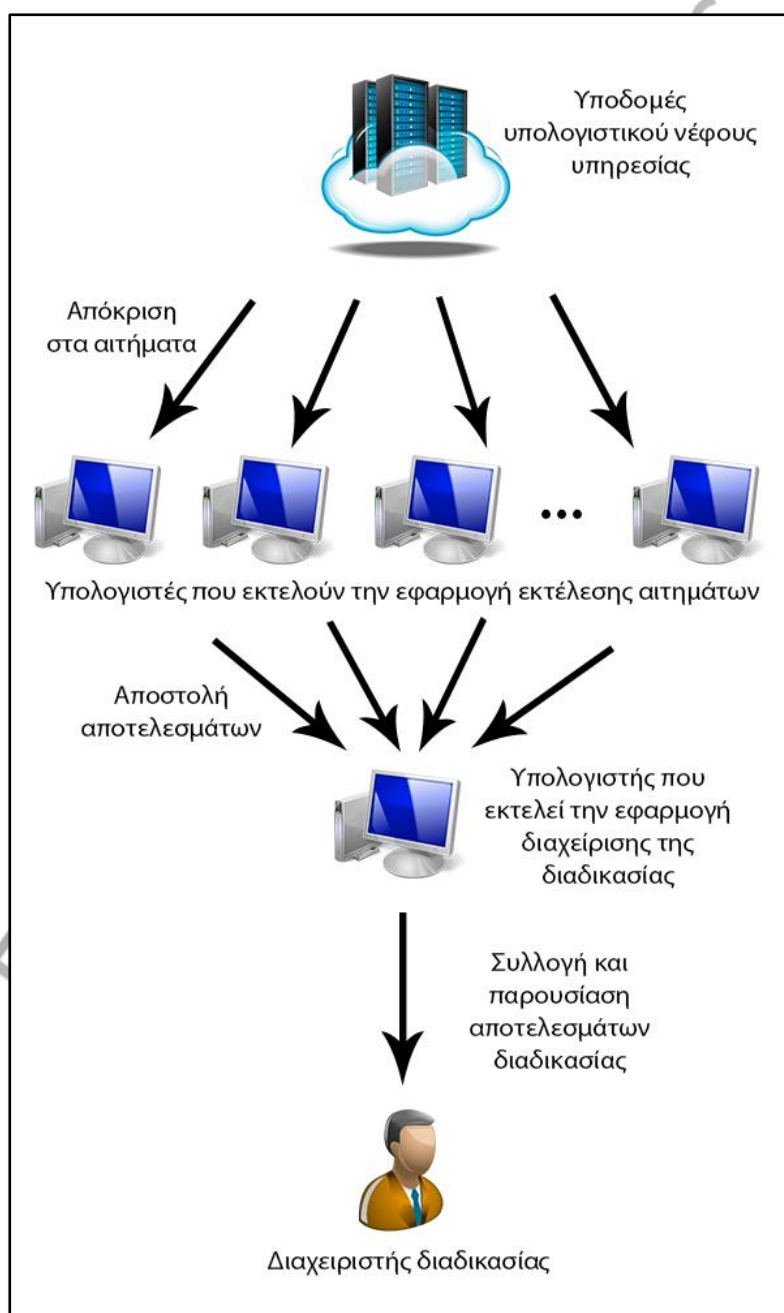
Για να γίνει πιο κατανοητή η δομή της εφαρμογής που αναπτύχθηκε θα παρουσιαστεί παρακάτω στα παρακάτω σχήματα. Αρχικά, στο Σχήμα 5.1, παρουσιάζεται η διαδικασία συλλογής δεδομένων, όπου ο διαχειριστής εισάγει τις παραμέτρους εκτέλεσης της

διαδικασίας, μέσω του υπολογιστή που είναι εγκατεστημένη η εφαρμογή που την διαχειρίζεται. Στην συνέχεια ξεκινάει η εκτέλεση της διαδικασίας, όπου η εφαρμογή στέλνει τις παραμέτρους εκτέλεσης, μέσω του πρωτοκόλλου TCP/IP, σε κάθε έναν από τους υπολογιστές που εκτελείται η εφαρμογή εκτέλεσης των αιτημάτων. Έτσι, ο κάθε υπολογιστής από αυτούς που εκτελούν την δεύτερη εφαρμογή (εφαρμογή εκτέλεσης των αιτημάτων), εκτελεί παράλληλα αιτήματα προς την υπηρεσία με βάση τις παραμέτρους που έχει δεχθεί. Τα αιτήματα εξυπηρετούνται από τις υποδομές του υπολογιστικού νέφους της υπηρεσίας.



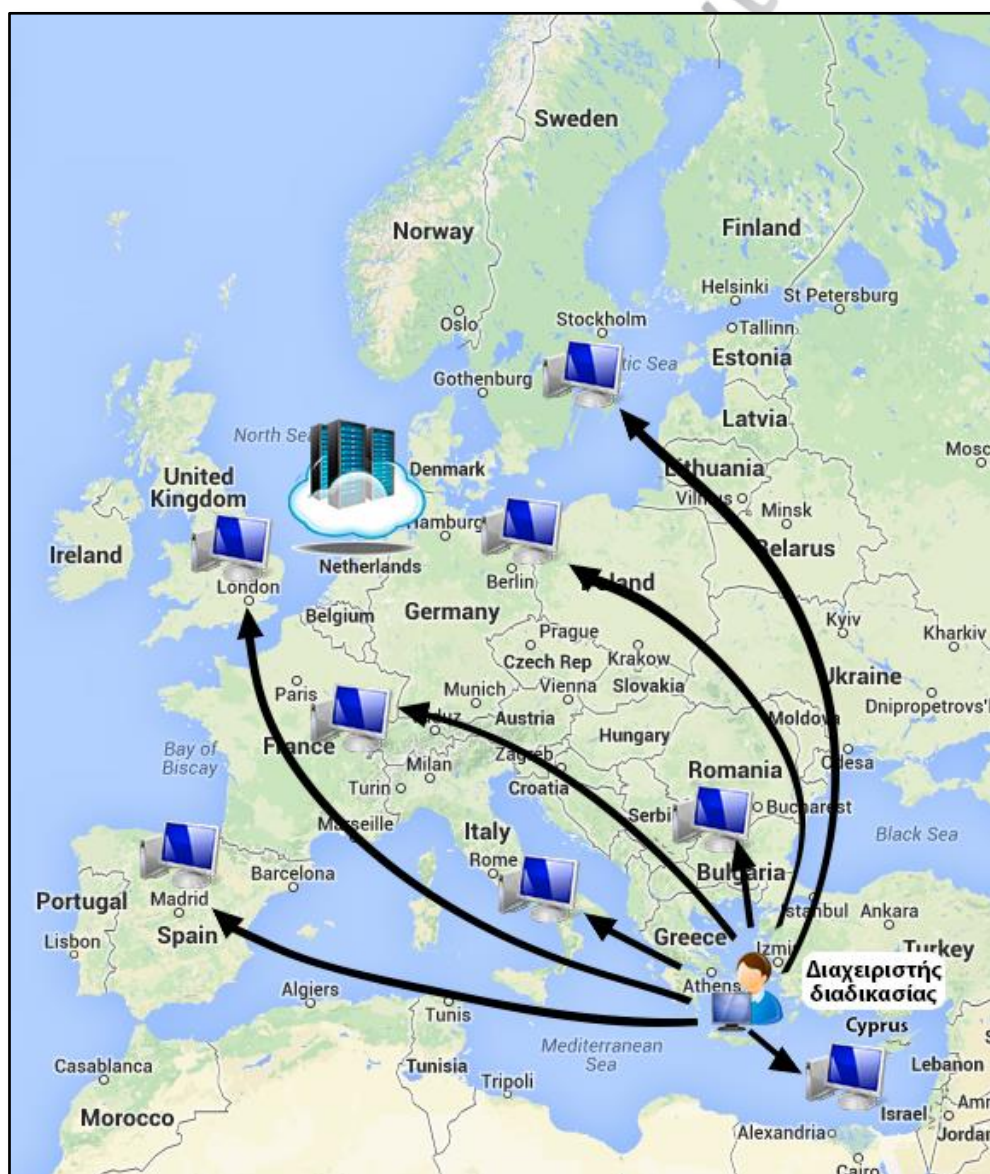
Σχήμα 5.1: Διαδικασία αποστολής παραμέτρων και αιτημάτων

Στο Σχήμα 5.2 παρουσιάζεται η συνέχεια της διαδικασίας, όπου ο εξυπηρετητής της υπηρεσίας, όπου βρίσκεται εγκατεστημένος στις υποδομές του υπολογιστικού νέφους, αποστέλλει τον κωδικό απόκρισης των αιτημάτων. Η εφαρμογή εκτέλεσης των αιτημάτων, που είναι εγκατεστημένη στους υπολογιστές της διαδικασίας, αφού λάβει την απόκριση της αίτησης από τον εξυπηρετητή της υπηρεσίας, συλλέγει και αποθηκεύει τα απαραίτητα δεδομένα (π.χ. χρόνος απόκρισης της υπηρεσίας, αριθμός των παράλληλων αιτημάτων που εκτελούνται). Στην συνέχεια όλα τα δεδομένα, αφού συλλεχθούν, αποστέλλονται στον υπολογιστή που εκτελείται η εφαρμογή διαχείρισης της διαδικασίας. Η εφαρμογή διαχείρισης, συλλέγει τα δεδομένα και τα αποθηκεύει ξεχωριστά για τον κάθε υπολογιστή που εκτέλεσε τα αιτήματα προς την υπηρεσία.



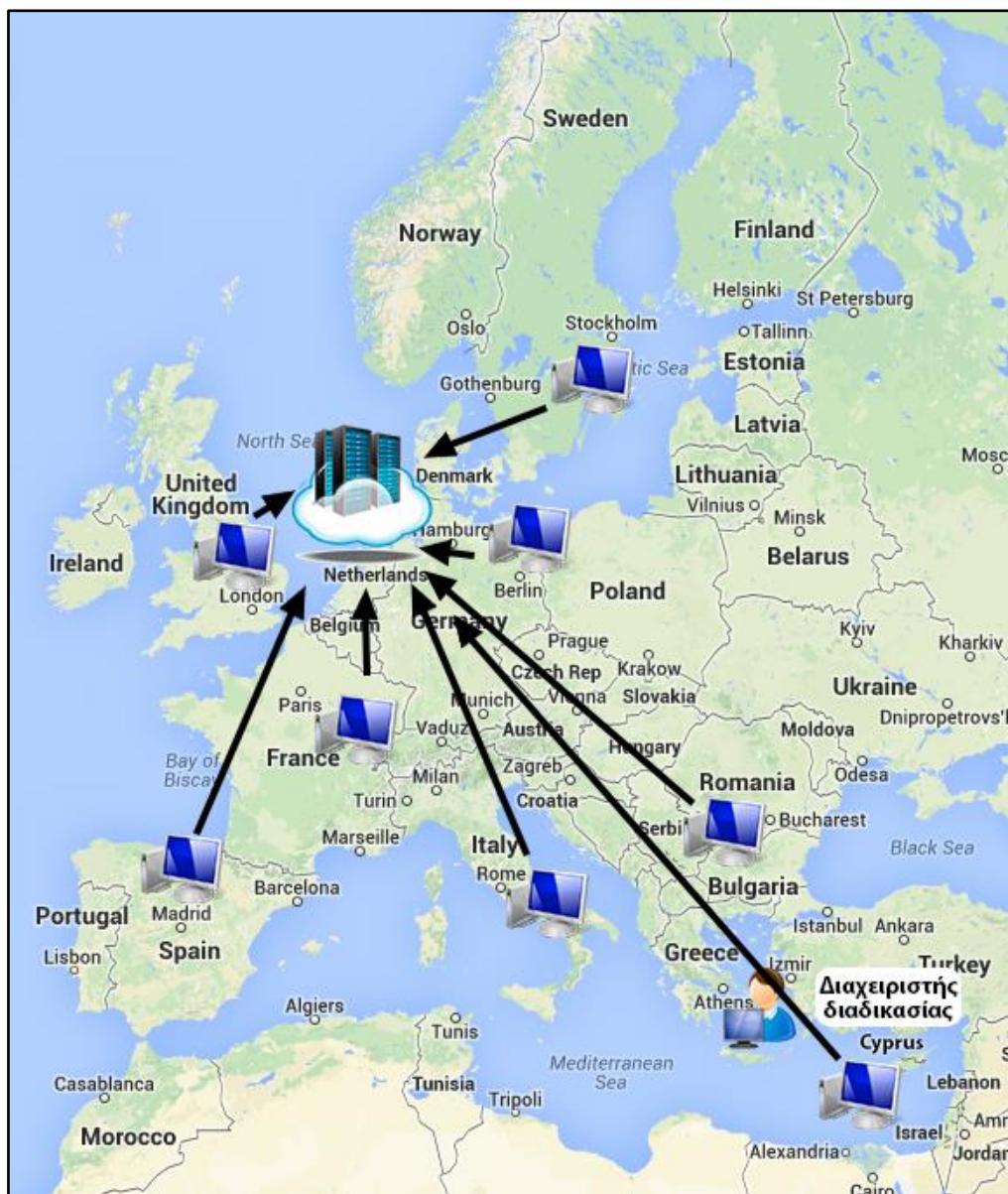
Σχήμα 5.2: Διαδικασία απόκρισης των αιτημάτων και συλλογής δεδομένων

Παραπάνω, παρουσιάστηκε η διαδικασία συλλογής των εμπειρικών δεδομένων, με την χρήση των εφαρμογών που αναπτύχθηκαν. Η αποστολή των αιτημάτων προς την υπηρεσία, πραγματοποιείται από την εφαρμογή εκτέλεσης των αιτημάτων, η οποία είναι εγκατεστημένη σε υπολογιστές σε διαφορετικές γεωγραφικές θέσεις. Ο λόγος που γίνεται αυτό, είναι επειδή μια από τις παραμέτρους που επηρεάζουν την απόδοση της υπηρεσίας, και μελετάται από την παρούσα εργασία, είναι η απόσταση των υπολογιστών που εκτελούν τα αιτήματα μέχρι τις υποδομές του υπολογιστικού νέφους, όπου βρίσκεται ο εξυπηρετητής της υπηρεσίας, που θα τα διαχειριστεί. Στα επόμενα σχήματα (Σχήμα 5.3, 5.4, 5.5, 5.6), γίνεται αναπαράσταση της γεωγραφικής διασποράς των υπολογιστών, όπου εκτελούν τα αιτήματα προς την υπηρεσία. Αρχικά στο Σχήμα 5.3, παρουσιάζεται η διαδικασία αποστολής των παραμέτρων εκτέλεσης, από την εφαρμογή διαχείρισης διαδικασίας, προς τους υπολογιστές που είναι εγκατεστημένη η εφαρμογή εκτέλεσης των αιτημάτων.



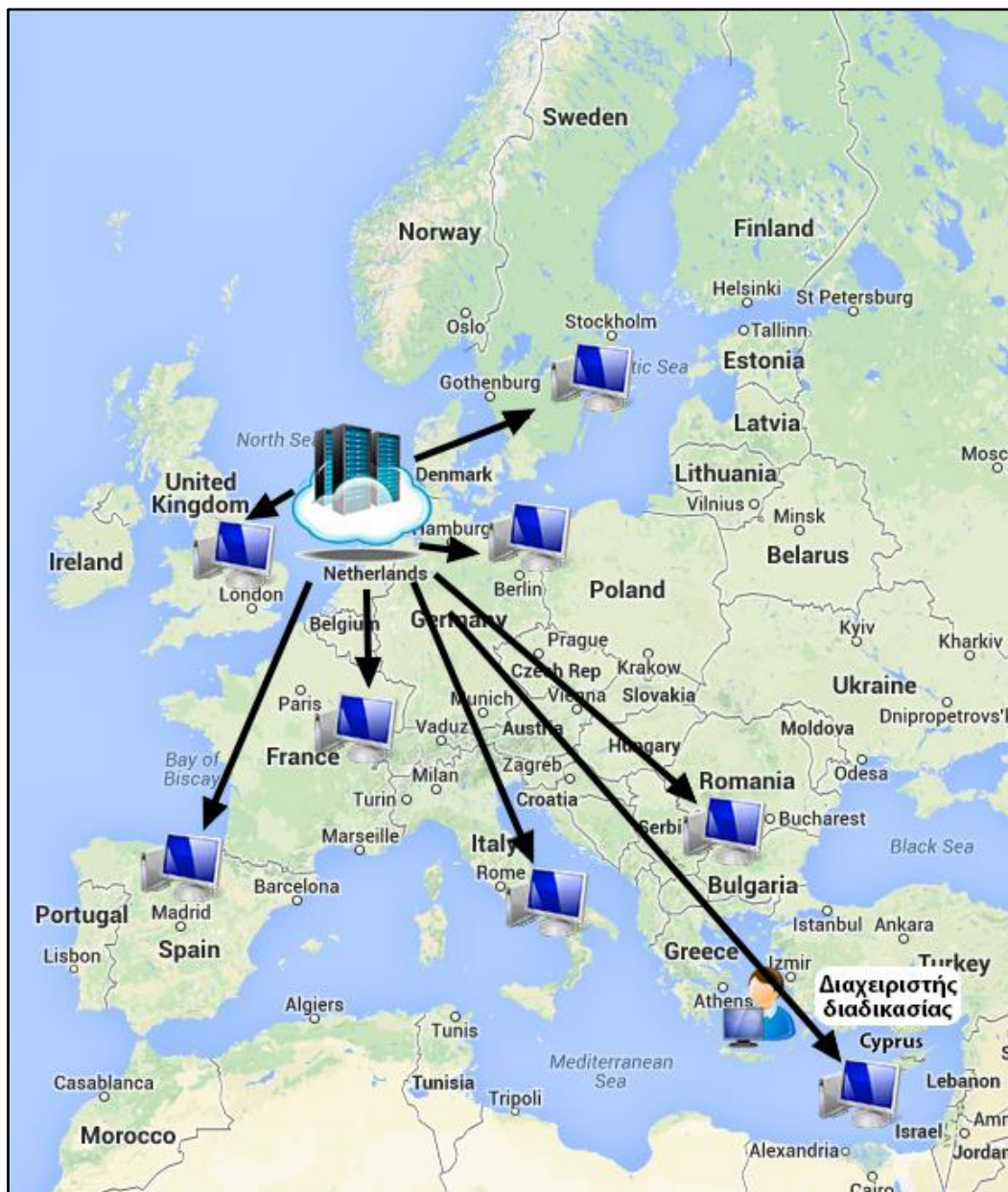
Σχήμα 5.3: Διαδικασία αποστολής παραμέτρων σε γεωγραφική απεικόνιση

Οι υπολογιστές που έχουν εγκατεστημένη την εφαρμογή εκτέλεσης των αιτημάτων, λαμβάνουν τις παραμέτρους εκτέλεσης, μέσω του πρωτοκόλλου TCP/IP, από την εφαρμογή διαχείρισης της διαδικασίας. Στην συνέχεια, ξεκινάει η αποστολή των παράλληλων αιτημάτων προς την υπηρεσία. Στο Σχήμα 5.4 παρουσιάζεται η αποστολή των αιτημάτων προς την υπηρεσία.



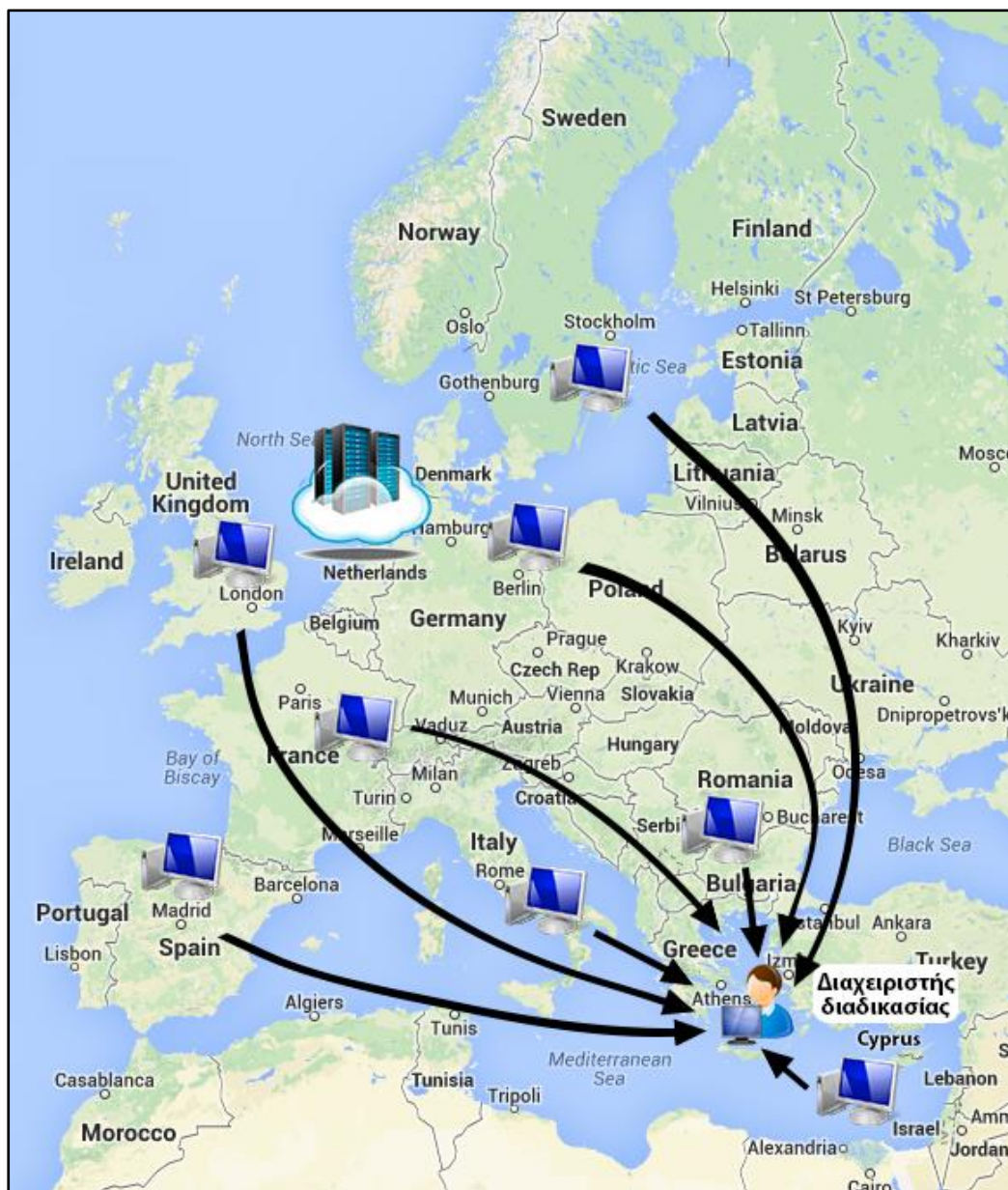
Σχήμα 5.4: Διαδικασία αποστολής παράλληλων αιτημάτων προς την υπηρεσία, σε γεωγραφική απεικόνιση

Λαμβάνοντας τα αιτήματα ο εξυπηρετητής της υπηρεσίας, τα επεξεργάζεται και απαντάει στους υπολογιστές που τα πραγματοποίησαν. Οι υπολογιστές που έχουν εγκατεστημένη την εφαρμογή εκτέλεσης των αιτημάτων, συλλέγουν τις πληροφορίες απόκρισης, μαζί με τις παραμέτρους εκτέλεσης της διαδικασίας (π.χ. αριθμός παράλληλων αιτημάτων προς την υπηρεσία).



Σχήμα 5.5: Διαδικασία απόκρισης στα αιτήματα, σε γεωγραφική απεικόνιση

Όταν ολοκληρώνεται η κάθε διαδικασία αποστολής και απόκριση των αιτημάτων, τα δεδομένα που έχουν συλλεχθεί, αποστέλλονται στην εφαρμογή διαχείρισης της διαδικασίας. Τα δεδομένα αυτά αποθηκεύονται ξεχωριστά για κάθε υπολογιστή που έχει εκτελέσει τα αιτήματα, και έχει την δυνατότητα ο διαχειριστής να τα προβάλλει.



Σχήμα 5.6: Διαδικασία αποστολής των δεδομένων προς την εφαρμογή διαχείρισης, σε γεωγραφική απεικόνιση

5.2.2 Ορίσματα Εκτέλεσης της Εφαρμογής

Η εφαρμογή που αναπτύχθηκε για την συλλογή των εμπειρικών δεδομένων, χωρίζεται σε δύο τμήματα, που μπορούν να θεωρηθούν ξεχωριστές εφαρμογές. Η πρώτη εφαρμογή δέχεται από τον χρήστη ορίσματα για την εκτέλεση της διαδικασίας. Τα ορίσματα αυτά αποστέλλονται μέσω του πρωτοκόλλου TCP/IP στην δεύτερη εφαρμογή, όπου βάσει αυτών, θα εκτελέσει τα αιτήματα, για την συλλογή των δεδομένων. Παρακάτω παρουσιάζονται τα ορίσματα που θέτει ο χρήστης, στην εφαρμογή διαχείρισης της διαδικασίας, για την εκτέλεση της:

- Οι δημόσιες διευθύνσεις IP των υπολογιστών που εκτελούν την εφαρμογή εκτέλεσης των αιτημάτων προς την υπηρεσία.
- Η πόρτα του πρωτοκόλλου TCP, μέσω της οποίας επικοινωνούν οι εφαρμογές της διαδικασίας.
- Η διεύθυνση της υπηρεσίας στην οποία θα εκτελεστούν τα αιτήματα (αφού πρόκειται για ιστοσελίδα θα είναι της μορφής “http://www.wikipedia.org”).
- Ο αριθμός των αιτημάτων προς την υπηρεσία, καθώς και ο ρυθμός εκτέλεσης τους.

Παραδείγματα παραμέτρου:

- **5,10,20,35,50,75,100**: Αυτό σημαίνει ότι θα αποσταλεί πρώτα στους υπολογιστές που εκτελούν τα αιτήματα ο αριθμός 5, και στην συνέχεια ο κάθε υπολογιστής θα εκτελέσει παράλληλα 5 αιτήματα. Όταν συλλεχθούν τα αποτελέσματα για τα 5 αιτήματα από όλους του υπολογιστές, θα αποσταλεί ο αριθμός, 10 όπου μόλις τα λάβει, ο κάθε υπολογιστής θα εκτελέσει παράλληλα 10 αιτήματα προς την υπηρεσία κ.ο.κ. μέχρι τα 100 παράλληλα αιτήματα.
- **5-100**: Αυτό σημαίνει ότι θα αποσταλεί αρχικά ο αριθμός 5 στους υπολογιστές που εκτελούν τα αιτήματα, στην συνέχεια ο 10, μετά ο 15, μέχρι το 100 με βήμα όσο είναι ο πρώτος αριθμός (δηλαδή 5,10,15,20,25,30,...,85,90,95,100)
- **10-100-5**: Αυτό σημαίνει ότι θα αποστέλλονται στους υπολογιστές που εκτελούν τα αιτήματα, οι αριθμοί 10,20,30 μέχρι το 100 και στην συνέχεια θα συνεχίζονται να στέλνονται οι αριθμοί 100,95,90,85,...,20,15,10,5. Δηλαδή θα αποσταλούν οι αριθμοί από το 10 μέχρι το 100 με βήμα 10 (όσο ο πρώτος αριθμός) και στην συνέχεια από το 100 μέχρι το 5 με βήμα 5 (όσο ο τελευταίος αριθμός).
- **1~250#100**: Αυτό σημαίνει ότι θα αποσταλούν 100 τυχαίοι αριθμοί από το 1 μέχρι το 250. Η εφαρμογή δηλαδή θα επιλέγει τυχαία έναν αριθμό κάθε φορά που θα τον αποστέλλει στους υπολογιστές που εκτελούν τα αιτήματα. Άρα για παράδειγμα μπορεί να αποσταλούν: 44,8,173,63,202,...,90, 245,11,58,122.
- **10-200-5, 80, 100, 200, 50~250#70**: Αυτό αποτελεί συνδυασμό όλων των προηγούμενων. Δηλαδή θα αποσταλούν οι αριθμοί από το 10 μέχρι το 200 με βήμα 10, μετά από το 200 μέχρι το 5 με βήμα 5, μετά ο αριθμός 80, μετά ο 100, μετά ο 200 και τέλος και τέλος 70 τυχαίοι αριθμοί από το 50 μέχρι το 250.

5.3: Διαδικασία Συλλογής Δεδομένων

Όπως αναφέρθηκε στην Ενότητα 5.2, για την συλλογή των δεδομένων αναπτύχθηκαν δύο εφαρμογές. Οι εφαρμογές αυτές χρησιμοποιήθηκαν από την παρούσα εργασία, ώστε να συλλεχθούν τα δεδομένα για την υπηρεσία ηλεκτρονικής εγκυκλοπαίδειας Wikipedia.org. Στην Ενότητα 5.3, παρουσιάζεται η προετοιμασία και ο ορισμός των παραμέτρων για την έναρξη εκτέλεσης της διαδικασίας. Επιπλέον προβάλλεται ένα δείγμα των δεδομένων που συλλέχθηκαν από την υπηρεσία Wikipedia.org.

