



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΟΛΥΔΙΑΣΤΑΤΟΙ ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΚΑΙ
ΔΙΑΧΕΙΡΙΣΗΣ ΓΝΩΣΗΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΙΣΟΜΟΡΦΙΚΗ
ΥΠΕΡΚΥΒΙΚΗ ΤΟΠΟΛΟΓΙΑ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΒΑΣΙΛΕΙΟΣ Γ. ΝΙΚΟΛΟΠΟΥΛΟΣ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού και Μηχανικού Η/Υ
Imperial College Λονδίνου

ΑΘΗΝΑ

Δεκέμβριος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΟΛΥΔΙΑΣΤΑΤΟΙ ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗΣ ΓΝΩΣΗΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΙΣΟΜΟΡΦΙΚΗ ΥΠΕΡΚΥΒΙΚΗ ΤΟΠΟΛΟΓΙΑ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Συμβουλευτική Επιτροπή : Βασίλειος Λούμος
Ελευθέριος Καγιάφας
Κωνσταντίνος Παπαδουσσέας

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 5η Δεκεμβρίου 2011

.....
Βασίλειος Λούμος
Καθηγητής ΕΜΠ

.....
Ελευθέριος Καγιάφας
Καθηγητής ΕΜΠ

.....
Κωνσταντίνος Παπαδουσσέας
Αναπληρωτής Καθηγητής ΕΜΠ

.....
Ιωάννης Ψαρράς
Καθηγητής ΕΜΠ

.....
Μαρία Ιωαννίδου
Καθηγήτρια ΕΜΠ

.....
Χρήστος Αναγνωστόπουλος
Επίκουρος Καθηγητής
Πανεπιστημίου Αιγαίου

.....
Παντελής Μπίσκας
Λέκτορας ΑΠΘ

.....
Βασίλειος Γ. Νικολόπουλος
Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Βασίλειος Γ. Νικολόπουλος, 2011
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

*Αφιερώνεται στην γυναίκα μου Ιωάννα,
που τόσα χρόνια δίπλα μου, μου έδινε πάντα την δύναμη
να γίνομαι όλο και καλύτερος και στον γιό μου Γιώργο
που ήταν εκεί πάντα δίπλα μου στην συγγραφή...*

ΕΥΧΑΡΙΣΤΙΕΣ

Μέσα σε όλα αυτά τα χρόνια σκληρής δουλειάς, πάντα υπάρχουν άνθρωποι που σε βοηθούν να ξεπεράσεις δυσκολίες και εμπόδια και άνθρωποι που σε εμπνέουν να δημιουργήσεις. Αυτός ο άνθρωπος που τα έπραξε και τα δύο; που με ενέπνευσε να δημιουργήσω και να φτάσω εδώ που είμαι, δεν είναι άλλος από τον Καθηγητή μου και Επιβλέποντα κ. Βασίλη Λούμο. Θέλω να τον ευχαριστήσω εκ βάθους, για την καθοδήγηση του, την έμπνευσή του, την δημιουργικότητά του, την όρεξή του και την συνολική του επιρροή που είχε στην δουλειά μου, το έργο μου και την εξέλιξή μου. Είναι κάτι που κάθε Πανεπιστημιακός δάσκαλος πρέπει να κάνει. Τον ευχαριστώ θερμά.

Επίσης, ιδιαίτερες ευχαριστίες στον Καθηγητή κ. Ε. Καγιάφα, όπου με τις δικές του υποδείξεις σε σημαντικά θέματα και σταθμούς της έρευνάς μου, όλα αυτά τα χρόνια, αλλά και της συνολικής μου πορείας στο Εργαστήριο Τεχνολογίας Πολυμέσων του ΕΜΠ, μου έβαζε πάντα νέους δημιουργικούς προβληματισμούς και τροφή για περαιτέρω σκέψη.

Θερμές ευχαριστίες και στον Καθηγητή κ. Παπαοδυσσέα και όλη την επιτροπή Καθηγητών, που μου έκανε την τιμή να είναι στην 7μελή.

Πολλά ευχαριστώ σε όλους τους ανθρώπους και ερευνητές που συνεργάστηκα όλα αυτά τα χρόνια στο εργαστήριο και στο ΕΜΠ και που σε κάθε συζήτηση υπήρχε μία νέα ιδέα, ένας νέος προβληματισμός που συνέβαλε θετικά στο ερευνητικό μου έργο.

Τέλος, χωρίς την συμπαράσταση της οικογένειάς μου και της γυναίκας μου Ιωάννας όλα αυτά τα χρόνια, που πάντα με προσγείωναν και με έφερναν στον σωστό δρόμο, δεν θα υπήρχε πραγματική και ουσιαστική δουλειά. Τους ευχαριστώ θερμά.

Δεκέμβριος 2011

Βασίλειος Νικολόπουλος

ΠΕΡΙΛΗΨΗ

Η παρούσα Διατριβή επικεντρώνεται στην έρευνα και εφαρμογή συγκεκριμένων καινοτόμων μεθοδολογιών διαχείρισης γνώσης που έχουν σαν απώτερο στόχο την θεμελίωση νέων αλγοριθμικών τεχνικών συσταδοποίησης σε πραγματικό χρόνο δομημένου ή μη-δομημένου περιεχομένου, με σκοπό την πιο αποτελεσματική και γρήγορη εύρεση και συσχέτιση πληροφορίας. Η φιλοσοφία βασίζεται σε μία νέα χρήση της Υπερκυβικής Τοπολογίας (Hypercube) και του αλγορίθμου υπερκυβικής δρομολόγησης του L. Valiant, αλλά και στην χρήση δυναμικών αναδρομικών μεθόδων συσταδοποίησης.

Ένα από τα βασικά προβλήματα και θέματα στην διαχείριση γνώσης και πληροφορίας, είναι ο μεγάλος αριθμός και όγκος δεδομένων, όπου βάσει της Web 2.0 φιλοσοφίας και των κοινωνικών δικτύων, αυξάνεται γεωμετρικά. Η μεθοδολογία και οι αλγοριθμικές τεχνικές που αναπτύχθηκαν στα πλαίσια της παρούσας διατριβής, πιλοτικά εφαρμόστηκαν στον τομέα της Ενέργειας, των Ευφών Δικτύων, του smart metering και των Ενεργειακών Πληροφοριακών Συστημάτων. Ο τομέας των ευφών δικτύων θα δημιουργήσει τα επόμενα χρόνια petabytes από δεδομένα, λόγω της εγκατάστασης έξυπνων μετρητών ηλεκτρικής ενέργειας. Η Διατριβή προτείνει μία νέα μεθοδολογία και μαθηματικό μοντέλο για ανάλυση δεδομένων σε πραγματικό χρόνο και συσχέτιση συγκριμένων δεικτών, με σκοπό την διαχείριση συμπερασμάτων και υπηρεσιών, ως εργαλείο λήψης απόφασης και συγχώνευση γνώσης

Ο αλγόριθμος και οι παραλλαγές του εφαρμόστηκαν σε δύο κατηγορίες πληροφοριακών πόρων: 1) στατιστικά δεδομένα (log statistical files) από LMS (Learning Management Systems), δηλ. Εκπαιδευτικά Πληροφοριακά Συστήματα για ανάλυση και εξόρυξη γνώσης σχετικά με διαδικασίες και αξιολόγηση τηλεκπαίδευσης (e-learning) αλλά και 2) σε ενεργειακά δεδομένα με απώτερο στόχο την δημιουργία ενός έξυπνου μεσιμικού (middleware) το οποίο, μέσω ενός cloud-based Ενεργειακού Πληροφοριακού Συστήματος να δημιουργήσει έξυπνα και προσαρμοστικά ενεργειακά προφίλ καταναλωτών και να παράσχει έξυπνες υπηρεσίες στην ενεργειακή αγορά. Οι εφαρμογές του αλγορίθμου στον τομέα της Ενέργειας αναλύονται και συγκρίνονται διεξοδικά, με βάση τις διάφορες μαθηματικές μεθοδολογίες κατανεμημένης συσταδοποίησης σε δυαδικό υπο-χώρο.

Λέξεις Κλειδιά: Ενεργειακή πληροφορική, υπερκύβος, διαχείριση γνώσης, k-means, συσταδοποίηση, ενεργειακά πληροφοριακά συστήματα, ευφυή δίκτυα, διαχείριση ενέργειας

ABSTRACT

The present research Thesis focuses on the scientific field of knowledge management and data fusion, by using specific algorithmic procedures based on hypercubic models and recursive k-means clustering.

One of the basic future problems of modern data analysis and knowledge management will be the huge amount and volume of generated data, from internet, social media and Web 2.0 technologies. Especially in the domain of Energy and Smart Grids, some petabytes of information will be created by the smart metering infrastructure. The increasing number and speed of internet connections can be seen as the tool for measuring electricity and consumers' behavior on energy consumption, giving an added value to consumer access and interactive web approach to energy and energy related solutions. An increasing number of consumers from all levels of the society, cultures, lifestyles and social status have continuous internet access to social networks and internet groups, through broadband connections.

In this research Thesis, a new methodology and algorithmic procedure is proposed, based on distributed hypercubic recursive clustering process (RKM: Recursive K- Means). A web-based knowledge system is proposed, using distributed cloud architecture and metering grids over ADSL broadband connections. By using some clustering algorithms and a web middleware, we analyse and observe energy profiles over time. The resulted clusters and centroids, are projected and statistically analysed over time, producing a centroid-locus. Hypercube topology was used for efficient data management and software agent-based parallel analysis. The system operates efficiently on a multi-tier cloud-based middleware that generates in real-time various service software components to the end consumers.

There are case studies on real Greek energy measurements and some Learning Management Systems (LMS) data, for the first time in Greece, indicating a compact and efficient distributed procedure that could analyse and produce adaptive personalised information services. The technique allows creation of dynamic clusters of customers, each with a representative centroid, which can lead to the discovery of links among different consumers, categories and then to the possibility of introducing new adaptive service strategies.

Key Words: Energy informatics, hypercube, clustering, k-Means, knowledge management, data fusion, energy information systems, smart grids, smart metering, energy analytics

ΠΕΡΙΕΧΟΜΕΝΑ

| | |
|---|-----|
| <i>ΕΥΧΑΡΙΣΤΙΕΣ</i> | 7 |
| <i>ΠΕΡΙΛΗΨΗ</i> | 9 |
| <i>ABSTRACT</i> | 10 |
| <i>ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ</i> | 13 |
| <i>ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ</i> | 14 |
| <i>0. ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΑΤΡΙΒΗΣ - ΣΥΝΕΙΣΦΟΡΑ</i> | 16 |
| 0.1 Εισαγωγή στο Πρόβλημα | 16 |
| 0.2 Υπάρχουσες Προσεγγίσεις και Ιστορική Αναδρομή | 16 |
| 0.3 Συνεισφορά Διατριβής και Μεθοδολογία | 19 |
| 0.4 Δομή Διατριβής | 20 |
| <i>1. ΔΙΑΧΕΙΡΙΣΗ ΓΝΩΣΗΣ & ΣΥΣΤΗΜΑΤΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ</i> | 22 |
| 1.1 Συστήματα Διαχείρισης Γνώσης και Λήψης Απόφασης | 22 |
| 1.2 Εξόρυξη Δεδομένων και Εφαρμογή σε Χρονοδυναμικά δεδομένα | 23 |
| <i>2. ΥΠΕΡΚΥΒΙΚΗ ΙΣΟΜΟΡΦΙΚΗ ΤΟΠΟΛΟΓΙΑ</i> | 28 |
| 2.1 Υπερκύβος και Ισομορφική Τοπολογία | 28 |
| 2.2 Τεχνικές Διαχείρισης Γνώσης μέσω Υπερκυβικής Δρομολόγησης | 31 |
| 2.3 Υπερκυβικοί Αλγόριθμοι Δρομολόγησης Γνωσιακών Πρακτόρων | 33 |
| <i>3. ΘΕΩΡΙΑ ΑΝΑΔΡΟΜΙΚΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ</i> | 48 |
| 3.1 Data Mining και Θεωρία Συσταδοποίησης (Clustering) | 48 |
| 3.2 Αλγόριθμος RKM (recursive K-Means) | 55 |
| 3.3 Ποιοτική Αλγεβρική Ανάλυση Κεντροειδών | 60 |
| 3.4 Εφαρμογή σε Δεδομένα Μαθησιακού LMS | 62 |
| <i>4. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΛΓΟΡΙΘΜΩΝ ΣΤΗΝ ΕΝΕΡΓΕΙΑ</i> | 67 |
| 4.1 Ενεργειακά Πληροφοριακά Συστήματα & Ευφυή Δίκτυα | 67 |
| 4.2 Ενεργειακή Πληροφορική και Ενεργειακή μοντελοποίηση | 76 |
| 4.3 Εφαρμογή Υπερκυβικών Αλγόριθμων Γνώσης σε Ενεργειακά Δεδομένα (AMR) | 79 |
| 4.4 Ενεργειακά Κοινωνικά Δίκτυα και Εφαρμογή του RKM | 100 |
| <i>5. ΣΥΜΠΕΡΑΣΜΑΤΑ</i> | 102 |
| 5.1 Συμπεράσματα Διατριβής | 102 |
| 5.2 Μελλοντικές Προεκτάσεις στην Έρευνα και Καινοτομία | 103 |
| <i>6. ΔΗΜΟΣΙΕΥΣΕΙΣ ΔΙΑΤΡΙΒΗΣ</i> | 104 |
| <i>7. REFERENCES</i> | 107 |

| | |
|---|-----|
| Εικόνα 1. Μακροσκοπική αρχιτεκτονική συστήματος Διαχείρισης Γνώσης για Ενέργεια | 23 |
| Εικόνα 2. Στάδια στατικής εξόρυξης Γνώσης | 24 |
| Εικόνα 3. Σχήμα 8-διάστατου υπερκύβου (2^3) με 8 κόμβους και ο ισομορφισμός του | 30 |
| Εικόνα 4. Σχήμα 16-διάστατου υπερκύβου (2^4) με 16 κόμβους | 30 |
| Εικόνα 5. Σχήμα ισομορφικού υπερκύβου βαθμού διάστασης $N=9$ | 31 |
| Εικόνα 6. Συνολική αρχιτεκτονική υπερκυβικής διαχείρισης γνώσης | 32 |
| Εικόνα 7. Στάδια κλασσικού υπερκυβικού routing των software agents | 34 |
| Εικόνα 8. Συνολικός αλγόριθμος δρομολόγησης υπερκυβικών πρακτόρων | 39 |
| Εικόνα 9. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε ντετερμινιστική δρομολόγηση | 40 |
| Εικόνα 10. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε ντετερμινιστική δρομολόγηση | 41 |
| Εικόνα 11. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε στοχαστική δρομολόγηση | 41 |
| Εικόνα 12. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε στοχαστική δρομολόγηση | 42 |
| Εικόνα 13. Απόσταση Hamming ντετερμινιστικού μοντέλου | 43 |
| Εικόνα 14. Συντελεστές CB και CR πρακτόρων (ντετερμινιστικό μοντέλο) | 44 |
| Εικόνα 15. Δυναμική απόσταση Hamming στοχαστικού μοντέλου (Valiant) | 45 |
| Εικόνα 16. Συντελεστές CB και CR πρακτόρων (ντετερμινιστικό μοντέλο) | 45 |
| Εικόνα 17. Κίνηση πρακτόρων ανά στάδια δρομολόγησης (στοχαστικό μοντέλο με $N=8$) | 46 |
| Εικόνα 18. Συντελεστής καθυστέρησης CB πρακτόρων σε 3D (στοχαστικό μοντέλο $N=8$) | 46 |
| Εικόνα 19. Διαδικασία Εξόρυξης Δεδομένων | 49 |
| Εικόνα 20. Αλγόριθμος RKM σε εφαρμογή στην ενέργεια, βάσει συγκεκριμένων KPIs | 58 |
| Εικόνα 21. Λειτουργική αρχιτεκτονική του RKM αλγορίθμου | 59 |
| Εικόνα 22. Αρχιτεκτονική ανάλυσης δεδομένων από Moodle LMS | 62 |
| Εικόνα 23. Διαδικασία ανάλυσης μαθησιακών δεδομένων | 63 |
| Εικόνα 24. Ανάλυση MO χρονικής παραμονής σε ενότητα από μαθητή | 64 |
| Εικόνα 25. Ανάλυση MO χρονικής παραμονής σε ενότητα από μαθητή | 65 |
| Εικόνα 26. Ανάλυση MO χρονικής παραμονής σε ενότητα από μαθητή | 65 |
| Εικόνα 27. Παίκτες που εμπλέκονται σε Ενεργειακά πληροφοριακά συστήματα | 67 |
| Εικόνα 28. Ενεργειακό Πληροφοριακό Σύστημα διαχείρισης Μετρήσεων AMR (MDM) | 69 |
| Εικόνα 29. Τμηματοποίηση Ευφών Δικτύων και υπο-συστημάτων υπηρεσιών | 74 |
| Εικόνα 30. Τεχνική αρχιτεκτονική για το Ελληνικό Case Study | 80 |
| Εικόνα 31. το σύστημα Energy Analytics που χρησιμοποιήθηκε για ανάλυση των μετρήσεων | 81 |
| Εικόνα 32. Παράδειγμα αιχμιακού φορτίου που ανιχνεύεται με τον δείκτη P_{av}/P_{max} | 87 |
| Εικόνα 33. Μηνιαίες περιοδικότητες που ανιχνεύονται με τον RKM σε συσχετισμένους κόμβους | 88 |
| Εικόνα 34. 3 δυναμικές συστάδες και πως κινούνται σε περίοδο 6 ημερών αναδρομικά | 89 |
| Εικόνα 35. Δυναμική αλλαγή πληθυσμών P_i | 90 |
| Εικόνα 36. Γεωμετρικός τόπος κεντροειδών (X,Y Variance) | 91 |
| Εικόνα 37. Αναδραστικά clusters για 50 κτίρια και 7 ημέρες ανάλυση | 94 |
| Εικόνα 38. Συσχέτιση 2 καταναλώσεων στη μονάδα της ημέρας, βάσει υπερκυβικών πρακτόρων | 96 |
| Εικόνα 39. Συντελεστές συσχέτισης μεγαλύτεροι από 0,6 μεταξύ πρώτου προφίλ και υπολοίπων | 98 |
| Εικόνα 40. Διάγραμμα συμπερασμού βάσει των ποιοτικών χαρακτηριστικών g , C , e και S | 99 |
| Εικόνα 41. Ενεργειακά Κοινωνικά Δίκτυα με συσχέτιση Ενεργειακού και κοινωνικού προφίλ | 100 |
| Εικόνα 42. Ενεργειακά Κοινωνικά Δίκτυα και ο συνδυασμός τους με τον RKM | 101 |

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

| | |
|--|------------|
| Πίνακας 1. Διάφορα μοντέλα Εξόρυξης γνώσης χρονο-δυναμικών δεδομένων | 25 |
| Πίνακας 2. Στατιστικά αποτελέσματα ντετερμινιστικού τρόπου δρομολόγησης | 47 |
| Πίνακας 3. Στατιστικά αποτελέσματα στοχαστικού τρόπου δρομολόγησης | 47 |
| Πίνακας 4. Ανάλυση δεικτών σε έναν συγκεκριμένο μαθητή από σμήνος πρακτόρων σε υπερκύβο | 66 |
| Πίνακας 5. Αποτελέσματα πρακτόρων σε έναν 6-διάστατο υπερκύβο για υπολογισμό KPIs | 86 |
| Πίνακας 6. Αποτελέσματα του RKM για 90 κτίρια και με 6 iteration ημερήσια, N=3 clusters | 87 |
| Πίνακας 7. Χρήση δεικτών από RKM για Ενεργειακή Στρατηγική και υπηρεσίες | 100 |

0. ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΑΤΡΙΒΗΣ - ΣΥΝΕΙΣΦΟΡΑ

0.1 Εισαγωγή στο Πρόβλημα

Ο ρόλος της εξαγωγής πληροφοριών, στα πλαίσια της ανάκτησης πληροφοριών και διαχείρισης γνώσης, είναι η αναγνώριση εξειδικευμένης πληροφορίας και η εξαγωγή γνώσης από μη δομημένα δεδομένα με μηχανικό (αυτόματο) τρόπο. Αντίθετα με την κλασσική ανάκτηση πληροφοριών, σύμφωνα με την οποία η αναζήτηση γίνεται με βάση συγκεκριμένες λέξεις-κλειδιά και το αποτέλεσμα περιλαμβάνει μόνο κείμενα στα οποία βρίσκεται (ενδεχομένως) η χρήσιμη πληροφορία, η εξόρυξη πληροφοριών στοχεύει ακριβώς στην αναγνώριση της χρήσιμης μόνο πληροφορίας και το περιβάλλον (context) στο οποίο αυτή εμφανίζεται. Ειδικά και όταν το σύστημά μας απαιτεί και συσχέτισμό πληροφορίας σε πραγματικό χρόνο ή διασύνδεση ειδικών μαθηματικών δεικτών (KPIs) μεταξύ τους, τότε το πρόβλημα είναι μεγαλύτερο, όσον αφορά την αποτελεσματικότητα των μεθόδων, την αποδοτικότητά τους και την υπολογιστική πολυπλοκότητα τους.