5.3.1 Ορισμός Παραμέτρων Εκτέλεσης της Διαδικασίας Συλλογής Δεδομένων

Για την συλλογή των δεδομένων απόκρισης της υπηρεσίας από υπολογιστές σε διαφορετικές χιλιομετρικές αποστάσεις από τις υποδομές υπολογιστικού νέφους της υπηρεσίας, η δεύτερη εφαρμογή (εφαρμογή εκτέλεσης αιτημάτων), εγκαταστάθηκε σε υπολογιστές από τρεις διαφορετικές πόλεις. Οι εγκαταστάσεις της υποδομής του υπολογιστικού νέφους της υπηρεσίας ηλεκτρονικής εγκυκλοπαίδειας Wikipedia.org, για αιτήματα που προέρχονται από την Ευρώπη, βρίσκονται στην πόλη του Άμστερνταμ της Ολλανδίας [59]. Οι πόλεις, λοιπόν, που επιλέχθηκαν και εγκαταστάθηκαν στους υπολογιστές τους η εφαρμογή ήταν:

- Αθήνα, 2166 χιλιόμετρα από τις υποδομές.
- Ρώμη, 1300 χιλιόμετρα από τις υποδομές,
- Άμστερνταμ, στην ίδια πόλη που είναι οι υποδομές.

Για τον προσδιορισμό των αριθμών των παράλληλων αιτημάτων από τους υπολογιστές, έγιναν διάφορες προσεγγίσεις. Αρχικά ορίστηκε ένα σύνολο από προκαθορισμένους αριθμούς. Δηλαδή από τον αριθμό 2, που αντιστοιχεί σε 2 παράλληλα αιτήματα από τον κάθε υπολογιστή προς την υπηρεσία, μέχρι τον αριθμό 100 με βήμα 2 (2,4,6,8,...,96,98,100). Για την αύξηση των αριθμών των παράλληλων αιτημάτων, εκτός από το προηγούμενο σύνολο προκαθορισμένων αριθμών, δοκιμάστηκαν και οι αριθμοί από το 3 μέχρι το 150 με βήμα 3 (3,6,9,...,144,147,150), από 4 μέχρι το 200 με βήμα 4 (4,8,12,...,242,246,250) και από 5 μέχρι το 250 με βήμα 5 (5,10,15,...,240,245,250).

Στην συνέχεια για να αποδειχθεί η μη επιρροή της συλλογής των δεδομένων από επιπλέον παραμέτρους, εκτός από αυτές που είχαν ορισθεί, δοκιμάστηκε η επιλογή μεγαλύτερων συνόλων αριθμών. Εκτός από τους προκαθορισμένους αριθμούς από 2 μέχρι 100 με βήμα 2, 3 μέχρι 150 με βήμα 3, 4 μέχρι 200 με βήμα 4, 5 μέχρι 250 με βήμα 5, προστέθηκαν οι ίδιοι ξανά αριθμοί αλλά με ανάποδο βήμα μετά την εκτέλεση του πρώτου συνόλου. Επομένως υπήρχαν τα σύνολα από 2 μέχρι 100 με βήμα 2 και στην συνέχεια από το 100 μέχρι το 2 με βήμα -2 (2,4,6,8,...,96,98,100,100,98,96,...,8,6,4,2). Ακόλουθα και για τα άλλα σύνολα από 3 μέχρι 150 με βήμα 3 και από 150 μέχρι 3 με βήμα -3 κ.ο.κ. Έτσι αν υπάρχουν παρόμοιες τιμές του χρόνου απόκρισης για τους ίδιους αριθμούς παράλληλων αιτημάτων, τότε δεν επηρεάζονται τα αποτελέσματα από άλλες παραμέτρους. Πράγματι, όπως θα παρουσιαστεί και στο επόμενο κεφάλαιο με την ανάλυση των δεδομένων, οι χρόνοι

απόκρισης της υπηρεσίας για τους ίδιους αριθμούς παράλληλων αιτημάτων, ήταν σχεδόν ίδιοι.

Πραγματοποιώντας όμως την διαδικασία συλλογής των δεδομένων, με τον τρόπο που αναφέρθηκε στην προηγούμενη παράγραφο, δημιουργήθηκε ένα βασικό πρόβλημα. Οι παράμετροι που επηρεάζουν την απόδοση της υπηρεσίας θα πρέπει να είναι ανεξάρτητες μεταξύ τους. Αυτό σημαίνει ότι δεν θα μπορεί να προβλεφθεί η τιμή της μιας παραμέτρου, με βάση την τιμή κάποια άλλης. Στην προκειμένη όμως περίπτωση, γνωρίζοντας τον αριθμό των παράλληλων αιτημάτων, που εκτελέστηκε από τους υπολογιστές, μπορεί να υπολογιστεί για κάθε χρονική στιγμή (η τρίτη παράμετρος που έχει οριστεί) ποιος είναι ο αριθμός των παράλληλων αιτημάτων που εκτελέστηκε. Για να αποφευχθεί κάτι τέτοιο, και να υπάρχουν τρεις ανεξάρτητες παράμετροι, θα πρέπει ο αριθμός των παράλληλων αιτημάτων να είναι τυχαίος κάθε χρονική στιγμή. Για τον λόγο αυτό, στον προσδιορισμό των παραμέτρων της διαδικασίας εκτέλεσης στην εφαρμογή διαχείρισης, ο χρήστης ορίζει μόνο τα όρια για τους τυχαίους αριθμούς που θα παράγει η εφαρμογή.

Συνεπώς, έχοντας τυχαίο πλήθος παράλληλων αιτημάτων που θα εκτελούνται κάθε χρονική στιγμή προς την υπηρεσία, και έχοντας εγκαταστήσει την εφαρμογή εκτέλεσης των αιτημάτων, σε υπολογιστές από τουλάχιστον τρεις διαφορετικές πόλεις της Ευρώπης, μπορεί να εκτελεστεί η διαδικασία συλλογής των δεδομένων. Ο λόγος που στην διαδικασία συλλογής των δεδομένων, θα πρέπει να περιλαμβάνονται τουλάχιστον τρεις υπολογιστές από διαφορετικές γεωγραφικές θέσεις, είναι η δημιουργία αντικειμενικών δεδομένων της παραμέτρου της απόστασης των αιτημάτων από τις υποδομές της υπηρεσίας.

5.3.2 Παρουσίαση Δείγματος Δεδομένων που Συλλέχθηκαν

Στην ενότητα αυτή παρουσιάζεται ένα δείγμα των δεδομένων που συλλέχθηκαν. Η αποθήκευση των δεδομένων γίνεται σε αρχεία κειμένου. Η κάθε εγγραφή των αρχείων αυτών, περιέχει τις ακόλουθες πληροφορίες:

- **Αριθμός δειγματοληψίας:** Η κάθε λήψη παραμέτρων, αποτελεί και μια δειγματοληψία που πρόκειται να εκτελεστεί. Εκτός από την διεύθυνση της υπηρεσίας που θα αποσταλούν τα αιτήματα, ορίζεται και ο αριθμός των παράλληλων αιτημάτων που θα εκτελεστούν.
- **Πλήθος παράλληλων αιτημάτων:** Είναι ο αριθμός των παράλληλων αιτημάτων που εκτελούσε ο υπολογιστής, την χρονική στιγμή που καταγράφηκαν τα δεδομένα για το κάθε αίτημα.
- **Χρόνος απόκρισης:** Είναι ο χρόνος απόκρισης για το συγκεκριμένο αίτημα. Η μονάδα μέτρησης του χρόνου αυτού είναι χιλιοστά του δευτερολέπτου (milliseconds).
- **Σφάλμα αιτήματος:** Αποτελεί το αναγνωριστικό της εγγραφής, για το αν το αίτημα είχε την αναμενόμενη απόκριση ή όχι. Οι τιμές του είναι 0 σε περίπτωση που είχε την αναμενόμενη απόκριση και 1 σε περίπτωση που δεν είχε. Έτσι, σε περίπτωση που ο κωδικός απόκρισης του πρωτοκόλλου HTTP είναι 200, τότε η εφαρμογή δέχθηκε την αναμενόμενη απόκριση και η τιμή του πεδίου είναι 0. Σε περίπτωση σφάλματος, για παράδειγμα, ο χρόνος αναμονής για την εξυπηρέτηση του

αιτήματος ολοκληρώθηκε χωρίς να δεχτεί κάποιο αίτημα, ή το αίτημα δεν μπορούσε να εξυπηρετηθεί για κάποιον άλλον λόγο, τότε η τιμή του πεδίου είναι 1.

- **Αριθμός αιτήματος:** Αποτελεί τον αριθμό του κάθε αιτήματος που εκτελέστηκε. Σε περίπτωση εκτέλεσης 100 παράλληλων αιτημάτων, τότε το κάθε αίτημα θα έχει έναν αριθμό από το 0 μέχρι το 99.
- **Ημερομηνία και ώρα εκτέλεσης:** Αποτελεί την χρονική στιγμή που εκτελέστηκε το κάθε αίτημα.

Ένα δείγμα μιας διαδικασίας συλλογής δεδομένων, από ένα υπολογιστή που είχε εγκατεστημένη την εφαρμογή εκτέλεσης αιτημάτων είναι:

4,131,1062,0,9,13/02/2014,11:55:34
4,131,1078,0,64,13/02/2014,11:55:34
4,131,1078,0,96,13/02/2014,11:55:34
4,131,1094,0,72,13/02/2014,11:55:34
4,131,1094,0,83,13/02/2014,11:55:34
4,131,1172,0,35,13/02/2014,11:55:34
4,131,1265,0,17,13/02/2014,11:55:34
4,131,1328,0,32,13/02/2014,11:55:34
4,131,1359,0,15,13/02/2014,11:55:34

Στο συγκεκριμένο δείγμα, η τιμή 4 στην αρχή, αποτελεί τον αριθμό της δειγματοληψίας. Το συγκεκριμένο λοιπόν, αποτελεί την τέταρτη αποστολή παραμέτρων για εκτέλεση. Η επόμενη τιμή, που είναι το 131, αποτελεί το αριθμό των παράλληλων αιτημάτων που εκτελούσε η εφαρμογή εκείνη την χρονική στιγμή, για τις συγκεκριμένες εγγραφές. Άρα η τέταρτη αποστολή παραμέτρων για εκτέλεση είχε σαν αριθμό δημιουργίας παράλληλων αιτημάτων το 131. Η επόμενη τιμή, 1062, 1078,..., 1359, αποτελεί τον χρόνο απόκρισης για το κάθε αίτημα, σε χιλιοστά του δευτερολέπτου. Στη συνέχεια, η τιμή που στο συγκεκριμένο δείγμα είναι μηδέν, σημαίνει ότι δεν παρουσιάστηκε κάποιο μην αναμενόμενο αποτέλεσμα στην απόκριση της υπηρεσίας για το κάθε αίτημα. Η επόμενη τιμή (9,64,96,...,32,15) αποτελεί τον αριθμό του κάθε αιτήματος. Τέλος, οι τελευταίες δύο τιμές είναι η ημερομηνία και ώρα εκτέλεσης της διαδικασίας συλλογής των δεδομένων για τα συγκεκριμένα αιτήματα.

5.4: Συμπεράσματα

Στο Κεφάλαιο 5, παρουσιάστηκε η διαδικασία συλλογής των εμπειρικών δεδομένων. Στα πλαίσια εκτέλεσης της διαδικασίας αυτής, αναπτύχθηκε εφαρμογή όπου προσομοιώνει τις αιτήσεις των χρηστών για χρήση της υπηρεσίας. Σκοπός είναι η καταγραφή της συμπεριφοράς της υπηρεσίας που μελετάται, κατά την διάρκεια εξυπηρέτησης των παράλληλων αιτημάτων που δέχεται. Έτσι η εφαρμογή, πέρα από την προσομοίωση των αιτημάτων για χρήση της υπηρεσίας, καταγράφει και αποθηκεύει τις τιμές για την απόδοση της υπηρεσίας, καθώς και τις παραμέτρους που την επηρεάζουν.

Πράγματι η εφαρμογή αυτή, συλλέγει όλες τις απαραίτητες πληροφορίες για την συμπεριφορά της απόδοσης της υπηρεσίας, και των παραμέτρων που ορίστηκαν (π.χ.

αριθμός παράλληλων αιτημάτων κάθε χρονική στιγμή). Από τα δεδομένα που συλλέχθηκαν, μπορούν να εξαχθούν κάποια χρήσιμα συμπεράσματα, για την μορφή του υποδείγματος της σχέσης των παραμέτρων, ξεχωριστά, με την απόδοση της υπηρεσίας (παρουσίαση τους στο Κεφάλαιο 6). Για την εξαγωγή των συμπερασμάτων αυτών, αλλά και για την προετοιμασία τους ώστε να αναλυθούν από το εργαλείο Eureka, που θα παραχθούν μοντέλα πρόβλεψης, θα πρέπει πρώτα να επεξεργαστούν. Στο Κεφάλαιο 6, γίνεται παρουσίαση της επεξεργασίας που πραγματοποιήθηκε πάνω στα δεδομένα που συλλέχθηκαν από την εφαρμογή.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 6: Επεξεργασία Δεδομένων

6.1: Εισαγωγή

Στο Κεφάλαιο 5 παρουσιάστηκε η διαδικασία συλλογής των δεδομένων, με την ανάπτυξη εφαρμογής. Τα δεδομένα αυτά αφού συλλεχθούν θα πρέπει να επεξεργαστούν, πριν την τελική τους ανάλυση. Ο λόγος είναι για να εξαχθούν κάποια χρήσιμα συμπεράσματα, σχετικά την διαδικασία συλλογής των δεδομένων, και για την μορφή του υποδείγματος που προκύπτει από την αναπαράσταση των παραμέτρων σε διαγράμματα δύο διαστάσεων. Βάσει της ανάλυσης του υποδείγματος, και με την χρήση μαθηματικών εννοιών, στο Κεφάλαιο 7, παρουσιάζεται η καταλληλότερη προσέγγιση για την δημιουργία των μοντέλων πρόβλεψης.

Επιπλέον, τα δεδομένα που έχουν συλλεχθεί πριν την ανάλυση τους από το εργαλείο Eureka, θα πρέπει να είναι σε συγκεκριμένη μορφή. Αυτό απαιτεί μια επεξεργασία, ώστε να είναι έτοιμα για το επόμενο στάδιο. Έτσι, τόσο για την επεξεργασία τους για την εξαγωγή συμπερασμάτων (που αναφέρθηκαν παραπάνω), όσο και για την μορφοποίηση τους, έγινε χρήση τους εργαλείου “R Project for Statistical Computing” [60]. Όλα αυτά τα εξαγόμενα συμπεράσματα και δεδομένα χρησιμοποιήθηκαν στο Κεφάλαιο 7, για την παραγωγή των μοντέλων πρόβλεψης.

Στο παρόν κεφάλαιο θα αναλυθεί σε βάθος η διαδικασία επεξεργασίας των δεδομένων που συλλέχθηκαν, με τη χρήση κατάλληλων εργαλείων. Αρχικά θα γίνει παρουσίαση του εργαλείου της γλώσσας R, το οποίο και χρησιμοποιήθηκε. Αποτελεί ένα χρήσιμο εργαλείο για την μελέτη στατιστικών και άλλων δεδομένων, παρόμοιο με το GNU Octave που χρησιμοποιήθηκε στις εργασίες της σχετικής βιβλιογραφίας (Ενότητα 2.4). Στην συνέχεια, αναφέρεται αναλυτικά ο τρόπος επεξεργασίας των δεδομένων, και κάποια χρήσιμα συμπεράσματα που προκύπτουν. Τέλος, παρουσιάζονται τα διαγράμματα δύο διαστάσεων, για την κάθε παράμετρο σε σχέση με την απόδοση της υπηρεσίας. Η προβολή των διαγραμμάτων είναι απαραίτητη για την εξαγωγή συμπερασμάτων, σχετικά με την μορφή του υποδείγματος των παραμέτρων με την απόδοση της υπηρεσίας. Τα συμπεράσματα που προκύπτουν προβάλλονται αναλυτικά στην Ενότητα 6.3.

6.2: Παρουσίαση Εργαλείου “R Project for Statistical Computing”

Η γλώσσα R είναι μια ταχύτατα αναπτυσσόμενη γλώσσα στατιστικών υπολογισμών και η πιο κοινή γλώσσα έρευνας στα πανεπιστήμια γύρω από στατιστικές εφαρμογές. Είναι ελεύθερα διαθέσιμη από το διαδίκτυο και η υποστήριξή της γίνεται μέσω της εθελοντικής συνεισφοράς πολλών ανθρώπων ανά τον κόσμο, οι οποίοι είναι και υπεύθυνοι για την ανάπτυξη της. Η ιστοσελίδα [52] περιέχει περαιτέρω πληροφορίες καθώς και συνδέσμους για τα σχετικά προγράμματα που αφορούν την αποθήκευση και εκτέλεση του προγράμματος σε διάφορα λειτουργικά συστήματα. Σημειωτέο, ότι η R μπορεί να τρέξει σε περιβάλλον Linux, Mac OS και Windows.

Η R εφαρμόζει μια διάλεκτο της γλώσσας S η οποία είναι μια διερμηνέας γλώσσα προγραμματισμού. Αυτό σημαίνει ότι οι εντολές διαβάζονται και μετά εκτελούνται αμέσως. Αντίθετα, η C και η Fortran είναι γλώσσες προγραμματισμού στις οποίες ολοκληρωμένα προγράμματα μεταφράζονται με τη βοήθεια ενός μεταγλωττιστή στην κατάλληλη γλώσσα μηχανής. Το μεγάλο πλεονέκτημα των διερμηνέων γλωσσών προγραμματισμού είναι ότι επιτρέπουν σταδιακή ανάπτυξη. Με άλλα λόγια, μια συνάρτηση μπορεί να δημιουργηθεί, να εκτελεσθεί και μετά να δημιουργηθεί μια καινούργια συνάρτηση η οποία καλεί την προηγούμενη κ.ο.κ.

Για την εισαγωγή δεδομένων και, γενικότερα, για την απόδοση τιμών σε μεταβλητές χρησιμοποιούμε συνήθως το συνδυασμό "<-" (assignment symbol), που "φαίνεται" σαν βέλος από δεξιά προς τα αριστερά. Η λειτουργία του συμβόλου αυτού είναι να υπολογίσει την έκφραση που δίνεται δεξιά του και να την αποδώσει στη μεταβλητή που βρίσκεται αριστερά του, χωρίς να τυπωθεί το αποτέλεσμα. Άλλος τρόπος για την αντιστοίχιση τιμών σε μεταβλητές είναι το ίσο "=", ή το αντίστροφο βέλος. Σημειώνεται ότι στην S δεν υπάρχουν απλοί αριθμοί, αλλά μόνον διανύσματα, συναρτήσεις, λίστες και άλλα αντικείμενα (objects). Οι απλοί αριθμοί θεωρούνται ως διανύσματα διαστάσεως ένα. Οι κύριες μορφές των δεδομένων, που υπάρχουν στην R, είναι οι ακόλουθες:

- Διάνυσμα (vector)
- Πίνακας (matrix)
- Πίνακας μεγαλύτερης διάστασης (array)
- Λίστα (list)
- Παράγοντας (factor)
- Χρονοσειρές (time series)
- Πλαίσιο δεδομένων (data frame)

Τα διανύσματα είναι ο κύριος τρόπος εισόδου και εξόδου σταθερών ή μεταβλητών ποσοτήτων στην R. Τα στοιχεία των διανυσμάτων μπορεί να είναι αριθμητικές ή λογικές τιμές, ή συμβολοσειρές (strings), χωρίς όμως να αναμειγνύονται. Δεν μπορεί δηλαδή το ένα στοιχείο του διανύσματος να είναι αριθμός και το άλλο συμβολοσειρά. Για την περίπτωση των αριθμητικών τιμών δεν ενδιαφέρει αν είναι ακέραιες, πραγματικές ή και μιγαδικές ούτε και αν είναι απλής ή διπλής ακρίβειας. Η εισαγωγή των δεδομένων στα διανύσματα της γλώσσας, μπορεί να γίνει είτε από το πληκτρολόγιο του χρήστη, είτε από αποθηκευμένα αρχεία (πχ. CSV, Excel, κλπ.). Οι βασικές αριθμητικές πράξεις γίνονται με τη βοήθεια των συμβόλων που βρίσκονται στο ακόλουθο σχήμα:

Σύμβολο	Πράξη
+	Πρόσθεση
-	Αφαίρεση
*	Πολλαπλασιασμός
/	Διαίρεση
^	Υψωση σε δύναμη
%%	Ακέραια διαίρεση
%	Υπόλοιπο διαίρεσης

Σχήμα 6.1: Πίνακας μαθηματικών πράξεων στην γλώσσα R

Οι εντολές για μαθηματικές πράξεις χρησιμοποιούνται όχι μόνο σε αριθμούς, αλλά και σε διανύσματα και πίνακες. Οι περισσότεροι υπολογισμοί με διανύσματα και πίνακες γίνονται κατά στοιχείο, υποθέτοντας ότι οι πίνακες έχουν τις ίδιες διαστάσεις. Στις πράξεις με διανύσματα, αν το ένα διάνυσμα είναι μικρότερης διάστασης από το άλλο, τότε τα στοιχεία του μικρότερου διανύσματος επαναλαμβάνονται κυκλικά, έτσι ώστε τα δύο διανύσματα να έχουν στο τέλος ίσες διαστάσεις. Μαθηματικοί υπολογισμοί μεταξύ διανυσμάτων και πινάκων δεν έχουν συνήθως τα αναμενόμενα αποτελέσματα και θα πρέπει να χρησιμοποιούνται με μεγάλη προσοχή.

Τα γραφήματα είναι πολύ χρήσιμα για την οπτική αναπαράσταση των δεδομένων και καθοδηγούν τον στατιστικό στην διαδικασία της μοντελοποίησης και αξιολόγησης της ανάλυσης. Έτσι, τα δεδομένα που έχουν αποθηκευτεί στα διανύσματα, μπορούν εύκολα να μετατραπούν με τις κατάλληλες εντολές σε διάφορα γραφήματα που προσφέρει η R. Τα πιο απλά γραφήματα είναι οι γραφικές παραστάσεις που συσχετίζονται με μονοδιάστατη τυχαία μεταβλητή, οι γραφικές παραστάσεις συναρτήσεων και οι γραφικές παραστάσεις χρονοσειρών. Η βασική εντολή για γραφική παράσταση είναι η εντολή "plot", η οποία έχει πολλές δυνατότητες και μπορεί να πάρει διάφορες γραφικές παραμέτρους για ορίσματα.

Στην παρούσα εργασία έγινε χρήση των διανυσμάτων στα οποία φορτώθηκαν τα δεδομένα από τα αρχεία συλλογής τους. Στην συνέχεια με την χρήση κάποιων βασικών συναρτήσεων που είναι ενσωματωμένες στο εργαλείο της γλώσσας R, επεξεργάστηκαν τα δεδομένα έτσι ώστε να έρθουν σε μορφή που μπορούν να χρησιμοποιηθούν από το επόμενο στάδιο. Επιπλέον έγινε χρήση των γραφημάτων που προσφέρει το εργαλείο, για την εξαγωγή συμπερασμάτων, όπου αναφέρονται στην συνέχεια του κεφαλαίου. Στην Ενότητα 6.3 παρουσιάζονται τα αποτελέσματα της επεξεργασίας των δεδομένων με την χρήση της γλώσσας R.

6.3: Διαδικασία Επεξεργασίας Δεδομένων

6.3.1 Παρουσίαση Επεξεργασίας Με Χρήση της Γλώσσας R

Η γλώσσα R, μέσω του εργαλείου της, χρησιμοποιείται για την επεξεργασία δεδομένων και την εκτέλεση στατιστικών υπολογισμών. Προσφέρει μια πληθώρα από έτοιμες συναρτήσεις, που βοηθούν τον χρήστη της να επεξεργαστεί τα δεδομένα του με ελάχιστες εντολές. Επιπλέον προσφέρει μεγάλη ποικιλία στην απεικόνιση των δεδομένων σε γραφικές παραστάσεις, μιας ή περισσότερων διαστάσεων. Έτσι, με την χρήση των δυνατοτήτων της συγκεκριμένης γλώσσας, επεξεργάστηκαν και τα δεδομένα για την παρούσα εργασία.

Τα δεδομένα που συλλέχθηκαν, και ένα δείγμα τους παρουσιάστηκε στο Κεφάλαιο 5, έχουν πολύ μεγάλο όγκο πληροφορίας. Ένα χαρακτηριστικό παράδειγμα είναι ότι οι εγγραφές των δεδομένων για τα αιτήματα που εκτελέστηκαν σε κάθε υπολογιστή κυμαίνονται περίπου στις 12000. Αν υποθέσουμε ότι η διαδικασία εκτελέστηκε σε τουλάχιστον τρεις υπολογιστές, τότε οι εγγραφές των δεδομένων είναι περίπου 36000. Όπως είναι λογικό η επεξεργασία αυτών των δεδομένων είναι απαραίτητη, ώστε να μειωθεί ο όγκος τους, κρατώντας όμως τις απαραίτητες πληροφορίες. Επιπλέον, σε ένα τόσο μεγάλο όγκο δεδομένων, υπήρχαν και πληροφορίες όπου περιείχαν ακραίες τιμές, για τις παραμέτρους που καταγράφηκαν. Αυτό μπορεί να οφείλεται, είτε σε παράγοντες που δεν μπορούν να προβλεφθούν (π.χ. καθυστέρηση στην αποστολή του αιτήματος από τον υπολογιστή που εκτελεί τα αιτήματα, λόγω κάποιας άλλης διεργασίας που εκτελέστηκε στιγμιαία), είτε λόγω παραγόντων της υποδομής του υπολογιστικού νέφους που εξυπηρετεί τα αιτήματα (π.χ. αποθήκευση δεδομένων στην προσωρινή μνήμη, για την ταχύτερη εξυπηρέτηση των αιτημάτων). Έτσι για τους παραπάνω λόγους, αποφασίστηκε η μελέτη της μέσης τιμής απόκρισης, για την κάθε χρονική στιγμή εκτέλεσης της διαδικασίας.

Αρχικά τα δεδομένα εισήχθησαν στο εργαλείο επεξεργασίας της γλώσσας R. Χωρίστηκαν ανάλογα την πόλη στην οποία βρισκόταν ο υπολογιστής που εκτέλεσε τα αιτήματα. Στην συνέχεια αφαιρέθηκαν τα δεδομένα τα οποία είχαν σφάλματα. Όπως αναφέρθηκε στα ορίσματα της εφαρμογής που αναπτύχθηκε (Ενότητα 5.2.2), υπάρχει συγκεκριμένο πεδίο σε κάθε εγγραφή που δηλώνει αν δημιουργήθηκε κάποιο σφάλμα ή όχι. Έγινε σχετική μελέτη για το αν το μέγεθος των εσφαλμένων εγγραφών ήταν πολύ μεγάλο ή όχι. Σε περίπτωση που ήταν πολύ μεγάλο θα έπρεπε να διερευνηθεί ο λόγος ο οποίος ήταν και να πραγματοποιηθεί ξανά η διαδικασία συλλογής των δεδομένων. Στην συγκεκριμένη τελική συλλογή δεδομένων που έγινε για την εργασία, τα σφάλματα ήταν από 0% μέχρι 0,6% των συνολικών εγγραφών.

Στην συνέχεια, αποθηκεύτηκαν οι τιμές για τα πεδία του χρόνου απόκρισης, των συνολικών παράλληλων αιτημάτων που εκτελέστηκαν και της χρονικής στιγμής εκτέλεσης τους, σε ξεχωριστά διανύσματα - μεταβλητές της γλώσσας. Τα διανύσματα αυτά επεξεργάστηκαν με την κατάλληλες εντολές, της γλώσσας R, για την εξαγωγή των μέσων τιμών του χρόνου απόκρισης των αιτημάτων, για την κάθε χρονική στιγμή. Τα αποτελέσματα της διαδικασίας αυτής αποθηκεύτηκαν σε δύο νέα διανύσματα.

Το ένα διάνυσμα είχε τις μέσες τιμές του χρόνου απόκρισης για το σύνολο εκτέλεσης των παράλληλων αιτημάτων. Για παράδειγμα η μέση τιμή απόκρισης της υπηρεσίας όταν εκτελέστηκαν 150 παράλληλα αιτήματα από συγκεκριμένο υπολογιστή, ήταν 1289 χιλιοστά του δευτερολέπτου. Συνεπώς, το διάνυσμα περιείχε ένα σύνολο από τιμές για το πλήθος των παράλληλων αιτημάτων που εκτελέστηκαν την κάθε χρονική στιγμή, σε αντιστοιχία με την μέση τιμή απόκρισης για τα αιτήματα αυτά.

Το δεύτερο διάνυσμα περιείχε τις μέσες τιμές του χρόνου απόκρισης για την κάθε χρονική στιγμή εκτέλεσης. Για παράδειγμα την χρονική στιγμή εκτέλεσης, από την έναρξη της διαδικασίας, «20 δευτερόλεπτα» η μέση τιμή απόκρισης της υπηρεσίας ήταν «0,428 δευτερόλεπτα». Έτσι, το διάνυσμα, περιέχει ένα σύνολο από τις χρονικές περιόδους εκτέλεσης της διαδικασίας, σε αντιστοιχία με τον μέσο χρόνο απόκρισης της κάθε χρονικής στιγμής.