Πιο ειδικά, δεδομένου του μεγάλου όγκου ενεργειακών πληροφοριών που παράγονται και διακινούνται σήμερα (κύριο χαρακτηριστικό των ευφυών δικτύων και του φαινομένου big data) το ζητούμενο στις μέρες μας είναι όχι απλώς η κατοχή της πληροφορίας – ο οποιοσδήποτε σήμερα μπορεί να έχει πρόσβαση σε σχετικά οποιαδήποτε πληροφορία – αλλά η διαχείριση της πληροφορίας και ο εντοπισμός της «σχετικής» πληροφορίας σε πραγματικό χρόνο. Έτσι, ενώ με μια κλασσική μηχανή αναζήτησης ο ενδιαφερόμενος θα λάβει ως απάντηση ένα σύνολο περιεχομένου που ενδεχομένως περιέχει την απάντηση που περιμένει, η εξόρυξη συσχετισμένων πληροφοριών στοχεύει στην σύγκριση πληροφορίας, την συσχέτισή της στο πεδίο του χρόνου καθώς και την συσταδοποίησή της σε κοινές ομάδες.

Υπάρχει λοιπόν έντονο πεδίο έρευνας και ανάλυσης / πρότασης νέων μεθοδολογιών ανάλυσης πληροφορίας σε πραγματικό χρόνο (stream mining) καθώς και συσχέτισής της στο πεδίο του χρόνου με ενδογενή και εξωγενή παράγοντες. Επίσης, η οπτικοποίηση αυτής της πληροφορίας αλλά και των αποτελεσμάτων, αποτελεί επίσης πεδίο έρευνας (πχ. Dynamic infographics)

0.2 Υπάρχουσες Προσεγγίσεις και Ιστορική Αναδρομή

Οι επιδόσεις ανάλογων συστημάτων που έχουν παρουσιαστεί κατά καιρούς σε συνέδρια και ερευνητικές εργασίες αναφέρονται σε έρευνα πάνω σε συγκεκριμένους τομείς (κείμενα

γραμμένα για συγκεκριμένους θεματικούς χώρους, αριθμητικές τιμές, δεδομένα μαθηματικά, κλπ) και διαφέρουν ανάλογα με το είδος της πληροφορίας που στοχεύουν να αναγνωρίσουν. Η σχετική έρευνα έχει γίνει πιο εντατική τα τελευταία χρόνια. Συνεχώς προτείνονται νέες μέθοδοι κατασκευής ανάλογων αρχιτεκτονικών και συστημάτων οι οποίες στοχεύουν σε καλύτερες επιδόσεις και στην απλοποίηση της επέκτασης της λειτουργίας υπαρχόντων συστημάτων σε νέους τομείς και νέες γλώσσες εφαρμογής. Έμφαση δίνεται κυρίως στην ανάπτυξη ειδικών συστημάτων μέσω αυτόματης εκμάθησης της ανάλυσης της πληροφορίας και των ιδιαιτεροτήτων του χώρου από το ίδιο το σύστημα.

Η έρευνα για την υποστήριξη αποφάσεων ξεκίνησε το 1971. Ο Scott Morton [2] έδειξε ότι οι managers επωφελήθηκαν από την χρήση ενός συστήματος διαχείρισης αποφάσεων μέσω υπολογιστή. Σύμφωνα με τους Klein και Methlie [1], τα πρώτα άρθρα DSS (Decision Support System) δημοσιεύτηκαν από διδακτορικούς φοιτητές ή καθηγητές σε επαγγελματικές σχολές που είχαν πρόσβαση σε συστήματα χρήσης υπολογιστών για πρώτη φορά: Project MAC στην σχολή Sloan, το Dartmouth Time Sharing Systems στην σχολή Tuck και στην Γαλλία στην σχολή HEC.

Μετά το 1990 οι ερευνητές εξέταζαν τις πιθανότητες για μια μελλοντική γενιά των DSS [8,10,11]. Διάφοροι ακαδημαϊκοί ερευνητές και developers λογισμικού διαπίστωσαν ότι το Παγκόσμιο Διαδίκτυο και οι τεχνολογίες Διαδικτύου δημιούργησαν νέες ευκαιρίες για την κατασκευή και ανάπτυξη συστημάτων υποστήριξης αποφάσεων. Χρησιμοποιώντας το Διαδίκτυο ως έναν μηχανισμό των DSS ενισχύθηκε η αρχική υπόθεση ως προς την ανάπτυξη των DSS. Μετά το 1995 (και το 1998) ο Bhargava και ο Krishnan παρουσίασαν το πρώτο από μια σειρά άρθρων στο Δίκτυο Αποφάσεων (Decision Net), μια ηλεκτρονική αγορά συστημάτων υποστήριξης αποφάσεων βασισμένα σε υπολογιστές [6,9]. Τον Αύγουστο του 2001, οι Power και Bhargava [7] κατέθεσαν την αρχική Έρευνα των Συστημάτων Υποστήριξης Αποφάσεων. Ο στόχος ήταν το να παρέχουν ένα χρήσιμο σημείο εκκίνησης για την πρόσβαση σε υλικό βασισμένο στο Δίκτυο και που αφορούσε τον σχεδιασμό, την ανάπτυξη, την αξιολόγηση και την αξιοποίηση των Συστημάτων Υποστήριξης Αποφάσεων. Κάποια άλλα άρθρα [3,4,5,12,13,14,15,16,18] που παρουσιάστηκαν σε διάφορα συνέδρια εξέτασαν επίσης την χρήση του Δικτύου και του Διαδικτύου για την υποστήριξη αποφάσεων σε διάφορα πεδία.

Πιο συγκεκριμένα στον χώρο της διαχείρισης ενεργειακής γνώσης, η χρήση του Internet ως μέσου για συλλογή ενεργειακών δεδομένων, έδωσε εργαλεία για την ανάλυση και αποθήκευση τέτοιων δομών στο διαδίκτυο. Τα τελευταία χρόνια, ενεργειακά δεδομένα που

προέρχονται από τηλε-μετρούμενους ενεργειακούς μετρητές σε ηλεκτρονική μορφή (Automatic Meter Readings - AMR) μπορούν να αποθηκευτούν και να αναλυθούν βάσει κάποιων διαδικτυακών συστημάτων διαχείρισης γνώσης και λήψης απόφασης. Την άνοιξη του 2000, τα πρώτα reports κάνανε την εμφάνισή τους, σχετικά με συστήματα διαχείρισης ενεργειακής εξοικονόμησης (Demand Reduction Programs - DRP) που υλοποιούνταν εκείνο τον καιρό στο UK από εταιρίες Ενέργειας (Κοινής Ωφελείας - Utilities) [17]. Αυτή η διαδικασία, όπου υπάρχει ένα σύστημα για ανάλυση ενεργειακών δεδομένων, οργάνωσή τους και ανάλυσή τους, παροχή απόφασης και σχετικών ενεργειακών υπηρεσιών, βάσει DSS μηχανισμών, ονομάζεται Ενεργειακό Πληροφοριακό Σύστημα (Energy Information System - EIS) [46]. Κάποιες συστηματικές προσπάθειες έγιναν στο UK [19, 20], σύμφωνα με αυτές έγινε μία πρώτη κατηγοριοποίηση των καταναλωτών βάσει του ενεργειακού τους προφίλ και του λογαριασμού τους. Άλλη προσπάθεια αναφέρεται στο [21, 22], όπου κάποιοι δείκτες εξήχθησαν από κάποιες καμπύλες φορτίου ενεργειακών καταναλωτών.

Επίσης, αρκετές αναφορές υπάρχουν και σε επίπεδο κτιριακής διαχείρισης, όπου υπάρχουν εφαρμογές διαχείρισης γνώσης και δεδομένων σε ενεργειακή κατανάλωση κτιρίων [23,24] και συστημάτων εμπορίας ενέργειας, συσχετισμένα με τις υπηρεσίες ευφυών δικτύων [25, 26,27]. Σχετικά με συγκεκριμένες εφαρμογές και αναφορές σε ενεργειακή εξόρυξη δεδομένων, υπάρχουν software εργαλεία τα οποία προσφέρουν σχετικά καλά αποτελέσματα, ειδικά σε θέματα τιμολογιακής πολιτικής [28], βελτιστοποίηση ενεργειακών στρατηγικών με ανάλυση προφίλ [29,30] και υπηρεσιών διαχείρισης ζήτησης (Demand Response - DR) και ανάλυσης προφίλ φορτίου καταναλωτή (Energy profiling) [31,32]. Πρόσφατα, υπάρχει και σχετική έρευνα στον τομέα της πληροφορικής για την ανίχνευση ασταθειών στην ηλεκτρική κατανάλωση και οπτικοποίηση κρυμμένων τάσεων (frauds detection, smart grid safety) σε καμπύλες ενεργειακών φορτίων καταναλωτών [33,34,35,36].

Παρόλα αυτά, οι παραπάνω τεχνικές δεν απέδωσαν τα μέγιστα και δεν ήταν αποτελεσματικές, όταν έπρεπε να συσχετίσουν διάφορους δείκτες και αν προβούν σε μία κοινή απόφαση (fused decision). Σε τέτοιου είδους περιπτώσεις, η λύση που προτείνεται στην παρούσα διατριβή μπορεί να δώσει εντυπωσιακά αποτελέσματα, στον τομέα του data stream mining [37,38] για την ανάλυση και γνωσιακή διαχείριση μεγάλων βάσεων δεδομένων με δεδομένα πραγματικού χρόνου

0.3 Συνεισφορά Διατριβής και Μεθοδολογία

Η συγκεκριμένη διατριβή βρίσκεται στον ερευνητικό χώρο της προχωρημένης Διαχείρισης Γνώσης και εφαρμογών Ενεργειακών Πληροφοριακών Συστημάτων με σκοπό την βέλτιστη παροχή έξυπνων ψηφιακών ενεργειακών υπηρεσιών.

Η μεθοδολογία που ακολουθήθηκε είναι η εξής:

- Σε πρώτο στάδιο, αναλύεται ο ισομορφικός υπερκύβος [82,87] ως βασική δομή διαφορικής τοπολογίας και χρησιμοποιείται ως το βασικό μοντέλο και σύστημα δρομολόγησης των κατανεμημένων παράλληλων πρακτόρων (software agents) που θα αναλύουν τα δεδομένα σε πραγματικό χρόνο.
- Το επόμενο στάδιο είναι η σύνδεση των υπερκυβικών δομών και κόμβων με την μέτρηση και ανάλυση των δεδομένων. Σε αυτό ρόλο παίζουν οι κατανεμημένοι υπερκυβικοί αλγόριθμοι, οι οποίοι εκτελούν τον αλγόριθμο RKM σε κάθε υπερκυβικό κόμβο, βάσει του στοχαστικού αλγορίθμου L. Valiant [40,41] και βάσει των δεδομένων που έχει ο κάθε υπερκυβικός κόμβος
- Το τελικό στάδιο είναι η ανάλυση των αποτελεσμάτων και οι γραφικές παραστάσεις που μας δείχνουν δυναμικές συσχετίσεις και αυτά εκφράζονται με τον γεωμετρικό τόπο των κεντροειδών (centroids) αλλά και τα λοιπά διαγράμματα των διαφόρων δεικτών (εντροπία, πληθυσμούς, συστάδες, κλπ)
- Τέλος, αναφέρονται και άλλες εφαρμογές (ενεργειακά κοινωνικά δίκτυα, κλπ) που μπορεί να έχει η συγκεκριμένη εφαρμογή του κατανεμημένου υπερκυβικού αλγορίθμου για δυναμικές συσταδοποιήσεις (RKM)

Η πρωτοτυπία και συνεισφορά στην επιστήμη της παρούσας διατριβής εντοπίζεται στα ακόλουθα σημεία:

- Στην εισαγωγή μίας νέας φιλοσοφίας και αλγοριθμικής προσέγγισης στην ανάλυση δεδομένων (large volume data) σε πραγματικό χρόνο
- Στην πρόταση νέων αναλυτικών διαδικασιών μοντελοποίησης της γνώσης, με αρχές διαφορικής τοπολογίας και αριθμητικής άλγεβρας.
- Στην χρησιμοποίηση των Κοινωνικών Δικτύων σαν βάση ανάλυσης της Ενεργειακής Συμπεριφοράς (Ενεργειακά Κοινωνικά Δίκτυα)
- Στην σύγκλιση πληροφορικής, αλγορίθμων διαχείρισης γνώσης και ενέργειας για παροχή συσχετισμένων έξυπνων ψηφιακών υπηρεσιών, φιλοσοφίας Web 2.0

- Αναλυτικότερα, η διατριβή προτείνει ένα γενικό προσαρμοστικό μοντέλο για την ανάλυση ενεργειακών μετρήσεων σε πραγματικό χρόνο και παροχή υπηρεσιών.
- Την χρήση αριθμητικής Γραμμικής άλγεβρας για ανάλυση ειδικών δεικτών διαχείρισης γνώσης και συμπερασμού
- Την ολοκλήρωση και τεκμηρίωση εννοιών και υπηρεσιών της σύγχρονης Ενεργειακής Πληροφορικής

Παράλληλα, η εργασία εμφανίζει σημεία πρωτοτυπίας τόσο ως αρχική σύλληψη, όσο και ως προς την τελική αξιοποίηση των γνώσεων του εν λόγω πεδίου της μοντελοποίησης και της βέλτιστης διαχείρισης ενεργειακών πόρων / μετρήσεων για διαχείριση υπηρεσιών. Αποτέλεσμα αυτού είναι τα Δύο Εθνικά Διπλώματα Ευρεσιτεχνίας που απονεμήθηκαν στον κ. Νικολόπουλο από τον OBI (Οργανισμό Βιομηχανικής Ιδιοκτησίας) και ένα Αμερικανικό Δίπλωμα Ευρεσιτεχνίας (US Provisional), σχετικά με την αλγοριθμική ανάλυση ενεργειακών πόρων σε πραγματικό χρόνο και την έξυπνη και προσαρμοστική παροχή ψηφιακών ενεργειακών υπηρεσιών.

Επίσης το πλήθος των βραβείων καινοτομίας και επιχειρηματικότητας για την εφαρμογή των παραπάνω ερευνητικών αποτελεσμάτων της Διατριβής ως καινοτόμες εμπορικές υπηρεσίες ενεργειακών πληροφοριακών συστημάτων.

0.4 Δομή Διατριβής

Η διατριβή περιλαμβάνει πέντε (5) βασικά κεφάλαια:

- Στο πρώτο κεφάλαιο (Κεφάλαιο 1) παρουσιάζονται τα βασικά πρότυπα και θεωρίες διαχείρισης γνώσης και συστήματα λήψης απόφασης, ειδικά για εφαρμογή σε ενεργειακά δεδομένα. Επίσης, γίνεται αναφορά σε δημοσιεύσεις για προχωρημένα θέματα web-based DSS, data mining καθώς και σε θέματα stream mining σε μεγάλα data sets .
- Στο δεύτερο κεφάλαιο (Κεφάλαιο 2) η διατριβή αναλύει σε βάθος την υπερκυβική τοπολογία (hypercube), από την πλευρά της διαφορικής τοπολογίας (ισομορφισμός, θεωρία γράφων, κλπ), εξηγεί και χρησιμοποιεί κάποια θεωρήματα της άλγεβρας για το πώς η συγκεκριμένη τοπολογία θα χρησιμοποιηθεί για την μοντελοποίηση και αναπαράσταση γνώσης (δεδομένα) και αναφέρεται εκτενώς στις διαδικασίες δρομολόγησης κατανεμημένων software agents μέσα σε αυτή. Επίσης, αναφέρεται

όλο το μαθηματικό υπόβαθρο στο οποίο θα βασιστεί ο δυναμικός αλγόριθμος RKM (Recursive K- Means) για την αναδρομική συσταδοποίηση των μετρούμενων ενεργειακών φορτίων και δεικτών που υπάρχουν σε κάθε υπερκυβικό κόμβο.

- Στο τρίτο κεφάλαιο (Κεφάλαιο 3) αναπτύσσεται η θεωρία και η αλγοριθμική προσέγγιση του Αναδρομικού Αλγόριθμου Υπεκρυβικής Συσταδοποίησης (Recursive K-Means) ο οποίος εκτελείται καταναμημένα από software agents σε μία υπερκυβική δομή. Η ειδική δομή του αλγορίθμου δημιουργεί αναδρομικά κεντροειδή και μεταβαλλόμενα metrics (population, entropy, centroid locus) τα οποία και αναλύονται σε καθορισμένα χρονικά διαστήματα.
- Στο τέταρτο κεφάλαιο (Κεφάλαιο 4) παρουσιάζονται αναλυτικά οι εφαρμογές των παραπάνω διαδικασιών και αλγορίθμων, στον τομέα της Ενέργειας και πιο συγκεκριμένα των ψηφιακών ενεργειακών υπηρεσιών και εφαρμογών ευφών δικτύων (smart grids). Σχεδιάστηκε και δημιουργήθηκε ένα προχωρημένο Ενεργειακό Πληροφοριακό σύστημα (EMIR [45]) και ειδικό μεσισμικό (middleware) όπου μετρούμενες ενεργειακές καταναλώσεις από διάφορες υποδομές (σπίτια, μεγάλα κτίρια) αναλύονται, με πολύ σημαντικά αποτελέσματα για παροχή προσωποποιημένων Web 2.0 ενεργειακών υπηρεσιών σε πραγματικό χρόνο.
- Στο πέμπτο και τελευταίο κεφάλαιο (Κεφάλαιο 5) πραγματοποιείται ο συνολικός απολογισμός της προτεινόμενης μεθοδολογίας και των αλγορίθμων στην Ενέργεια και τις ενεργειακές υπηρεσίες και πιστοποιείται η επίτευξη των στόχων της παρούσας διατριβής. Τέλος, παρουσιάζονται προοπτικές για περαιτέρω έρευνα σχετικά με τη θεματολογία που πραγματεύεται η παρούσα διατριβή, ειδικά στον τομέα της διαχείρισης γνώσης σε μεγάλα energy data sets και των προχωρημένων Ενεργειακών Πληροφοριακών Συστημάτων με εφαρμογές σε Ευφυή δίκτυα (Smart grids) και Κοινωνικά Ενεργειακά Δίκτυα (social energy networks)

1. ΔΙΑΧΕΙΡΙΣΗ ΓΝΩΣΗΣ & ΣΥΣΤΗΜΑΤΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ

1.1 Συστήματα Διαχείρισης Γνώσης και Λήψης Απόφασης

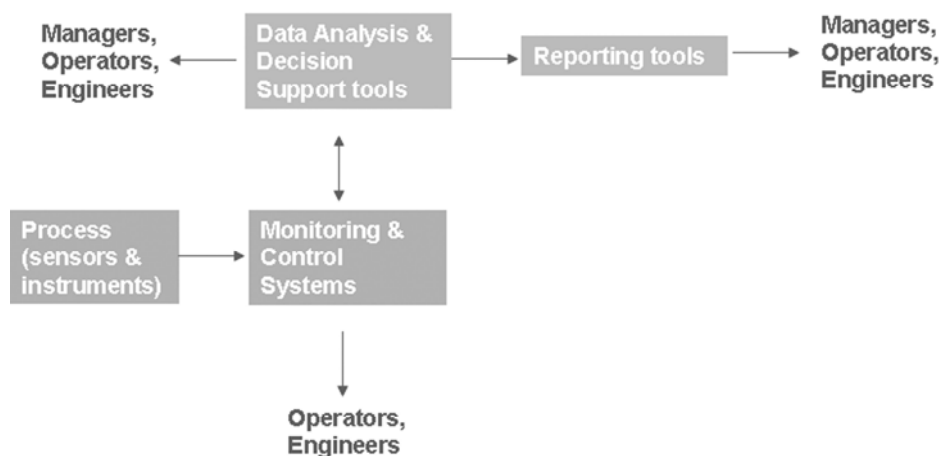
Οι παγκόσμιες τεχνολογίες έχουν μετατρέψει το σχεδιασμό, την ανάπτυξη, την εφαρμογή προγραμμάτων και την ανάπτυξη σε θέσεις ισχύος των συστημάτων υποστήριξης αποφάσεων. Αυτό το άρθρο αξιολογεί και συνοψίζει πρόσφατες τεχνολογικές εξελίξεις, την τρέχουσα χρήση των Web-based Decision Support Systems (DSS) καθώς και των τάσεων στην ανάπτυξη τέτοιων συστημάτων.

Η χρήση του Web-based υπολογισμού για να παρέχει επιδείξεις προϊόντων ή για να χρησιμοποιήσει εφαρμογές DSS για πρόσβαση εξ' αποστάσεως παραμένει λιγότερο κοινότυπη. Η ακαδημαϊκή λογοτεχνία σε ότι αφορά το διαδικτυακό DSS, ασχολείται διεξοδικά με εφαρμογές και πραγματοποιήσεις και μόνο μερικά άρθρα εξετάζουν αρχιτεκτονικά ζητήματα ή παρέχουν οδηγίες σχεδιασμού βασισμένες σε εμπειρικές ενδείξεις..

Τα μοντέρνα συστήματα ανάλυσης γνώσης και υποστήριξης αποφάσεων (Decision Support Systems - DSS) παρέχουν στους χρήστες τους μια μεγάλη γκάμα επιλογών/ικανοτήτων. Το τρέχον DSS διευκολύνει μια μεγάλη επιλογή από επιβεβλημένες αποφάσεις συμπεριλαμβανομένου και της συλλογής και ανάλυσης πληροφοριών, την κατασκευή μοντέλων, την ανάλυση ευαισθησίας, την συνεργασία, την εναλλακτική αξιολόγηση, και την λήψη αποφάσεων.

Συχνά τα DSS κατασκευάζονται και χρησιμοποιούνται για ανάλυση ad hoc, αλλά σε ολοένα αυξανόμενο ρυθμό, η υποστήριξη αποφάσεων ολοκληρώνεται/ενοποιείται σε διαδικασίες επιχειρήσεων και πληροφοριακά συστήματα. Στο κοντινό παρελθόν, το Παγκόσμιο Διαδίκτυο World Wide Web διευκόλυνε, φρόντισε και προώθησε μια ευρεία ανάκαμψη στην χρήση των τεχνολογιών αποφάσεων για να στηρίξουν τα καθήκοντα της υποστήριξης λήψεων αποφάσεων.

Το Παγκόσμιο Διαδίκτυο, είναι τώρα οι κύριες τεχνολογίες που επιτρέπουν την ολοκλήρωση υποστήριξης αποφάσεων μέσω υπολογιστών. Εξαιτίας του αναπτυσσομένου ενδιαφέροντος στο Διαδίκτυο, υπάρχουν πολλές συνεχιζόμενες προσπάθειες για να αναπτυχθεί και να ολοκληρωθεί το Web-based DSS [7] σε διάφορες περιοχές, όπως η υγεία, οι ιδιωτικές επιχειρήσεις, η κυβέρνηση και η εκπαίδευση.



Εικόνα 1. Μακροσκοπική αρχιτεκτονική συστήματος Διαχείρισης Γνώσης για Ενέργεια

Ένα Web-based DSS προσφέρει υποστήριξη αποφάσεων, χρησιμοποιώντας έναν «λεπτό και εύχρηστο για τον πελάτη» Web browser interface που μπορεί να ενοποιήσει την πλευρά του πελάτη σε ότι αφορά τις τεχνολογίες υπολογιστή όπως είναι τα Java applets ή τα JavaScript. Η υποστήριξη αποφάσεων Web-based έχει προοδεύσει πολύ από τότε που πρώτο-συζητήθηκε η ιδέα στο 3^ο Διεθνές Συνέδριο της Διεθνούς Κοινωνίας για την Υποστήριξη Συστημάτων Αποφάσεων (ISDSS) στο Χονγκ-Κονγκ το 1995 [3].

1.2 Εξόρυξη Δεδομένων και Εφαρμογή σε Χρονοδυναμικά δεδομένα

Η αποτελεσματική διαχείριση του τεράστιου όγκου δεδομένων που υπάρχει αυτή την στιγμή στο διαδίκτυο, αποτελεί ιδανικό κίνητρο για την έρευνα τεχνικών εξόρυξης γνώσης από μη-δομημένα δεδομένα. Ο όρος εξόρυξη γνώσης από δεδομένα (data mining) αναφέρεται στην εφαρμογή τεχνικών μηχανικής μάθησης (machine learning) σε μεγάλο όγκο δεδομένων για την ανακάλυψη γνώσης (knowledge discovery) από τα δεδομένα αυτά [43]. Η εξόρυξη γνώσης από δεδομένα του ιστού καλείται Web Mining [63,64,69,73,80,81] και αποτελεί σημείο συνάντησης ερευνητικών περιοχών όπως η μηχανική μάθηση, οι βάσεις δεδομένων, η επεξεργασία φυσικής γλώσσας, η ανάκτηση πληροφορίας, η εξαγωγή πληροφορίας και η μοντελοποίηση χρηστών.

Τα τελευταία χρόνια, πλήθος δημοσιεύσεων βγήκαν σχετικά με θέματα εξόρυξης γνώσης και συγκεκριμένων τεχνικών για την μεγαλύτερη επιτυχία [76]. Συνήθως, ανάλογα με το είδος των δεδομένων (κείμενο, δεδομένα φυσικών μεταβλητών, multimedia [53], κλπ) υπάρχει και η ανάλογη μαθηματική διαδικασία και προσέγγιση. Με χρήση της στατιστικής και λογικής ανάλυσης αλλά και με ανίχνευση προτύπων (pattern) γίνεται εξαγωγή συμπερασμάτων. Από τη συστηματική ανάλυση των, η εξόρυξη πληροφοριών μπορεί να προσφέρει δυνατότητες:

- Πρόβλεψης συμβάντων
- Κατηγοριοποίησης αντικειμένων και προσώπων μέσω αναγνώρισης προτύπων
- Συσταδοποίηση αντικειμένων κα προσώπων σε ομάδες βάση χαρακτηριστικών ιδιοτήτων
- Συσχετισμός συμβάντων
- Εύρεσης σχέσεων διαδοχής μεταξύ συμβάντων

Η εξόρυξη δεδομένων έχει διάφορα στάδια υλοποίησης [70]. Αρχικά τα δεδομένα συλλέγονται υφίστανται μια προ-επεξεργασία και αποθηκεύονται σε μια βάση δεδομένων. Στη συνέχεια τα δεδομένα υφίστανται επεξεργασία εφαρμόζοντας έξυπνους αλγόριθμους για να προκύψουν χρήσιμα συμπεράσματα, όπως μοντέλα συμπεριφοράς και συσχετίσεις μεταβλητών.



Εικόνα 2. Στάδια στατικής εξόρυξης Γνώσης

Η εξόρυξη γνώσης επάνω σε δυναμικά δεδομένα (stream) εμπεριέχει 3 βασικά στάδια: το 1^ο στάδιο της εισαγωγής και συγχρονισμού των δεδομένων εισόδου (input stream), το 2^ο στάδιο της εφαρμογής του αλγορίθμου ανάλυσης και εξόρυξης σε πραγματικό χρόνο (real time) και το 3^ο στάδιο της εξαγωγής συμπερασμάτων.

Υπάρχουν αρκετές αναφορές σε διάφορα μοντέλα και τεχνικές stream mining [52,54,55,56,57,74,75],

| Algorithm | Mining Task | Approach | Status |
|---------------------------------|-------------------|--|--|
| VFKM | K-Means | Sampling and reducing the number of passes at each step of the algorithm | Implemented and tested. |
| VFDT | Decision Trees | Sampling and reducing the number of passes at each step of the algorithm | Implemented and tested. |
| Approximate Frequent Counts | Frequent itemsets | Incremental Pruning and update of itemsets with each block of transactions | Implemented and tested. |
| FP- Stream | Frequent itemsets | Incremental Pruning and update of itemsets with each block of transactions and time-sensitive patterns extension | Implemented and tested. |
| Concept-Drifting Classification | Classification | Ensemble classifiers | Implemented and tested. |
| AWSOM | Prediction | Incremental Wavelets | Implemented and tested (This algorithm is designed to run on a sensor). The implementation is not on a sensor. |
| Approximate K-median | K-Median | Sampling and reducing the number of passes at each step of the | Analytical Study |

Πίνακας 1. Διάφορα μοντέλα Εξόρυξης γνώσης χρονο-δυναμικών δεδομένων

Παραδοσιακά, η ανάπτυξη μεθόδων ανάλυσης κατηγορικών δεδομένων υστερεί χρονικά των αντίστοιχων εξελίξεων για τα συνεχή δεδομένα. Όμως, οι απαιτήσεις της εποχής καθιστούν αναγκαία την εξέλιξη του κλάδου και την ανάπτυξη αλγορίθμων για τη διαχείριση τεράστιων συνόλων δεδομένων με κατηγορικές μεταβλητές. Είναι γνωστοί οι τύποι δεδομένων που συγκαταλέγονται στα κατηγορικά δεδομένα (δίτιμα, διακριτά, διατάξιμα, ονοματικά, διαστηματικά).

Σύμφωνα με τους Han και Kamber [76], μία ειδική ομάδα μεθόδων που συμπεριλαμβάνεται στις μεθόδους συσταδοποίησης, είναι η **συσταδοποίηση κατηγορικών δεδομένων**.

Οι αλγόριθμοι της ομάδας αυτής αναπτύσσονται ειδικά για δεδομένα όπου δε μπορεί να υπολογιστεί η ευκλείδεια απόσταση ή άλλες ανάλογες αποστάσεις που βασίζονται σε αριθμητικά δεδομένα.

Υπάρχουν διάφορες προσεγγίσεις για την αντιμετώπιση του ζητήματος της συσταδοποίησης κατηγορικών δεδομένων, οι περισσότερες εκ των οποίων στηρίζονται σε διαχωριστικές ή ιεραρχικές μεθόδους. Οι χαρακτηριστικότεροι αλγόριθμοι που έχουν κατασκευαστεί έτσι ώστε να διαχειρίζονται αποτελεσματικά τα κατηγορικά δεδομένα είναι οι **k-modes** και **ROCK** (*RObust Clustering using linKs*).

Ο αλγόριθμος **k-modes** [72] φέρεται να είναι ο πρώτος αλγόριθμος που δημοσιεύτηκε για τη διαχείριση ενός πολύ μεγάλου συνόλου δεδομένων με κατηγορικές μεταβλητές. Πρόκειται για μια επέκταση του γνωστού αλγορίθμου kmeans, που αφορά στη συσταδοποίηση συνεχών δεδομένων. Επομένως, είναι ένας διαιρετικός αλγόριθμος, όπως και ο kmeans. Στόχος του είναι η ανακάλυψη συστάδων, ενώ υιοθετεί νέες έννοιες, όπως την αντικατάσταση των κέντρων των συστάδων με τα «modes». Επίσης, εισάγει ένα νέο μέτρο ανομοιοτήτας για την εξέταση των κατηγορικών δεδομένων.

Ο **ROCK** [67,75] είναι ένας αντιπροσωπευτικός ιεραρχικός αλγόριθμος συσταδοποίησης κατηγορικών δεδομένων. Η νέα έννοια που εισάγει είναι τα «links» (δεσμοί ή σύνδεσμοι), ώστε να μετρήσει την ομοιότητα / εγγύτητα ανάμεσα σε ζεύγη σημείων. Έτσι, η μέθοδος συσταδοποίησης του ROCK βασίζεται σε μη μετρικά κριτήρια ομοιότητας, τα οποία είναι εφαρμόσιμα σε κατηγορικά σύνολα δεδομένων.

Ένα κοινό των δύο αυτών αλγορίθμων είναι ότι απαιτούν ως παράμετρο εισόδου τον αριθμό των συστάδων. Όμως, ο ROCK δουλεύει εντελώς διαφορετικά από ότι ο k-modes, όχι μόνο επειδή είναι ιεραρχικός, αλλά επειδή δουλεύει και σε δείγματα των δεδομένων. Βέβαια, ο k-modes έχει καλύτερες δυνατότητες κλιμάκωσης από τον ROCK, ενώ το μόνο αρνητικό του ROCK είναι ότι τα αποτελέσματα βασίζονται κατά πολύ στη δειγματοληψία που διενεργεί (Andritsos [79]). Τα αδύνατα σημεία του k-modes, είναι ότι δε μπορεί να χειριστεί το θόρυβο και τις έκτροπες παρατηρήσεις, καθώς και ότι δε μπορούν να χειριστεί συστάδες αυθαίρετου σχήματος.

Δύο άλλοι σημαντικοί αλγόριθμοι συσταδοποίησης κατηγορικών δεδομένων είναι ο **STIRR** (*Sieving Through Iterated Reinforcement*) και ο **CACTUS** (*Clustering Categorical Data Using Summaries*). Ο STIRR (Gibson [71]) είναι ιδιαίτερα γρήγορος και αποδοτικός, ενώ

προσφέρει τη χαρτογράφηση κατηγορικών δεδομένων σε ένα μη-γραμμικό δυναμικό σύστημα, μέσω επαναληπτικής προσέγγισης. Ο CACTUS (Ganti [74]) είναι η βελτίωση του STIRR και δεν επηρεάζεται από την υπερκάλυψη συστάδων. Με βάση αποτελέσματα πειραματισμών, ο CACTUS είναι πιο αποτελεσματικός από τον STIRR όσον αφορά το χρόνο εκτέλεσης και τον αριθμό των μεταβλητών που μπορεί να χειριστεί. Όμως, έχει το μειονέκτημα ότι δε μπορεί να χειριστεί σύνολα δεδομένων με αυξανόμενο πλήθος διαστάσεων.

2. ΥΠΕΡΚΥΒΙΚΗ ΙΣΟΜΟΡΦΙΚΗ ΤΟΠΟΛΟΓΙΑ

2.1 Υπερκύβος και Ισομορφική Τοπολογία

Μία υπερκυβική δομή αποτελεί ένα χώρο όπου υπήρξαν αρκετές δημοσιεύσεις, σχετικά με θέματα δρομολόγησης και ανάλυσης. Στην συγκεκριμένη διατριβή, η υπερκυβική δομή [87] χρησιμοποιείται για την δημιουργία μιας τοπολογικής βάσης, ικανής και αναγκαίας να υποστηρίξει μία παράλληλη δρομολόγηση ευφών πρακτόρων, οι οποίοι θα έχουν σαν στόχο την συλλογή και συσχέτιση δεδομένων και την πολυδιάστατη και παράλληλη εφαρμογή του επαναληπτικού αλγορίθμου RKM (Recursive K-Means).

Ορισμός 2.1.1 Ένας υπερκύβος είναι μία τοπολογική δομή μη-κατευθυνόμενου γράφου G που ανήκει στο σύνολο V ο οποίος εμπεριέχει ένα πεπερασμένο σύνολο από **κόμβους** $\{x_1, x_2 \dots x_n\}$ και από ένα υποσύνολο E το οποίο ορίζεται από το καρτεσιανό εξωτερικό γινόμενο $V \times V$ και εμπεριέχει τις **ακμές**.

Ο αριθμός των κόμβων ενός υπερκύβου ονομάζεται **διάσταση** n και αποτελεί την τάξη του γράφου. Μπορούμε να πούμε ότι ο μη-κατευθυνόμενος γράφος μπορεί να οριστεί ως μία ακμή υπερκύβου $\{x, y\}$ μεταξύ δύο κόμβων x, y .