Τα δύο αυτά διανύσματα χρησιμοποιήθηκαν στην συνέχεια για την απεικόνιση τους σε διαγράμματα για την εξαγωγή συμπερασμάτων. Επιπλέον, δημιουργήθηκε ένα τρίτο διάνυσμα με τα περιεχόμενα των δυο άλλων διανυσμάτων που αναφέρθηκαν προηγουμένως. Το τρίτο διάνυσμα, περιείχε πληροφορίες για την κάθε χρονική στιγμή, όπου είχε εκτελεστεί συγκεκριμένος αριθμός παράλληλων αιτημάτων, έχοντας συγκεκριμένη μέση τιμή απόκρισης της υπηρεσίας. Για την μετέπειτα επεξεργασία του διανύσματος, από το Eureka για την παραγωγή των μοντέλων πρόβλεψης, πραγματοποιήθηκε εξαγωγή του σε μορφή CSV (αρχείο κειμένου με οριοθετημένα πεδία).

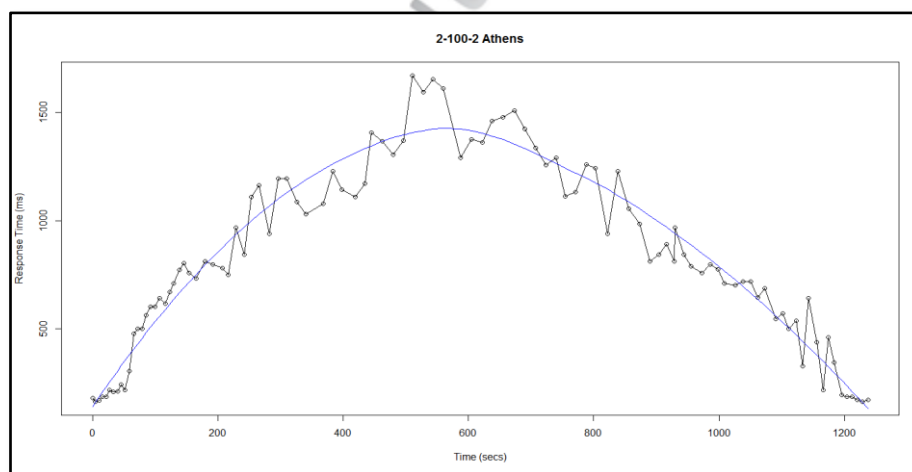
Τα δύο πρώτα διανύσματα, με τις μέσες τιμές του χρόνου απόκρισης της υπηρεσίας, σε αντιστοιχία με αριθμό των παράλληλων αιτημάτων και της χρονικής στιγμής εκτέλεσης τους, απεικονίστηκαν στο σύστημα αξόνων δύο διαστάσεων. Αρχικά η απεικόνιση αυτή ήταν χρήσιμη για την απόδειξη ότι σε διαφορετικές χρονικές στιγμές εκτέλεσης, σε περίπτωση που ο αριθμός των παράλληλων αιτημάτων είναι ίδιος, τότε δεν υπάρχει μεγάλη αύξηση στον χρόνο απόκρισης της υπηρεσίας. Επιπλέον αποδεικνύεται ότι η εκτέλεση της διαδικασίας συλλογής των δεδομένων σε διαφορετικές χρονικές περιόδους, δηλαδή σε διαφορετική μέρα και ώρα, δεν επηρεάζει τον χρόνο απόκρισης της υπηρεσίας. Αυτό σημαίνει ότι παρόλο που η υπηρεσία χρησιμοποιείται από μεγάλο αριθμό χρηστών της, και διαφέρει κατά την διάρκεια της ημέρας, η υποδομή του υπολογιστικού νέφους, είναι αρκετή για υποστηρίξει τα όρια της απόδοσης της με βάση της ποιότητα της υπηρεσίας. Επιπλέον αυτό βοηθάει στην μελέτη για την δημιουργία των μοντέλων πρόβλεψης για την συγκεκριμένη υπηρεσία. Μερικά παραδείγματα των γραφικών παραστάσεων για την εξαγωγή των παραπάνω συμπερασμάτων, παρουσιάζονται στην επόμενη ενότητα.

6.3.2 Παρουσίαση Διαγραμμάτων των Δεδομένων, για την Εξαγωγή Συμπερασμάτων

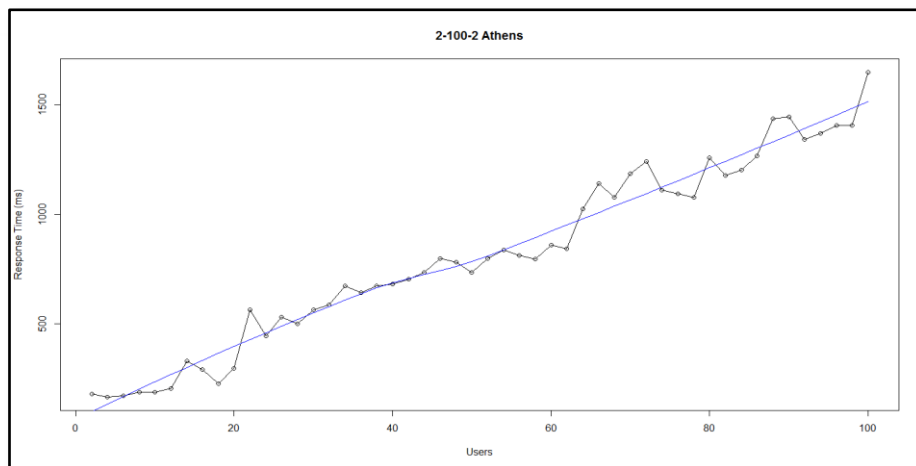
Με την χρήση του εργαλείου της γλώσσας R, μέσω των συναρτήσεων που προσφέρει για την δημιουργία γραφικών παραστάσεων, απεικονίστηκαν τα δεδομένα που συλλέχθηκαν. Ο λόγος ήταν η εξαγωγή κάποιων συμπερασμάτων, χρήσιμων στην ανάλυση της συμπεριφοράς της υπηρεσίας. Στα ακόλουθα σχήματα (Σχήμα 6.2, 6.3, 6.4, 6.5, 6.6, 6.7) παρουσιάζονται τα διαγράμματα για την συλλογή των δεδομένων που πραγματοποιήθηκε

από υπολογιστή που ήταν εγκατεστημένος στην Αθήνα. Η διαδικασία για το Σχήμα 6.2 και 6.3, περιλαμβάνει κλιμάκωση των αιτημάτων, ξεκινώντας από δύο αιτήσεις την πρώτη χρονική στιγμή, και αυξάνεται κατά δύο μέχρι τις εκατό αιτήσεις. Στην συνέχεια, μειώνονται οι αιτήσεις από τις εκατό, κατά δύο μέχρι τις δύο αιτήσεις, όπου και ολοκληρώνεται η διαδικασία. Αντίστοιχα για το Σχήμα 6.4 και 6.5, οι αιτήσεις αρχικά ήταν τρεις, αυξήθηκαν κλιμακωτά, κατά τρεις, μέχρι τις εκατό πενήντα, και στην συνέχεια μειώθηκαν κλιμακωτά, μέχρι τις τρεις. Παρόμοια διαδικασία απεικονίζεται και στο Σχήμα 6.6. και 6.7, ξεκινώντας με τέσσερις αιτήσεις μέχρι τις διακόσες, και κλιμακωτή μείωσή τους μέχρι τις τέσσερις. Για τις διαδικασίες που αναφέρθηκαν παραπάνω, στα Σχήματα 6.2, 6.4, 6.6, παρουσιάζονται τα δεδομένα για τις παραμέτρους του χρόνου εκτέλεσης της διαδικασίας, και του χρόνου απόκρισης της. Αντίστοιχα, στα Σχήματα 6.3, 6.5, 6.7, απεικονίζονται σε διάγραμμα τα δεδομένα για το πλήθος των παράλληλων αιτημάτων και του χρόνου απόκρισης.

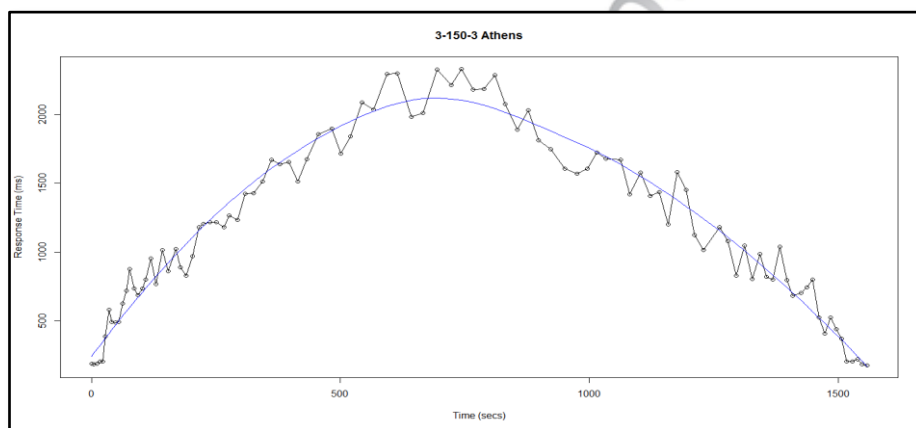
Η συγκεκριμένη πρακτική συλλογής των εμπειρικών δεδομένων, δεν κρίθηκε σωστή για την παραγωγή μοντέλων πρόβλεψης, καθώς όλες οι μεταβλητές που επηρεάζουν το χρόνο απόκρισης της υπηρεσίας δεν είναι ανεξάρτητες (στις πραγματικές συνθήκες οι παράμετροι είναι ανεξάρτητες μεταξύ τους). Αυτό οφείλεται στο γεγονός ότι με βάση την χρονική στιγμή εκτέλεσης της υπηρεσίας, μπορεί να προβλεφθεί ο αριθμός των παράλληλων αιτημάτων προς την υπηρεσία. Αυτή η συλλογή όμως, πραγματοποιήθηκε για την εξαγωγή του συμπεράσματος ότι ο χρόνος απόκρισης δεν διαφέρει πολύ, όταν ο αριθμός των παράλληλων αιτημάτων είναι ίδιος. Υπάρχει μια μικρή αυξομείωση που οφείλεται στον χρόνο εκτέλεσης των αιτημάτων προς την υπηρεσία.



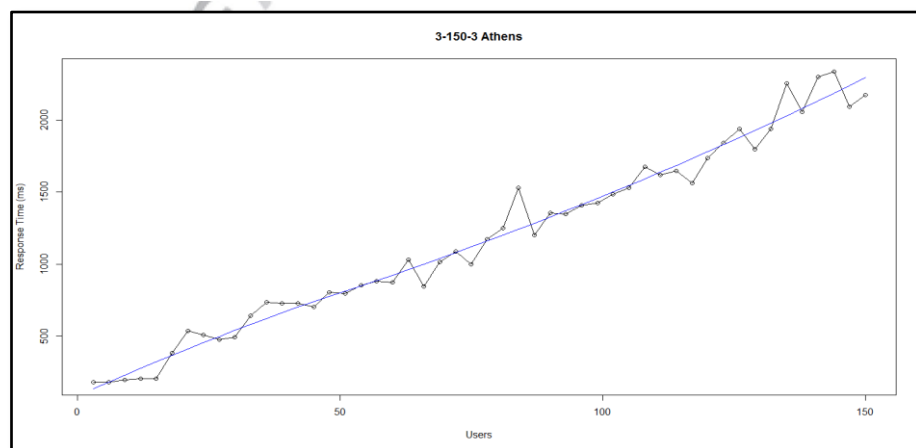
Σχήμα 6.2: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά δύο



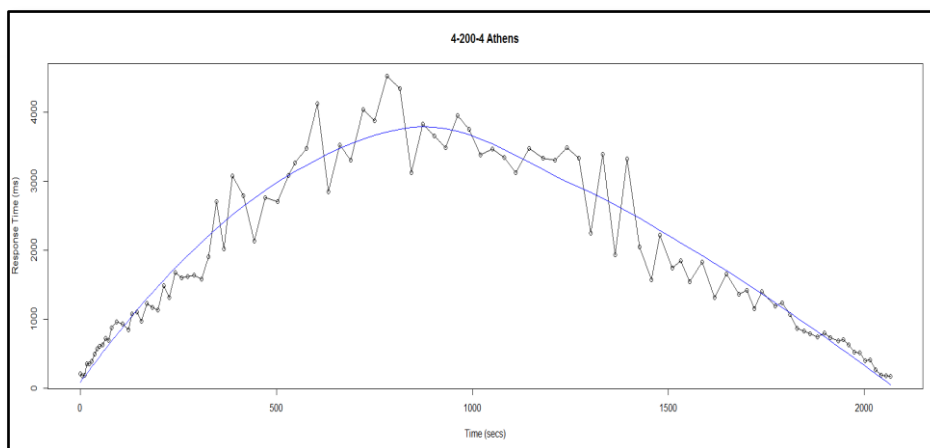
Σχήμα 6.3: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά δύο



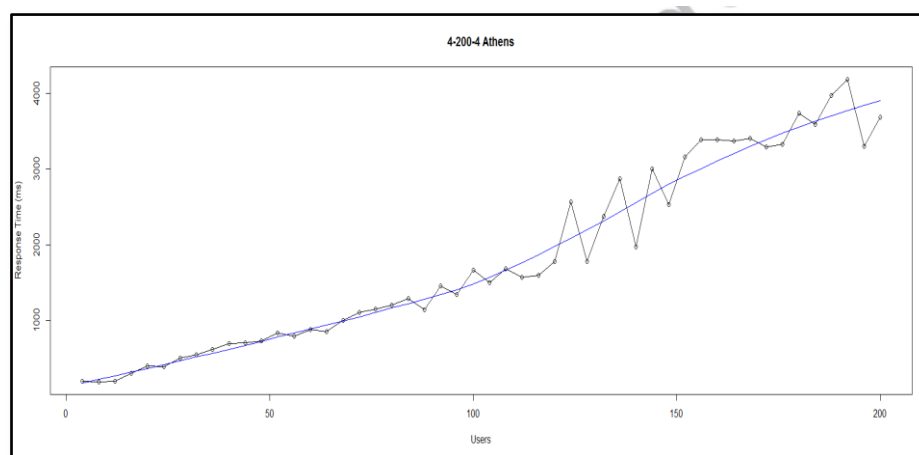
Σχήμα 6.4: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τρία



Σχήμα 6.5: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τρία



Σχήμα 6.6: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τέσσερα



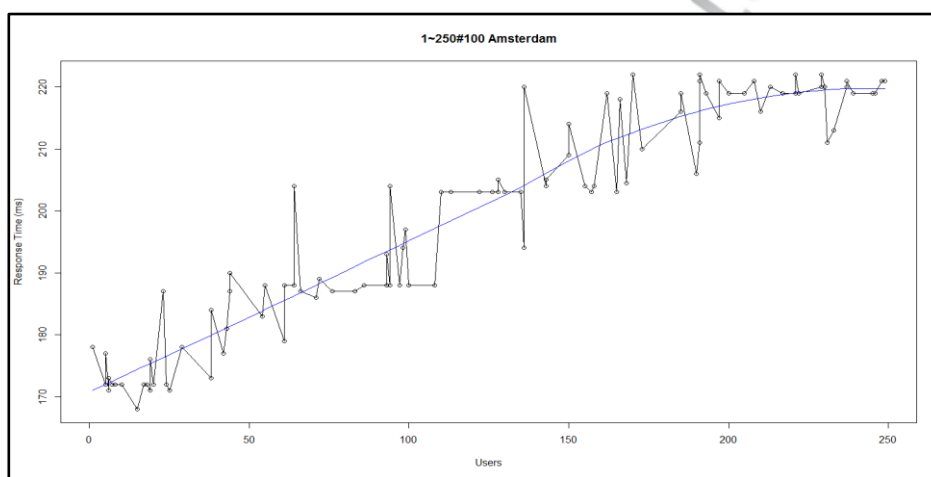
Σχήμα 6.7: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με κλιμακωτή αύξηση και μείωση των παράλληλων αιτημάτων, κατά τέσσερα

Ένα ακόμα χρήσιμο συμπέρασμα που μπορεί να εξαχθεί από την απεικόνιση των δεδομένων σε διαγράμματα αξόνων δύο διαστάσεων, είναι η επιρροή των παραμέτρων στο χρόνο απόκρισης της υπηρεσίας. Ένα παράδειγμα, είναι αυτή της χιλιομετρικής απόστασης του υπολογιστή που εκτελεί τα αιτήματα, από τις υποδομές του υπολογιστικού νέφους της υπηρεσίας. Πράγματι, όπως παρουσιάζεται στα διαγράμματα παρακάτω, όσο αυξάνεται η χιλιομετρική απόσταση των υπολογιστών που εκτελούν τα αιτήματα προς την υπηρεσία, τόσο αυξάνεται και ο χρόνος απόκρισης. Αντίστοιχα, όσο αυξάνεται ο αριθμός των παράλληλων αιτημάτων προς την υπηρεσία, τόσο αυξάνεται και ο χρόνος απόκρισης.

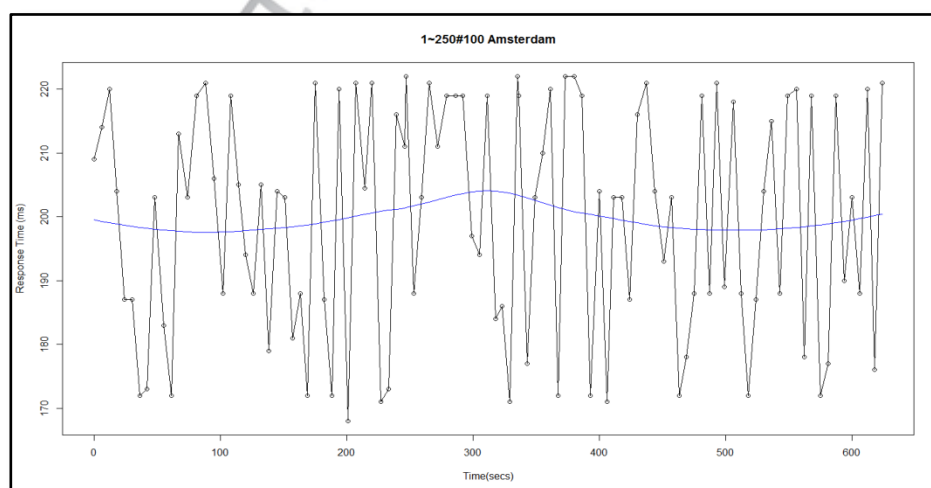
Τα Σχήματα 6.8, 6.9, 6.10, 6.11, 6.12, 6.13 όπου ακολουθούν, έχουν προκύψει από την επεξεργασία των δεδομένων για τυχαίους αριθμούς παράλληλων αιτημάτων προς την υπηρεσία. Η συλλογή των δεδομένων πραγματοποιήθηκε με αυτό τον τρόπο, για να υπάρχει ανεξαρτησία μεταξύ των μεταβλητών. Για την απόδειξη του τελευταίου, μαζί με τα διαγράμματα για τους αριθμούς των παράλληλων αιτημάτων ως προς το χρόνο απόκρισης

της υπηρεσίας, παρουσιάζονται και τα διαγράμματα του χρόνου εκτέλεσης της διαδικασίας ως προς το χρόνο απόκρισης. Ο λόγος που δεν υπάρχει σχετική ομαλότητα στην γραφική παράσταση του χρόνο απόκρισης προς τον χρόνο εκτέλεσης, είναι επειδή ο αριθμός των παράλληλων αιτημάτων είναι τυχαίος κάθε φορά, και έτσι επηρεάζεται ο χρόνος απόκρισης της υπηρεσίας. Η διαδικασία συλλογής των δεδομένων, πραγματοποιήθηκε σε τρεις διαφορετικές πόλεις της Ευρώπης (Άμστερνταμ, Αθήνα και Ρώμη), με τυχαίο αριθμό αιτήσεων κάθε χρονική στιγμή, που εκτελέστηκαν προς την υπηρεσία Wikipedia.org, όπου οι υποδομές του υπολογιστικού νέφους, βρίσκονται στο Άμστερνταμ της Ολλανδίας.

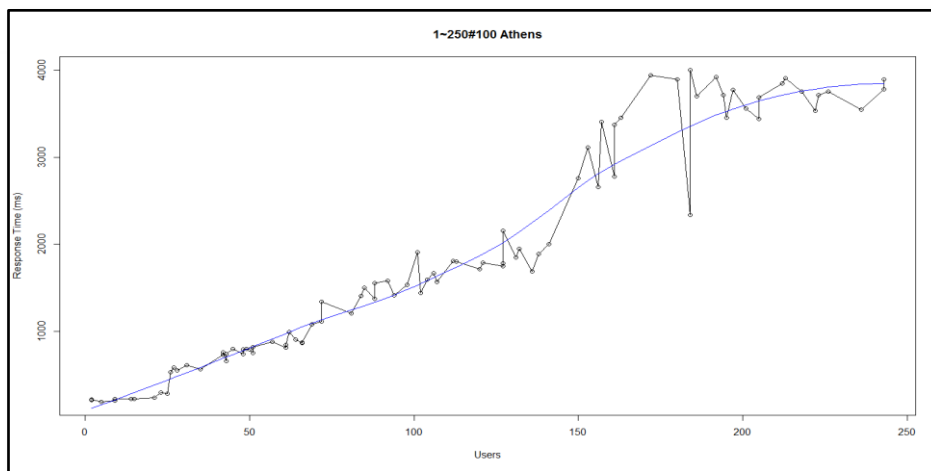
Στην συγκεκριμένη διαδικασία, στο Άμστερνταμ καταγράφηκε μέγιστος χρόνος απόκρισης περίπου 220 χιλιοστά του δευτερολέπτου, στην Ρώμη που απέχει 1300 χιλιόμετρα, περίπου 3000 χιλιοστά του δευτερολέπτου και στην Αθήνα που απέχει 2166 χιλιόμετρα, περίπου 4000 χιλιοστά του δευτερολέπτου, για μέχρι 250 παράλληλα αιτήματα προς την υπηρεσία.



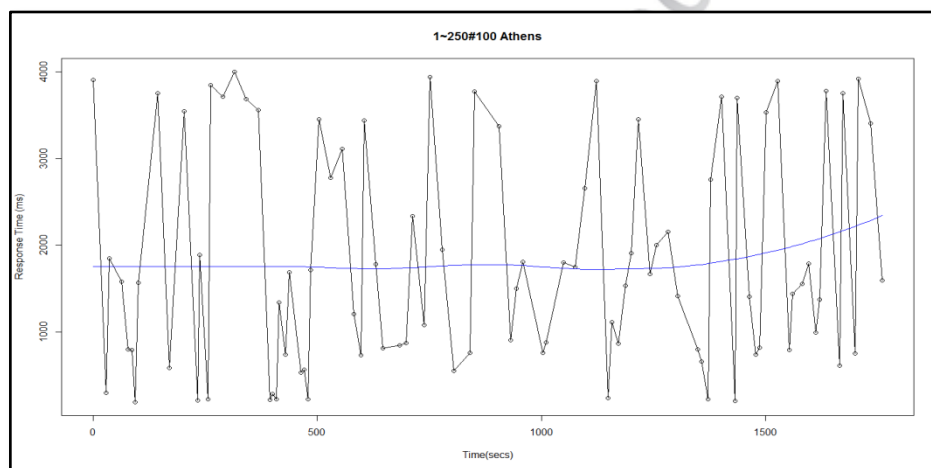
Σχήμα 6.8: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων (από Άμστερνταμ)



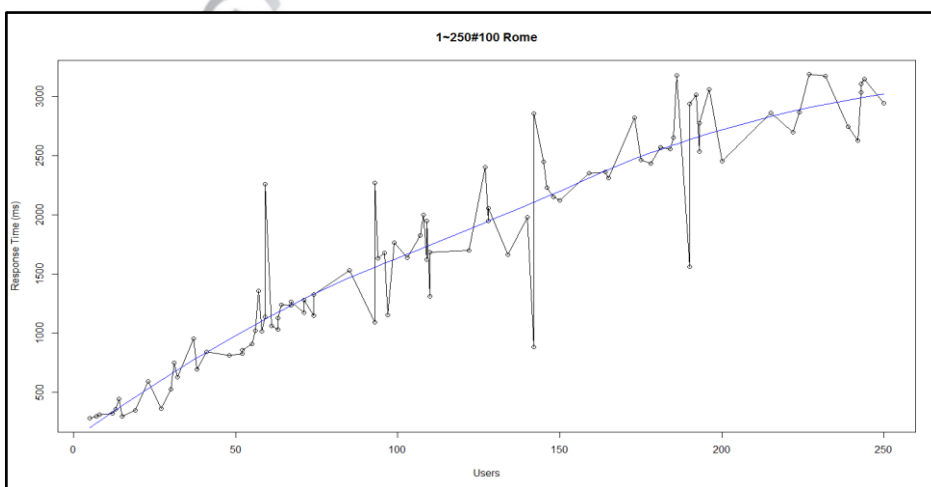
Σχήμα 6.9: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Άμστερνταμ)



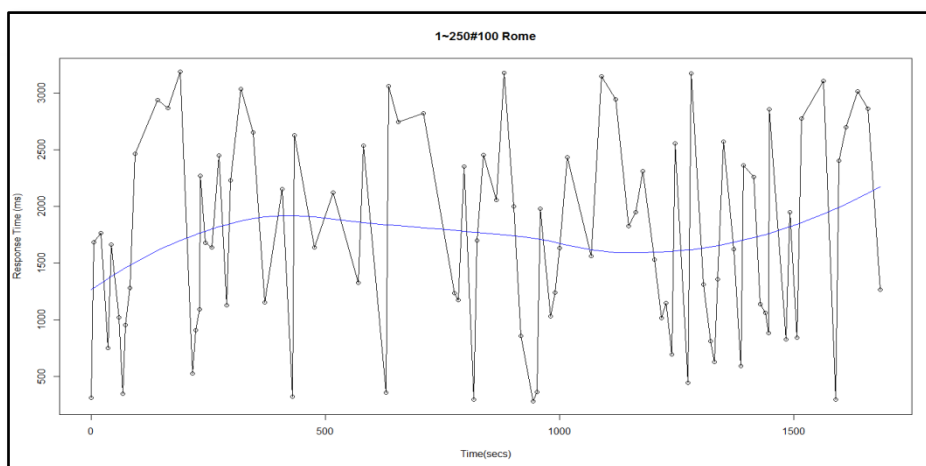
Σχήμα 6.10: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Αθήνα)



Σχήμα 6.11: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Αθήνα)



Σχήμα 6.12: Διάγραμμα απεικόνισης δεδομένων (αριθμός παράλληλων αιτημάτων – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Ρώμη)



Σχήμα 6.13: Διάγραμμα απεικόνισης δεδομένων (χρόνος εκτέλεσης – χρόνος απόκρισης), με τυχαία αύξηση και μείωση των παράλληλων αιτημάτων (από Ρώμη)

Ένα τελευταίο χρήσιμο συμπέρασμα που προέκυψε από την γραφική απεικόνιση των δεδομένων είναι ο καθορισμός του μέγιστου αριθμού των παράλληλων αιτημάτων προς την υπηρεσία. Σκοπός της παρούσας εργασίας, είναι η δημιουργία μοντέλων πρόβλεψης της απόδοσης μιας υπηρεσίας, βάσει των παραμέτρων κλιμάκωσής της. Λόγω έλλειψης όμως της γνώσης σχετικά με τις υποδομές του υπολογιστικού νέφους που χρησιμοποιούνται από την υπηρεσία, δεν μπορεί να προβλεφθεί τότε αυξάνονται οι πόροι για την εξυπηρέτηση των αιτημάτων. Αν όμως γίνει αυτό, τότε τα μοντέλα πρόβλεψης που θα έχουν παραχθεί δεν θα είναι σωστά, καθώς η τιμές του χρόνου απόκρισης θα αντιστοιχούν σε ορισμένες περιπτώσεις, σε διαφορετική χρήση πόρων.

Μέσα από την γραφική απεικόνιση των παράλληλων αιτημάτων, ως προς τον χρόνο απόκρισης, που παρουσιάστηκαν στα προηγούμενα διαγράμματα, μπορεί να βρεθεί το σημείο όπου πιθανότατα αυξάνονται οι πόροι για την εξυπηρέτηση των αιτημάτων. Εκτός όμως από την αύξηση των πόρων, μπορεί να υπάρχουν και άλλοι άγνωστοι παράγοντες που βοηθούν στην ταχύτερη απόκριση στα αιτήματα. Παρατηρείται, λοιπόν, με βάση τα προηγούμενα διαγράμματα των τριών πόλεων, ότι όταν ο αριθμός των παράλληλων αιτημάτων ξεπερνάει το διακόσια, τότε δεν αυξάνεται αναλόγως, η τιμή του χρόνου απόκρισης. Επομένως, είναι πολύ πιθανό να γίνεται κάποια αύξηση στους πόρους ή οποιαδήποτε άλλη κίνηση που βοηθάει στην ταχύτερη εξυπηρέτηση των αιτημάτων. Έτσι, τα δεδομένα που πρέπει να ληφθούν υπόψη και να επεξεργαστούν από το επόμενο στάδιο για την παραγωγή των μοντέλων πρόβλεψης, είναι αυτά για μέχρι διακόσια παράλληλα αιτήματα προς την υπηρεσία.