Η διάσταση ενός υπερκύβου αντιπροσωπεύει ο μέγεθος του παραλληλισμού και ουσιαστικά μας ορίζει τον αριθμό των μικρο-επεξεργαστών ή ευφών πρακτόρων που μπορούν να τρέξουν παράλληλα σε μία υπερκυβική δομή.

Ορισμός 2.1.2 Ο αριθμός των ακμών που αναχωρούν από έναν υπερκυβικό κόμβο, ονομάζεται βαθμός N του υπερκύβου και περιγράφεται από την σχέση:

$$N = \log_n n \Leftrightarrow 2^N \quad (2.1.1)$$

Όπως μπορούμε να παρατηρήσουμε από την εξίσωση 2.1.1, η υπερκυβική τοπολογία ακολουθεί την δυαδική λογική και η σχέση μεταξύ της υπερκυβικής διάστασης και του βαθμού N ανήκει στον δυαδικό υποχώρο B

Ορισμός 2.1.3 Ο αριθμός των ακμών που αναχωρούν από έναν υπερκυβικό κόμβο, ονομάζεται βαθμός N του υπερκύβου

Ένα υπερκυβικό πλέγμα, θα αποτελέσει την τοπολογική βάση για την δρομολόγηση και διαχείριση της κατανεμημένης συσταδοποίησης, η οποία θα βασιστεί στην ισομορφική του δομή.

Ορισμός 2.1.4 Δύο κόμβοι $\{x_1, x_2\}$ σε ένα υπερκυβικό πλέγμα, ο οποίοι συνδέονται από μία ακμή $\{\hat{x}_1, \hat{x}_2\}$ ονομάζονται γειτονικοί κόμβοι. Οι κόμβοι και η τοπολογία του υπερκύβου ανήκουν σε έναν δυαδικό χώρο B^N με διάσταση N , η οποία πάντα συμπίπτει με τον βαθμό N του υπερκύβου. Συνεπώς, οι διαδοχικοί κόμβοι $\{x_1, x_2\}$ και για $\forall \{x_1, x_2\} \in B^N$, ανήκουν στον δυαδικό χώρο, όπου βάσει αυτού θα γίνεται και η δρομολόγηση των πρακτόρων

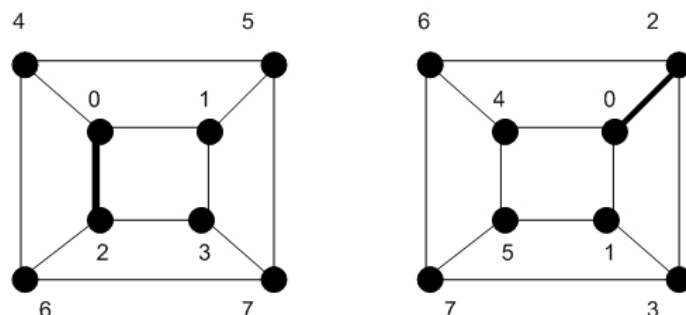
Ορισμός 2.1.5 Ορίζουμε την υπερκυβική απόσταση $D_{x_1 \rightarrow x_2}$ την απόσταση μεταξύ δύο υπερκυβικών κόμβων, η οποία πρέπει να διανύσει ένας ευφυής πράκτορας, ακολουθώντας τις διαδοχικές ακμές $\{\hat{x}_1, \hat{x}_2\}$. Η απόσταση αυτή, επειδή βρισκόμαστε τον δυαδικό χώρο B^N , είναι η ίδια με την απόσταση Hamming H στα ψηφιακά δυαδικά συστήματα.

Ορισμός 2.1.6 Ορίζουμε τον ισομορφισμό ενός υπερκύβου φ , όταν υπάρχουν δύο γράφοι $G = (V, E)$ και $G' = (V', E')$ και ένας αλγεβρικός αμφίδρομος τελεστής τέτοιος που $\varphi: V \rightarrow V'$ και περιγράφεται από την αλγεβρική σχέση:

$$xy \in E \Leftrightarrow \varphi(x)\varphi(y) \in E', \forall \{x, y\} \in V \quad (2.1.2)$$

Η έννοια και η ιδιότητα του ισομορφισμού σε έναν υπερκύβο είναι ιδιαίτερα σημαντική, καθώς μπορεί να αυτοπροσαρμοστεί τοπολογικά με οποιονδήποτε τρόπο, και να μην έχουμε διατάραξη του κανόνα δρομολόγησης και της τοπολογικής βάσης. Συνεπώς, εάν δημιουργήσουμε μία τυχαία αλγεβρική μετάθεση (permutation) στους κόμβους και τις ακμές του υπερκύβου, η διαδικασία δρομολόγησης δεν αλλάζει καθόλου και ουσιαστικά τα υπερκυβικά κεντροειδή και οι πράκτορες με την πληροφορία που μεταφέρουν δεν θα έχουν καμία αλλαγή στον αλγόριθμό τους. Ο ισομορφισμός μας επιτρέπει να αναδιοργανώνουμε το υπερκυβικό πλέγμα, ανάλογα με τα προβλήματα γνώσης και συσταδοποίησης που θέλουμε να επιλύσουμε.

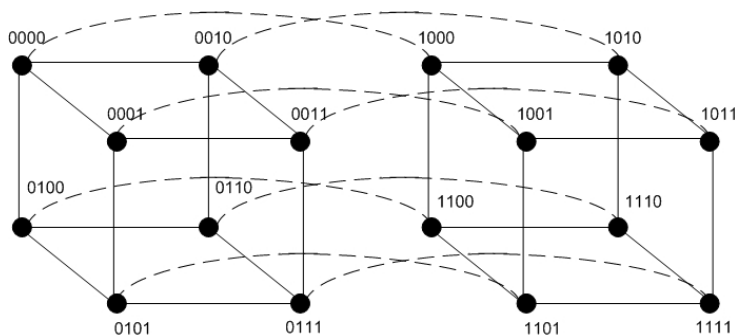
Ένας ισομορφικός υπερκύβος φαίνεται στο παρακάτω σχέδιο.



Εικόνα 3. Σχήμα 8-διάστατου υπερκύβου (2^3) με 8 κόμβους και ο ισομορφισμός του

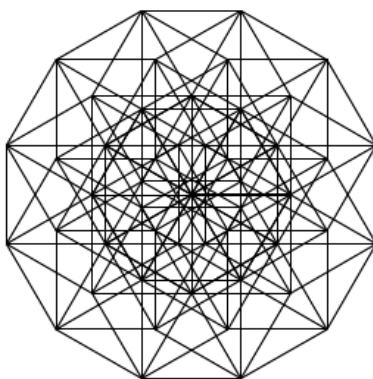
Όπως παρατηρούμε, ο υπερκύβος G_H με διάσταση $n = 8$ και βαθμό $N = 3$ μπορεί να αναδιαμορφωθεί, βάσει αυτομορφισμού και της ιδιότητας του ισομορφισμού φ σε $N = 3$ τρόπου δημιουργώντας $G_H(N - 1)$ γράφους και έχοντας ο καθένας $\frac{N}{2}$ ακμές μεταξύ των κόμβων. Στην παραπάνω εικόνα, ο ευφυής πράκτορας μπορεί και στις δύο περιπτώσεις να εισαχθεί στον πλέγμα και να αρχίσει την συσταδοποίηση, χωρίς να εξαρτάται η εισαγωγή του ή η δρομολόγησή του από τον κόμβο εισόδου ή από το πλέγμα.

Στο παρακάτω σχέδιο, βλέπουμε έναν 16διάστατο υπερκύβο με 16 κόμβους και βαθμό $N = 4$. Όπως παρατηρούμε, σε κάθε κόμβο έχει αποδοθεί μία δυαδική ταμπέλα, η οποία ακολουθεί τους κανόνες του δυαδικού χώρου B^N



Εικόνα 4. Σχήμα 16-διάστατου υπερκύβου (2^4) με 16 κόμβους

Ορισμός 2.1.6 Σε έναν υπερκύβο, διάστασης n και βαθμού N , οι κόμβοι ακολουθούν μία δυαδική ονομασία $\in \{0,1\}^N \in B^N$ με αποτέλεσμα δύο διαδοχικοί γειτονικοί κόμβοι $\{x_1, x_2\}$ να απέχουν μεταξύ τους ένα δυαδικό bit. Ο παραπάνω ορισμός είναι εξαιρετικά σημαντικός, καθώς καθορίζει την διαδικασία δρομολόγησης σε μία υπερκυβική δομή, βάσει ντετερμινιστικών και στοχαστικών μεθόδων που θα δούμε παρακάτω. Συνεπώς, ο υπερκυβικό πλέγμα έχει διάσταση n , βαθμό N , σύνολο κόμβων 2^N και σύνολο ακμών $N \cdot 2^{N-1}$.



Εικόνα 5. Σχήμα ισομορφικού υπερκύβου βαθμού διάστασης $N=9$

2.2 Τεχνικές Διαχείρισης Γνώσης μέσω Υπερκυβικής Δρομολόγησης

Όπως αναφέρθηκε και παραπάνω, στην διατριβή χρησιμοποιείται η υπερκυβική δομή για την συσταδοποίηση δεδομένων, τα οποία βρίσκονται αποθηκευμένα σε υπερκυβικούς κόμβους. Η χρήση του υπερκυβικού πλέγματος, ως ενός μεσισμικού για την βέλτιστη δρομολόγηση των πρακτόρων και την εφαρμογή της αναδρομικής συσταδοποίησης (επόμενο κεφάλαιο) δημιούργησε εντυπωσιακά αποτελέσματα, τα οποία και θα παρουσιασθούν στον Κεφάλαιο 4.

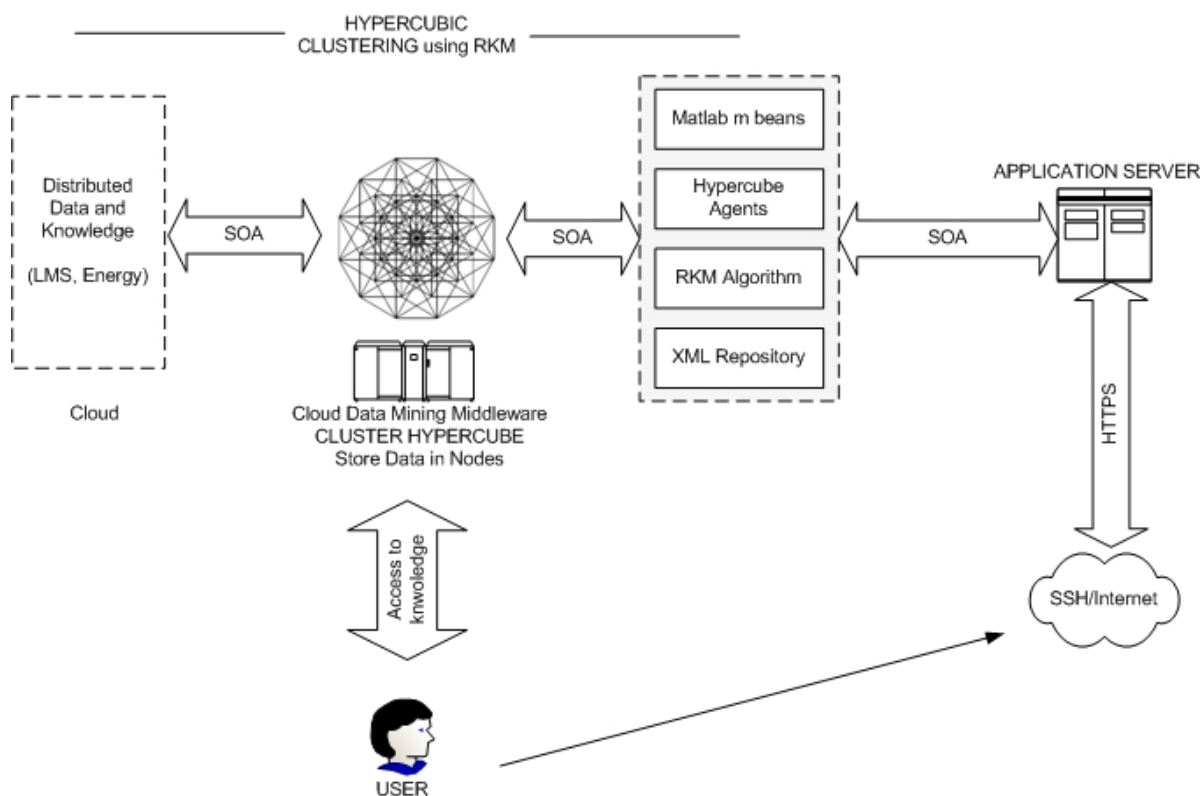
Η ουσία είναι ότι στην συγκεκριμένη διατριβή, ακολουθείται μια εντελώς ριζοσπαστική προσέγγιση, όπου τα δεδομένα και η γνώση και τα διάφορα metrics (KPIs - Key Performance Indicators) αποθηκεύεται σε υπερκυβικές συστάδες (κόμβοι υπερκύβου) και είναι προσβάσιμη από τους πράκτορες, οι οποίοι ακολουθούν συγκεκριμένο αλγόριθμο δρομολόγησης αλλά και συγκεκριμένο αλγόριθμο συσταδοποίησης (RKM) από τα δεδομένα που αλιεύουν στα διαφορετικά hypercubic nodes. Η συγκεκριμένη διαδικασία ουσιαστικά

καταργεί τις βάσεις δεδομένων, όταν η πληροφορία που μας ενδιαφέρει είναι περιγραφική, μη-δομημένη και αναφέρεται σε συσχετίσεις δεικτών και δυναμικές αλλαγές στο πεδίο του χρόνου. Στους συγκεκριμένους αλγόριθμους, αυτό που μας ενδιαφέρει είναι η **ανάλυση της δυναμικότητας των δεικτών και όχι η στατική τους φύση** (δυναμικές συστάδες). Επομένως, η συγκεκριμένη τεχνική είχε εντυπωσιακά αποτελέσματα στην ανάλυση και περιγραφή των αποτελεσμάτων και με γραφικό τρόπο (γεωμετρικός τόπος) αλλά και στην συνολική απόδοση της υπολογιστικής πολυπλοκότητας.

ΒΑΣΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ

Η βασική αρχιτεκτονική του συστήματος, βασίστηκε ασφαλώς στο διαδίκτυο και στην φιλοσοφία των Web based DSS συστημάτων. Επίσης, οι νέες τεχνικές του cloud computing χρησιμοποιήθηκαν επίσης, για την μέγιστη απόδοση του συστήματος και την ευκολία προγραμματισμού και εκτέλεσης των προγραμμάτων και αλγορίθμων.

Η γενική αρχιτεκτονική φαίνεται παρακάτω:



Εικόνα 6. Συνολική αρχιτεκτονική υπερκυβικής διαχείρισης γνώσης

Όπως μπορούμε να δούμε, η αρχιτεκτονική χρησιμοποιεί έναν κατανεμημένο υπερκύβο, συγκεκριμένης διάστασης, όπου σε κάθε virtual web node, λαμβάνονται και υπολογίζονται διάφορα metrics και δείκτες από τον εκάστοτε κόμβο και παράλληλα εφαρμόζεται υπερκυβική συσταδοποίηση από τους πράκτορες, βάσει SOA (Service Oriented Architecture) και Web services δια-λειτουργικότητας. Η αρχιτεκτονική είναι δυναμική, καθώς ο επαναληπτικός RKM αλγόριθμος εκτελείται ανά τακτά χρονικά διαστήματα (sec or mins).

Η άντληση των δεδομένων και ο υπολογισμός των δεικτών γίνεται με ηλεκτρονικούς τρόπους, (SOA, energy metering, Web 2.0 services) και εν συνεχεία εκτελείται ο αλγόριθμος της συσταδοποίησης για την αποθήκευση των αποτελεσμάτων σε αλγεβρικές δομές. Βάση της δομής μπορεί κάλλιστα να γίνει και ανεύρεση πληροφορίας, είτε από τον κόμβο κατευθείαν είτε από τις αλγεβρικές γραμμικές μήτρες που αποθηκεύουν και αναλύουν (πχ. Ιδιοτιμές, similarity measure, etc) τα αποτελέσματα της υπερκυβικής συσταδοποίησης

2.3 Υπερκυβικοί Αλγόριθμοι Δρομολόγησης Γνωσιακών Πρακτόρων

Μπαίνοντας σε βάθος στην τεχνική και διαδικασία της υπερκυβικής δρομολόγησης [83,84,85,86] μέσα στο υπερκυβικό πλέγμα, αναφέρουμε κάποιες σημαντικές έννοιες και ορισμούς. Η βασική δρομολόγηση μέσα στον υπερκύβο, γίνεται από τους software agents (πράκτορες) οι οποίοι ακολουθούν μία αλγεβρική μετάθεση (permutation) και περνώντας από διαδοχικούς υπερκυβικούς κόμβους, ξεκινούν από έναν αρχικό κόμβο x_1 και καταλήγουν σε έναν τελικό κόμβο x_2 , και υπολογίζουν διάφορα metrics και δείκτες τα οποία λειτουργούν σαν input στον αναδρομικό υπερκυβικό αλγόριθμο συσταδοποίησης. Στο πέρασμα του κάθε πράκτορα από τους διαδοχικούς κόμβους και έχοντας σαν output τα metrics (KPI), σε κάθε κόμβο εφαρμόζεται ο αλγόριθμος συσταδοποίησης K-means, ο οποίος είναι επαναληπτικός και εφαρμόζεται από όλους τους πράκτορες και όλους τους κόμβους. Τα αποτελέσματα αποθηκεύονται στην μνήμη του πράκτορα και συσχετίζονται με τα υπόλοιπα αποτελέσματα από όλους τους πράκτορες.

Ορισμός 2.3.1 Σε έναν υπερκύβο, ορίζουμε την απόσταση Hamming $H(x_1, x_2)$ μεταξύ δύο υπερκυβικών κόμβων ως τον αριθμό των δυαδικών bits τα οποία διαφέρουν μεταξύ των δύο δυαδικών ονομασιών των κόμβων και που ανήκουν στον δυαδικό χώρο B^N .