6.4: Συμπεράσματα

Στο Κεφάλαιο 6 παρουσιάστηκε η διαδικασία επεξεργασίας των δεδομένων, που συλλέχθηκαν από την εφαρμογή που αναπτύχθηκε (Κεφάλαιο 5), για την υπηρεσία της ηλεκτρονικής εγκυκλοπαίδειας Wikipedia.org. Η επεξεργασία των δεδομένων έγινε με την χρήση του εργαλείου της γλώσσας R. Η συγκεκριμένη γλώσσα, χρησιμοποιείται για την επεξεργασία δεδομένων και την εκτέλεση στατιστικών υπολογισμών. Προσφέρει μια

πληθώρα από συναρτήσεις που βοηθούν στην εύκολη επεξεργασία των δεδομένων. Επίσης, περιέχει μεθόδους για την απεικόνιση των δεδομένων σε διαγράμματα μιας ή περισσότερων διαστάσεων. Έτσι, με την χρήση των δυνατοτήτων της συγκεκριμένης γλώσσας, επεξεργάστηκαν και τα δεδομένα της παρούσας εργασίας.

Η επεξεργασία των εμπειρικών δεδομένων ήταν απαραίτητη, καθώς ο όγκος τους ήταν αρκετά μεγάλος ώστε να χρησιμοποιηθεί στην συνέχεια, για την παραγωγή των μοντέλων πρόβλεψης. Επιπλέον, σε ένα τόσο μεγάλο όγκο δεδομένων, υπήρχαν και πληροφορίες όπου περιείχαν ακραίες τιμές, για τις παραμέτρους που μελετώνται. Αυτό μπορεί να οφείλεται, είτε σε παράγοντες που δεν μπορούν να προβλεφθούν (π.χ. καθυστέρηση στην αποστολή του αιτήματος από τον υπολογιστή που εκτελεί τα αιτήματα, λόγω κάποιας άλλης διεργασίας που εκτελέστηκε στιγμιαία), είτε λόγω παραγόντων της υποδομής του υπολογιστικού νέφους που εξυπηρετεί τα αιτήματα (π.χ. αποθήκευση δεδομένων στην προσωρινή μνήμη, για την ταχύτερη εξυπηρέτηση των αιτημάτων). Για τους παραπάνω λόγους, αποφασίστηκε η μελέτη της μέσης τιμής απόκρισης, για την κάθε χρονική στιγμή εκτέλεσης της διαδικασίας. Συνεπώς, για κάθε χρονική στιγμή της διαδικασίας, προέκυψε ένα σύνολο δεδομένων, που περιείχε τα παράλληλα αιτήματα που εκτελέστηκαν προς την υπηρεσία, και τη μέση τιμή του χρόνου απόκρισής τους. Το σύνολο των δεδομένων αυτών, αντιπροσωπεύει την συμπεριφορά της υπηρεσίας, στην εξυπηρέτηση των αιτημάτων κάθε χρονική στιγμή. Τα δεδομένα αυτά, χρησιμοποιήθηκαν από το επόμενο στάδιο, όπου αναλύονται από το εργαλείο Eureqa, για την δημιουργία των μοντέλων πρόβλεψης.

Επίσης, ένας ακόμα λόγος για την επεξεργασία των δεδομένων, ήταν η εξαγωγή χρήσιμων συμπερασμάτων. Αυτό έγινε με την προβολή των δεδομένων σε διαγράμματα δύο διαστάσεων (Ενότητα 6.3.2). Αρχικά, τα συμπεράσματα αυτά, αφορούσαν την ορθότητα της διαδικασίας συλλογής των εμπειρικών δεδομένων. Δηλαδή, αν τα δεδομένα ήταν αντιπροσωπευτικά της συμπεριφοράς της υπηρεσίας στα αιτήματα χρήσης της, και επιπλέον αν υπήρχε επιρροή άλλων παραγόντων (πέρα από αυτών που καταγράφονται οι τιμές τους), στην απόδοση της υπηρεσίας. Έτσι, προέκυψε ότι η αρχική προσέγγιση εκτέλεσης της διαδικασίας, με κλιμακωτή αυξομείωση των παράλληλων αιτημάτων προς την υπηρεσία, δεν ήταν η κατάλληλη. Ο λόγος ήταν, ότι οι παράμετροι για τις οποίες συλλέγονται τα δεδομένα, πρέπει να είναι ανεξάρτητες μεταξύ τους (όπως συμβαίνει στις πραγματικές συνθήκες). Αντιθέτως, βάσει την αρχικής προσέγγισης, υπήρχε εξάρτηση μεταξύ του χρόνου εκτέλεσης της διαδικασίας και του πλήθους των παράλληλων αιτημάτων προς της υπηρεσία, καθώς μπορούσε να προβλεφθεί το πλήθος των παράλληλων αιτημάτων, για κάθε χρονική στιγμή. Στην συνέχεια, για την συλλογή των εμπειρικών δεδομένων, χρησιμοποιήθηκε τυχαίος αριθμός παράλληλων αιτημάτων, για κάθε χρονική στιγμή.

Από την αρχική όμως προσέγγιση, προέκυψε ένα χρήσιμο συμπέρασμα, για την αντικειμενικότητα της συμπεριφοράς της υπηρεσίας. Καθώς υπήρχε συγκεκριμένη αυξομείωση στην τιμή της παραμέτρου των παράλληλων αιτημάτων, υπήρχαν χρονικές στιγμές, όπου το πλήθος τους ήταν ίδιο. Έτσι, αν δεν υπήρχε όμοια αυξομείωση του χρόνου απόκρισης μεταξύ των ίδιων τιμών της παραμέτρου των παράλληλων αιτημάτων, λόγω της παραμέτρου του χρόνου εκτέλεσης, θα σήμαινε ότι υπάρχουν και άλλοι παράγοντες που επηρεάζουν την απόδοση της υπηρεσίας. Προέκυψε λοιπόν το συμπέρασμα, από την

μελέτη των διαγραμμάτων, ότι υπήρχε όμοια αυξομείωση της μέσης τιμής του χρόνου απόκρισης της υπηρεσίας, μεταξύ των χρονικών στιγμών που είχαν τις ίδιες τιμές για την παράμετρο των παράλληλων αιτημάτων. Συνεπώς, η καταγραφή της συμπεριφοράς της υπηρεσίας, που έγινε στα πλαίσια της διαδικασίας συλλογής των εμπειρικών δεδομένων, θεωρείται αντιπροσωπευτική.

Στο Κεφάλαιο 7, παρουσιάζεται η ανάλυση των επεξεργασμένων δεδομένων, από το εργαλείο Eureqa, για την παραγωγή των μοντέλων πρόβλεψης. Με την χρήση συγκεκριμένων μαθηματικών εννοιών, συγκρίνονται τα μοντέλα πρόβλεψης της συμπεριφοράς της υπηρεσίας, που προκύπτουν από το Eureqa. Με αυτό τον τρόπο γίνεται η επιλογή του καταλληλότερου, το οποίο στην συνέχεια θα ελεγχθεί για να αξιολογηθεί, με βάση τα δεδομένα των πραγματικών συνθηκών.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 7: Προτεινόμενο Μοντέλο Δυναμικής Πρόβλεψης

7.1: Εισαγωγή

Στο Κεφάλαιο 5 παρουσιάστηκε η διαδικασία συλλογής των εμπειρικών δεδομένων, για την υπηρεσία της ηλεκτρονικής εγκυκλοπαίδειας Wikipedia.org. Στην συνέχεια, όπως αναφέρθηκε στο Κεφάλαιο 6, τα δεδομένα που είχαν συλλεχθεί, επεξεργάστηκαν για την μείωση του όγκου τους και για την εξαγωγή χρήσιμων συμπερασμάτων, σχετικά με την διαδικασία συλλογής και την αντικειμενικότητα της συμπεριφοράς της υπηρεσίας, που καταγράφηκε.

Σκοπός του Κεφαλαίου 7, είναι η ανάλυση των επεξεργασμένων δεδομένων, για την παραγωγή του προτεινόμενο μοντέλο πρόβλεψης της απόδοσης της υπηρεσίας Wikipedia.org, σε σχέση με τις παραμέτρους που ορίστηκαν. Κριτήριο απόδοσης της υπηρεσίας είναι ο χρόνος απόκρισης της. Συνεπώς, το μοντέλο πρόβλεψης θα έχει ως στόχο την πρόβλεψη του χρόνου απόκρισης της υπηρεσίας, με βάση τα αιτήματα των χρηστών, την γεωγραφική θέση απ' όπου προέρχονται και ο χρόνος κατά τον οποίον εκτελούνται τα αιτήματα προς την υπηρεσία.

Ύστερα από μελέτη της συμπεριφοράς της υπηρεσίας Wikipedia.org και ανάλυση των παραμέτρων που επηρεάζουν την απόδοσή της, βγήκαν χρήσιμα συμπεράσματα σχετικά με την βαρύτητα τους, ως προς την επιρροή τους στον χρόνο απόκρισης της υπηρεσίας. Τα συμπεράσματα αυτά, μπορούν να βοηθήσουν στην αύξηση της απόδοσης της υπηρεσίας, πραγματοποιώντας αλλαγές στις υποδομές του υπολογιστικού νέφους.

Αρχικά, στην Ενότητα 7.2, γίνεται μια παρουσίαση των μαθηματικών εννοιών που σχετίζονται με την ανάλυση των δεδομένων και την αξιολόγηση των μοντέλων πρόβλεψης. Βάσει των εννοιών αυτών, αναλύεται στην Ενότητα 7.3 η μορφή του υποδείγματος, για τα συγκεκριμένα δεδομένα. Με την χρήση του εργαλείου Eureqa, παρουσιάζεται στην Ενότητα 7.4, η διαδικασία εύρεσης των μοντέλων πρόβλεψης της απόδοσης της υπηρεσίας της Wikipedia.org. Στην Ενότητα 7.5 μελετώνται και συγκρίνονται τα μοντέλα που έχουν προκύψει για την επιλογή του καταλληλότερου. Τέλος, στην Ενότητα 7.6, παρουσιάζονται τα συμπεράσματα που προέκυψαν από την ανάλυση του κεφαλαίου.

7.2: Παρουσίαση Μαθηματικών Εννοιών.

Στην ενότητα αυτή αναφέρονται κάποιες βασικές μαθηματικές έννοιες, έτσι ώστε να γίνει πιο κατανοητός ο τρόπος διερεύνησης του μοντέλου πρόβλεψης που παρουσιάζεται στην Ενότητα 7.3. Όλες οι έννοιες που θα αναφερθούν, βασίζονται στην ανάλυση που παρουσιάζεται στο βιβλίο Οικονομετρίας, του καθηγητή Ηλία Τζαβαλή [61]. Οι έννοιες που αναφέρονται, σχετίζονται με την μελέτη υποδειγμάτων, που προκύπτουν από τα δεδομένα που συλλέγονται για την δημιουργία μοντέλων πρόβλεψης. Έτσι, ξεκινώντας από το απλό γραμμικό υπόδειγμα, για την κατανόηση της έννοιας του υποδείγματος και της χρήσης του,

στην συνέχεια παρουσιάζονται οι εκτιμήσεις που μπορούν να γίνουν για τους συντελεστές τους. Αφού γίνουν κατανοητές οι έννοιες αυτές, αναφέρεται το πολλαπλό γραμμικό υπόδειγμα και οι προβλέψεις που μπορούν να γίνουν χρησιμοποιώντας το. Τέλος, γίνεται παρουσίαση των εκτιμήσεων που μπορούν να γίνουν πάνω σε μη γραμμικά υποδείγματα. Συνεπώς, σκοπός της ενότητας αυτής, είναι η κατανόηση των υποδειγμάτων και της προβλεπτικής τους ικανότητας.

Το υπόδειγμα που παράχθηκε, βάσει των επεξεργασμένων δεδομένων (Κεφάλαιο 6), είναι ένα μη γραμμικό πολλαπλό υπόδειγμα. Εξαρτημένη τιμή του υποδείγματος είναι ο χρόνος απόκρισης της υπηρεσίας, και ανεξάρτητες μεταβλητές του, είναι το πλήθος των παράλληλων αιτημάτων προς την υπηρεσία, η γεωγραφική απόσταση απ' όπου προέρχονται και ο χρόνος εκτέλεσης τους.

7.2.1 Γραμμικό Υπόδειγμα

Το πιο απλό υπόδειγμα που χρησιμοποιείται στην ανάλυση, για μοντελοποίηση παρατηρήσεων σε μοντέλα πρόβλεψης, είναι το απλό γραμμικό υπόδειγμα ή όπως αναφέρεται εναλλακτικά, γραμμική παλινδρόμηση μιας ανεξάρτητης μεταβλητής (linear single regression model). Αυτό ορίζεται ως εξής:

$$y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i, \quad i = 1, 2, \dots, N.$$

Όπου οι y_i και x_i συμβολίζουν τις παρατηρήσεις ή τις τιμές της εξαρτημένης και ανεξάρτητης μεταβλητής του υποδείγματος. Αντίστοιχα σε ένα δείγμα $i=1,2,\dots,N$ παρατηρήσεων, οι β_1 και β_2 αποτελούν τις άγνωστες προς εκτίμηση παραμέτρους του υποδείγματος και η ε_i συμβολίζει τις παρατηρήσεις του διαταρακτικού όρου του υποδείγματος. Οι παράμετροι β_1 και β_2 του υποδείγματος, συχνά αναφέρονται και ως συντελεστές του υποδείγματος και ο διαταρακτικός όρος αναφέρετε επίσης και ως σφάλμα παλινδρόμησης.

Ο διαταρακτικός όρος θεωρείται ότι αποτελεί μια τυχαία μεταβλητή, της οποίας οι τιμές δεν είναι παρατηρήσιμες. Αυτές οφείλονται σε τυχαίους, μη συστηματικούς παράγοντες που επηρεάζουν τις μεταβολές των τιμών της εξαρτημένης μεταβλητής, πέρα από τις μεταβολές στις τιμές της ανεξάρτητης μεταβλητής x_i και επομένως δεν μπορούν να ερμηνευτούν εύκολα. Για τον λόγο αυτό μπορούμε να υποθέσουμε ότι οι τιμές του διαταρακτικού όρου, για όλες τις παρατηρήσεις ενός δείγματος, θεωρούνται ότι αντιπροσωπεύουν τιμές τυχαίων μεταβλητών που προέρχονται από κατανομές με αναμενόμενη (μέση) τιμή που ισούται με το μηδέν.

Το γραμμικό υπόδειγμα μπορεί να διαχωριστεί σε δύο μέρη: το $\beta_1 + \beta_2 * x_i$ και το διαταρακτικό όρο ε_i . Το πρώτο μέρος ($\beta_1 + \beta_2 * x_i$) αποτελεί το ερμηνευμένο ή προβλεπόμενο μέρος του. Αυτό εξηγεί τις συστηματικές μεταβολές της εξαρτημένης μεταβλητής y_i με βάση εκείνες των τιμών της ανεξάρτητης x_i . Έτσι, πολλές φορές η μεταβλητή x_i αναφέρεται ως ερμηνευτική μεταβλητή ή παλινδρομητής του υποδείγματος, ενώ η y_i ως ερμηνευόμενη ή παλινδρονούμενη μεταβλητή.

Οι τιμές της μεταβλητής x_i δεν θα πρέπει να αλλάζουν μεταξύ δειγματοληψιών, καθώς μεταχειρίζονται ως σταθερές, όπου η κατανομή τους είναι γνωστή και δεν αλλάζει από δείγμα σε δείγμα. Οπότε η αναμενόμενη μέση τιμή θα ισούται με την αντίστοιχη της παρατήρησης στο δείγμα. Με βάση τα προηγούμενα, για τις τιμές της ανεξάρτητης μεταβλητής μπορούμε να εξηγήσουμε το ερμηνευμένο μέρος του υποδείγματος ($\beta_1 + \beta_2 * x_i$), ως της αναμενόμενη (μέση) τιμή της εξαρτημένης μεταβλητής y_i , που αντιστοιχεί σε μια τιμή x_i της ανεξάρτητης μεταβλητής.

Όπως αναφέρθηκε παραπάνω, η μέση επίδραση των τυχαίων παραγόντων που αντανακλώνται στις τιμές του διαταρακτικού όρου και επηρεάζουν τις μεταβολές της εξαρτημένης μεταβλητής y_i είναι μηδέν. Η υπόθεση αυτή κρίνεται απαραίτητη για τη σωστή εξειδίκευση και αμερόληπτη εκτίμηση του υποδείγματος, καθώς θεωρείται ότι το ερμηνευτικό μέρος του υποδείγματος περιλαμβάνει όλους τους συστηματικούς παράγοντες που επηρεάζουν τις μεταβολές της y_i . Έτσι οι τιμές του διαταρακτικού όρου ε_i αποτελούν τυχαίες αποκλίσεις των τιμών της εξαρτημένης μεταβλητής γύρω από τις μέσες ή προβλεπόμενες τιμές τους που αντιστοιχούν σε τιμές της ανεξάρτητης μεταβλητής.

7.2.2 Εκτίμηση των Συντελεστών του Γραμμικού Υποδείγματος με τη Μέθοδο των Ελαχίστων Τετραγώνων (LS).

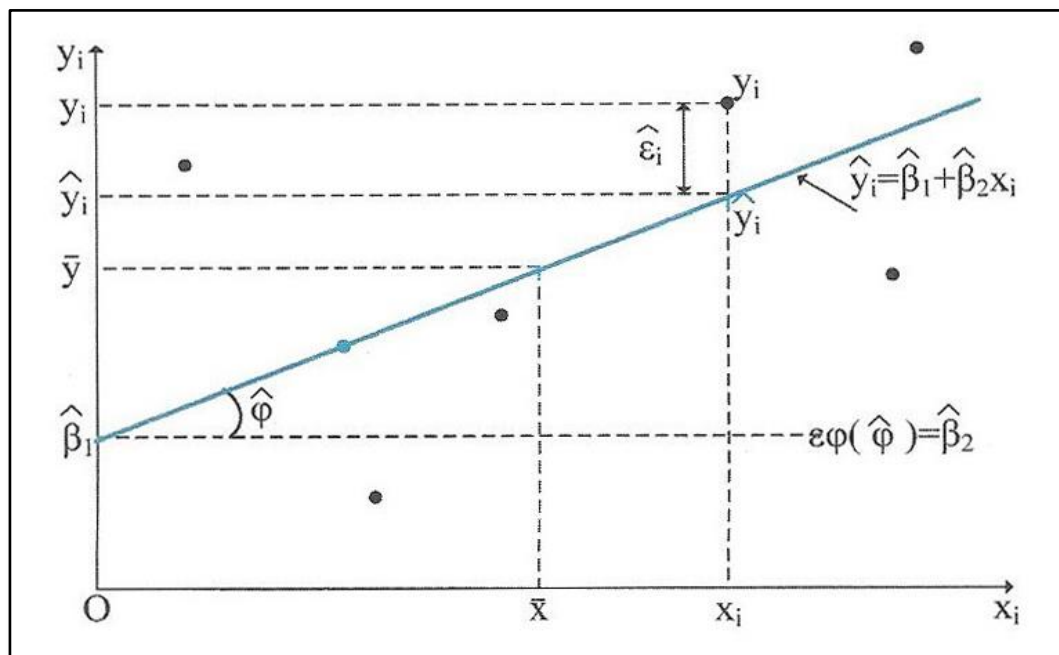
Μια από τις πλέον διαδεδομένες μέθοδοι εκτίμησης των συντελεστών του γραμμικού υποδείγματος με βάση κάποιο δείγμα παρατηρήσεων, είναι αυτή των ελαχίστων τετραγώνων (least squares – LS). Η εκτίμηση των συντελεστών του γραμμικού υποδείγματος, που παρουσιάστηκε προηγουμένως, β_1 και β_2 με την μέθοδο των ελαχίστων τετραγώνων στηρίζεται στην αρχή ότι οι εκτιμήσεις (εκτιμημένες τιμές) της εξαρτημένης μεταβλητής y_i , που θα προκύψουν από την εκτίμηση των συντελεστών β_1 και β_2 , θα πρέπει να έχουν την ελάχιστη δυνατή απόσταση από τις παρατηρούμενες τιμές τους, δηλαδή τις y_i , για όλες τις παρατηρήσεις του δείγματος $i=1,2,\dots,N$.

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 * x_i, \quad i = 1, 2, \dots, N.$$

Όπου με καπέλο “^” παριστάνονται οι εκτιμήσεις (εκτιμημένες τιμές) των συντελεστών β_1 και β_2 (δηλαδή $\hat{\beta}_1$ και $\hat{\beta}_2$) του υποδείγματος στο δείγμα καθώς και της εξαρτημένης μεταβλητής y_i (δηλαδή \hat{y}_i). Οι τελευταίες υπολογίζονται χρησιμοποιώντας τις εκτιμήσεις των συντελεστών $\hat{\beta}_1$ και $\hat{\beta}_2$ και τις τιμές των παρατηρήσεων της ανεξάρτητης μεταβλητής x_i .

Στο Σχήμα 7.1, παρουσιάζεται η εκτιμημένη σχέση του υποδείγματος $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 * x_i$ μαζί με τις παρατηρούμενες τιμές της εξαρτημένης μεταβλητής y_i . Στον κάθετο άξονα δίνονται οι παρατηρούμενες (y_i) και οι εκτιμημένες (\hat{y}_i) τιμές της μεταβλητής αυτής, ενώ στον οριζόντιο δίνονται οι τιμές της ανεξάρτητης μεταβλητής x_i . Στο εσωτερικό του σχήματος παρουσιάζονται ζεύγη παρατηρήσεων της εξαρτημένης και ανεξάρτητης μεταβλητής του υποδείγματος. Στο ίδιο σχήμα επίσης, ορίζονται οι αποστάσεις μεταξύ των παρατηρούμενων και εκτιμημένων τιμών της εξαρτημένης μεταβλητής υποδείγματος y_i και \hat{y}_i ως εξής: $\varepsilon_i = y_i - \hat{y}_i$. Οι αποστάσεις αυτές αποτελούν τα κατάλοιπα (residuals) της εκτίμησης του γραμμικού υποδείγματος. Αυτά θεωρούνται ως εκτιμήσεις των μη

παρατηρούμενων τιμών του διαταρακτικού όρου ε_i , που αντιστοιχούν στις $i=1,2,\dots,N$ παρατηρήσεις του δείγματος. Οι τιμές των καταλοίπων $\hat{\varepsilon}_i$ μπορεί να είναι θετικές ή αρνητικές, ανάλογα με το αν οι παρατηρούμενες τιμές της μεταβλητής y_i βρίσκονται πάνω ή κάτω από τις εκτιμήσεις τους \hat{y}_i , που βρίσκονται στην γραμμή $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 * x_i$. Έχοντας ορίσει τα κατάλοιπα του υποδείγματος ως τις αποστάσεις $\hat{\varepsilon}_i = y_i - \hat{y}_i$, στην συνέχεια θα παρουσιαστεί η μέθοδος εκτίμησης των ελαχίστων τετραγώνων.



Σχήμα 7.1: Σχέση ανάμεσα στις παρατηρούμενες (y_i), τις εκτιμημένες (\hat{y}_i) τιμές της εξαρτημένης μεταβλητής και τα κατάλοιπα $\hat{\varepsilon}_i = y_i - \hat{y}_i$

Η μέθοδος των ελαχίστων τετραγώνων παρέχει εκτιμήσεις των συντελεστών β_1 και β_2 του γραμμικού υποδείγματος, που ελαχιστοποιούν το άθροισμα των τετραγώνων των καταλοίπων (residual sum of squares – RSS) που ορίζεται ως εξής: $RSS(\hat{\beta}_1, \hat{\beta}_2) \equiv \sum_{i=1}^N \hat{\varepsilon}_i^2$. Με βάση αυτή την αρχή, επιτυγχάνει να μειώσει τις αποστάσεις μεταξύ των παρατηρούμενων και εκτιμημένων τιμών της εξαρτημένης μεταβλητής y_i , στο ελάχιστο δυνατό σημείο, έτσι ώστε η εκτιμημένη σχέση του υποδείγματος $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 * x_i$ να περνά διαμέσου των παρατηρούμενων τιμών της μεταβλητής y_i στο δείγμα. Επιπλέον, η μέθοδος των ελαχίστων τετραγώνων επιλέγει να ελαχιστοποιήσει το άθροισμα των τετραγώνων των καταλοίπων $\hat{\varepsilon}_i^2$ και όχι των επιπέδων τους $\hat{\varepsilon}_i$, γιατί αν ελαχιστοποιούσε το τελευταίο, οι αρνητικές τιμές των $\hat{\varepsilon}_i$ θα απαλείφονταν με τις θετικές τους, με συνέπεια να μη μπορεί να βρεθούν μοναδικές εκτιμήσεις (ή λύσεις) για τους συντελεστές του υποδείγματος.

7.2.3 Το Πολλαπλό Γραμμικό Υπόδειγμα

Στην ενότητα αυτή θα παρουσιαστεί το πολλαπλό (multiple) γραμμικό υπόδειγμα. Το υπόδειγμα αυτό περιλαμβάνει περισσότερες από μια ανεξάρτητες μεταβλητές (έστω K μαζί με την σταθερά που παίρνει ως τιμή την μονάδα) και αποτελεί επέκταση του απλού

υποδείγματος που παρουσιάστηκε στην προηγούμενη ενότητα. Έτσι το πολλαπλό γραμμικό υπόδειγμα ορίζεται ως εξής:

$$y_i = \beta_1 + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \dots + \beta_K * x_{iK} + \varepsilon_i, i = 1, 2, \dots, N$$

Όπου $x_{i1} = 1, x_{i2}, x_{i3}, \dots, x_{iK}$ αποτελούν τις ανεξάρτητες μεταβλητές του υποδείγματος και $\beta_1, \beta_2, \beta_3, \dots, \beta_K$ τους K-συντελεστές του που αντιστοιχούν στις ανεξάρτητες μεταβλητές. Για το υπόδειγμα αυτό υποθέτουμε ότι ισχύουν οι κλασικές υποθέσεις, που στηρίχθηκε και το απλό γραμμικό υπόδειγμα. Μια αντιστοιχία του πολλαπλού γραμμικού υποδείγματος με αυτό του απλού γραμμικού υποδείγματος παρουσιάζεται παρακάτω:

$$y_1 = \beta_1 + x_1 * \beta_2 + \varepsilon_1$$

$$y_2 = \beta_1 + x_2 * \beta_2 + \varepsilon_2$$

$$y_3 = \beta_1 + x_3 * \beta_2 + \varepsilon_3$$

$$\vdots \quad \quad \quad \vdots$$

$$y_N = \beta_1 + x_N * \beta_2 + \varepsilon_N$$

Κάθε μια από τις οποίες αντιστοιχεί σε μια από τις παρατηρήσεις $i=1,2,3,\dots,N$ του δείγματος. Επειδή οι εκτιμητές των συντελεστών του πολλαπλού γραμμικού υποδείγματος καθώς και οι διακυμάνσεις τους αποτελούν πολύπλοκες συναρτήσεις, οι οποίες δυσκολεύουν τους αλγεβρικούς χειρισμούς, γίνεται χρήση της γραμμικής άλγεβρας. Η χρήση αυτή διευκολύνει σημαντικά την εξαγωγή των ιδιοτήτων των εκτιμητών της μεθόδου των ελαχίστων τετραγώνων, για τους συντελεστές του πολλαπλού γραμμικού υποδείγματος. Έτσι με την χρήση της γραμμικής άλγεβρας σε ένα πολλαπλό γραμμικό υπόδειγμα K-ανεξάρτητων μεταβλητών το:

$$y_i = \beta_1 + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \dots + \beta_K * x_{iK} + \varepsilon_i$$

μπορεί να παρασταθεί για όλες τις παρατηρήσεις τους δείγματος i ως εξής:

$$\begin{matrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} \\ (NX1) \end{matrix} = \begin{matrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ (NX1) \end{matrix} \beta_1 + \begin{matrix} \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{N2} \end{bmatrix} \\ (NX1) \end{matrix} \beta_2 + \begin{matrix} \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \\ \vdots \\ x_{N3} \end{bmatrix} \\ (NX1) \end{matrix} \beta_3 + \dots + \begin{matrix} \begin{bmatrix} x_{1K} \\ x_{2K} \\ x_{3K} \\ \vdots \\ x_{NK} \end{bmatrix} \\ (NX1) \end{matrix} \beta_K + \begin{matrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_K \end{bmatrix} \\ (NX1) \end{matrix}$$

ή ως:

$$y = (NX1)\beta_1 + (NX1)\beta_2 + (NX1)\beta_3 + \dots + (NX1)\beta_K + \varepsilon =$$

$$[1 \ x_2 \ \dots \ x_K] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix} + \varepsilon = \begin{matrix} X & \beta \\ (NXK) & (NXK) \end{matrix} + \begin{matrix} \varepsilon \\ (NXK) \end{matrix}$$

Όπου το $\begin{matrix} \beta \\ (NXK) \end{matrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix}$ αποτελεί το διάνυσμα των K συντελεστών του υποδείγματος, ενώ το

$\begin{matrix} X \\ (NXK) \end{matrix} = [1 \ x_2 \ \dots \ x_K] = \begin{bmatrix} 1 & x_{12} & \dots & x_{1K} \\ 1 & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N2} & \dots & x_{NK} \end{bmatrix}$ είναι η μήτρα που έχει ως στήλες τις τις

παρατηρήσεις των K ανεξάρτητων μεταβλητών του υποδείγματος μαζί με τη σταθερά. Επομένως, όπως αναφέρθηκε και παραπάνω, ο γενικός τύπος του εκτιμητή $\hat{\beta}$ της μεθόδου των ελαχίστων τετραγώνων, του διανύσματος των συντελεστών του υποδείγματος β θα παραμείνει αμετάβλητος.