Η απόσταση Hamming $H(x_1, x_2)$ είναι μία γραμμική απόσταση και περιγράφεται από τον τύπο:

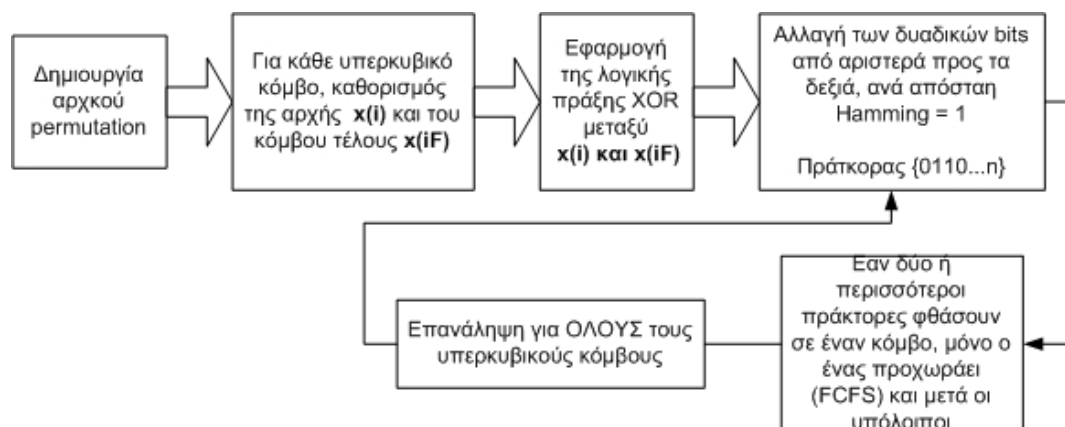
$$H(x_1, x_2) = \sum_{i=0}^{N-1} h_i \in R^1 \quad (2.3.1)$$

Όπου ο τελεστής h είναι ίσος με 1 εάν το i -οστό bit είναι διαφορετικό και ίσος με 0 εάν το i -οστό bit μεταξύ δύο κόμβων είναι ίδιο. Οπότε το άθροισμα των διαφορετικών bit σε μία δυαδική ταυτότητα, είναι η απόσταση hamming.

Πάντα η μέγιστη απόσταση Hamming μεταξύ 2 κόμβων δεν θα υπερβαίνει την διάσταση του υπερκύβου και η απόσταση μεταξύ δύο γειτονικών κόμβων είναι πάντα 1.

Ορισμός 2.3.2 Σε έναν υπερκύβο, ορίζουμε μία μοναδική δυαδική ονομασία / ταμπέλα στον δυαδικό χώρο B^N με $x_i = \{b_{N-1}^{x_i}, b_{N-2}^{x_i}, \dots, b_1^{x_i}, b_0^{x_i}\}^N$ όπου $0 \leq i \leq n-1$ και $n = 2^N$ όπου το $b \in \{1,0\}$ και ξεχωρίζει μοναδικά έναν κόμβο στο πλέγμα.

Συνεπώς, η διαδικασία δρομολόγησης $D_{x1 \rightarrow x2}$, θα ακολουθεί μία συγκεκριμένη διαδικασία, η οποία θα λαμβάνει υπόψιν την απόσταση μεταξύ δύο υπερκυβικών κόμβων και τις δυαδικές τους ταμπέλες. Οι δύο τρόποι δρομολόγησης που συγκρίνονται, είναι ο ντετερμινιστικός τρόπος του “Bit-Fixing” και ο στοχαστικός του Les Valiant [41]. Για την εφαρμογή των αλγορίθμων και του συστήματος στα ενεργειακά και μαθησιακά δεδομένα, ακολουθήθηκε ο στοχαστικός αλγόριθμος, δεδομένου ότι έχει καλύτερα αποτελέσματα που φαίνονται και στα διαγράμματα παρακάτω, από την εκτέλεση ενός 64διάστατου υπερκύβου (simulation).



Εικόνα 7. Στάδια κλασσικού υπερκυβικού routing των software agents

Στην Εικόνα 7, βλέπουμε την συνολική κλασσική μεθοδολογία δρομολόγησης των υπερκυβικών πρακτόρων, σε ένα πλέγμα υπερκύβου. Οι δύο μεθοδολογίες που συγκρίθηκαν ήταν η κλασσική ντετερμινιστική και η στοχαστική (Les Valiant).

BIT-FIXING ΔΡΟΜΟΛΟΓΗΣΗ

Η δρομολόγηση βάσει του ντετερμινιστικού αλγόριθμου “bit fixing” [84] είναι μία ευρέως διαδεδομένη τεχνική και μεθοδολογία, για ανταλλαγή μηνυμάτων σε ένα υπερκυβικό πλέγμα. Έστω ότι έχουμε έναν υπερκύβο G_N διάστασης n και βαθμού N και θέλουμε να εφαρμόσουμε κάποιους software agents εντός του πλέγματος. Ο αλγόριθμος bit-fixing, ακολουθεί την διαδρομή από έναν κόμβο $x_1 \rightarrow x_2$ αλλάζοντας κάθε φορά ένα δυαδικό bit από αριστερά προς τα δεξιά στο MSB (Most Significant Bit) μέχρι να φτάσει στο τελικό binary tag το οποίο είναι ο ίδιο με την δυαδική ταμπέλα.

Έστω λοιπόν ο κόμβος εισόδου των πρακτόρων $x_i = \{b_{N-1}^{x_i}, b_{N-2}^{x_i}, \dots, b_1^{x_i}, b_0^{x_i}\}^N$ και ο κόμβος τέλους $x_j = \{b_{N-1}^{x_j}, b_{N-2}^{x_j}, \dots, b_1^{x_j}, b_0^{x_j}\}^N$ και ο πράκτορας θα πρέπει να διασχίσει την ενδιάμεση διαδρομή στο υπερκυβικό πλέγμα. Για την εφαρμογή της ντετερμινιστικής μεθόδου, εκτελούμε την λογική πράξη XOR (\oplus) μεταξύ των δύο κόμβων για να ορίσουμε και την απόσταση Hamming ($H(x_i, x_j)$) μεταξύ των δύο αλλά και την ακριβή διαδρομή που θα ακολουθήσει ο πράκτορας. Τα στάδια δρομολόγησης είναι προκαθορισμένα και ισούνται με την απόσταση Hamming συν κάποια καθυστέρηση που μπορεί να έχουμε από παράλληλη παρουσία πρακτόρων σε κόμβους.

Συνεπώς:

$$Steps_D = H_{(x_i \rightarrow x_j)} + DELAYS \quad (2.3.2)$$

Οι βασικοί κανόνες δρομολόγησης των πρακτόρων μέσα στο υπερκυβικό πλέγμα είναι οι εξής:

- Οι κινήσεις των πρακτόρων μέσα στον υπερκύβο γίνονται παράλληλα
- Μέσα σε έναν κόμβο εκτελείται μόνο μία υπολογιστική διαδικασία (KPI) σε ένα κβάντο χρόνου

- Η διαδικασία εκτέλεσης ακολουθεί την φιλοσοφία First Come First Served (FCFS) και η επικοινωνία στις ακμές είναι αμφίδρομη και παράλληλη
- Κάθε κόμβος αναγνωρίζεται με μία μοναδική δυαδική ID ταμπέλα ($0 \rightarrow n-1$)
- Η παράλληλη μεταγωγή των πρακτόρων στο υπερκυβικό πλέγμα ονομάζεται Αλγεβρική Μετάθεση Π (**Permutation**) και το ορίζουμε σαν αλγεβρικό τελεστή Π

Ορισμός 2.3.3 Σε έναν υπερκύβο, ορίζουμε μία μοναδική αλγεβρική μετάθεση Π , ως τελεστή μετάθεσης των πρακτόρων. Συνεπώς σε κάθε κόμβο i , υπάρχει ένας αριθμός πρακτόρων M_i με διεύθυνση που ορίζεται ως $\Pi_i = x_{iF}$

Παράδειγμα: Έστω υπερκύβος $G(5)$ με βαθμό διάστασης 5 (και κόμβους $n = 2^5$) και έστω ότι ένας πράκτορας ξεκινά από τον κόμβο $x_i = \{b_4^{x_i}, b_3^{x_i}, \dots, b_1^{x_i}, b_0^{x_i}\}^5 = \{0,0,0,0,0\}^5$ προς τον κόμβο $x_{iF} = \{b_4^{x_i}, b_3^{x_i}, \dots, b_1^{x_i}, b_0^{x_i}\}^5 = x_{iF} = \{1,0,1,1,1\}^5$. Μετά την εφαρμογή της λογικής πράξης XOR μεταξύ των δύο κόμβων έχουμε:

$x_i \oplus x_{iF} \Leftrightarrow \{1,0,1,1,1\}$ οπότε αρχίζοντας από το MSB έχουμε την ντετερμινιστική διαδρομή του πράκτορα:

$$\{0,0,0,0,0\} \rightarrow \{1,0,0,0,0\} \rightarrow \{1,0,1,0,0\} \rightarrow \{1,0,1,1,0\} \rightarrow \{1,0,1,1,1\}$$

Όπως είδαμε ο πράκτορας έκανε μία απόσταση 4 επιπέδων, ακριβώς ίση με την απόσταση Hamming:

$$H(x_i, x_{iF}) = \sum_{i=0}^4 h_i = 4$$

Η ανάλυση πολυπλοκότητας μας δίνεται από τον τύπο (Borodin-Hopcroft):

$$c \left(\Omega \left(\sqrt{\frac{2^N}{\log_2 n}} \right) \right) \quad (2.3.3)$$

Όπου βάσει της απόδειξης στο [40] η πλήρη εφαρμογή του bit-fixing σε μία αλγεβρική μετάθεση Π , το ανώτατο όριο στην πολυωνυμική πολυπλοκότητα είναι το παραπάνω στη χειρότερη περίπτωση

LES VALIANT ΔΡΟΜΟΛΟΓΗΣΗ

Η δρομολόγηση κατά τον Les Valiant [41] αποτελεί μία παραλλαγμένη δομή του bit-fixing, στο ότι χωρίζουμε πια την διαδικασία δρομολόγησης των πρακτόρων σε δύο φάσεις (Phase A & Phase B). Κατά την πρώτη φάση, η αλγεβρική μετάθεση εμπεριέχει και μία τυχαία μεταβλητή, οπότε υπάρχει πια και η στοχαστική συνιστώσα. Οι πράκτορες στέλνονται σε έναν ενδιάμεσο τυχαίο προορισμό, με μία φιλοσοφία που ακολουθεί την κανονική κατανομή και εν συνεχεία από τον ενδιάμεσο προορισμό τους αποστέλλονται στην τελική τους διεύθυνση.

Έστω ένας πράκτορας που θα αρχίσει την δρομολόγησή του από έναν υπερκυβικό κόμβο $x_i = \{b_{N-1}^{x_i}, b_{N-2}^{x_i}, \dots, b_1^{x_i}, b_0^{x_i}\}^N$. Ακολουθώντας μία κανονική κατανομή (Normal distribution, N), τυχαίος κόμβος \hat{x}_i επιλέγεται και βάσει του “bit fixing” ο πράκτορας δρομολογείται προς τα εκεί, σε μία ΦΑΣΗ Α (Phase A)., Εν συνεχεία, ακολουθώντας ντετερμινιστική διαδικασία, ο πράκτορας δρομολογείται από το \hat{x}_i προς το x_{if} σε μία ΦΑΣΗ Β (Phase B). Η ενδιάμεση στοχαστική διαδικασία εξασφαλίζει έναν καλύτερο και πιο αποδοτικό/γρήγορο τρόπο παράλληλης δρομολόγησης των πρακτόρων με μία στατιστική πιθανότητα εμπλοκής σε κόμβους πολύ πιο μικρή από την κλασσική διαδικασία. Αυτό αποδεικνύεται και στην ανάλυση πολυπλοκότητας στο [41].

Βάσει λοιπόν του [41], μέσα στον υπερκύβο G υπάρχει μία στοχαστική διαδικασία, όπου χρησιμοποιεί $O(nN)$ τυχαία bits, τέτοια ώστε δεδομένου του θ , με μία πιθανότητα μεγαλύτερη από $1 - \frac{2}{n^\theta}$ κάθε πράκτορας φθάνει στον προορισμό του, μέσα στην εκάστοτε ΦΑΣΗ (Α ή Β) σε $(2\theta + 4)N$ στάδια στο μέγιστο.

Έτσι, έχοντας μια τυχαία μεταβλητή $H_{i,j}$ που λαμβάνει την τιμή 1 στις ακμές $i(x_i)$ προς \hat{x}_i και από το $\hat{j}(x_j)$ προς το \hat{x}_j και οι δύο μεταβλητές συνυπάρχουν σε έναν κόμβο υπερκυβικό.

Τότε θέτουμε την καθυστέρηση στον κόμβο (Delay, D) λόγω της επιλογής του κόμβου να υπολογίσει ξεχωριστά τους δύο πράκτορες, ως:

$$\text{delay} \leq \sum_{i=0}^{n-1} H_{ij} \quad (2.3.4)$$

και μέσω γραμμικής συσχέτισης, τις μαθηματικές προσδοκίες:

$$E[\text{Delay}_i] \leq E\left[\sum_{i=0}^{n-1} H_{ij}\right] = \sum_{i=0}^{n-1} E[H_{ij}] \quad (2.3.5)$$

Εάν εφαρμόσουμε το άνω όριο του Chernov [41], στο $\sum_{i=0}^{n-1} [H_{ij}]$ με $\theta > 0$ και μ είναι η μαθηματική προσδοκία του H_{ij} με ακμή το a_k , έχουμε:

$$P[H_{ij} > (1 + \theta)\mu] < \left[\frac{a_k^\theta}{(1 + \theta)^{1+\theta}}\right]^\mu < \left[\frac{a_k}{(1 + \theta)}\right]^{(1+\theta)\mu} \quad (2.3.6)$$

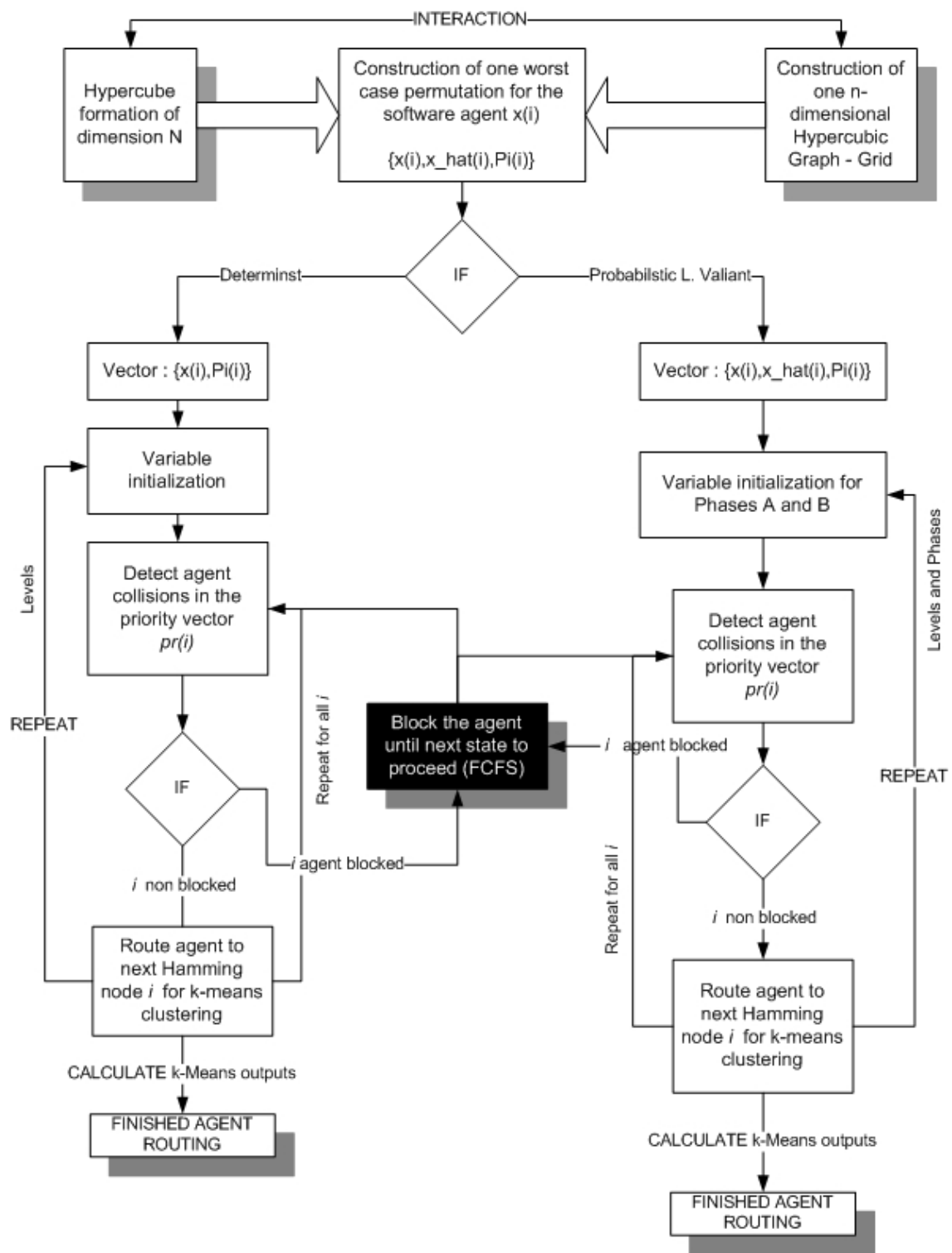
Για $1+\theta > 2$ ακ έχουμε :

$$P[H_{ij} > (1 + \theta)\mu] < 2^{-(1+\theta)\mu} \Leftrightarrow P[H_{ij} > \Delta] < 2^{-\Delta} \quad (2.3.7)$$

Έτσι για παράδειγμα, για $\Delta=3$ η ανισότητα $\Delta/\mu > 2$ ακ είναι αληθινή και μπορούμε να πούμε ότι $P[H_{ij} > 3N] < 2^{-3N}$ όπου η προσδοκία μ του H_{ij} είναι φραγμένη από $N/2$.

Αποδεικνύουμε λοιπόν ότι η πιθανότητα η μεταβλητή $(\theta+1)N$ να είναι μεγαλύτερη από $2^{-(\theta+1)N}$ και συνεπώς η πιθανότητα ενός από δύο πράκτορες να είναι σε καθυστέρηση σε έναν κόμβο, μεγαλύτερης του $(\theta+1)N$ είναι μεγαλύτερη από $2^{-\theta N}$.

Εν συνεχεία, με μία πιθανότητα μικρότερη από $1/n^\theta$ ένας πράκτορας φθάνει στον ενδιαμέσο προορισμό του (PHASE A) σε $(\theta+2)N$ στάδια στο μέγιστο, συν την καθυστέρηση (delay). Συνεπώς, και για τις 2 φάσεις της στοχαστικής δρομολόγησης, με μία πιθανότητα $2/n^\theta$ κάθε πράκτορας φθάνει στον τελικό του προορισμό σε $(2\theta+4)N$ στάδια.

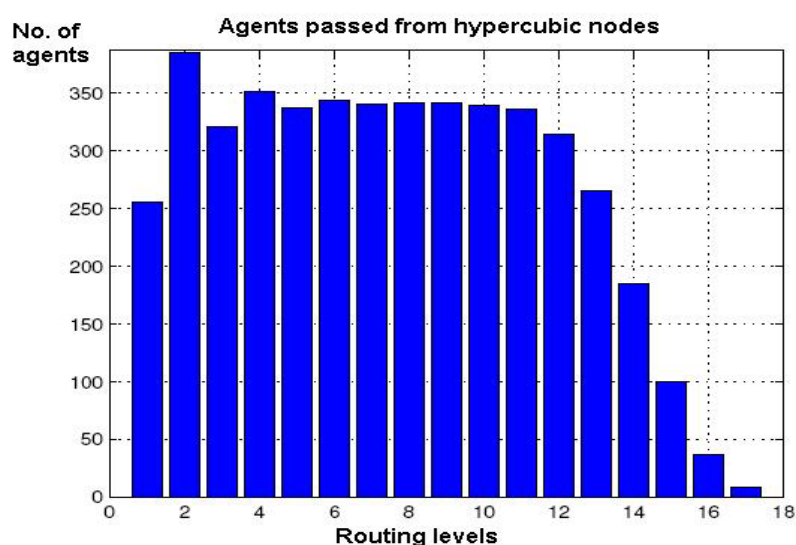


Εικόνα 8. Συνολικός αλγόριθμος δρομολόγησης υπερκυβικών πρακτόρων

ΕΞΟΜΟΙΩΣΗ ΥΠΕΡΚΥΒΙΚΩΝ ΠΡΑΚΤΟΡΩΝ

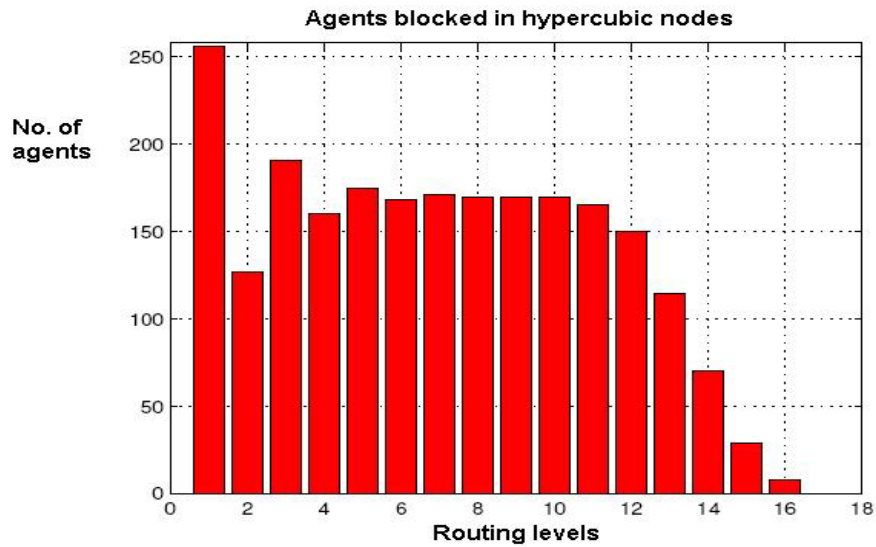
Χρησιμοποιώντας Matlab [44] και ολοκληρωμένο πληροφοριακό σύστημα, έγινε εξομοίωση της υπερκυβικής δραστηριότητας και του routing των πρακτόρων μέσα στον υπερκύβο. Τα αποτελέσματα επιβεβαίωσαν την πιθανοτική θεωρία.

Στο παρακάτω γράφημα έχουμε τα αποτελέσματα ενός υπερκύβου με βαθμό διάστασης $N=9$ και 2^9 πράκτορες σε ντετερμινιστική δρομολόγηση. Όπως παρατηρούμε, η πλήρης δρομολόγηση σε μία Αλγεβρική μετάθεση μέγιστης απόστασης Hamming (οι πράκτορες έχουν σαν κόμβο τέλους το πιο απομακρυσμένο κόμβο από την αρχική τους θέση) τελειώνει σε 17 στάδια και το επίπεδο της παράλληλης δρομολόγησης από τους υπερκυβικούς κόμβους φθάνει σε επίπεδα 350 πρακτόρων / στάδιο. Μετά το 12 στάδιο οι πράκτορες αρχίζουν και φθάνουν στον προορισμό τους, οπότε τα επίπεδα δρομολόγησης στους κόμβους μειώνεται αισθητά, μέχρι την τελική φάση



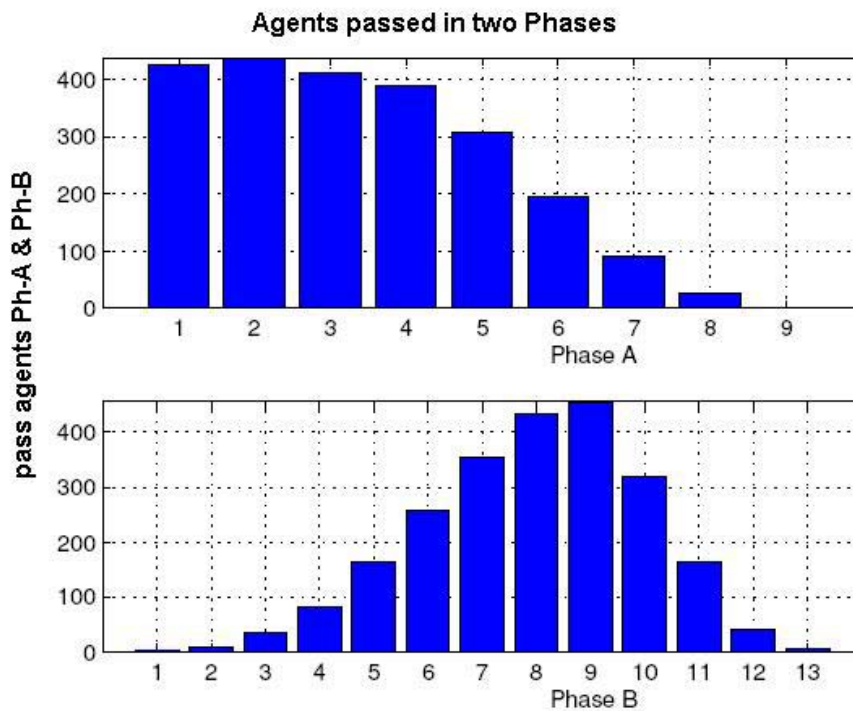
Εικόνα 9. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε ντετερμινιστική δρομολόγηση

Εάν αναλύσουμε τώρα τους πράκτορες που μπλοκαρίστηκαν σε κάποιον κόμβο, βάσει του αλγορίθμου FCFS, βλέπουμε τα επίπεδα της καθυστέρησης. Παρατηρούμε ότι με την έναρξη της δρομολόγησης είχαμε ένα μέγιστο στην καθυστέρηση (congestion σε περίπου 255 πράκτορες) λόγω του ότι το αρχικό στάδιο του ντετερμινιστικού αλγορίθμου προκαλεί μεγάλες καθυστερήσεις. Αυτό είναι κάτι που διορθώνει ο στοχαστικό αλγόριθμος.



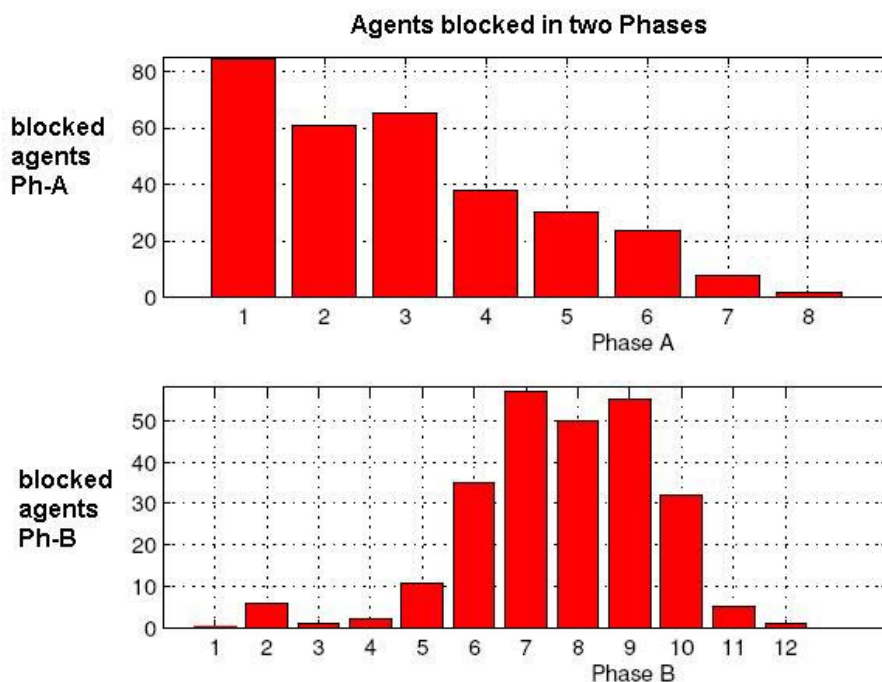
Εικόνα 10. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε ντετερμινιστική δρομολόγηση

Εν συνεχεία, παρατηρούμε ότι υπάρχει μία σταθερή τιμή στην καθυστέρηση στους κόμβους, λόγω και τους ντετερμινιστικού αλγόριθμου αλλά και του αλγόριθμου FCFS. Εάν αλλάξει ο αλγόριθμος εξυπηρέτησης στους υπερκυβικούς κόμβους, είναι σίγουρο ότι η διαδικασία θα μπορεί να επιταχυνθεί αρκετά, σε συνδυασμός και με την στοχαστική προσέγγιση.



Εικόνα 11. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε στοχαστική δρομολόγηση

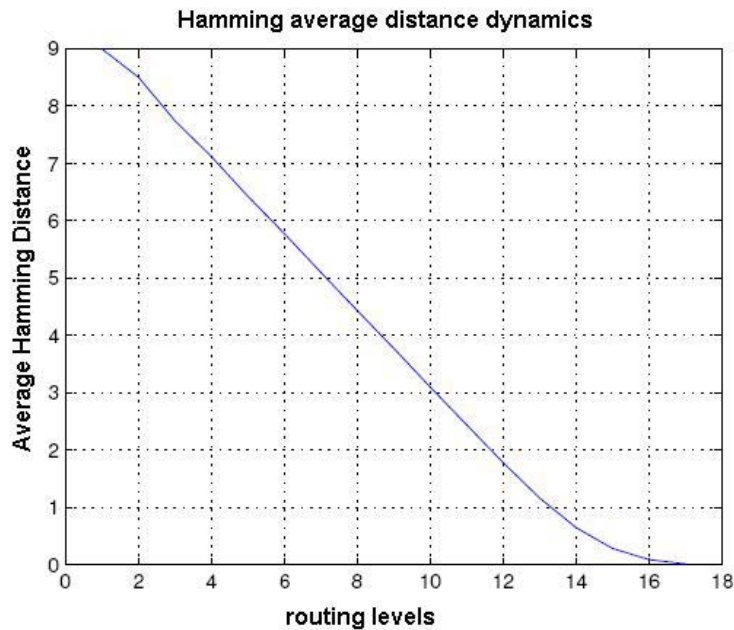
Εάν τώρα εφαρμόσουμε στον ίδιο υπερκύβο την στοχαστική δρομολόγηση σε δύο στάδια (PHASE A - PHASE B) παρατηρούμε στα γραφήματα 11 και 12 ότι η διαδικασία ολοκληρώνεται σε 12 στάδια αντί 17 και υπάρχει μεγάλη επιτάχυνση. Το επίπεδο των κόμβων που δρομολογούνται ανά στάδιο έχει αυξηθεί σε μέσο όρο (πάνω από 400 πράκτορες) και το επίπεδο των καθυστερήσεων έχει μειωθεί αισθητά κάτω από 90 και ειδικά στην δεύτερη φάση κάτω από 60.



Εικόνα 12. Εξομοίωση 9-διάστατου υπερκύβου με 2^9 πράκτορες σε στοχαστική δρομολόγηση

Η παραπάνω ανάλυση μας δείχνει το πόσο πιο αποδοτικός είναι ο στοχαστικός αλγόριθμος για την δρομολόγηση των πρακτόρων στον υπερκύβο και την εκτέλεση αναλύσεων σε κάθε κόμβο. Όσο οι πράκτορες αυξάνονται όσο και η γενική πολυπλοκότητα, τόσο πιο αποδοτικός γίνεται ο στοχαστικός αλγόριθμος έναντι του ντετερμινιστικού. Αυτό ασφαλώς βοηθάει πάρα πολύ στην εφαρμογή της μεθόδου και του αλγορίθμου στην ανάλυση μεγάλων δεδομένων (big data) όπου η πολυπλοκότητα αυξάνεται εκθετικά αλλά και το επίπεδο της ανάλυσης που πρέπει να γίνει.

Ειδικά, όταν έχουμε χρονοδυναμικά δεδομένα (πχ. ενέργεια) η ανάλυση θα πρέπει να είναι σε πραγματικό χρόνο, ώστε να έχουμε πλήρη εικόνα των μεταβολών αλλά και της δυναμικότητας των δεικτών, οι οποίοι αναλύονται και υπολογίζονται από τον εκάστοτε υπερκυβικό πράκτορα σε κάθε κόμβο.



Εικόνα 13. Απόσταση Hamming ντετερμινιστικού μοντέλου

Εάν αναλύσουμε τώρα το πως κινείται γραμμικά η μέση απόσταση Hamming, που ορίζεται ως:

$$CH_j = \frac{\sum_{i=1}^n H(x_i, x_{iF})_i}{n} \quad (2.3.8)$$

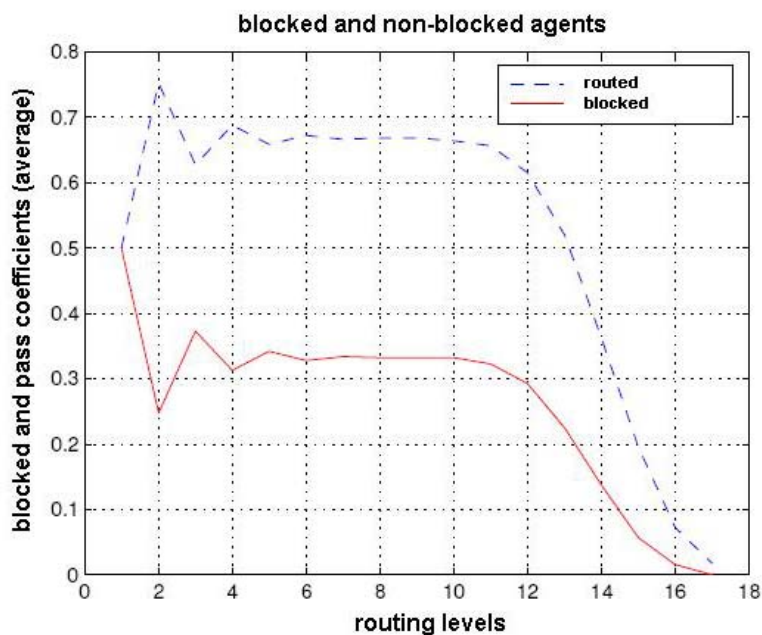
Όπου j : τα στάδια και i : οι πράκτορες, θα παρατηρήσουμε ότι η μέση απόσταση Hamming κάθε πράκτορα από τον αρχικό του κόμβο προς τον τελικό, έχει μια γραμμική μορφή, αναδεικνύοντας μια σταθερή ταχύτητα δρομολόγησης του υπερκυβικού νέφους πρακτόρων. Αυτή η σταθερή ταχύτητα φανερώνει και την γραμμικότητα του ντετερμινιστικού αλγόριθμου, η οποία όμως τον καθιστά πιο αργό και λιγότερο ελαστικό από τον στοχαστικό.

Εάν αναλύσουμε και τον συντελεστή καθυστέρησης (CB) και δρομολόγησης (CR), που ορίζονται ως εξής:

$$CR_j = \frac{\sum_{i=1}^n (M_{ri})_i}{n} \quad (2.3.9) \quad \text{και} \quad CB_j = \frac{\sum_{i=1}^n (M_{bi})_i}{n} \quad (2.3.10)$$

οι

πράκτορες παρατηρούμε το εξής γράφημα:

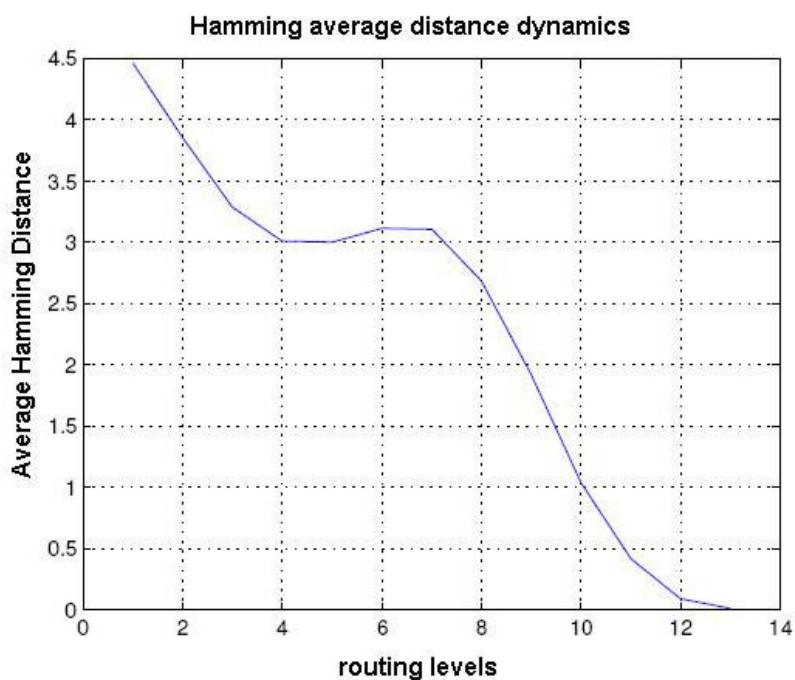


Εικόνα 14. Συντελεστές CB και CR πρακτόρων (ντετερμινιστικό μοντέλο)

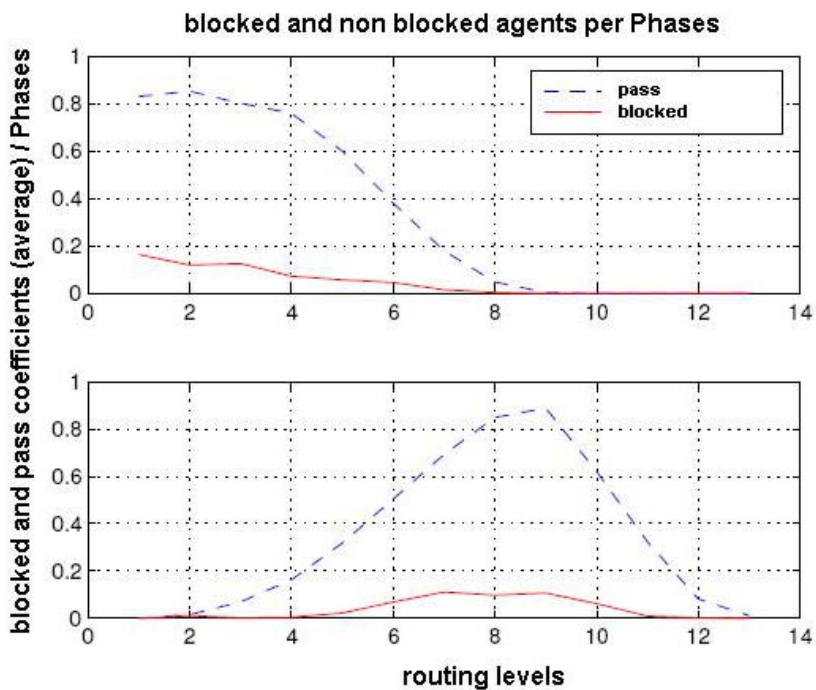
Η παραπάνω ανάλυση μας δείχνει πως οι δύο συντελεστές κινούνται στον χρόνο. Βλέπουμε λοιπόν μια γραμμική ισορροπία μεταξύ του 3^{ου} και 12^{ου} σταδίου, που ευθύνεται για την γραμμικότητα που αναλύσαμε πριν, στην απόσταση Hamming.