7.2.4 Προβλέψεις και Προβλεπτική Ικανότητα του Γραμμικού Υποδείγματος

Όπως αναφέρθηκε στις Ενότητες 7.2.1, 7.2.2, 7.2.3, τα γραμμικά υποδείγματα βοηθούν στην εκτίμηση και έλεγχο υποθέσεων. Επιπλέον, όμως ένας ακόμα σκοπός τους είναι η χρήση τους για παροχή προβλέψεων. Χρησιμοποιώντας το απλό υπόδειγμα ως παράδειγμα, μπορεί να θεωρηθεί πρόβλεψη η εκτίμηση της εξαρτημένης μεταβλητής ενός γραμμικού υποδείγματος, που αντιστοιχεί σε μια τιμή της ανεξάρτητης μεταβλητής και προκύπτει με βάση τις εκτιμήσεις της μεθόδου των ελαχίστων τετραγώνων, των συντελεστών του υποδείγματος, για κάποιο άλλο δείγμα παρατηρήσεων.

Η δυνατότητα ενός υποδείγματος να παρέχει μεγάλης ακρίβειας προβλέψεις είναι σημαντική για τους αναλυτές που ενδιαφέρονται να προβλέψουν, για παράδειγμα, τα μελλοντικά επίπεδα ζήτησης μιας υπηρεσίας από τους χρήστες. Γενικά, η παροχή ακριβών προβλέψεων είναι χρήσιμη για την λήψη βέλτιστων αποφάσεων και διαχείρισης μελλοντικών καταστάσεων.

Στην ενότητα αυτή θα παρουσιαστεί πως υπολογίζεται η πρόβλεψη ενός απλού ή πολλαπλού γραμμικού υποδείγματος, καθώς και το διάστημα πρόβλεψης. Αυτό γίνεται στηριζόμενοι στον εκτιμητή ελαχίστων τετραγώνων για την εκτίμηση των συντελεστών του υποδείγματος. Επειδή οι εκτιμητές αυτοί παρέχουν αμερόληπτες και αποτελεσματικές εκτιμήσεις των συντελεστών του υποδείγματος, οι προβλέψεις που θα προκύπτουν μέσω αυτών, αποδεικνύεται ότι θα έχουν το ελάχιστο δυνατό σφάλμα πρόβλεψης. Επίσης θα παρουσιαστεί ένα μέτρο προβλεπτικής (ή ερμηνευτικής) ικανότητας του υποδείγματος, το οποίο προσφέρει την δυνατότητα να αξιολογηθεί, για πόσο κοντά είναι οι προβλέψεις ενός

υποδείγματος, για την εξαρτημένη μεταβλητή στις πραγματικές τιμές αυτής, στο δείγμα. Το μέτρο αυτό ονομάζεται συντελεστής προσδιορισμού (coefficient of determination) και συμβολίζεται ως R^2 . Χρησιμοποιείται δε ευρέως στη σύγκριση υποδειγμάτων και την επιλογή των πιο κατάλληλων μεταβλητών σε ένα υπόδειγμα. Ο συντελεστής αυτός αποτελεί το τετράγωνο του συντελεστή συσχέτισης μεταξύ των προβλέψεων ενός υποδείγματος για την εξαρτημένη μεταβλητή και των πραγματικών της τιμών στο δείγμα.

Πρόβλεψη με Χρήση Απλού Γραμμικού Υποδείγματος

Για να παρουσιαστεί ο τρόπος υπολογισμού της πρόβλεψης για την εξαρτημένη μεταβλητή y_i ενός υποδείγματος, θεωρείται αρχικά ότι η μεταβλητή αυτή προσδιορίζεται από το ακόλουθο απλό γραμμικό υπόδειγμα:

$$y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$$

Αν υποθεθεί ότι η x_0 αποτελεί μια τιμή της ανεξάρτητης μεταβλητής του γραμμικού υποδείγματος, για την οποία θέλουμε να βρούμε την πρόβλεψη της εξαρτημένης μεταβλητής του υποδείγματος y_i , η οποία για την συνέχεια θα συμβολίζεται ως y_0 . Η πρόβλεψη αυτή της μεταβλητής y_i αναφέρεται ως πρόβλεψη στο σημείο «0» της ανεξάρτητης μεταβλητής x_i (εννοώντας την τιμή της x_0). Η τιμή αυτή της ανεξάρτητης μεταβλητής μπορεί να μην είναι κάποια γνωστή τιμή του δείγματος των παρατηρήσεων, αλλά κάποια άλλη τιμή. Αν η x_0 δεν ανήκει στο δείγμα, τότε η πρόβλεψη y_0 θα αναφέρεται ως εκτός του δείγματος, ενώ αν ανήκει, ως εντός του δείγματος. Έτσι με βάση το απλό γραμμικό υπόδειγμα, η αληθινή τιμή της μεταβλητής y_i στο σημείο «0» δίνεται ως εξής:

$$y_0 = \beta_1 + \beta_2 * x_0 + \varepsilon_0$$

Ενώ η πρόβλεψη του δείγματος ως η ακόλουθη μέση τιμή της στο σημείο «0»:

$$E(y_0) = \beta_1 + \beta_2 * x_0$$

καθώς $E(\varepsilon_0) = 0$. Μια αμερόληπτη εκτίμηση της πρόβλεψης αυτής της εξαρτημένης μεταβλητής, δίνεται αν αντικατασταθούν οι συντελεστές του υποδείγματος της εκτιμήσεις ελαχίστων τετραγώνων τους με βάση κάποιο δείγμα παρατηρήσεων, δηλαδή:

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 * x_0$$

Όπου οι $\hat{\beta}_1$ και $\hat{\beta}_2$ αποτελούν τους εκτιμητές ελαχίστων τετραγώνων των συντελεστών β_1 και β_2 του υποδείγματος. Η εκτίμηση αυτή της εξαρτημένης μεταβλητής στο σημείο «0», \hat{y}_0 , αναφέρεται ως πρόβλεψη της y_0 στο δείγμα ή απλά ως πρόβλεψη. Επίσης η σχέση $\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 * x_0$, με βάση την οποία υπολογίστηκε η πρόβλεψη, αναφέρεται ως εκτιμητής της πρόβλεψης, καθώς βασίζεται στους εκτιμητές $\hat{\beta}_1$ και $\hat{\beta}_2$. Η διαφορά ανάμεσα στην πρόβλεψη \hat{y}_0 και την αληθινή τιμή της y_0 , αναφέρεται ως σφάλμα πρόβλεψης και γράφεται ως:

$$\hat{y}_0 - y_0 = (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) * x_0 - \varepsilon_0$$

Όπως δείχνει ο παραπάνω ορισμός, το σφάλμα πρόβλεψης είναι ανάλογο των καταλοίπων του εκτιμημένου υποδείγματος.

Πρόβλεψη με Χρήση Πολλαπλού Γραμμικού Υποδείγματος

Ανάλογα με το απλό γραμμικό υπόδειγμα, μπορεί να υπολογιστεί και να προβλεφθεί η τιμή της εξαρτημένης μεταβλητής, για το πολλαπλό γραμμικό υπόδειγμα, το οποίο περιλαμβάνει K ανεξάρτητες μεταβλητές:

$$y_i = \beta_1 + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \dots + \beta_K * x_{iK} + \varepsilon_i, i = 1, 2, \dots, N$$

Αν το διάνυσμα των τιμών των ανεξάρτητων μεταβλητών του υποδείγματος στο σημείο της πρόβλεψης «0» ορίζεται ως το διάνυσμα γραμμής

$$\begin{matrix} x'_0 \\ (K \times 1) \end{matrix} = [1, x_{02}, \dots, x_{0K}],$$

τότε η πρόβλεψη ελαχίστων τετραγώνων του υποδείγματος για την ανεξάρτητη μεταβλητή στο σημείο αυτό, υπολογίζεται ως εξής:

$$\begin{matrix} \hat{y}_0 \\ (1 \times 1) \end{matrix} = [1, x_{02}, \dots, x_{0K}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} = x'_0 \hat{\beta},$$

Όπου το $\hat{\beta}$ αποτελεί τον εκτιμητή ελαχίστων τετραγώνων του διανύσματος των συντελεστών του υποδείγματος β .

Το σφάλμα της πρόβλεψης \hat{y}_0 για το πολλαπλό υπόδειγμα ορίζεται ως:

$$\hat{y}_0 - y_0 = x'_0 \hat{\beta} - (x'_0 \beta + \varepsilon_0) = x'_0 (\hat{\beta} - \beta) - \varepsilon_0$$

Ορισμός του Συντελεστή Προσδιορισμού R^2

Για να κριθεί το κατά πόσο ένα γραμμικό υπόδειγμα έχει την ικανότητα να προβλέπει ικανοποιητικά τις μεταβολές, της εξαρτημένης μεταβλητής του υποδείγματος, για οποιαδήποτε τιμή των ανεξάρτητων μεταβλητών του, θα πρέπει να οριστεί ένα μέτρο που θα αποτελεί το συντελεστή προσδιορισμού R^2 . Επειδή υψηλή προβλεπτική ικανότητα σημαίνει και μεγάλη ικανότητα προσαρμογής του υποδείγματος στα δεδομένα του δείγματος, ο συντελεστής αυτός αναφέρεται και ως συντελεστής προσαρμογής του υποδείγματος.

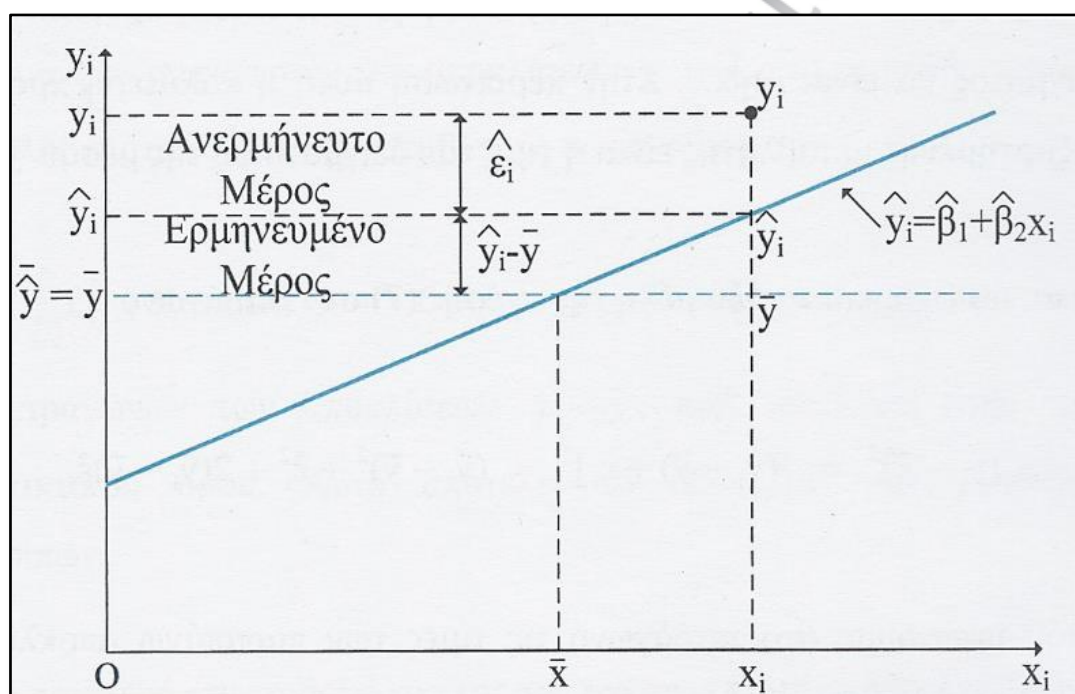
Για να προσδιοριστεί ο συντελεστής R^2 , ορίζεται η τιμή της αντιπροσωπευτικής παρατήρησης της εξαρτημένης μεταβλητής y_i , ως το άθροισμα της τιμής που προβλέπεται (εκτιμάται) με βάση το υπόδειγμα χρησιμοποιώντας τον εκτιμητή ελαχίστων τετραγώνων και αυτό που παραμένει ανερμήνευτο και ορίζει τα κατάλοιπα. Δηλαδή ως:

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

Αφαιρώντας και από τα δύο μέλη της τελευταίας σχέσης το δειγματικό μέσο της μεταβλητής y_i , \bar{y} , η συνάρτηση γίνεται:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + \hat{\varepsilon}_i$$

Όπως φαίνεται, στη τελευταία σχέση και από το Σχήμα 7.2, η απόκλιση της τιμής της εξαρτημένης μεταβλητής y_i από το μέσο της \bar{y} , $(y_i - \bar{y})$, αποτελείται από ένα μέρος που προβλέπεται (ή ερμηνεύεται) από το υπόδειγμα $(\hat{y}_i - \bar{y})$ και το κατάλοιπο $\hat{\varepsilon}_i$. Όσο μεγαλύτερο είναι το ερμηνευμένο μέρος $(\hat{y}_i - \bar{y})$, τόσο μεγαλύτερη θα είναι και η ικανότητα του γραμμικού υποδείγματος να προβλέπει τις μεταβολές της εξαρτημένης μεταβλητής y_i , με βάση εκείνες της ανεξάρτητης μεταβλητής x_i , για κάθε παρατήρηση i του δείγματος. Όταν $(\hat{y}_i - \bar{y}) = 0$ ή $(\hat{y}_i = \bar{y})$, τότε η προβλεπτική ικανότητα του υποδείγματος θα είναι μηδέν. Στην περίπτωση αυτή η καλύτερη πρόβλεψη για τις τιμές της εξαρτημένης μεταβλητής είναι η τιμή του δείγματος της μέσου \bar{y} .



Σχήμα 7.2: Ερμηνευμένο και ανερμηνευμένο μέρος του δείγματος

Υψώνοντας τα δύο μέλη της προηγούμενης σχέσης στο τετράγωνο, προκύπτει:

$$(y_i - \bar{y})^2 = [(\hat{y}_i - \bar{y}) + \hat{\varepsilon}_i]^2 = (y_i - \bar{y})^2 + \hat{\varepsilon}_i^2 + 2(\hat{y}_i - \bar{y})\hat{\varepsilon}_i$$

Ο λόγος που υψώθηκαν στο τετράγωνο οι τιμές των παραπάνω αποκλίσεων για την κατασκευή του συντελεστή R^2 είναι για να αποφευχθεί η απαλοιφή των θετικών με τις αρνητικές τιμές των αποκλίσεων $y_i - \bar{y}$ μεταξύ τους (ή των προβλεπόμενων τους τιμών $\hat{y}_i - \bar{y}$), όταν αθροίζονται και έτσι αυτός να μην αποτελεί ένα αξιόπιστο μέτρο της προβλεπτικής ικανότητας του υποδείγματος. Παίρνοντας το άθροισμα της παραπάνω σχέσης για όλες τις παρατηρήσεις του δείγματος συνεπάγεται:

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N \hat{\varepsilon}_i^2 + 2 \sum_{i=1}^N (\hat{y}_i - \bar{y})\hat{\varepsilon}_i = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N \hat{\varepsilon}_i^2$$

καθώς ισχύει ότι $\sum_{i=1}^N (\hat{y}_i - \bar{y})\varepsilon_i = 0$, λόγω της ιδιότητας των καταλοίπων να είναι ορθογώνια των προβλεπόμενων τιμών και ότι αθροίζουν στο μηδέν, η τελευταία σχέση μπορεί να οριστεί και ως εξής:

$$TSS = ESS + RSS,$$

όπου $TSS = \sum_{i=1}^N (y_i - \bar{y})^2$ (Total Sum of Squares) ή αλλιώς Συνολικό Άθροισμα Τετραγώνων των αποκλίσεων του y_i από το δειγματικό τους μέσο \bar{y} .

$ESS = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$ (Explained Sum of Squares), που αποτελεί το ερμηνευμένο άθροισμα των τετραγώνων των αποκλίσεων $y_i - \bar{y}$ από το υπόδειγμα με βάση τις μεταβολές της ανεξάρτητης μεταβλητής x_i .

Τέλος, $RSS = \sum_{i=1}^N \hat{\varepsilon}_i^2$ (Residual Sum of Squares) που αποτελεί το ανερμήνευτο άθροισμα των τετραγώνων των αποκλίσεων $y_i - \bar{y}$, που οφείλεται στην ύπαρξη του διαταρακτικού όρου. Αυτό αποτελεί το άθροισμα των τετραγώνων των καταλοίπων.

Διαιρώντας το αριστερό και το δεξιό σκέλος της τελευταίας σχέσης με το άθροισμα TSS δίνει την ακόλουθη σχέση:

$$\frac{TSS}{TSS} = \frac{ESS}{TSS} + \frac{RSS}{TSS} \Rightarrow R^2 \equiv \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Που αποτελεί τον ορισμό του συντελεστή προσδιορισμού R^2 . Αντικαθιστώντας τους ορισμούς των αθροισμάτων TSS, ESS και RSS, στην τελευταία σχέση ο συντελεστής R^2 γράφεται πιο αναλυτικά ως εξής:

$$R^2 \equiv 1 - \frac{\sum_{i=1}^N \hat{\varepsilon}_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Εξ ορισμού, ο συντελεστής R^2 που δίνεται από την προηγούμενη σχέση, μετρά το ποσοστό του αθροίσματος των τετραγώνων των αποκλίσεων $y_i - \bar{y}$ για όλες τις παρατηρήσεις του δείγματος που ερμηνεύεται με βάση τις ανεξάρτητες μεταβλητές του γραμμικού υποδείγματος. Οι τιμές που παίρνει αυτός βρίσκονται στο ανοιχτό διάστημα (0,1), δηλαδή ισχύει $0 < R^2 < 1$. Όταν η τιμή του R^2 είναι κοντά στην μονάδα, τότε αυτό σημαίνει ότι η προβλεπτική ικανότητα του υποδείγματος είναι πολύ υψηλή. Στην περίπτωση αυτή το ανερμήνευτο άθροισμα των τετραγώνων των αποκλίσεων RSS είναι πολύ μικρό σε σχέση με το TSS. Το αντίθετο συμβαίνει αν η τιμή του R^2 είναι κοντά στο μηδέν.

Η χρήση του συντελεστή προσδιορισμού R^2 στη μέτρηση της προβλεπτικής ικανότητας του υποδείγματος παρουσιάζει ένα μειονέκτημα. Αν το δείγμα των παρατηρήσεων είναι αρκετά μικρό, τότε αυξάνοντας τον αριθμό των ερμηνευτικών μεταβλητών, η τιμή του R^2 μπορεί να αυξηθεί σημαντικά και μάλιστα, να τείνει προς την μονάδα, πλασματικά. Δηλαδή, χωρίς να δείχνει την πραγματική ερμηνευτική ικανότητα του υποδείγματος. Αυτό οφείλεται στο μικρό αριθμό των βαθμών ελευθερίας που είναι διαθέσιμοι στην εκτίμηση των συντελεστών του υποδείγματος. Το πρόβλημα αυτό παρουσιάζεται όταν ο αριθμός των

παρατηρήσεων του δείγματος είναι μικρός σχετικά με εκείνο των συντελεστών υποδείγματος. Μπορεί δε να διαπιστωθεί εύκολα από τη σχέση του ορισμού του συντελεστή R^2 , $R^2 = 1 - \frac{RSS}{TSS}$. Όπως δείχνει η σχέση αυτή, μια σημαντική μείωση των βαθμών ελευθερίας θα προκαλέσει σημαντική μείωση του ανερμήνευτου αθροίσματος τετραγώνων RSS (δηλαδή $\sum_{i=1}^N \hat{\varepsilon}_i^2$) σε σχέση με το συνολικό TSS (δηλαδή $\sum_{i=1}^N (y_i - \bar{y})^2$) με συνέπεια να οδηγήσει σε σημαντική αύξηση της τιμής του συντελεστή R^2 .

7.2.5 Εκτίμηση Μη Γραμμικών Υποδειγμάτων

Στην μέχρι τώρα σχετικά σύντομη ανάλυση παρουσιάστηκαν υποδείγματα, στα οποία η εξαρτημένη μεταβλητή y_i θεωρήθηκε ότι αποτελεί γραμμική συνάρτηση ενός συνόλου ανεξάρτητων μεταβλητών (έστω K), που συλλέγονται στο διάνυσμα $x_i = (x_{i1}, x_{i2}, \dots, x_{iK})'$, για κάθε παρατήρηση i . Πιο συγκεκριμένα, η γραμμικότητα των υποδειγμάτων αυτών ορίστηκε ως προς τους συντελεστές του υποδείγματος βήτα, που συλλέγονται στο διάνυσμα $\beta = (\beta_1, \beta_2, \dots, \beta_K)'$.

Συνέπεια του παραπάνω ορισμού της γραμμικότητας των υποδειγμάτων είναι ότι οι συντελεστές τους μπορούν να εκτιμηθούν στηριζόμενοι σε έναν γραμμικό, αμερόληπτο και αποτελεσματικό εκτιμητή, όπως αυτόν των ελαχίστων τετραγώνων. Κάτω από τις κλασικές υποθέσεις και την υπόθεση ότι ο διαταρακτικός όρος ακολουθεί την κανονική κατανομή, αποδείχθηκε ότι ο εκτιμητής αυτός ακολουθεί την κανονική κατανομή για πεπερασμένα δείγματα παρατηρήσεων. Όπως τόσο στην πράξη, όσο και στην θεωρία, πολλές φορές συναντώνται υποδείγματα των οποίων η συναρτησιακή σχέση που συνδέει την εξαρτημένη μεταβλητή με τις ανεξάρτητες, δεν είναι γραμμική στις παραμέτρους. Ένα τέτοιο παράδειγμα αποτελεί το ακόλουθο υπόδειγμα:

$$y_i = \beta_1 + \beta_2 x_i^2 + \varepsilon_i \quad \text{με} \quad f(x_i, \beta) = \beta_1 + \beta_2 x_i^2$$

Ή το υπόδειγμα

$$y_i = \beta x_{i1} + \beta^2 x_{i2} + \varepsilon_i \quad \text{με} \quad f(x_i, \beta) = \beta x_{i1} + \beta^2 x_{i2}$$

Όπου $f(x_i, \beta)$ αποτελεί τη συναρτησιακή σχέση που υπάρχει ανάμεσα στην εξαρτημένη y_i και τις ανεξάρτητες μεταβλητές του υποδείγματος, που συλλέγονται στο διάνυσμα $x_i = (x_{i1}, x_{i2}, \dots, x_{iK})'$. Με άλλα λόγια, η συνάρτηση αυτή καθορίζει το ερμηνευμένο μέρος του υποδείγματος. Μια από της μεθόδους εκτίμησης των συντελεστών ενός μη γραμμικού υποδείγματος είναι η μέθοδος ελαχίστων τετραγώνων. Για να ευκολότερη κατανόηση θεωρείται το ακόλουθο υπόδειγμα:

$$y_i = f(x_i, \beta) + \varepsilon_i = \beta x_{i1} + \beta^2 x_{i2} + \varepsilon_i$$

Όπου η συναρτησιακή σχέση $f(x_i, \beta) = \beta x_{i1} + \beta^2 x_{i2}$ καθορίζεται μη γραμμικά μόνο από ένα συντελεστή, τον β . Αντίστοιχα αυτά που θα αναφερθούν ισχύουν και για παραπάνω από ένα συντελεστές.

Σύμφωνα με την αρχή της μεθόδου αυτής, ο εκτιμητής του συντελεστή β του υποδείγματος, που συνήθως συμβολίζεται με $\hat{\beta}$, θα πρέπει να ελαχιστοποιεί το άθροισμα των τετραγώνων των καταλοίπων αυτού ε_i . Δηλαδή, αυτός βρίσκεται ως λύση του ακόλουθου προβλήματος:

$$\underset{\hat{\beta}}{\text{ελαχ}} \text{RSS}(\hat{\beta}) \equiv \underset{\hat{\beta}}{\text{ελαχ}} \sum_{i=1}^N [y_i - f(x_i, \hat{\beta})]^2$$

Όπου $f(x_i, \hat{\beta}) = \hat{\beta}x_{i1} + \hat{\beta}^2x_{i2}$ και $\varepsilon_i = y_i - f(x_i, \hat{\beta})$ αποτελούν τα κατάλοιπα του υποδείγματος. Για τη λύση του προβλήματος αυτού ως προς τον εκτιμητή $\hat{\beta}$ απαιτείται η ικανοποίηση της ακόλουθης συνθήκης πρώτης τάξης:

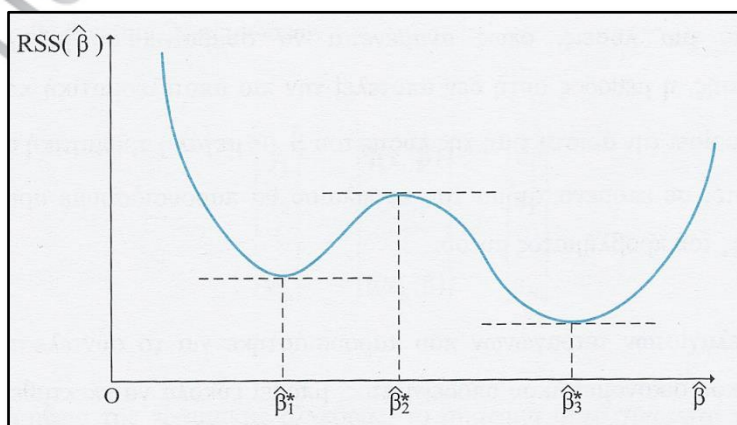
$$\frac{\partial \text{RSS}(\hat{\beta})}{\partial \hat{\beta}} = \frac{\partial}{\partial \hat{\beta}} \sum_{i=1}^N [y_i - f(x_i, \hat{\beta})]^2 = 0$$

Με την χρήση αλγεβρικών πράξεων η προηγούμενη σχέση καταλήγει στην ακόλουθη σχέση:

$$2\hat{\beta}^3 \sum_{i=1}^N x_{i2}^2 + 3\hat{\beta}^2 \sum_{i=1}^N x_{i1}x_{i2} + \hat{\beta} \left(\sum_{i=1}^N x_{i1}^2 - 2 \sum_{i=1}^N x_{i2}y_i \right) - \sum_{i=1}^N x_{i1}y_i = 0 \quad (1)$$

Η τελευταία σχέση αποτελεί μια πολυωνυμική συνάρτηση του εκτιμητή $\hat{\beta}$ τρίτου βαθμού. Έτσι η λύση της θα δίνει τρεις ξεχωριστές λύσεις για τον $\hat{\beta}$. Οι λύσεις αυτές δεν αποτελούν γραμμικές συναρτήσεις των παρατηρήσεων της ανεξάρτητης μεταβλητής y_i , όπως συνέβαινε για τον εκτιμητή ελαχίστων τετραγώνων. Για το λόγο αυτό, ο εκτιμητής $\hat{\beta}$ αναφέρεται συχνά και ως μη γραμμικός εκτιμητής ελαχίστων τετραγώνων (non linear least squares – NLLS).

Το παραπάνω δείχνει καθαρά τα προβλήματα που υπάρχουν όταν εφαρμόζεται η μέθοδος ελαχίστων τετραγώνων για να εκτιμηθούν οι συντελεστές ενός μη γραμμικού υποδείγματος. Για να γίνει πιο κατανοητό στο Σχήμα 7.3, παρουσιάζεται η συνάρτηση των ελαχίστων τετραγώνων $\text{RSS}(\hat{\beta})$ για διάφορες τιμές του $\hat{\beta}$.



Σχήμα 7.3: Εκτίμηση συντελεστών μη γραμμικού υποδείγματος

Όπως δείχνει το σχήμα αυτό, υπάρχουν τρία σημεία όπου η πρώτη παράγωγος $\frac{\partial RSS(\hat{\beta})}{\partial \hat{\beta}}$ ισούται με μηδέν και επομένως, η συνάρτηση $RSS(\hat{\beta})$ ελαχιστοποιείται. Τα σημεία αυτά απεικονίζουν τις τρεις δυνατές λύσεις του μη γραμμικού εκτιμητή ελαχίστων τετραγώνων $\hat{\beta}$, που προκύπτουν λύνοντας την πολυωνυμική εξίσωση (1). Έστω ότι αυτές συμβολίζονται ως $\hat{\beta}_1^*$, $\hat{\beta}_2^*$ και $\hat{\beta}_3^*$ αντίστοιχα, ξεκινώντας από την αρχή των αξόνων. Από τις λύσεις αυτές, εκείνη που αντιστοιχεί στο σημείο $\hat{\beta}_3^*$ αποτελεί την λύση του εκτιμητή ελαχίστων τετραγώνων για την οποία η συνάρτηση $RSS(\hat{\beta})$ λαμβάνει την ελάχιστη τιμή της μεταξύ όλων των λύσεων της. Η λύση αυτή αναφέρεται ως ολικό ελάχιστο (global minimum) της συνάρτησης $RSS(\hat{\beta})$. Σε αντίθεση με τη λύση αυτή, εκείνη που αντιστοιχεί στο σημείο $\hat{\beta}_1^*$ αποτελεί ένα τοπικό ελάχιστο (local minimum) της συνάρτησης $RSS(\hat{\beta})$, καθώς μόνο στην γειτονιά γύρω από το σημείο αυτό αυτή ελαχιστοποιείται.