Προχωρώντας και στην ανάλυση της στοχαστικής προσέγγισης, παρατηρούμε από το γράφημα 15 την μη-γραμμική φύση του γραφήματος και της δυναμικής απόστασης hamming. Ειδικά, στην μεταβατική περίοδο από την φάση A στην φάση B, παρατηρούμε μία στιγμιαία ανωμαλία στην όλη διαδικασία, μεταξύ του 4^{ου} και 7^{ου} σταδίου. Παρόλα αυτά όπως βλέπουμε, λόγω της μεγαλύτερης κλίσης της γραφικής παράστασης (μεγαλύτερη ταχύτητα δρομολόγησης) η διαδικασία ολοκληρώνεται σε λιγότερα στάδια από την ντετερμινιστική προσέγγιση. Αυτό, επίσης φαίνεται και από τις εικόνες 16 και 17, όπου οι συντελεστές δρομολόγησης και καθυστέρησης και η συνολική κίνηση πρακτόρων έχουν μία σαφή διαφορά στον χρόνο, με τον συντελεστή καθυστέρησης να είναι αισθητά μικρότερος στην B ΦΑΣΗ, κάτι που φαίνεται και στον συντελεστή hamming. Τέλος, στο 3-διάστατο γράφημα στην εικόνα 18, βλέπουμε και τρι-σδιάστατα την κατανομή της καθυστέρησης μέσα στο υπερκυβικό πλέγμα. Παρατηρούμε ότι στην φάση A, η καθυστέρηση είναι αιχμακική αλλά βραχύς ενός στην φάση B η καθυστέρηση είναι πιο κατανεμημένη αλλά μικρής έντασης. **Δηλαδή ο αλγόριθμος βρίσκει ή κοντινές ακμές με μεγάλη σχετικά καθυστέρηση ή**

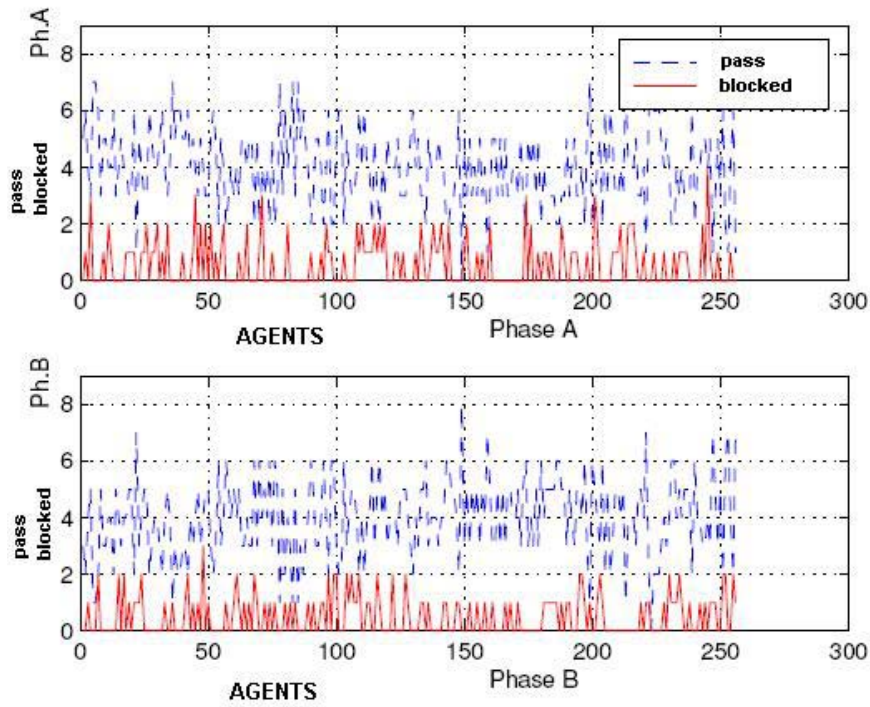
μακρινές ακμές με μικρή καθυστέρηση και κατά συνέπεια μεγάλη ταχύτητα δρομολόγησης.



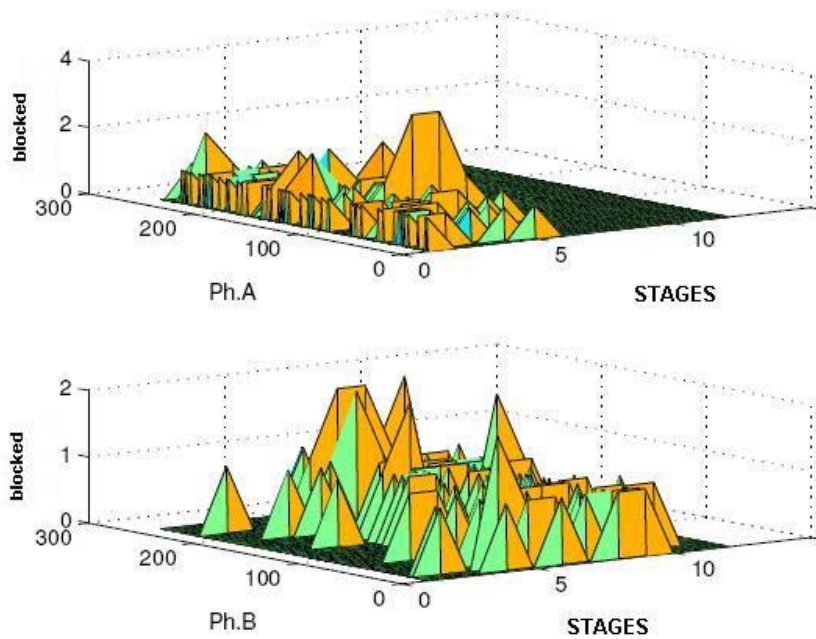
Εικόνα 15. Δυναμική απόσταση Hamming στοχαστικού μοντέλου (Valiant)



Εικόνα 16. Συντελεστές CB και CR πρακτόρων (ντετερμινιστικό μοντέλο)



Εικόνα 17. Κίνηση πρακτόρων ανά στάδια δρομολόγησης (στοχαστικό μοντέλο με N=8)



Εικόνα 18. Συντελεστής καθυστέρησης CB πρακτόρων σε 3D (στοχαστικό μοντέλο N=8)

Αν συνοψίσουμε τα στατιστικά αποτελέσματα σε 2 πίνακες, έναν για την ντετερμινιστική προσέγγιση και έναν για την στοχαστική προσέγγιση στην υπερκυβική δρομολόγηση των πρακτόρων στο πλέγμα, έχουμε:

| ΓΕΝΙΚΑ ΣΤΑΔΙΑ | 17 |
|--|---|
| Συνολικοί πράκτορες με καθυστέρηση | 2245 |
| Συνολικοί πράκτορες με δρομολόγηση | 4608 |
| Μέση τιμή Συντελεστή Δρομολόγησης CR | 0,56 |
| Μέση τιμή Συντελεστή Καθυστέρησης CB | 0,28 |
| Μέγιστη καθυστέρηση / πράκτορα | 8 στάδια |
| Εφαπτομένη της μέσης απόστασης Hamming CH | -0.55 / στάδιο |
| Πρώτος πράκτορας που έφτασε στον προορισμό του | 11 στάδια |
| Μέγιστη καθυστέρηση | 260 πράκτορες στο 1 ^ο στάδιο |
| Μέγιστη δρομολόγηση | 385 πράκτορες στο 2 ^ο στάδιο |

Πίνακας 2. Στατιστικά αποτελέσματα ντετερμινιστικού τρόπου δρομολόγησης

| ΓΕΝΙΚΑ ΣΤΑΔΙΑ | 13 |
|--|---|
| Συνολικοί πράκτορες με καθυστέρηση | 567 |
| Συνολικοί πράκτορες με δρομολόγηση | 4608 |
| Μέση τιμή Συντελεστή Δρομολόγησης CR – ΦΑΣΗ Α | 0,68 |
| Μέση τιμή Συντελεστή Δρομολόγησης CR – ΦΑΣΗ Β | 0,57 |
| Μέση τιμή Συντελεστή Καθυστέρησης CB – ΦΑΣΗ Α | 0,10 |
| Μέση τιμή Συντελεστή Καθυστέρησης CB – ΦΑΣΗ Β | 0,08 |
| Μέγιστη καθυστέρηση / πράκτορα | 4 στάδια |
| Εφαπτομένη της μέσης απόστασης Hamming CH | -1,4 / στάδιο |
| Πρώτος πράκτορας που έφτασε στον προορισμό του | 9 στάδια |
| Μέγιστη καθυστέρηση – ΦΑΣΗ Α | 84 πράκτορες στο 1 ^ο στάδιο |
| Μέγιστη καθυστέρηση – ΦΑΣΗ Β | 58 πράκτορες στο 7 ^ο στάδιο |
| Μέγιστη δρομολόγηση – ΦΑΣΗ Α | 422 πράκτορες στο 2 ^ο στάδιο |
| Μέγιστη δρομολόγηση – ΦΑΣΗ Β | 448 πράκτορες στο 9 ^ο στάδιο |

Πίνακας 3. Στατιστικά αποτελέσματα στοχαστικού τρόπου δρομολόγησης

Από τους παραπάνω πίνακες, είναι σαφής η αποτελεσματικότητα της στοχαστικής μεθόδου, όπως άλλωστε είχε προβλεφθεί και στην μαθηματική ανάλυση πιο πριν. Εάν προσαρμοσθεί και η τοπική προσέγγιση εξυπηρέτησης (FCFS) τα αποτελέσματα θα είναι εντυπωσιακά.

Επίσης, το άνω φράγμα του Borodin-Horogoft $\Omega\left(\sqrt{\frac{2^N}{\log 2_n}}\right)$ είναι πάντα αλήθεια (17 στάδια)

3. ΘΕΩΡΙΑ ΑΝΑΔΡΟΜΙΚΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

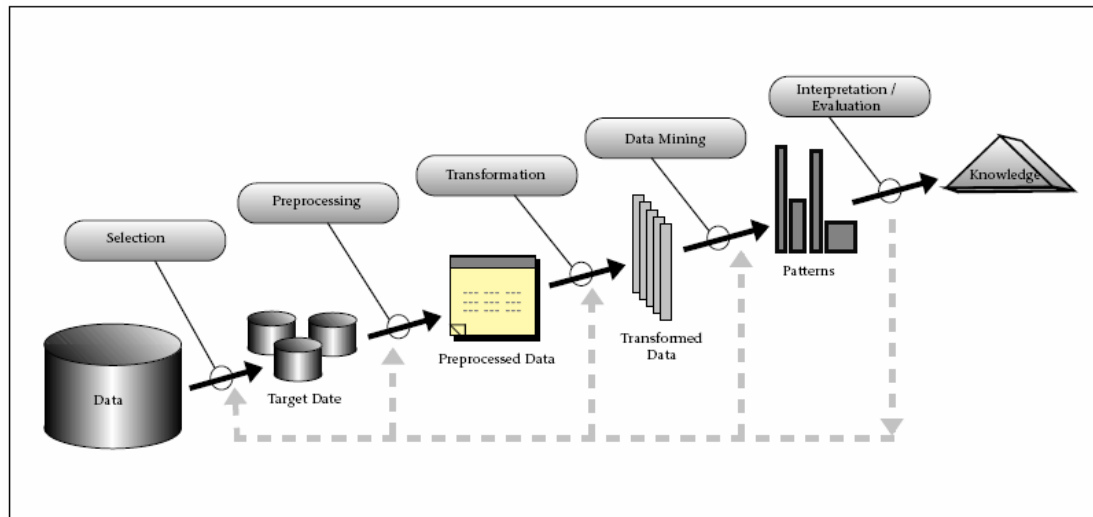
3.1 Data Mining και Θεωρία Συσταδοποίησης (Clustering)

Η εξόρυξη δεδομένων (data mining) [77,78] που αναφέρεται επίσης και ως εξόρυξη γνώσης (knowledge discovery) σε βάσεις δεδομένων, σημαίνει μία διαδικασία μη τετριμμένης εξαγωγής των υπονοούμενων, προηγουμένως άγνωστων, ενδεχομένως χρήσιμων και πλήρως κατανοητών πληροφοριών (όπως κανόνες γνώσης, περιορισμούς, κανονικότητες) από βάσεις δεδομένων.

Με την ανάκτηση γνώσης από βάσεις δεδομένων, μπορούν να προκύψουν ενδιαφέρουσες περιοδικότητες, ή υψηλού επιπέδου πληροφορία από τα σχετικά σύνολα δεδομένων δίνοντας την ευκαιρία για παραπέρα εξέταση τους από διαφορετικές οπτικές γωνίες. Ας έρθουμε όμως στην εφαρμογή των αποτελεσμάτων της εξόρυξης δεδομένων: Η αποκτηθείσα γνώση μπορεί να εφαρμοσθεί στη διαχείριση πληροφορίας, στην επεξεργασία ερωτημάτων (query) για άντληση πληροφοριών, στη λήψη αποφάσεων, σε διαδικασίες ελέγχου και σε πληθώρα εφαρμογών. Επιπλέον η χρησιμότητα των τεχνικών εξόρυξης δεδομένων είναι εμφανής σε πολλές ανερχόμενες εφαρμογές που σχετίζονται με υπηρεσίες παροχής πληροφοριών, όπως υπηρεσίες μέσω διαδικτύου, όπου απαιτείται βαθύτερη κατανόηση της συμπεριφοράς των καταναλωτών έτσι ώστε να βελτιστοποιηθούν οι παρεχόμενες υπηρεσίες σε πραγματικό χρόνο. Μία από αυτές τις «ανερχόμενες εφαρμογές» είναι και η παροχή ενεργειακών υπηρεσιών μέσω διαδικτύου κατόπιν επεξεργασίας και εξόρυξης γνώσης από real-time ενεργειακές μετρήσεις. Βέβαια τα ενεργειακά δεδομένα είναι πολύ γενικός όρος και ενδεχομένως να περιλαμβάνουν πολλά διαφορετικά μετρήσιμα μεγέθη, η επεξεργασία των οποίων θα πρέπει να μπορεί να διαφοροποιείται προσαρμοζόμενη στις εκάστοτε ανάγκες. Εξάλλου σύμφωνα με τους [59,60,61,62], η εξόρυξη γνώσης οφείλει να είναι διαδραστική (interactive) διαδικασία με πολλαπλά αφαιρετικά επίπεδα. Εφόσον είναι εξαιρετικά δύσκολο να προβλεφθεί πάντα ποια ακριβώς θα μπορούσαν να είναι τα εξαγόμενα συμπεράσματα από μία βάση δεδομένων, κάθε υψηλού επιπέδου προσπάθεια για εξόρυξη δεδομένων από μία βάση θα πρέπει να αντιμετωπίζεται ως δοκιμή που ενδεχομένως να περικλείει σημαντική γνώση για βαθύτερη μελέτη των δεδομένων. Η αλληλεπίδραση στην εξόρυξη δεδομένων επιτρέπει στο μελετητή να διαφοροποιεί δυναμικά την εστίαση στα δεδομένα, σταδιακά να εμβαθύνει σε αυτά αλλά και αποστασιοποιημένος να παρατηρεί τα δεδομένα και τα αποτελέσματα των τεχνικών εξόρυξης από διαφορετικές οπτικές γωνίες.

Η εξόρυξη δεδομένων αποτελεί ένα στάδιο στη διαδικασία απόκτησης γνώσης, η οποία αποτελείται από εφαρμογή της ανάλυσης δεδομένων και την εφεύρεση αλγορίθμων, οι

οποίοι έπειτα από τους υπαρκτούς περιορισμούς στην υπολογιστική αποτελεσματικότητα, παράγουν έναν συγκεκριμένο αριθμό από πρότυπα (patterns) ή μοντέλα για τα δεδομένα. Αρκεί να αναλογιστεί κανείς ότι ο χώρος των προτύπων είναι συχνά άπειρος, και ότι η απαρίθμηση των προτύπων περιλαμβάνει κάποιου είδους αναζήτηση σε αυτό τον χώρο. Δηλαδή εντελώς πρακτικά, τα υπολογιστικά όρια θέτουν σοβαρούς περιορισμούς στον εύρος του υποχώρου που προκύπτει από έναν αλγόριθμο εξόρυξης δεδομένων.



Εικόνα 19. Διαδικασία Εξόρυξης Δεδομένων

Μέθοδοι Εξόρυξης Δεδομένων:

- Η *κατάταξη (classification)* είναι μία διαδικασία μάθησης η οποία χαρτογραφεί (κατατάσσει) τα δεδομένα σε μία από τις προκαθορισμένες κατηγορίες
- Η *κατηγοριοποίηση (Clustering)* είναι μία συνηθισμένη μέθοδος περιγραφικής διαδικασίας κατά την οποία κάποιος προσπαθεί να προσδιορίσει ένα καθορισμένο σύνολο από κατηγορίες ή κλάσεις ή συστάδες έτσι ώστε να περιγράφονται τα δεδομένα. Οι συστάδες μπορεί να είναι αμοιβαία αποκλειόμενες ή να αποτελούν επικαλυπτόμενες κλάσεις. Τα συνηθέστερα παραδείγματα εφαρμογής της συσταδοποίησης για εξόρυξη γνώσης περιλαμβάνουν την εύρεση ομογενών πληθυσμών από πελάτες σε βάσεις δεδομένων με αντικείμενο το μάρκετινγκ [39].
- Η *περίληψη (Summarization)* περιλαμβάνει μεθόδους για εύρεση συμπαγούς περιγραφής υποσυνόλων των δεδομένων.

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, η βασική διαδικασία ανάλυσης της δυναμικής πληροφορίας σε αυτή την διατριβή, είναι ένας εξελιγμένος αλγόριθμος **K-Means** [42,], όπου χρησιμοποιούνται σχετικά βάρη και αναδραστική μεθοδολογία, για την εφαρμογή του συγκεκριμένου αλγόριθμου συσταδοποίησης, επάνω από υπερκυβικούς κόμβους, οι οποίοι εμπεριέχουν ως input τους δείκτες, βάσει των οποίων θα τρέξει το clustering.