Από την ανάλυση του παραπάνω σχήματος γίνεται αντιληπτό ότι ένας τρόπος για να βρεθεί η άριστη τιμή του εκτιμητή ελαχίστων τετραγώνων $\hat{\beta}$, για μη γραμμικά υποδείγματα στους συντελεστές τους, θα ήταν να παρασταθεί γραφικά η συνάρτηση των ελαχίστων τετραγώνων $RSS(\hat{\beta})$ ως προς διάφορες τιμές του εκτιμητή $\hat{\beta}$ και στην συνέχεια, να επιλεγεί η τιμή του $\hat{\beta}$ που αντιστοιχεί στο ολικό ελάχιστο της συνάρτησης αυτής ως η τιμή του εκτιμητή ελαχίστων τετραγώνων. Όμως η γραφική μέθοδος αυτή δεν είναι πάντα πρακτική, ιδιαίτερα όταν το υπόδειγμα έχει παραπάνω από ένα συντελεστές και καταλήγει σε περισσότερες από μια λύσεις, όπως αναμένεται να συμβαίνει σε μη γραμμικά υποδείγματα.

Τέλος, η μέθοδος ελαχίστων τετραγώνων που παρουσιάστηκε για το συντελεστή β του απλού μη γραμμικού υποδείματος, μπορεί εύκολα να επεκταθεί και για πιο πολύπλοκα υποδείγματα, που αποτελούνται από K συντελεστές. Στην περίπτωση αυτή, αλλά και γενικότερα, ο αριθμός των ανεξάρτητων μεταβλητών κάποιου μη γραμμικού υποδείματος δεν είναι απαραίτητο να ισούται με εκείνο των συντελεστών του υποδείματος, όπως συμβαίνει στα γραμμικά υποδείγματα. Αυτός μπορεί να είναι μικρότερος ή μεγαλύτερος.

7.3: Τεκμηρίωση Μορφής Υποδείματος

Για την δημιουργία του μοντέλου πρόβλεψης, θα πρέπει για αρχή να προσδιοριστεί η μορφή του υποδείματος. Αν μπορεί να αντιμετωπιστεί σαν γραμμικό ή μη γραμμικό υπόδειγμα. Επιπλέον θα πρέπει να καθοριστεί το αν είναι ένα απλό ή πολλαπλό υπόδειγμα. Έτσι με τον προσδιορισμό των παραπάνω και την χρήση της προβλεπτικής ικανότητας των υποδειμάτων, θα δημιουργηθεί ένα ή περισσότερα μοντέλα πρόβλεψης. Στην συνέχεια για τα μοντέλα αυτά θα γίνει αξιολόγηση με την βοήθεια του εκτιμητή ελαχίστων τετραγώνων και άλλων δεικτών. Στην ενότητα αυτή, θα παρουσιαστεί η ανάλυση όλων των παραπάνω με την χρήση κατάλληλων προγραμμάτων για την κατασκευή ενός ή περισσότερων προβλεπτικών μοντέλων ή λύσεων, με την καλύτερη δυνατή προσέγγιση.

7.3.1 Καθορισμός Πολλαπλότητας Υποδείγματος

Για τον καθορισμό της μορφής του υποδείγματος, θα πρέπει να αναπαρασταθεί στο σύστημα αξόνων x, y . Για αρχή όμως, θα μελετηθεί η πολλαπλότητα του, για το αν είναι απλό ή πολλαπλό υπόδειγμα. Οι μεταβλητές που μελετώνται από την παρούσα εργασία, και επηρεάζουν την επίδοση της υπηρεσίας είναι:

- Το πλήθος των παράλληλων αιτημάτων που εξυπηρετούνται
- Η γεωγραφική απόσταση από όπου αποστέλλονται τα αιτήματα μέχρι τις υποδομές του υπολογιστικού νέφους της υπηρεσίας
- Ο χρόνος κατά τον οποίο εκτελείται η δειγματοληψία.

Αυτό συνεπάγεται ότι η συνάρτηση που θα αναπαριστά την τιμή του χρόνου απόκρισης της υπηρεσίας, θα είναι της μορφής:

$$y_i = \beta_1 + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \beta_4 * x_{i4} + \varepsilon_i, \quad i = 1, 2, 3, \dots, N$$

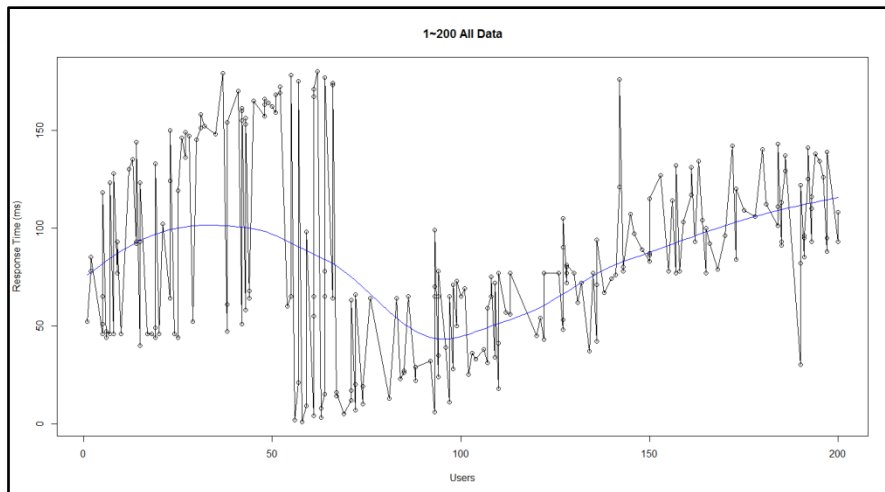
Όπου η $x_{i1} = 1$, και οι x_{i2}, x_{i3}, x_{i4} αποτελούν τις ανεξάρτητες μεταβλητές του υποδείγματος, τα $\beta_1, \beta_2, \beta_3, \beta_4$ τους συντελεστές του, που αντιστοιχούν στις ανεξάρτητες μεταβλητές, και η y_i αποτελεί την εξαρτημένη μεταβλητή, για τα N σημεία του υποδείγματος.

Άρα, με βάση αυτά που αναφέρθηκαν στην Ενότητα 7.2, το συγκεκριμένο υπόδειγμα θα είναι πολλαπλό με τρεις ανεξάρτητες μεταβλητές. Οι μεταβλητές αυτές είναι ανεξάρτητες καθώς δεν επηρεάζει η μια την άλλη, ούτε μπορεί η τιμή της μιας να προβλεφθεί με βάση την τιμή κάποια άλλης μεταβλητής.

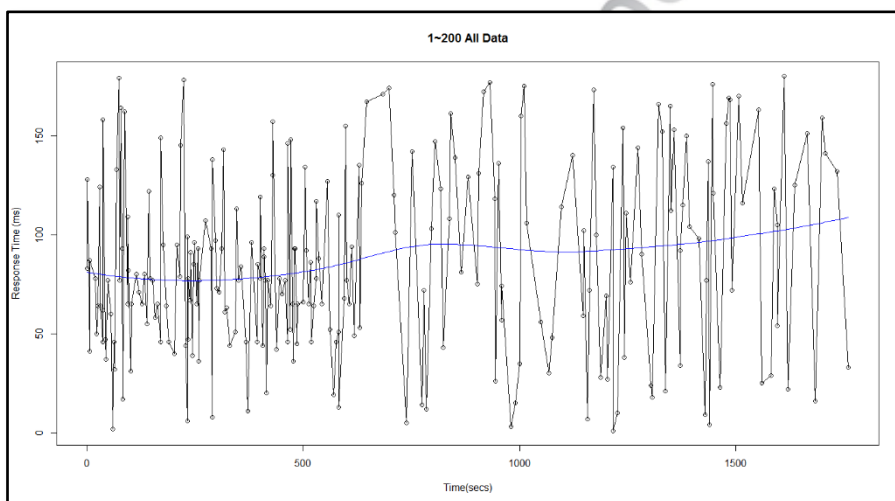
7.3.2 Καθορισμός Γραμμικότητας του Υποδείγματος

Για τον προσδιορισμό της γραμμικότητας ή μη, του υποδείγματος, θα πρέπει να αναπαρασταθεί στο σύστημα αξόνων x, y, z . Για τον λόγο αυτό, αλλά και για την επεξεργασία και ανάλυση των δειγματοληψιών χρησιμοποιήθηκε το εργαλείο της γλώσσας προγραμματισμού R.

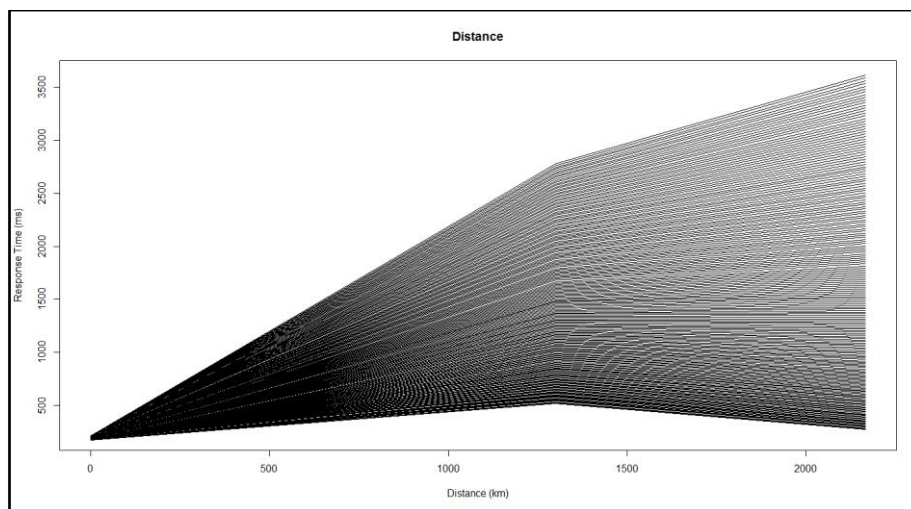
Στην συνέχεια παρουσιάζονται τα Σχήματα 7.4, 7.5, 7.6, που προέκυψαν από τα αποτελέσματα επεξεργασίας και ανάλυσης των δεδομένων των δειγματοληψιών, πάνω στο σύστημα αξόνων x, y , για κάθε μια από τις ανεξάρτητες μεταβλητές (αριθμός αιτημάτων, χιλιομετρική απόσταση, χρόνος εκτέλεσης), με την εξαρτημένη (χρόνος απόκρισης). Περισσότερες πληροφορίες σχετικά με την επεξεργασία και ανάλυση τους, έγινε αναφορά στο Κεφάλαιο 6.



Σχήμα 7.4: Αναπαράσταση της μεταβλητής των αιτημάτων των χρηστών με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα



Σχήμα 7.6: Αναπαράσταση της μεταβλητής του χρόνου εκτέλεσης με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα



Σχήμα 7.5: Αναπαράσταση της μεταβλητής της χιλιομετρικής απόστασης με την μεταβλητή του χρόνου απόκρισης, στο σύστημα αξόνων, για όλα τα δεδομένα

Από την ανάλυση των διαγραμμάτων, είναι φανερό ότι πρόκειται για ένα μη γραμμικό υπόδειγμα. Ο λόγος που παρουσιάστηκαν ξεχωριστά οι ανεξάρτητες μεταβλητές με την εξαρτημένη, στα τρία προηγούμενα διαγράμματα, είναι επειδή δεν μπορούν να αναπαρασταθούν οι τέσσερις μεταβλητές σε διάγραμμα πάνω από τρεις διαστάσεις. Η καμπυλωτή γραμμή που είναι σχεδιασμένη στα διαγράμματα, αποτελεί μια συνάρτηση που δημιουργήθηκε με την χρήση των έτοιμων συναρτήσεων της γλώσσας R πάνω στα δεδομένα, για να δείξει μια προσεγγιστική μη γραμμική συνάρτηση. Έτσι είναι ξεκάθαρο, ότι δεν πρόκειται για ένα γραμμικό υπόδειγμα, αλλά για ένα μη γραμμικό.

7.4: Εύρεση Μοντέλων Πρόβλεψης με Χρήση του Eureqa

Η εφαρμογή Eureqa[53] αποτελεί ένα εργαλείο που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα υπολογιστών[62] και μαζί με την χρήση μαθηματικών μοντέλων, εκτελεί δοκιμές στις τιμές των συντελεστών, που αντιστοιχούν στις ανεξάρτητες μεταβλητές, δημιουργώντας έτσι μοντέλα πρόβλεψης. Πιο αναλυτικά, έχοντας μια συνάρτηση του τύπου:

$$y_i = f(x_{i1}, x_{i2}, x_{i3}) \quad i = 1, 2, 3, \dots, N$$

και ορίζοντας συγκεκριμένες γνωστές συναρτήσεις για δοκιμή μεταξύ των ανεξάρτητων μεταβλητών, ξεκινάει μια διαδικασία εύρεσης των μοντέλων πρόβλεψης πάνω στα δεδομένα. Τα δεδομένα χωρίζονται σε δεδομένα εκμάθησης, δηλαδή ένα ποσοστό των συνολικών δεδομένων όπου χρησιμοποιούνται για την δημιουργία του μοντέλου πρόβλεψης (διαδικασία εκπαίδευσης), και σε δεδομένα για την δοκιμή των μοντέλων πρόβλεψης (διαδικασία γενίκευσης ή επικύρωσης). Χαρακτηριστικό παράδειγμα χωρισμού των δεδομένων είναι το 70% αυτών να χρησιμοποιούνται στην διαδικασία εκπαίδευσης και το 30% αυτών στην διαδικασία γενίκευσης ή επικύρωσης.

Στην παρούσα εργασία, μετά από την εισαγωγή των επεξεργασμένων δεδομένων στο Eureqa, συμπληρώθηκε η συναρτησιακή σχέση των ανεξάρτητων μεταβλητών (αριθμός παράλληλων αιτημάτων, γεωγραφική απόσταση τους απ' όπου προέρχονται και χρόνος εκτέλεσης των αιτημάτων), με την εξαρτημένη (χρόνος απόκρισης). Αυτή ήταν:

$$responseTime = f(time, users, distance)$$

Στην συνέχεια επιλέχθηκαν από την λίστα, τα μαθηματικά μοντέλα για την εύρεση των μοντέλων πρόβλεψης. Καθώς το Eureqa κάνει χρήση τεχνητών νευρωνικών δικτύων, σε κάθε κόμβο του δικτύου δοκιμάζεται ένα μαθηματικό μοντέλο, πάνω στη συνάρτηση που του δόθηκε για την σχέση μεταξύ των ανεξάρτητων και της εξαρτημένης μεταβλητής. Δοκιμάζοντας λοιπόν, περισσότερα μαθηματικά μοντέλα και συνδυάζοντας τα μεταξύ τους, επιτυγχάνει την δημιουργία μοντέλων πρόβλεψης με την καλύτερη προσέγγιση. Μετά την εισαγωγή και τον προσδιορισμό της σχέσης και των μαθηματικών μοντέλων, ξεκινάει η διαδικασία εύρεσης των μοντέλων πρόβλεψης.

Κατά την διαδικασία παραγωγής και δοκιμής των μοντέλων πρόβλεψης, αλλά και μετά την λήξη της, παρουσιάζονται τα μοντέλα ταξινομημένα με βάση τον αριθμό που δείχνει την προσέγγιση στο υπόδειγμα. Επίσης παρουσιάζονται πληροφορίες σχετικά με τον εκτιμητή

ελαχίστων τετραγώνων, τον συντελεστή συσχέτισης, τη μέγιστη και την μέση απόκλιση της προβλεπόμενης τιμής της εξαρτημένης μεταβλητής, με βάση το μοντέλο πρόβλεψης από την τιμή του υποδείγματος και άλλες ακόμα χρήσιμες πληροφορίες.

Η διαδικασία εύρεσης μοντέλων πρόβλεψης, είναι μια διαδικασία που χρειάζεται αρκετούς πόρους από το σύστημα που εκτελείται. Ένα χαρακτηριστικό παράδειγμα είναι ότι διαδικασία αυτή, για το συγκεκριμένο υπόδειγμα της παρούσας εργασίας, χρειάστηκε κοντά 3 εβδομάδες συνεχόμενης εκτέλεσης, σε έναν υπολογιστή με 4 φυσικούς πυρήνες χρονισμένους στα 3,4GHz και 11GB μνήμης RAM, για να ολοκληρωθεί. Έτσι λόγω της ανάγκης για αρκετούς πόρους από το σύστημα που εκτελεί την διαδικασία αυτή, η εταιρεία που έχει αναπτύξει την εφαρμογή Eureqa, η Nutonian, προσφέρει την ενοικίαση υποδομών υπολογιστικού νέφους για την εκτέλεση της. Μπορεί λοιπόν ο χρήστης να επιλέξει τους πόρους που επιθυμεί και όλη η διαδικασία εκτελείται στο υπολογιστικό νέφος, αρκετά πιο σύντομα (ανάλογα τους πόρους). Με αυτόν τον τρόπο μπορεί να εκτελούνται παράλληλα πολλές διαδικασίες, παρακολουθώντας ο χρήστης την πρόοδο τους, από την εφαρμογή Eureqa που είναι εγκατεστημένη στον υπολογιστή του.

Στην Ενότητα 7.3.4, παρουσιάζονται τα μοντέλα πρόβλεψης, που προέκυψαν ύστερα από την ανάλυση που πραγματοποιήθηκε στα δεδομένα, από το εργαλείο Eureqa. Τα μοντέλα αυτά, θα συγκριθούν με βάση μαθηματικές έννοιες, ώστε να επιλεγεί το καταλληλότερο. Οι μαθηματικές έννοιες που χρησιμοποιήθηκαν, αναφέρονται συνοπτικά στην αρχή της ενότητας.

7.5: Σύγκριση και Επιλογή Καταλληλότερου Μοντέλου Πρόβλεψης

Στις προηγούμενες ενότητες προσδιορίστηκε η μορφή του υποδείγματος. Πρόκειται, λοιπόν, για ένα πολλαπλό μη γραμμικό υπόδειγμα. Στην ενότητα αυτή παρουσιάζονται τα μοντέλα πρόβλεψης που προέκυψαν, για το συγκεκριμένο υπόδειγμα. Για την παραγωγή τους έγινε χρήση του εργαλείου Eureqa. Από τα συγκεκριμένα μοντέλα πρόβλεψης, θα επιλεγεί το καταλληλότερο. Έτσι παρακάτω παρουσιάζονται όλα τα μοντέλα, συγκρίνοντας τον εκτιμητή ελαχίστων τετραγώνων και άλλους δείκτες για το κάθε ένα από αυτά, επιλέγοντας αυτό που έχει την καλύτερη προσέγγιση.

Αρχικά, γίνεται υπενθύμιση των μαθηματικών εννοιών που αναφέρθηκαν στην ενότητα 7.2.2, για την εκτίμηση των συντελεστών ενός υποδείγματος με την μέθοδο των ελαχίστων τετραγώνων. Οι έννοιες αυτές θα χρησιμοποιηθούν για την σύγκριση των μοντέλων πρόβλεψης.

- **Κατάλοιπα υποδείγματος:** αποτελούν τις εκτιμήσεις των μη παρατηρούμενων τιμών του διαταρακτικού όρου ε_i ή αλλιώς τις αποστάσεις μεταξύ των παρατηρούμενων και εκτιμώμενων τιμών. Στα μαθηματικά συμβολίζεται με $\sum_{i=1}^N \varepsilon_i = \sum_{i=1}^N (y_i - \hat{y}_i)$ ή $\varepsilon_i = y_i - \hat{y}_i$, όπου y_i είναι παρατηρούμενες τιμές του υποδείγματος και \hat{y}_i είναι οι εκτιμώμενες τιμές του.

- **Μέση τιμή καταλοίπων υποδείγματος:** στα μαθηματικά συμβολίζεται με $\frac{1}{N} \sum_{i=1}^N |\hat{\epsilon}_i|$ ή $\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$.
- **Άθροισμα των τετραγώνων των καταλοίπων:** αποτελεί την αρχή της μεθόδου των ελαχίστων τετραγώνων και στα μαθηματικά συμβολίζεται με $\sum_{i=1}^N \hat{\epsilon}_i^2$ ή $\sum_{i=1}^N (y_i - \hat{y}_i)^2$.
- **Μέση τιμή αθροίσματος των τετραγώνων των καταλοίπων:** στα μαθηματικά συμβολίζεται με $\frac{1}{N} \sum_{i=1}^N \hat{\epsilon}_i^2$ ή $\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$.
- **Μέγιστη απόκλιση:** αποτελεί την μέγιστη απόκλιση μεταξύ των παρατηρούμενων και εκτιμώμενων τιμών ή αλλιώς το μέγιστο κατάλοιπο του υποδείγματος ($\max(\hat{\epsilon}_i = y_i - \hat{y}_i)$).
- **Αριθμός συντελεστών:** αποτελεί το πλήθος των συντελεστών για το μοντέλο πρόβλεψης. Δηλαδή το K στους συντελεστές $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K$ του προβλεπόμενου μοντέλου.
- **Εκτιμητής των ελαχίστων τετραγώνων (R^2):** αποτελεί τον εκτιμητή της μεθόδου των ελαχίστων τετραγώνων όπου παίρνει τιμές από 0 μέχρι 1. Όσο πιο κοντά στο 1 είναι, τόσο οι τιμές της εξαρτημένης μεταβλητής του μοντέλου πρόβλεψης ταυτίζονται με τις τιμές της εξαρτημένης μεταβλητής του υποδείγματος.

Αποτελέσματα Διαδικασίας Εύρεσης Μοντέλων Πρόβλεψης

Ύστερα από την υπενθύμιση των μαθηματικών εννοιών που χρησιμοποιήθηκαν στην σύγκριση, παρουσιάζονται τα μοντέλα πρόβλεψης. Για λόγους παρουσίασης, εμφανίζονται παρακάτω τα τρία καταλληλότερα μοντέλα. Για την καλύτερη παρουσίαση των μοντέλων, έγινε αντιστοιχία των παραμέτρων – μεταβλητών με ακρωνύμια. Παρακάτω αναφέρεται αυτή η αντιστοιχία.

ΧΑΥ = Χρόνος Απόκρισης Υπηρεσίας (σε χιλιοστά του δευτερολέπτου).

ΑΑ = Αριθμός Αιτήσεων κάθε χρονική στιγμή.

ΓΑΑ = Γεωγραφική Απόσταση Αιτήσεων σε χιλιόμετρα από τις υποδομές του υπολογιστικού νέφους.

ΧΣΕΑ = Χρονική Στιγμή Εκτέλεσης Αιτημάτων (σε δευτερόλεπτα).

εφ = εφαπτομένη.

Logistic = σιγμοειδής συνάρτηση.

$$\text{logistic}(x) = \frac{1}{1 + e^{-x}}$$

Μοντέλο 1:

Μοντέλο	$ΧΑΥ = 174.371577745135 + 13.5508835543719 * AA + 1007.35698496204 * \text{logistic}(AA * \Gamma AA - 324899.814186988) + \epsilon\phi(99.1024505244415 + \chi\Sigma EA) + \epsilon\phi(170.397542761089 * \chi\Sigma EA) - 13.3276281789849 * AA * \Gamma AA^{(-7.668502577257 * AA / (\Gamma AA - 1007.35698496204))}$
Εκτιμητής R ²	0.97719983
Μέγιστη απόκλιση	750.47336
Μέση τιμή τετραγώνων καταλοίπων	23714.158
Μέση τιμή καταλοίπων	92.471422

Μοντέλο 2:

Μοντέλο	$ΧΑΥ = 167.83185640811 + 13.5212205552888 * AA - 13.2506591729765 * AA / \Gamma AA^2 + 1019.14270610974 * \text{logistic}(AA * \Gamma AA - 324899.800027787) + \epsilon\phi(99.1024505244415 + \chi\Sigma EA) + \epsilon\phi(170.397542761089 * \chi\Sigma EA)$
Εκτιμητής R ²	0.97620706
Μέγιστη απόκλιση	761.22634
Μέση τιμή τετραγώνων καταλοίπων	24746.73
Μέση τιμή καταλοίπων	95.112789

Μοντέλο 3:

Μοντέλο	$ΧΑΥ = 169.150525219296 + 13.4948700215329 * AA + 0.292514925774821 * \epsilon\phi(168.2169581 + \chi\Sigma EA) - 13.2376816988943 * AA / \Gamma AA^2 + 1020.8945686668 * \text{logistic}(0.000255021407781689 * AA * \Gamma AA - 82.6559957116165)$
Εκτιμητής R ²	0.96558743
Μέγιστη απόκλιση	1295.6955
Μέση τιμή τετραγώνων καταλοίπων	35792.072
Μέση τιμή καταλοίπων	102.09298

Επιλογή Καταλληλότερου Μοντέλου Πρόβλεψης

Η εφαρμογή Eureka για το συγκεκριμένο υπόδειγμα παρήγαγε συνολικά τριάντα ένα μοντέλα πρόβλεψης. Προηγουμένως παρουσιάστηκαν τα τρία καλύτερα, ταξινομημένα με βάση την μεγαλύτερη τιμή του εκτιμητή των ελαχίστων τετραγώνων. Και τα τρία μοντέλα πρόβλεψης που παρουσιάστηκαν είχαν σχετικά μεγάλη ακρίβεια, καθώς η τιμή του εκτιμητή σε κάθε ένα από τα μοντέλα αυτά ήταν αρκετά κοντά στην μονάδα (όσο πιο κοντά στην μονάδα, τόσο περισσότερο πλησιάζει την πραγματική τιμή). Ως καλύτερη λύση ή καλύτερο μοντέλο πρόβλεψης για τον υπόδειγμα, βάσει των αποτελεσμάτων του Eureka, και των τιμών των μαθηματικών εννοιών, είναι το πρώτο. Για την επιλογή του έγινε σύγκριση των κριτηρίων: (i) εκτιμητής ελαχίστων τετραγώνων, (ii) μέγιστη απόκλιση, (iii) μέση τιμή τετραγώνων καταλοίπων και (iv) μέση τιμή καταλοίπων. Παρακάτω παρουσιάζεται αναλυτικά η συνάρτηση του μοντέλου πρόβλεψης της απόδοσης της υπηρεσίας Wikipedia.org.

$$XAY = \alpha + \beta * AA + \gamma * \left(\frac{1}{1 + e^{-1*(AA*ΓAA-\delta)}} \right) + \varepsilon\varphi(\varepsilon + XΣEA) + \varepsilon\varphi(\zeta + XΣEA) - \eta * AA * \frac{\Gamma AA^{-\theta*AA}}{\Gamma AA - \iota}$$

Όπου:

XAY = Χρόνος Απόκρισης Υπηρεσίας.

AA = Αριθμός Αιτήσεων κάθε χρονική στιγμή.

ΓAA = Γεωγραφική Απόσταση Αιτήσεων σε χιλιόμετρα από τις υποδομές του υπολογιστικού νέφους.

XΣEA = Χρονική Στιγμή Εκτέλεσης Αιτημάτων.

Και οι τιμές των παραμέτρων α, β, γ, δ, ε, ζ, η, θ είναι:

$$\alpha = 174,371577745135,$$

$$\beta = 13,5508835543719,$$

$$\gamma = 1007,35698496204,$$

$$\delta = 324899,814186988,$$

$$\varepsilon = 99,1024505244415,$$

$$\zeta = 170,397542761089,$$

$$\eta = 13,3276281789849,$$

$$\theta = 7,668502577257,$$

$$\iota = 1007,35698496204.$$

Αντικαθιστώντας τις παραπάνω τιμές (στρογγυλοποιημένες στα δύο δεκαδικά) στις παραμέτρους του μοντέλου έχουμε:

$$XAY = 174,372 + 13,5509 * AA + 1007.36 * \left(\frac{1}{1 + e^{-1*(AA*ΓAA-324900)}} \right) + \varepsilon\varphi(99,1025 + XΣEA) + \varepsilon\varphi(170,398 + XΣEA) - 13,3276 * AA * \frac{\Gamma AA^{-7,6685*AA}}{\Gamma AA - 1007,36}$$

7.6: Έλεγχος Μοντέλου Πρόβλεψης

7.6.1 Επιλογή Κριτηρίων Ελέγχου

Στην Ενότητα 7.5, παρουσιάστηκαν τα μοντέλα πρόβλεψης που παράχθηκαν με την ανάλυση των δεδομένων, μέσω του εργαλείου Eureqa. Για την παραγωγή τους, το Eureqa, χρησιμοποιεί αλγορίθμους μηχανικής μάθησης και τεχνητών νευρωνικών δικτύων υπολογιστών. Τα μοντέλα αυτά, στην συνέχεια αξιολογήθηκαν, βάσει συγκεκριμένων κριτηρίων που ορίστηκαν με την χρήση μαθηματικών εννοιών (π.χ. εκτιμητής ελαχίστων τετραγώνων). Από την αξιολόγησή τους, επιλέχθηκε το καταλληλότερο, αυτό δηλαδή που προσεγγίζει καλύτερα την πραγματικότητα στην συμπεριφορά της υπηρεσίας. Σκοπός λοιπόν του μοντέλου αυτού, είναι η πρόβλεψη του χρόνου απόκρισης της υπηρεσίας Wikipedia.org, βάσει συγκεκριμένων παραμέτρων κλιμάκωσής της. Οι παράμετροι, λοιπόν, που μελετήθηκαν είναι, (i) το πλήθος των παράλληλων αιτημάτων για χρήση της υπηρεσίας, (ii) η γεωγραφική θέση απ' όπου προέρχονται τα αιτήματα και (iii) η διάρκεια εκτέλεσης των αιτημάτων προς την υπηρεσία.

Σκοπός της Ενότητας 7.6, είναι ο έλεγχος του μοντέλου πρόβλεψης που επιλέχθηκε. Ο έλεγχος αυτός, θα γίνει με βάση συγκεκριμένα κριτήρια. Τα κριτήρια που επιλέχθηκαν είναι (i) η μέγιστη απόκλιση των προβλεπόμενων δεδομένων, της απόκρισης της υπηρεσίας και (ii) η μέση τιμή των καταλοίπων (δηλαδή η μέση τιμή όλων των αποκλίσεων). Ο λόγος που επιλέχθηκαν τα δύο αυτά κριτήρια, οφείλεται στο γεγονός ότι προβάλλουν την απόκλιση των δεδομένων του μοντέλου πρόβλεψης, από την πραγματικότητα.

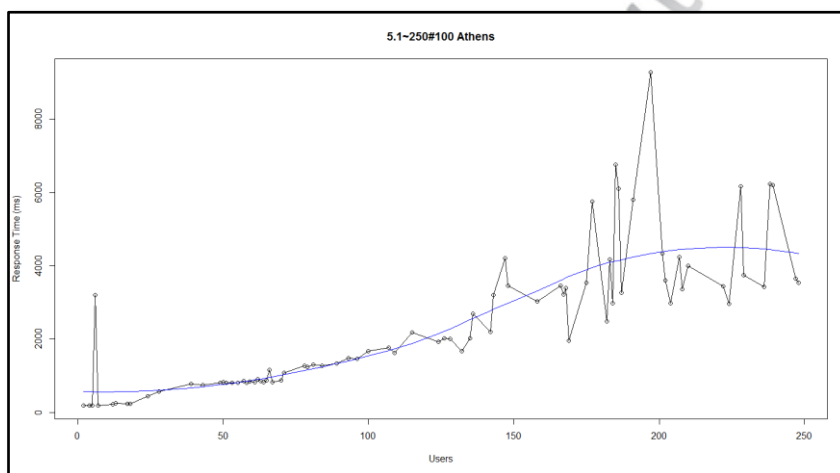
Το επιλεγμένο μοντέλο, όπως και τα υπόλοιπα που προέκυψαν από τον εργαλείο Eureqa, αξιολογήθηκαν με βάση τα κριτήρια που ορίστηκαν παραπάνω, όπως αναφέρεται στην Ενότητα 7.5 στην παρουσίαση τους. Η αξιολόγηση αυτή έγινε από το Eureqa, καθώς χωρίζει τα δεδομένα που δέχεται σαν όρισμα, σε δύο τμήματα. Το ένα τμήμα, χρησιμοποιείται για την δημιουργία των μοντέλων πρόβλεψης, και το δεύτερο για την δοκιμή του μοντέλου. Αν και το ποσοστό των δεδομένων που κατέχει το κάθε τμήμα, μπορεί να οριστεί παραμετρικά, το εργαλείο ως προεπιλεγμένα ποσοστά δίνει 70% στο τμήμα για την δημιουργία των μοντέλων και το υπόλοιπο 30%, στην δοκιμή τους. Συνεπώς, έχει γίνει ένας πρώτος έλεγχος πάνω στα δεδομένα, έχοντας μικρή απόκλιση (μέση τιμή απόκλισης 92,47 χιλιοστά του δευτερολέπτου).

7.6.2 Διαδικασία Ελέγχου του Μοντέλου Πρόβλεψης

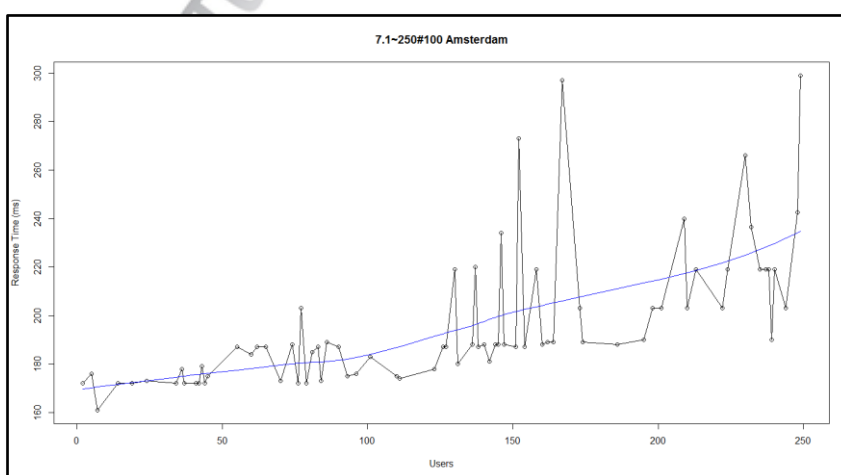
Για τον έλεγχο του μοντέλου πρόβλεψης, ακολουθήθηκαν κάποια κοινά βήματα με αυτά που έγιναν για την παραγωγή του. Αρχικά πραγματοποιήθηκε η διαδικασία συλλογής των εμπειρικών δεδομένων, από τις τρεις πόλεις που είχε γίνει και αρχικά η διαδικασία. Ο λόγος επιλογής των τριών αυτών πόλεων (Άμστερνταμ, Αθήνας και Ρώμης) έγινε επειδή ποικίλει η χιλιομετρική τους απόσταση από τις υποδομές του υπολογιστικού νέφους της υπηρεσίας Wikipedia.org. Το γεγονός αυτό επιτρέπει την καλύτερη μελέτη της παραμέτρου της απόστασης από τις υποδομές, που χρησιμοποιείται στο μοντέλο πρόβλεψης. Ο υπολογιστής που βρισκόταν στο Άμστερνταμ, απέιχε ελάχιστα χιλιόμετρα από τις υποδομές. Ο υπολογιστής στην Αθήνα, απέιχε 2166 χιλιόμετρα, και ο αντίστοιχος στην

Ρώμη, 1300 χιλιόμετρα. Επιπλέον, κατά την μελέτη της εργασίας, δεν υπήρχε δυνατότητα συλλογής δεδομένων από περισσότερες πόλεις. Η συλλογή των εμπειρικών δεδομένων, έγινε με την χρήση της εφαρμογής που αναπτύχθηκε για τον σκοπό αυτό (Κεφάλαιο 5).

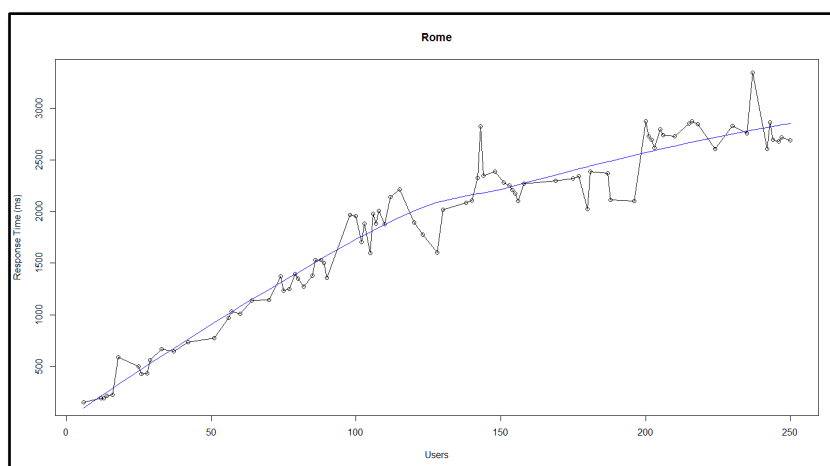
Μετά την ολοκλήρωση της διαδικασίας συλλογής των εμπειρικών δεδομένων, τα δεδομένα επεξεργάστηκαν για την διασφάλιση της ορθότητας εκτέλεσης της διαδικασίας (έλεγχος πλήθους σφαλμάτων να είναι κάτω του 2% του συνόλου). Επιπλέον, μέσω της επεξεργασίας, μορφοποιήθηκαν τα δεδομένα για να απεικονιστούν σε διαγράμματα δύο διαστάσεων. Τα δεδομένα που συλλέχθηκαν, παρουσιάζονται, μετά την επεξεργασία τους στα Σχήματα 7.7, 7.8, 7.9. Τα δεδομένα αυτά αφορούν τα αιτήματα που εκτελέστηκαν από την Αθήνα, Άμστερνταμ και Ρώμη. Το Σχήμα 7.7 εμφανίζει τα δεδομένα της Αθήνας. Αντίστοιχα στο Σχήμα 7.8, παρουσιάζονται αυτά του Άμστερνταμ, και στο Σχήμα 7.9 τα δεδομένα της Ρώμης.



Σχήμα 7.7: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Αθήνα)



Σχήμα 7.8: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Άμστερνταμ)

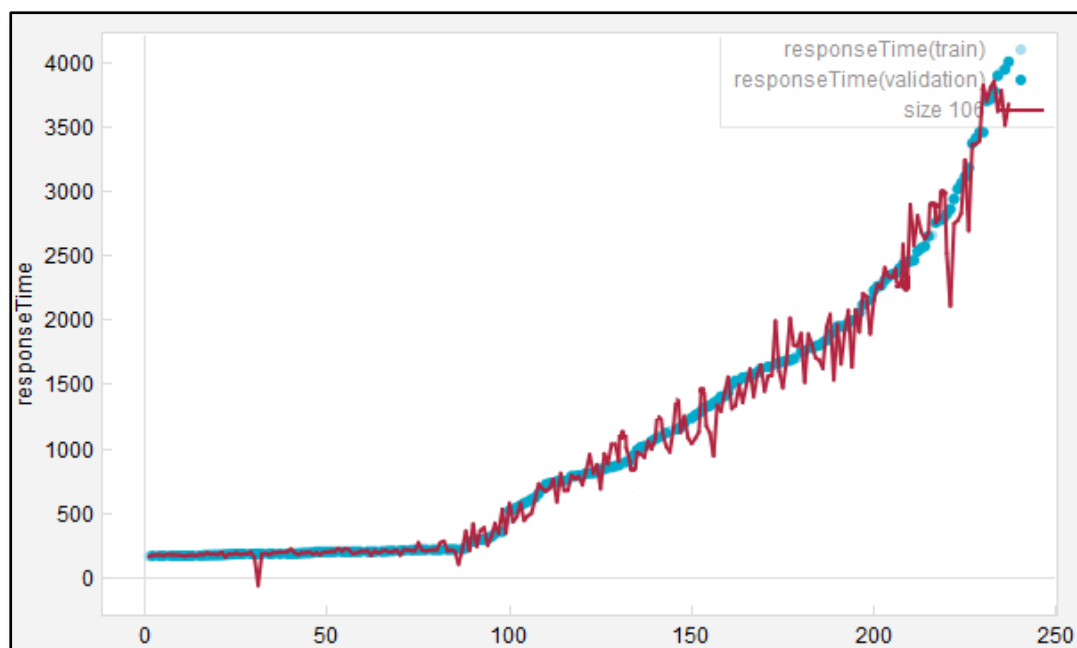


Σχήμα 7.9: Απεικόνιση τιμών παραμέτρου παράλληλων αιτημάτων και χρόνου απόκρισης (Ρώμη)

Αρχικά, από τα σχήματα, παρατηρείται ότι, εκτός από τις ακραίες τιμές, οι μέγιστες τιμές του χρόνου απόκρισης, κυμαίνονται στα ίδια επίπεδα με τα αντίστοιχα δεδομένα που είχαν συλλεχθεί για την παραγωγή των μοντέλων (Αθήνα μέχρι 4000 χιλιοστά, Άμστερνταμ μέχρι 300 χιλιοστά και Ρώμη μέχρι 3000 χιλιοστά του δευτερολέπτου). Υπενθυμίζεται σε αυτό το σημείο, ότι η συλλογή δεδομένων έγινε με τυχαίο αριθμό παράλληλων αιτημάτων κάθε χρονική στιγμή.

Επόμενο βήμα, του ελέγχου του μοντέλου πρόβλεψης, είναι η δοκιμή των προβλεπόμενων τιμών του χρόνου απόκρισης της υπηρεσίας, με αυτές που συλλέχθηκαν. Πράγματι, οι αποκλίσεις είναι παρόμοιες με αυτές που είχε εμφανίσει το Eureka, συγκρίνοντας το μοντέλο με το 30% των εισαχθέντων δεδομένων. Η μέγιστη απόκλιση είναι 724,5 χιλιοστά του δευτερολέπτου, ενώ η μέση τιμή απόκλισης (ή μέση τιμή των καταλοίπων) είναι 85,18 χιλιοστά. Συνεπώς το μοντέλο πρόβλεψης που επιλέχθηκε, περιέχει μικρό αριθμό μέσου σφάλματος ή του μέσου διαταρακτικού όρου, προβλέποντας τιμές μέχρι και για 4000 χιλιοστά του δευτερολέπτου. Στον παρακάτω πίνακα, αναφέρονται τα αποτελέσματα ελέγχου του μοντέλου πρόβλεψης, βάσει των κριτηρίων που είχαν τεθεί. Επιπλέον, στο Σχήμα 7.10 παρουσιάζονται σε διάγραμμα, οι τιμές του χρόνου απόκρισης που συλλέχθηκαν, σε σχέση με τις προβλεπόμενες. Είναι φανερό ότι το μοντέλο, προβλέπει τιμές για τον χρόνο απόκρισης της υπηρεσίας, που είναι πολύ κοντά με την πραγματικότητα. Συνεπώς το μοντέλο αυτό, θεωρείται αξιόπιστο και κατάλληλο για την δυναμική πρόβλεψη του χρόνου απόκρισης της υπηρεσίας Wikipedia.org, βάσει των παραμέτρων κλιμάκωσης της.

Αποτελέσματα Ελέγχου του Μοντέλου Πρόβλεψης	
Μέγιστη απόκλιση	724,5 χιλιοστά του δευτερολέπτου
Μέση τιμή καταλοίπων	85,18 χιλιοστά του δευτερολέπτου



Σχήμα 7.10: Απεικόνιση τιμών χρόνου απόκρισης της υπηρεσίας Wikipedia.org, και τιμών του μοντέλου δυναμικής πρόβλεψης

7.7: Συμπεράσματα

Στο παρόν κεφάλαιο, παρουσιάστηκε η διαδικασία ανάλυσης των δεδομένων για την δημιουργία των μοντέλων πρόβλεψης. Τα δεδομένα προηγουμένως είχαν επεξεργαστεί για έρθουν στην επιθυμητή μορφή, που θα μπορούσαν να αναλυθούν ευκολότερα. Έτσι, αρχικά τεκμηριώθηκε η μορφή του υποδείγματος, για τα δεδομένα που συλλέχθηκαν. Πρόκειται για ένα πολλαπλό, μη γραμμικό υπόδειγμα. Η πολλαπλότητα οφείλεται στον πλήθος των ανεξάρτητων παραμέτρων, που επηρεάζουν την εξαρτημένη μεταβλητή της απόδοσης της υπηρεσίας (Ενότητα 7.3.1). Η μη γραμμική μορφή του, παρατηρήθηκε ύστερα από απεικόνιση των δεδομένων σε διαγράμματα δύο αξόνων (Ενότητα 7.3.2).

Στην συνέχεια, έγινε εισαγωγή των δεδομένων στο εργαλείο Eureqa. Χρησιμοποιώντας το Eureqa, τεχνικές μηχανικής μάθησης και νευρωνικών δικτύων υπολογιστών, παρήγαγε ένα σύνολο από μοντέλα πρόβλεψης. Η διαδικασία αυτή ήταν αρκετά χρονοβόρα (είχε διάρκεια τριών εβδομάδων), λόγω των δοκιμών που γίνονται για την παρουσίαση των καταλληλότερων μοντέλων. Έτσι στην Ενότητα 7.5, παρουσιάστηκαν τα μοντέλα πρόβλεψης, από τα οποία επιλέχτηκε το καταλληλότερο. Για την επιλογή του, αναφέρθηκαν τα κριτήρια, όπου αφορούσαν την απόκλιση των προβλεπόμενων τιμών του χρόνου απόκρισης της υπηρεσίας, από αυτές που είχαν συλλεχθεί.

Το επιλεγμένο μοντέλο, ελέγχθηκε για την αξιοπιστία του, στην πρόβλεψη των τιμών του χρόνου απόκρισης της υπηρεσίας Wikipedia.org. Για τον λόγο αυτό, συλλέχθηκαν εκ νέου εμπειρικά δεδομένα από τρεις πόλεις της Ευρώπης (Αθήνα, Άμστερνταμ και Ρώμη). Τα δεδομένα αυτά στην συνέχεια επεξεργάστηκαν για να συγκριθούν με τις τιμές του μοντέλου πρόβλεψης. Το μοντέλο θεωρήθηκε αξιόπιστο, καθώς υπήρχε μικρή απόκλιση

από τα δεδομένα που συλλέχθηκαν. Για την ακρίβεια, η μέγιστη απόκλιση ήταν 724 χιλιοστά του δευτερολέπτου, ενώ η μέση τιμή απόκλισης ήταν στα 85,18 χιλιοστά. Επομένως, το μοντέλο δυναμικής πρόβλεψης που επιλέχθηκε, περιέχει μικρό αριθμό μέσου σφάλματος, προβλέποντας τιμές για τον χρόνο απόκρισης, μέχρι και για 4 δευτερόλεπτα.

Στο Κεφάλαιο 8 παρουσιάζεται η επίτευξη των αντικειμενικών στόχων που είχαν οριστεί στην Ενότητα 1.5. Επιπλέον, αναφέρονται τα οφέλη της παρούσας εργασίας, για την ερευνητική θεματική περιοχή που επιλέχθηκε. Τέλος, παρουσιάζονται τα χρήσιμα συμπεράσματα που προέκυψαν, αλλά και μελλοντικές επεκτάσεις που μπορούν να γίνουν, πάνω στο προτεινόμενο μοντέλο επίλυσης.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 8: Συμπεράσματα και Μελλοντική Έρευνα

8.1: Εισαγωγή

Τα τελευταία 10 χρόνια, τεχνολογίες όπως της νεφούπολογιστικής, που σχετίζονται με την εικονικοποίηση των υπολογιστικών πόρων, έχουν γίνει αρκετά δημοφιλείς, λόγω της ευρείας χρήσης τους. Για τον λόγο αυτό, το υπολογιστικό νέφος έχει γίνει αντικείμενο μελέτης και ανάλυσης, τόσο σε ακαδημαϊκό όσο και σε επιχειρηματικό επίπεδο. Μεγάλο κομμάτι της έρευνας, στρέφεται προς τη κατεύθυνση της μεγιστοποίησης της απόδοσης των εφαρμογών, που χρησιμοποιούν υποδομές υπολογιστικού νέφους, λαμβάνοντας υπόψη περιορισμούς κόστους. Συνεπώς, οι εφαρμογές πρέπει να κάνουν χρήση, όσο το δυνατόν, τους ελάχιστους αλλά και απαραίτητους πόρους, την κάθε χρονική στιγμή. Αυτό βέβαια πρέπει να γίνεται με τέτοιο τρόπο, έτσι ώστε να πληρούνται οι απαιτήσεις για τα επίπεδα απόδοσης της εφαρμογής.

Καθώς όμως, ο φόρτος χρήσης των εφαρμογών δεν είναι σταθερός, οι ανάγκες τους σε πόρους μεταβάλλονται δυναμικά, και εξαρτώνται από παράγοντες όπως: (i) είδος υπηρεσίας (π.χ. υπηρεσίες υποβολής φορολογικών δηλώσεων που χρησιμοποιούνται συγκεκριμένες χρονικές περιόδους ανά έτος), (ii) ώρες αιχμής δικτύου, (iii) μερίδα πελατών που απευθύνεται η εφαρμογή (π.χ. φοιτητές, εργαζόμενους, ανέργους, κλπ). Οστόσο η προετοιμασία που χρειάζεται για να προσφερθούν δυναμικά οι πόροι αυτοί, δεν είναι χρονικά αμελητέα. Επομένως, ο πάροχος της υπηρεσίας πρέπει να μπορεί να προβλέπει δυναμικά και με ακρίβεια κάθε χρονική στιγμή τις προκύπτουσες ανάγκες για πόρους, ώστε να ανταπεξέλθει στον προκείμενο φόρτο εργασίας.

Η ανάγκη, για δυναμική πρόβλεψη των αναγκών της εφαρμογής με σκοπό την μεγιστοποίηση της χρήσης των πόρων, έχει στρέψει μεγάλο κομμάτι της ερευνητικής προσπάθειας προς τη δημιουργία δυναμικών μοντέλων πρόβλεψης. Βέβαια, από την στιγμή που το υπολογιστικό νέφος, έχει γίνει εμπορικό προϊόν, πολλοί πάροχοι, μέσα από τα τμήματα έρευνας και ανάπτυξης τους, πραγματοποιούν συνεχείς μελέτες για την ελαχιστοποίηση του προβλήματος. Έχουν δημιουργηθεί έτσι, μηχανισμοί δυναμικής πρόβλεψης που κατηγοριοποιούν τις εφαρμογές, με βάση συγκεκριμένα κριτήρια (π.χ. μέγεθος δεδομένων που επεξεργάζονται και πόρους που χρησιμοποιούν σε κάθε φάση του κύκλο εκτέλεσής τους), μεταφέροντας τις πληροφορίες που παράγουν, σε μηχανισμούς προετοιμασίας και προσφοράς των πόρων στις εφαρμογές. Άλλες μελέτες, που παρουσιάστηκαν στη βιβλιογραφία του Κεφάλαιο 2, προσεγγίζουν την δημιουργία των μοντέλων πρόβλεψης, ξεχωριστά για κάθε τύπο υπηρεσίας (π.χ. ιστοσελίδα ηλεκτρονικού καταστήματος στην εργασία [48]). Οι μελέτες αυτές, απευθύνονται κυρίως σε συγκεκριμένους τύπους υπηρεσιών, ορίζοντας τις υποδομές του υπολογιστικού νέφους που χρησιμοποιούν. Δημιουργούνται με αυτό τον τρόπο, δυναμικά μοντέλα πρόβλεψης, για συγκεκριμένο τύπο εφαρμογών, έχοντας καλύτερα αποτελέσματα στην πρόβλεψη της απόδοσής τους.

Σκοπός της παρούσας εργασίας ήταν η δημιουργία ενός μοντέλου δυναμικής πρόβλεψης της απόδοσης μιας υπηρεσίας, εγκατεστημένης σε υποδομές υπολογιστικού νέφους, βάσει των παραμέτρων κλιμάκωσής της. Αρχικά λοιπόν, στο Κεφάλαιο 3, προσδιορίστηκαν τα κριτήρια επιλογής της υπηρεσίας που μελετήθηκε. Τα κριτήρια αυτά, είχαν ως σκοπό την επιλογή συγκεκριμένου τύπου εφαρμογών. Για παράδειγμα, η εφαρμογή που θα επιλεγόταν, θα έπρεπε να κάνει χρήση υποδομών υπολογιστικού νέφους. Ενώ πολλές υπηρεσίες πληρούσαν τα προκαθορισμένα κριτήρια, θεωρήθηκε σκόπιμο να μελετηθεί η υπηρεσία της ηλεκτρονικής εγκυκλοπαίδειας, Wikipedia.org. Ο λόγος ήταν ότι διαχειρίζεται μεγαλύτερο όγκο δεδομένων από αυτές που αξιολογήθηκαν.

Στο Κεφάλαιο 4, προσδιορίστηκε η ερευνητική μεθοδολογία που χρησιμοποιήθηκε. Έγινε χρήση λοιπόν, της μεθοδολογίας συλλογής εμπειρικών δεδομένων, βασιζόμενη στην συσχετιστική και την πειραματική προσέγγιση. Ο λόγος που επιλέχθηκε η συσχετιστική προσέγγιση της μεθοδολογίας, είναι η συνδιακύμανση που υπάρχει, μεταξύ των μεταβλητών. Ο συνδυασμός της με την πειραματική προσέγγιση, έχει ως αποτέλεσμα τον μερικό έλεγχο των μεταβλητών (π.χ. έλεγχος παράλληλων αιτημάτων προς την υπηρεσία).

Στην συνέχεια, παρουσιάστηκε στο Κεφάλαιο 5, η συλλογή των εμπειρικών δεδομένων για την υπηρεσία της Wikipedia.org. Για την συλλογή τους, αναπτύχθηκε μια εφαρμογή που χωριζόταν σε δυο ξεχωριστά τμήματα. Το πρώτο τμήμα της, διαχειριζόταν την διαδικασία, προσδιορίζοντας την υπηρεσία που θα γινόταν η εκτέλεση των αιτημάτων, με ποιο ρυθμό και από ποιους υπολογιστές. Το δεύτερο τμήμα, δεχόταν τις παραμέτρους, μέσω του πρωτοκόλλου TCP/IP, από το πρώτο, και εκτελούσε τα αιτήματα προς την επιλεγμένη υπηρεσία. Το δεύτερο τμήμα της υπηρεσίας, βρισκόταν εγκατεστημένο, σε τρεις υπολογιστές, ο καθένας σε μια διαφορετική πόλη της Ευρώπης (Αθήνα, Άμστερνταμ και Ρώμη).

Αφού συλλέχθηκαν τα δεδομένα, επεξεργάστηκαν με την βοήθεια του εργαλείου της γλώσσας R (Κεφάλαιο 6). Τα συμπεράσματα που προέκυψαν από την επεξεργασία αυτή, σχετιζόντουσαν με την ορθότητα εκτέλεσης της διαδικασίας συλλογής τους. Επιπλέον, η απεικόνιση των δεδομένων, σε διαγράμματα δύο διαστάσεων, βοήθησε στην μελέτη τους για δημιουργία των μοντέλων πρόβλεψης. Λόγω του όγκου των δεδομένων, αλλά και των εσφαλμένων τιμών που περιείχαν, μέσω της επεξεργασίας τους, μορφοποιήθηκαν με τέτοιο τρόπο ώστε να εισαχθεί η χρήσιμη πληροφορία στο επόμενο στάδιο για την παραγωγή των μοντέλων πρόβλεψης.

Με την βοήθεια του εργαλείου Eureka, και των δεδομένων που είχαν επεξεργαστεί, δημιουργήθηκαν μοντέλα δυναμικής πρόβλεψης της απόδοσης της υπηρεσίας Wikipedia.org. Τα μοντέλα αυτά, συγκρίθηκαν βάσει συγκεκριμένων κριτηρίων, για την επιλογή του καταλληλότερου. Στην συνέχεια, το επιλεγμένο μοντέλο ελέγχθηκε για να διαπιστωθεί η αξιοπιστία του, στην πρόβλεψη της απόδοσης της υπηρεσίας. Συλλέχθηκαν λοιπόν, δεδομένα για την υπηρεσία, όπου συγκρίθηκαν με τις τιμές που προέβλεπε το μοντέλο. Πράγματι το μοντέλο θεωρήθηκε αξιόπιστο, καθώς η απόκλισή του από την πραγματικότητα είχε μέση τιμή τα 85,18 χιλιοστά του δευτερολέπτου.

Στο παρόν κεφάλαιο, στην Ενότητα 8.2, παρουσιάζονται οι στόχοι της παρούσας εργασίας που επιτεύχθηκαν, όπως είχαν οριστεί στην Ενότητα 1.5. Η Ενότητα 8.3, αναφέρεται στις

γνώσεις που αποκτήθηκαν, μέσω της μελέτης που πραγματοποιήθηκε για την επίλυση της του συγκεκριμένου ερευνητικού προβλήματος. Επιπλέον οι Ενότητες 8.4 και 8.5, παρουσιάζουν την χρησιμότητα της παρούσας εργασίας και τις μελλοντικές επεκτάσεις της.

8.2: Επίτευξη Στόχων Εργασίας

Στο Κεφάλαιο 1, αναφέρθηκαν οι αντικειμενικοί στόχοι, που τέθηκαν για την εκτέλεση της παρούσας ερευνητικής εργασίας. Στην παρούσα ενότητα πραγματοποιείται ανασκόπηση των στόχων, παρουσιάζοντας τα αποτελέσματα για την επίτευξή τους.

Αντικειμενικός στόχος 1: Ανασκόπηση σχετικής βιβλιογραφίας:

Ο συγκεκριμένος στόχος, επιτυγχάνεται μέσω της ανάλυσης που παρουσιάστηκε στο Κεφάλαιο 2. Αναφέρονται οι βασικές έννοιες του υπολογιστικού νέφους, οι οποίες χρησιμοποιήθηκαν για την κατανόηση του ερευνητικού προβλήματος. Επιπλέον, μέσω συγκεκριμένης μεθοδολογίας αναζήτησης, που συνδυάζει αυτές που προτάθηκαν από τον Cooper και τον Brereton, επιλέχθηκαν σχετικές μελέτες που έχουν γίνει. Έτσι, στην Ενότητα 2.4, έγινε κριτική ανάλυση των σχετικών ερευνητικών μελετών, επιλέγοντας παράλληλα μεθοδολογίες και τεχνικές που θα βοηθούσαν στην επίλυση του ερευνητικού προβλήματος της παρούσας εργασίας.

Αντικειμενικός στόχος 2: Δημιουργία μηχανισμού συλλογής των δεδομένων, που σχετίζονται με την απόδοση της υπηρεσίας και των παραμέτρων που την επηρεάζουν.

Στο Κεφάλαιο 5, παρουσιάζεται η επίτευξη του συγκεκριμένου στόχου. Για την συλλογή των δεδομένων, που αφορούν την απόδοση της υπηρεσίας και των παραμέτρων που την επηρεάζουν, δημιουργήθηκε εφαρμογή που χωρίζεται σε δύο τμήματα. Τα δύο αυτά τμήματα, θα μπορούσαν να θεωρηθούν δύο ξεχωριστές εφαρμογές, που επικοινωνούν για να ανταλλάξουν πληροφορίες. Η πρώτη εφαρμογή διαχειρίζεται όλη την διαδικασία, θέτοντας ο χρήστης τα ορίσματα για τη συλλογή των δεδομένων. Για την εκτέλεση των αιτημάτων προς την επιλεγμένη υπηρεσία, η πρώτη εφαρμογή επικοινωνεί με τη δεύτερη μέσω του πρωτοκόλλου TCP/IP. Για την εκτέλεση των αιτημάτων από διαφορετικές γεωγραφικές αποστάσεις από τις υποδομές της υπηρεσίας, η δεύτερη εφαρμογή είναι εγκατεστημένη, σε υπολογιστές που βρίσκονται σε διαφορετικές πόλεις. Μόλις καταγραφεί η συμπεριφορά της υπηρεσίας, από τη δεύτερη εφαρμογή, αποστέλλονται τα δεδομένα, στην πρώτη, όπου τα συλλέγει ξεχωριστά για τον κάθε υπολογιστή. Τα δεδομένα, που έχουν συγκεντρωθεί, αποθηκεύονται σε αρχεία κειμένου, στα οποία γίνεται διαχωρισμός μεταξύ των πληροφοριών, βάσει επιλεγμένου χαρακτήρα (κόμμα).

Αντικειμενικός στόχος 3: Ανάλυση των δεδομένων που προκύπτουν για την εξαγωγή συμπερασμάτων.

Ο στόχος αυτός επιτυγχάνεται στο Κεφάλαιο 6 και στο Κεφάλαιο 7. Τα δεδομένα που έχουν συλλεχθεί, επεξεργάζονται με τη χρήση του εργαλείου της γλώσσας R. Μέσω της ανάλυσής τους (Ενότητα 6.3), προκύπτουν συμπεράσματα σχετικά με την ορθότητα της διαδικασίας συλλογής τους. Αν δηλαδή, τα δεδομένα που συλλέχθηκαν δεν περιλαμβάνουν μεγάλο αριθμό σφαλμάτων, και αν καταγράφηκε σωστά η συμπεριφορά της υπηρεσίας. Ένα χρήσιμο συμπέρασμα που προέκυψε, ήταν η αρχικά λανθασμένη συλλογή των δεδομένων, όπου υπήρχε προκαθορισμένη κλιμακωτή αυξομείωση των παράλληλων αιτημάτων κάθε χρονική στιγμή. Συνεπώς, δεν υπήρχε ανεξαρτησία μεταξύ των παραμέτρων που επηρεάζουν την απόδοση της υπηρεσίας, αφού μπορούσε να προβλεφθεί η τιμή της παραμέτρου του πλήθους των παράλληλων αιτημάτων, βάσει αυτής του χρόνου εκτέλεσης των αιτημάτων προς την υπηρεσία. Επιπλέον, στην τεκμηρίωση της μορφής του υποδείγματος (Ενότητα 7.3), μέσω της ανάλυσης των δεδομένων, και την απεικόνισή τους σε διαγράμματα δύο αξόνων, προέκυψαν συμπεράσματα για το υπόδειγμα. Πρόκειται για ένα πολλαπλό μη γραμμικό υπόδειγμα, όπου για τη δημιουργία μοντέλου πρόβλεψης του, απαιτείται χρήση μη γραμμικών μοντέλων. Τα συγκεκριμένα μαθηματικά μοντέλα, επιλέχθηκαν για να επεξεργαστούν μέσω του εργαλείου Eureka, για την εξαγωγή των μοντέλων πρόβλεψης.

Αντικειμενικός στόχος 4: Δημιουργία μοντέλου δυναμικής πρόβλεψης της απόδοσης της υπηρεσίας.

Το Κεφάλαιο 7, είναι αφιερωμένο στην επίτευξη του συγκεκριμένου στόχου. Τα δεδομένα αφού έχουν επεξεργαστεί, εισάγονται στο εργαλείο Eureka, για την παραγωγή των μοντέλων πρόβλεψης. Η συγκεκριμένη διαδικασία, χρειάστηκε 3 εβδομάδες για να ολοκληρωθεί. Στην συνέχεια, με την χρήση επιλεγμένων κριτηρίων που βασίζονται σε μαθηματικές έννοιες, επιλέγεται το καταλληλότερο μοντέλο δυναμικής πρόβλεψης της απόδοσης της υπηρεσίας που μελετάται.

Αντικειμενικός στόχος 5: Έλεγχος και αξιολόγηση του μοντέλου δυναμικής πρόβλεψης.

Ο στόχος αυτός, επιτεύχθηκε, στο Κεφάλαιο 7, μέσω της ερευνητικής μεθοδολογίας που είχε επιλεχθεί στο Κεφάλαιο 4. Έτσι, με την συλλογή δεδομένων από την υπηρεσία, και σύγκρισή τους με τις τιμές που προβλέπει το επιλεγμένο μοντέλο, πραγματοποιήθηκε ο έλεγχός του. Πράγματι το μοντέλο θεωρήθηκε αξιόπιστο, καθώς η μέση τιμή απόκλισής του από τα δεδομένα, ήταν 85,18 χιλιοστά του δευτερολέπτου, προβλέποντας τιμές μέχρι και τεσσάρων δευτερολέπτων.

8.3: Συμπεράσματα Μελέτης

Η παρούσα ενότητα, αναφέρεται στα συμπεράσματα που προέκυψαν από την μελέτη που πραγματοποιήθηκε για την επίλυση του συγκεκριμένου ερευνητικού προβλήματος. Τα συμπεράσματα σχετίζονται με την μελέτη που έγινε πάνω στον δυναμική προσφορά των πόρων του υπολογιστικού νέφους στις υπηρεσίες, καθώς και με τη αναγκαιότητα δημιουργίας μοντέλων δυναμικής πρόβλεψης. Επιπλέον, βγήκαν κάποια χρήσιμα συμπεράσματα σχετικά με τις παραμέτρους που επηρεάζουν την απόδοση υπηρεσιών, που πληρούν συγκεκριμένα κριτήρια.

8.3.1 Συμπεράσματα Αναγκαιότητας Μοντέλων Πρόβλεψης

Αρχικά, μελετήθηκε η σημαντικότητα του δυναμικού διαμοιρασμού των κοινόχρηστων πόρων ενός υπολογιστικού νέφους, στις παρεχόμενες υπηρεσίες του, όταν ο φόρτος εργασίας τους δεν είναι σταθερός. Το υπολογιστικό νέφος, κάνει χρήση τεχνολογιών εικονικοποίησης που επιτρέπουν τον διαμοιρασμό των φυσικών πόρων, μέσω εικονικών μηχανών στις υπηρεσίες. Με αυτό τον τρόπο, οι πόροι προσφέρονται δυναμικά προς τις υπηρεσίες, ανάλογα τις ανάγκες τους κάθε χρονική στιγμή. Αν όμως δεν γίνει έγκαιρα αυτή η δυναμική προσφορά, τότε θα επηρεαστεί η απόδοση των υπηρεσιών. Αυτό σημαίνει ότι η υπηρεσία μπορεί να μην αποδίδει κατάλληλα, ή ακόμα και να μην μπορεί να χρησιμοποιηθεί. Επιπλέον, και η λειτουργία της δυναμικής αποδέσμευσης πόρων, όταν αυτοί δεν χρησιμοποιούνται, είναι απαραίτητη για τον σωστό διαμοιρασμό τους. Το βασικό όμως πρόβλημα, στην προσφορά των πόρων του υπολογιστικού νέφους, είναι ότι χρειάζονται προετοιμασία πριν προσφερθούν στις υπηρεσίες. Η προετοιμασία αυτή, δεν είναι χρονικά αμελητέα και ποικίλει ανάλογα την κλιμάκωση που απαιτείται. Η δέσμευση πόρων, έτσι ώστε να είναι διαθέσιμοι σε περίπτωση ανάγκης, δεν είναι λύση καθώς επιφέρει τη μη βέλτιστη χρήση αυτών, καθώς θα μπορούσαν να προσφερθούν σε άλλες υπηρεσίες.

Συνεπώς, η επίλυση του προβλήματος, παροχής των πόρων στις υπηρεσίες, την κατάλληλη χρονική στιγμή, είναι αναγκαία. Ο τρόπος είναι η δημιουργία μοντέλων δυναμικής πρόβλεψης της απόδοσης των υπηρεσιών, βάσει των παραμέτρων κλιμάκωσής τους. Με αυτό τον τρόπο, θα προβλέπεται η απόδοση της υπηρεσίας, όταν αλλάζουν οι παράμετροι που την επηρεάζουν. Έτσι, ο πάροχος της υπηρεσίας, ανάλογα την επιθυμητή απόδοση, που έχει ορίσει, θα μπορεί να αυξάνει ή να μειώνει τους πόρους προς αυτή, ανάλογα την χρήση της. Η επίλυση του συγκεκριμένου προβλήματος, έχει απασχολήσει ερευνητές τόσο από τον ακαδημαϊκό, όσο και από τον επιχειρηματικό κόσμο. Έχουν προταθεί διάφορα μοντέλα, που είτε κατηγοριοποιούν τις υπηρεσίες, βάσει συγκεκριμένων κριτηρίων, είτε γίνεται ξεχωριστή μελέτη τους. Τα αποτελέσματα που θα μπορούσε να έχει η επίλυση του προβλήματος, είναι η αύξηση της υιοθέτησης λύσεων νεφοϋπολογιστικής από τους παρόχους υπηρεσιών, και η επίτευξη καλύτερων επιπέδων απόδοσης των προσφερόμενων υπηρεσιών.

8.3.2 Συμπεράσματα Ανάλυσης Μοντέλου Δυναμικής Πρόβλεψης

Η παρούσα εργασία, είχε ως σκοπό την δημιουργία δυναμικού μοντέλου πρόβλεψης της απόδοσης μιας υπηρεσίας, βάσει συγκεκριμένων παραμέτρων κλιμάκωσής της. Οι παράμετροι που μελετήθηκαν ήταν:

Παράμετρος 1: Ο αριθμός των παράλληλων αιτημάτων από τους χρήστες προς την υπηρεσία.

Παράμετρος 2: Η γεωγραφική θέση απ' όπου εκτελούνται τα αιτήματα,

Παράμετρος 3: Ο χρόνος κατά τον οποίο εκτελούνται τα αιτήματα προς την υπηρεσία.

Από την ανάλυση των δεδομένων που συλλέχθηκαν, για την υπηρεσία Wikipedia.org, προέκυψαν κάποια χρήσιμα συμπεράσματα.

- Οι παράμετροι που επηρεάζουν περισσότερο την απόδοση της υπηρεσίας, είναι αυτή του πλήθους των παράλληλων αιτημάτων που εκτελούνται προς την υπηρεσία, και η παράμετρος της γεωγραφικής θέσης απ' όπου εκτελούνται τα αιτήματα. Συνεπώς η παράμετρος του χρόνου εκτέλεσης των αιτημάτων, δεν επηρεάζει την απόδοσή της όσο οι προηγούμενες δύο. Αυτό σημαίνει ότι θα υπάρχουν μικρές διαφορές στην απόδοση της υπηρεσίας, όταν αυτή εκτελείται σε διαφορετικές χρονικές στιγμές.
- Η δημιουργία του συγκεκριμένου μοντέλου δυναμικής πρόβλεψης, είχε πολύ καλά αποτελέσματα, στην πρόβλεψη των τιμών απόδοσης της υπηρεσίας Wikipedia.org. Η σύγκριση μεταξύ των δεδομένων που συλλέχθηκαν και αυτών που προέβλεψε το μοντέλο, είχε μέση απόκλιση τα 85,18 χιλιοστά του δευτερολέπτου, πραγματοποιώντας προβλέψεις, για την απόδοση της υπηρεσίας, μέχρι και για δύο δευτερόλεπτα. Επιπλέον η μέγιστη απόκλιση ήταν 724,5 χιλιοστά του δευτερολέπτου.

8.4: Μελλοντική Έρευνα

Η δημιουργία μοντέλου δυναμικής πρόβλεψης της απόδοσης της υπηρεσιών, όπως της Wikipedia.org, που παρουσιάστηκε στην παρούσα εργασία, έχει ως σκοπό την επίλυση του προβλήματος έγκαιρης προετοιμασίας των πόρων, ώστε να προσφερθούν στις υπηρεσίες την στιγμή που τους χρειάζονται. Η συγκεκριμένη προσέγγιση, που ακολουθήθηκε, δεν περιλάμβανε την μελέτη παραμέτρων που αφορούν τους φυσικούς πόρους του υπολογιστικού νέφους. Για παράδειγμα, την επεξεργαστική ισχύ που χρησιμοποιεί η εφαρμογή για την εξυπηρέτηση των αιτημάτων. Εκτός όμως από τις παραμέτρους του υλικού του υπολογιστικού νέφους, θα μπορούσαν να μελετηθούν και επιπλέον παράμετροι της εφαρμογής, όπως το μέγεθος των δεδομένων που επεξεργάζεται κάθε χρονική στιγμή, ή ο συνολικός τους όγκος. Έτσι, μια χρήσιμη επέκταση του προτεινόμενου μοντέλου, θα ήταν η προσθήκη επιπλέον παραμέτρων, εκτός από αυτών που επιλέχθηκαν, για τον καλύτερο προσδιορισμό των πόρων που είναι αναγκαίοι κάθε χρονική στιγμή.

Η μελέτη παραμέτρων του υλικού, εκτός από αυτούς του υπολογιστικού νέφους, θα μπορούσε να επεκταθεί και σε αυτές των συσκευών που εκτελούν τα αιτήματα προς την

υπηρεσία. Διαφορετικές δυνατότητες, στην επεξεργασία και εμφάνιση των αποτελεσμάτων της υπηρεσίας, έχουν συσκευές με υψηλή επεξεργαστική ισχύ, όπως υπολογιστές τελευταίας τεχνολογίας, και διαφορετικές δυνατότητες έχουν συσκευές κινητής τηλεφωνίας πριν από δύο ή τρία χρόνια. Ακόμα και παράμετροι που δεν αφορούν το υλικό των συσκευών, όπως η ταχύτητα σύνδεσης στο διαδίκτυο, θα μπορούσαν να ληφθούν υπόψη στην απόδοση της υπηρεσίας. Έτσι, ανάλογα των δυνατοτήτων του πελάτη που εκτελεί το αίτημα προς την υπηρεσία, θα μπορούσε να επιστρέφεται διαφορετικό αποτέλεσμα, με μικρότερο όγκο δεδομένων, ή σταδιακή αποστολή και εμφάνιση του συνόλου των δεδομένων. Με αυτό τον τρόπο, θα μπορούσε να επιτευχθεί η επιθυμητή απόδοση της υπηρεσίας.

Επίσης, θα μπορούσαν να μελετηθούν διαφορετικά σενάρια που αφορούν την πρόβλεψη της απόδοσης υπηρεσιών. Συγκεκριμένα, θα μπορούσαν να μελετηθούν με τον ίδιο τρόπο που παρουσιάστηκε, διαφορετικοί τύποι υπηρεσιών ή υπηρεσίες με βάση άλλα κριτήρια από αυτά που τέθηκαν στην παρούσα εργασία. Επιπλέον, ακόμα και για την ίδια την υπηρεσία που μελετήθηκε (Wikipedia.org), ή και για οποιαδήποτε άλλη υπηρεσία, θα μπορούσε να γίνει συλλογή δεδομένων από περισσότερες πόλεις. Έτσι, το μοντέλο θα είχε μεγαλύτερη ακρίβεια στην πρόβλεψη της απόδοσης της υπηρεσίας. Ακόμα όμως και για τις ίδιες ή περισσότερες πόλεις, θα μπορούσε να μελετηθεί το σενάριο, όπου κάποιες πραγματοποιούν μεγαλύτερο πλήθος αιτημάτων σε σχέση με τις υπόλοιπες (π.χ. αν υπάρχει διαφορά στο μοντέλο, όταν από την Αθήνα πραγματοποιούνται τρεις φορές περισσότερα αιτήματα από αυτά που εκτελούνται, την ίδια χρονική στιγμή, από Άμστερνταμ και Ρώμη).

Ακόμα, το λογισμικό ανάπτυξης, εποπτείας και λειτουργίας του υπολογιστικού νέφους, μπορεί να επηρεάσει την απόδοση των εφαρμογών, που εκτελούνται σε αυτό. Για παράδειγμα, μπορεί το μοντέλο πρόβλεψης να προκύψει διαφοροποιημένο, όταν για την παροχή υπηρεσιών χρησιμοποιείται λογισμικό Openstack ή vCloud [63]. Τέλος, μια μελλοντική έρευνα, θα μπορούσε να μελετήσει τις πιθανές σχέσεις μεταξύ των διάφορων παραμέτρων, εκτός της απόδοσης της υπηρεσίας. Αν δηλαδή, δεν υπάρχει πλήρης ανεξαρτησία μεταξύ των παραμέτρων, αλλά η αυξομείωση της μιας, επηρεάζει τις τιμές της άλλης.

Βιβλιογραφικές Αναφορές

1. Mell, P. and T. Grance, *The NIST definition of cloud computing*. National Institute of Standards and Technology, 2009. **53**(6): p. 50.
2. Dada, G. *Top 10 Common Uses for the Cloud for 2012*. 2012; Available from: <http://www.rackspace.com/blog/top-10-common-uses-for-the-cloud-for-2012/>.
3. Amazon Web Services, I. *Amazon web services*. 2014; Amazon Web Services (abbreviated AWS) is a collection of remote computing services (also called web services) that together make up a cloud computing platform, offered over the Internet by Amazon.com.]. Available from: <http://aws.amazon.com>.
4. Inc, R. *The open cloud company*. 2014; Rackspace Inc. is an IT hosting company based in Windcrest, Texas, USA, a suburb of San Antonio, Texas.]. Available from: <http://www.rackspace.com>.
5. Inc, G. *Google Cloud Platform*. 2014; Google App Engine is a platform as a service (PaaS) cloud computing platform for developing and hosting web applications in Google-managed data centers.]. Available from: <https://cloud.google.com/products/app-engine/>.
6. Inc, M. *Microsoft Azure*. 2014; Microsoft Azure is a cloud computing platform and infrastructure, created by Microsoft, for building, deploying and managing applications and services through a global network of Microsoft-managed datacenters. It provides both PaaS and IaaS services.]. Available from: <http://azure.microsoft.com/en-us/>.
7. Inc, I. *Cloud Services*. 2014; Founded in 2005, SoftLayer Technologies, Inc. is a dedicated server, managed hosting and cloud computing provider.]. Available from: <http://www.softlayer.com/>.
8. Inc, S. *SAP Cloud for Customer - Sales, Service, Marketing, and Social*. 2014; Available from: <http://www.sap.com/pc/tech/cloud.html>.
9. Weinhardt, C., et al., *Cloud computing—a classification, business models, and research directions*. Business & Information Systems Engineering, 2009. **1**(5): p. 391-399.
10. Rogers, O. and D. Cliff, *A financial brokerage model for cloud computing*. Journal of Cloud Computing, 2012. **1**(1): p. 1-12.
11. Buyya, R., R. Ranjan, and R.N. Calheiros, *Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services*, in *Algorithms and architectures for parallel processing*. 2010, Springer. p. 13-31.
12. Phillips, E. and D. Pugh, *How to get a PhD: A handbook for students and their supervisors*. 2010: McGraw-Hill International.
13. Erl, T., *Service-Oriented Architecture (SOA) Concepts, Technology and Design*. 2005.
14. Foster, I., C. Kesselman, and S. Tuecke, *The anatomy of the grid: Enabling scalable virtual organizations*. International journal of high performance computing applications, 2001. **15**(3): p. 200-222.
15. Youseff, L., M. Butrico, and D. Da Silva. *Toward a unified ontology of cloud computing*. in *Grid Computing Environments Workshop, 2008. GCE'08*. 2008. IEEE.
16. Chiang, A. and K. Wainwright, *Fundamental methods of mathematical economics*. McGraw-Hill, New York, 2005.
17. Plummer, D.C., et al., *Five refining attributes of public and private cloud computing*. Gartner Research, Paper G, 2009. **167182**.
18. Model, O.M.U., *Compute Infrastructure as a Service*. Tech. rep., Open Data Center Alliance (OCDA), 2012.

- http://www.opendatacenteralliance.org/docs/ODCA_Compute_IaaS_MasterUM_v1.0_Nov2012.pdf.
19. Schouten, E. *Rapid elasticity and the cloud*. 2012; Available from: <http://thoughtsoncloud.com/2012/09/rapid-elasticity-and-the-cloud/>.
 20. Wolski, R., *Cloud Computing and Open Source: Watching Hype meet Reality*, May 2011. 2013.
 21. Cohen, R. *Defining Elastic Computing*. 2009; Available from: <http://www.elasticvapor.com/2009/09/defining-elastic-computing.html>.
 22. Herbst, N.R., S. Kounev, and R. Reussner. *Elasticity in Cloud Computing: What It Is, and What It Is Not*. in *Proceedings of the 10th International Conference on Autonomic Computing (ICAC 2013)*, San Jose, CA. 2013.
 23. Wang, L., et al., *Cloud computing: a perspective study*. *New Generation Computing*, 2010. **28**(2): p. 137-146.
 24. Kyriazis, D., *Cloud Computing Service Level Agreements: Exploitation of Research Results*. EUROPEAN COMMISSION DIRECTORATE GENERAL COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY UNIT E2 - SOFTWARE AND SERVICES, CLOUD, 2013.
 25. Frey, S., C. Reich, and C. Lüthje. *Key Performance Indicators for Cloud Computing SLAs*. in *EMERGING 2013, The Fifth International Conference on Emerging Network Intelligence*. 2013.
 26. Official-Site, I., *ITIL® glossary and abbreviations*. *ITIL officialsite*. 2011.
 27. VMware. 2014; VMware, Inc. is an American software company that provides cloud and virtualization software and services, and was the first to successfully virtualize the x86 architecture.]. Available from: <http://www.vmware.com/>.
 28. Citrix Systems, I. *Citrix XenServer*. 2014; Xen is a hypervisor using a microkernel design, providing services that allow multiple computer operating systems to execute on the same computer hardware concurrently. The University of Cambridge Computer Laboratory developed the first versions of Xen.]. Available from: <http://www.citrix.com/products/xenserver/overview.html>.
 29. Microsoft. *Server and Cloud Platform*. 2014; Available from: <http://www.microsoft.com/en-us/server-cloud/cloud-os/modern-data-center.aspx>.
 30. A Vouk, M., *Cloud computing—issues, research and implementations*. *CIT. Journal of Computing and Information Technology*, 2008. **16**(4): p. 235-246.
 31. Chowdhury, N.M.K. and R. Boutaba, *Network virtualization: state of the art and research challenges*. *Communications Magazine, IEEE*, 2009. **47**(7): p. 20-26.
 32. Peltz, C., *Web services orchestration and choreography*. *Computer*, 2003. **36**(10): p. 46-52.
 33. Chinnici, R., et al., *Web services description language (wsdl) version 2.0 part 1: Core language*. W3C recommendation, 2007. **26**: p. 19.
 34. Cerami, E., *Web services essentials: distributed applications with XML-RPC, SOAP, UDDI & WSDL*. 2002: " O'Reilly Media, Inc."
 35. Li, R., et al., *SOAP: short oligonucleotide alignment program*. *Bioinformatics*, 2008. **24**(5): p. 713-714.
 36. He, H., *What is service-oriented architecture*. *Publicação eletrônica em*, 2003. **30**: p. 1-5.
 37. Dean, J. and S. Ghemawat, *MapReduce: simplified data processing on large clusters*. *Communications of the ACM*, 2008. **51**(1): p. 107-113.
 38. Cooper, H.M., *Synthesizing research: A guide for literature reviews*. Vol. 2. 1998: Sage.

39. Brereton, P., et al., *Lessons from applying the systematic literature review process within the software engineering domain*. Journal of systems and software, 2007. **80**(4): p. 571-583.
40. K. Koumaditis, M.T., *SOA Governance from a Healthcare Perspective*, in *IADIS International Conference e-Health 2013*. 2013.
41. Chen, Y.-J., et al., *Solid state drive*. 2012, Google Patents.
42. Kousiouris, G., T. Cucinotta, and T. Varvarigou, *The effects of scheduling, workload type and consolidation scenarios on virtual machine performance and their prediction through optimized artificial neural networks*. Journal of Systems and Software, 2011. **84**(8): p. 1270-1291.
43. Kousiouris, G., et al. *A service-oriented framework for gnu octave-based performance prediction*. in *Services Computing (SCC), 2010 IEEE International Conference on*. 2010. IEEE.
44. Kousiouris, G., et al., *Parametric Design and Performance Analysis of a Decoupled Service-Oriented Prediction Framework based on Embedded Numerical Software*. 2012.
45. Kousiouris, G., et al., *Dynamic, behavioral-based estimation of resource provisioning based on high-level application terms in Cloud platforms*. Future Generation Computer Systems, 2014. **32**: p. 27-40.
46. Zhu, Q. and G. Agrawal. *Resource provisioning with budget constraints for adaptive applications in cloud environments*. in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*. 2010. ACM.
47. Jie, Y., J. Qiu, and Y. Li. *A profile-based approach to just-in-time scalability for cloud applications*. in *Cloud Computing, 2009. CLOUD'09. IEEE International Conference on*. 2009. IEEE.
48. Islam, S., et al., *Empirical prediction models for adaptive resource provisioning in the cloud*. Future Generation Computer Systems, 2012. **28**(1): p. 155-162.
49. *Open Cloud Computing Interface*. Available from: <http://occi-wg.org/>.
50. *Open Virtualization Format Specification*. September 2009; Available from: <http://xml.coverpages.org/DMTF-OVF-v10-DSP0243.pdf>.
51. Misra, S.C. and A. Mondal, *Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment*. Mathematical and Computer Modelling, 2011. **53**(3): p. 504-521.
52. *The R Project for Statistical Computing*. 2014; Available from: <http://www.r-project.org/>.
53. Nutonian. *Eureka Desktop Tool*. 2014; [Eureka is a breakthrough technology that uncovers the intrinsic relationships hidden within complex data.]. Available from: <http://www.nutonian.com/products/eureka/>.
54. Δημόπουλος, Γ., *Μεθοδολογία Έρευνας*. 2007: Πανεπιστήμιο Κύπρου, Τμήμα Πληροφορικής.
55. Paine, C., *Data collection techniques used to study programming-and AESOP*.
56. Oracle. *Net Beans*. 2014; [Quickly and easily develop desktop, mobile and web applications with Java, HTML5, PHP, C/C++ and more. NetBeans IDE is FREE, open source, and has a worldwide community of users and developers.]. Available from: <https://netbeans.org/>.
57. Berners-Lee, T., R. Fielding, and H. Frystyk, *Hypertext transfer protocol--HTTP/1.0*. 1996, May.
58. Forouzan, B.A., *TCP/IP protocol suite*. 2002: McGraw-Hill, Inc.
59. Foundation, W. *Wikimedia servers*. 2014; [Hosting servers for wikipedia.org]. Available from: http://meta.wikimedia.org/wiki/Wikimedia_servers.

60. Ripley, B.D., *The R project in statistical computing*. MSOR Connections, 2001. **1**(1): p. 23-25.
61. Τζαβαλής, Η., *Οικονομετρία*. 2008: Εκδόσεις ΟΠΑ. 605.
62. Haykin, S.S., et al., *Neural networks and learning machines*. Vol. 3. 2009: Pearson Education Upper Saddle River.
63. Maiya, M., et al. *Quantifying manageability of cloud platforms*. in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*. 2012. IEEE.

Πανεπιστήμιο Πειραιώς