Η συσταδοποίηση είναι ένα σημαντικό εργαλείο στο data mining [58,65,66], και πολλές περιπτώσεις και δημοσιεύσεις χρησιμοποιούν αλγόριθμους συσταδοποίησης (όπως ο k-means) με σκοπό να κατηγοριοποιήσουν δεδομένα. Η ικανότητα αυτών των τεχνικών να κατηγοριοποιήσουν αυτόματα δεδομένα, ακολουθώντας συγκεκριμένα βήματα, επιτρέπει στην εφαρμογή να αποκαλύπτει κρυμμένα trends και τάση πληροφορίας καθώς και να συμπίσει μεγάλο όγκο δεδομένων και πληροφορίας σε πιο μικρά και οργανωμένα γκρουπ. Κάποια μειονεκτήματα αυτών των αλγορίθμων, είναι ότι απαιτούν μεγάλη υπολογιστική πολυπλοκότητα και φόρτο για την εφαρμογή τους. Ο K-means είναι ο πιο κοινά αποδεκτός, ο οποίος χρησιμοποιεί το τετράγωνο του λάθους (sqf error) ως κριτήριο.

Η συσταδοποίηση περιλαμβάνει το διαχωρισμό ενός συνόλου δεδομένων σε μη επικαλυπτόμενες ομάδες ή κλάσεις, με μέλη τα οποία να σχετίζονται πολύ περισσότερο μεταξύ τους απ' ό,τι με μέλη άλλων κλάσεων. Ο όρος «σχετίζονται» όταν αναφέρεται σε συσταδοποιημένα δεδομένα, συνήθως σημαίνει «βρίσκονται πιο κοντά σύμφωνα με κάποιο μέτρο ομοιότητας». Όταν ένα σύνολο δεδομένων είναι συσταδοποιημένο, κάθε μέλος αποδίδεται σε μία κλάση και κάθε κλάση προσδιορίζεται από ένα χαρακτηριστικό σημείο αναφοράς, συνήθως κάποιον μέσο όρο των μελών της κλάσης.

Η μείωση όγκου δεδομένων είναι συνδεδεμένη με την αντικατάσταση των συντεταγμένων κάθε μέλους μίας κλάσης με τις συντεταγμένες του σημείου αναφοράς της κλάσης αυτής. Τα συσταδοποιημένα δεδομένα λοιπόν καταλαμβάνουν λιγότερο χώρο αποθήκευσης και φυσικά μπορούν να χειριστούν ταχύτερα και αποτελεσματικότερα για εξαγωγή συμπερασμάτων σε σχέση με τα αρχικά δεδομένα. Η αξία κάθε επιμέρους αλγόριθμου συσταδοποίησης θα εξαρτηθεί από το πόσο πιστά αναπαριστούν τα σημεία αναφοράς τα αρχικά δεδομένα, αλλά και από το πόσο ταχύς είναι ο αλγόριθμος. Ένα κλασικό παράδειγμα συσταδοποίησης είναι ενοποίηση ενός συνόλου από βαθμολογίες μαθητών εκφρασμένες σε ποσοστά (%), σε πέντε κατηγορίες, μία για κάθε βαθμό A, B, C, D. Οι αρχικές βαθμολογίες είναι τα δεδομένα και το σημείο αναφοράς για κάθε κλάση είναι ο μέσος όρος βαθμολογιών των μελών της κλάσης. Τα πέντε γράμματα μπορούν να θεωρηθούν ως συμβολική αντικατάσταση του εκάστοτε αριθμητικού σημείου αναφοράς. Οι βαθμολογίες μπορούν να θεωρηθούν ως μονοδιάστατα

δεδομένα. Πολυδιάστατα δεδομένα μπορεί να περιγράφονται από ένα πλήθος μετρήσιμων συνιστωσών. Κάθε ξεχωριστό χαρακτηριστικό ή μέτρηση αποτελεί μία διάσταση, μία συνιστώσα. Η ενοποίηση μεγάλων , πολυδιάστατων συνόλων δεδομένων είναι ο κύριος σκοπός του τομέα της ανάλυσης κλάσεων. Θα περιγραφούν διαφορετικές μέθοδοι (αλγόριθμοι) συσταδοποίησης. Σε όλες τις μεθόδους ο επιθυμητός αριθμός κλάσεων προσδιορίζεται από το αρχικό στάδιο, ενώ το σημείο αναφοράς κάθε κλάσης είναι συνήθως το κεντροειδές της κλάσης. Στην περίπτωση των μονοδιάστατων δεδομένων το κεντροειδές είναι ο αριθμητικός μέσος όρος των τιμών των μελών που ανήκουν σε μία κλάση, ενώ για τα πολυδιάστατα σύνολα δεδομένων το κεντροειδές θα είναι της ίδιας διάστασης με τα αρχικά δεδομένα και κάθε συνιστώσα θα είναι ο μέσος όρος των αντίστοιχων συνιστωσών όλων των μελών της κλάσης.

Ίσως ο πιο απλός (εύκολα αντιληπτός) αυτοματοποιημένος τρόπος είναι να συνδυάζονται τα δεδομένα ανά ζεύγος μέχρι να φτάσουμε στον επιθυμητό αριθμό κλάσεων, μία μέθοδος που είναι συνυφασμένη με πολλά εμπειρικά στατιστικά πακέτα. Υπάρχουν όμως δύο τεράστια μειονεκτήματα σε αυτή τη μέθοδο με κύριο και απαγορευτικό το γεγονός ότι ο αλγόριθμος είναι αναποτελεσματικός για ανάλυση μεγάλων βάσεων δεδομένων λόγω της υπολογιστικής πολυπλοκότητας. Κάθε βήμα της διαδικασίας απαιτεί υπολογισμό των αποστάσεων μεταξύ δύο οποιωνδήποτε πιθανών ζευγών δεδομένων και στη συνέχεια σύγκριση όλων των αποστάσεων μεταξύ τους. Η δεύτερη δυσκολία έγκειται στο γεγονός ότι παρόλο που ο αλγόριθμος παράγει τον επιθυμητό αριθμό κλάσεων , τα κεντροειδή τους ενδεχομένως να μην είναι πάντα αντιπροσωπευτικά των επιμέρους μελών της κλάσης.

Από τη μία μπορεί η συσταδοποίηση να αποτελεί μία βολική μέθοδο οργάνωσης ενός μεγάλου συνόλου δεδομένων έτσι ώστε να είναι πιο εύκολα κατανοητό και να παρέχει πιο αποτελεσματική πληροφόρηση. Για παράδειγμα σε μία έρευνα αγοράς οι διαφορετικές ομάδες καταναλωτών μπορεί να παρέχουν πολύ ακριβή πληροφόρηση και περιγραφή των προτιμήσεων των καταναλωτών δείχνοντας έτσι το κατάλληλο προϊόν για τον κατάλληλο τύπο πελάτη.

Κάτι παρόμοιο επιχειρούμε να συμπεράνουμε και με την εφαρμογή του αλγορίθμου συσταδοποίησης στα δεδομένα με τις ενεργειακές καταναλώσεις, έτσι ώστε να προκύψουν ομοιότητες και διαφορές στις ενεργειακές συμπεριφορές.

Σε πολλές εφαρμογές ωστόσο η συσταδοποίηση ενδεχομένως να επιτελεί και κάποιον απώτερο σκοπό εκτός από την αντιπροσωπευτική απεικόνιση του συνόλου δεδομένων. Στον

τομέα της διαχείρισης ενέργειας που μας ενδιαφέρει, η κατηγοριοποίηση πελατών αλλά και κτιρίων ή ακόμα και ωρών μέσα στην ημέρα όσον αφορά την ενεργειακή τους κατανάλωση, οδηγεί σε μοντέλα για την πρόβλεψη φορτίου και αν πάμε ακόμη πιο μακριά σε προγραμματισμό των φορτίων, ενδεχομένως μετακίνηση (ολίσθηση) κάποιων αιχμών ζήτησης έτσι ώστε να αποφευχθεί πιθανή υπερφόρτωση δικτύου ή και να προβλεφθεί η μελέτη και εγκατάσταση επιπλέον μονάδων παραγωγής στο μέλλον.

Είναι λοιπόν σαφές ότι σχεδόν σε κάθε περίπτωση υπάρχει μία πληθώρα επιλογών για το ως προς τι θα γίνει η κατηγοριοποίηση ακόμη και σε ένα συγκεκριμένο σύνολο μετρήσεων. Οι διαφορετικοί τρόποι συσταδοποίησης θα δώσουν και διαφορετικά αποτελέσματα, δηλαδή διαφορετικά μέλη θα ανήκουν τώρα στην ίδια κλάση, αλλά στα αποτελέσματα των αλγορίθμων αυτών δεν υπάρχει σωστό και λάθος · το μόνο που κρίνει μία δεδομένη κατηγοριοποίηση είναι η χρησιμότητα και η αποτελεσματικότητά της.

Ας δούμε όμως τι είναι εκείνο που χαρακτηρίζει μία μέθοδο συσταδοποίησης ως αποτελεσματική και αντιπροσωπευτική. Ας θεωρήσουμε μία μοναδική κλάση με μέλη γύρω από το κεντροειδές ή το μέσο της. Εφόσον τα δεδομένα-μέλη είναι πυκνά κατανεμημένα γύρω από το κεντροειδές τότε αυτό θα είναι αντιπροσωπευτικό όλων των μελών της κλάσης αυτής. Το κλασικό μέτρο της διασποράς μίας ομάδας δεδομένων γύρω από το μέσο τους είναι η απόκλιση, το άθροισμα των τετραγώνων των αποστάσεων από το κέντρο. Αν τα δεδομένα είναι πυκνά κατανεμημένα γύρω από το κέντρο η απόκλιση θα είναι μικρή. Μία γενίκευση του παραπάνω στην οποία το κεντροειδές μπορεί να αντικατασταθεί από ένα άλλο σημείο αναφοράς για να εκτιμηθεί η γενική ποιότητα της συσταδοποίησης, χρησιμοποιείται ένα μέτρο σφάλματος, το οποίο αποτελεί άθροισμα όλων των αποκλίσεων.

$$E = \sum_{i=1}^k \sum_{j=1}^{n_i} \|x_{ij} - z_i\|^2, \quad (3.1.1)$$

όπου το x_{ij} είναι το j -στο μέλος της i -στης κλάσης και z_i είναι το σημείο αναφοράς της κλάσης αυτής, το n_i αποτελεί το πλήθος των μελών της κλάσης, ενώ το $\|x_{ij} - z_i\|$ σημαίνει την απόσταση μεταξύ του x_{ij} και του κέντρου z_i . Έτσι το μέτρο E δείχνει τη συνολική διασπορά των μελών κάθε κλάσης γύρω από το κέντρο της και πάντα ο στόχος μίας μεθόδου συσταδοποίησης θα είναι η ελαχιστοποίηση αυτού του σφάλματος.

Η ανάλυση κλάσεων είναι απαραίτητη για να προκύπτουν ομάδες μεταξύ των μελών ενός συνόλου δεδομένων, ενώ οι μέθοδοι συσταδοποίησης δε θα πρέπει να συγχέονται με τους

όρους διάκριση ή κατάταξη (ή με τον όρο επιβλεπόμενη μάθηση από την τεχνητή νοημοσύνη), όπου οι ομάδες είναι εκ των προτέρων γνωστές και απλώς επιδιώκουμε να βρούμε έναν τρόπο να επινοήσουμε κανόνες για να εκχωρήσουμε νέα μέλη στις υπάρχουσες ομάδες. Όπως είδαμε και παραπάνω ο όρος κατάταξη (classification) και συσταδοποίηση (clustering) αποτελούν διαφορετικές μεθόδους εξόρυξης δεδομένων.

Οι όροι **κλάση ,ομάδα ή συστάδα** (class, group, cluster) χρησιμοποιούνται συνήθως με τον ίδιο ακριβώς τρόπο χωρίς να υπάρχει κάποιος επίσημος ορισμός για να τους διαχωρίζει. Στην πραγματικότητα ένας επίσημος ορισμός θα ήταν όχι μόνο δύσκολο να προκύψει αλλά πιθανότατα θα οδηγούσε και σε λάθος συμπεράσματα. Το μόνο που χρειάζεται για να δώσει αξία και σημασία στο περιεχόμενο του όρου «κλάση» (“cluster”) είναι η χρησιμότητά του για τον παρατηρητή.

Οι *Cormack* [88] και *Gordon* [89] κάνουν μια προσπάθεια δώσουν ένα ορισμό για το είναι ένα **cluster** σύμφωνα με τους όρους της εσωτερικής συνοχής (ομοιογένεια, homogeneity) και εξωτερική απομόνωση (διαχωρισμός, separation). Τέτοιου είδους ιδιότητες μπορούν να απεικονιστούν σε απλά 2-διάστατα διαγράμματα και από την παρατήρησή τους να προκύψει ο ορισμός του όρου cluster. Προσπαθώντας όμως να δοθούν πιο επίσημοι ορισμοί που να καλύπτουν όλες τις περιπτώσεις καταλήγουμε στο συμπέρασμα ότι είναι σχεδόν αδύνατο να υπάρξουν μαθηματικά ακριβείς ορισμοί για την ομοιογένεια και την απομόνωση που με τη μορφή αριθμητικών δεικτών να οδηγούν σε συγκεκριμένα κριτήρια για το διαχωρισμό σε συστάδες. Ένα καλό κριτήριο για τη διαδικασία αναγνώρισης κλάσεων θα περιλάμβανε σίγουρα μία αποτίμηση των σχετικών αποστάσεων μεταξύ των μελών κάθε κλάσης.

Διάφοροι Αλγόριθμοι Συσταδοποίησης

Για να επιλεγεί ο βέλτιστος αλγόριθμος συσταδοποίησης θα πρέπει να τεθούν κάποια κριτήρια όπως περιγράφηκε ήδη και στη συνέχεια θα πρέπει να βρεθούν όλες οι πιθανές διαμερίσεις και αφού υπολογιστούν και συγκριθούν τα αντικειμενικά κριτήρια που έχουν τεθεί (όπως για παράδειγμα η συνολική διασπορά που περιγράφηκε ωρίτερα), να επιλεγεί εκείνος ο διαχωρισμός σε συστάδες που βελτιστοποιεί το κριτήριο. Ωστόσο κάτι τέτοιο είναι πρακτικά αδύνατο αφού η υπολογιστική πολυπλοκότητα ακόμη και για μεσαίου μεγέθους βάσεις δεδομένων είναι τεράστια. Συγκεκριμένα οι διαφορετικοί τρόποι για να διαμερίσει

κανείς n αντικείμενα σε k κατηγορίες είναι $N(n, k) = \frac{1}{k!} \sum_{i=1}^k (-1)^{k-i} \cdot \binom{k}{i} m^i$ [90]. Ας πούμε

για να χωρίσουμε 100 μέλη σε 5 κατηγορίες , υπάρχουν $6,6 \cdot 10^{67}$ διαφορετικοί τρόποι, για καθέναν από τους οποίους θα πρέπει να υπολογίζουμε και να συγκρίνουμε το κριτήριο που έχουμε θέσει.

K-Means

Ο αλγόριθμος *k-means* [68,42,91] είναι ο πιο ευρεία διαδεδομένος αλγόριθμος συσταδοποίησης, και το όνομά του προέρχεται από το σημείο αναφοράς που επιλέγεται ως κέντρο κάθε κλάσης, ως σταθμισμένος μέσος των μελών που ανήκουν στην κλάση. Ο αλγόριθμος απαιτεί αρχικοποίηση του πλήθους των κλάσεων και ξεκινά με ανάθεση των κέντρων των κλάσεων επιλέγοντας από το αρχικό δείγμα. Στη συνέχεια κάθε μέλος αποδίδεται στην πλησιέστερη κλάση καθώς το κεντροειδές της κλάσης μετατοπίζεται, συνυπολογίζοντας τα νέα μέλη. Η διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί το κριτήριο που έχει τεθεί με συνθηθέστερο κριτήριο την απαίτηση να μην υπάρχουν ποια μετακινήσεις μελών μεταξύ κλάσεων [91]. Ο αλγόριθμος έχει το πλεονέκτημα ότι έχει σαφή γεωμετρική και στατιστική ερμηνεία, αλλά και το μειονέκτημα ότι μπορεί να αντιμετωπίσει μόνο αριθμητικά δεδομένα και ότι είναι ευαίσθητο σε εξωτερικούς παράγοντες.

Kohonen Self-Organized Maps (SOM) [92]

Τα δίκτυα Kohonen αποτελούν ένα είδος νευρωνικού δικτύου που πραγματοποιεί συσταδοποίηση στα εισαγόμενα δεδομένα. Οι βασικές μονάδες είναι νευρόνια οργανωμένα συνήθως σε δύο επίπεδα : το επίπεδο εισόδου και το επίπεδο εξόδου, που συχνά αναφέρεται και ως χάρτης εξόδου. Όλα τα νευρόνια εισόδου είναι συνδεδεμένα με όλα τα νευρόνια εξόδου με διαφορετικά βάρη. Τα δεδομένα εισόδου εισάγονται στο αρχικό επίπεδο και δρομολογούνται προς την έξοδο μέσω των νευρώνων , τα οποία αποκρίνονται ανάλογα με τα βάρη στις συνδέσεις. Η πιο ισχυρή απόκριση είναι και ο «νικητής». Αρχικά όλα τα βάρη αποδίδονται τυχαία και κατά την πορεία της διαδικασίας εκπαιδεύονται ανάλογα με τα αποτελέσματα που προκύπτουν από κάθε διαφορετική είσοδο δεδομένων. Έτσι το νευρόνιο που υπερिशύει κάθε φορά διαμορφώνει έτσι τα βάρη που το συνδέουν με το προηγούμενο επίπεδο ώστε την επόμενη φορά που θα υπάρξει παρόμοιο δεδομένο στην είσοδο , να είναι σίγουρο το δεδομένο νευρόνιο που θα υπερिशύσει. Επίσης τα κοντινά νευρόνια προσαρμόζουν κι αυτά τα βάρη τους έτσι ώστε να αντιδράσουν πιο θετικά την επόμενη φορά. Η διαδικασία επαναλαμβάνεται αρκετές φορές για όλα τα δεδομένα ώσπου οι αλλαγές που παρατηρούνται στα βάρη να είναι πολύ μικρές. Το πιο ενδιαφέρον χαρακτηριστικό του συγκεκριμένου αλγορίθμου συσταδοποίησης είναι ότι από τη στιγμή που θα τελειώσει η όλη

διαδικασία της εκπαίδευσης του δικτύου, αυτό θα είναι ένας χάρτης που θα απεικονίζει το N-διαστατο χώρο των δεδομένων σε έναν διδιάστατο χώρο.

Στην συγκεκριμένη διατριβή, χρησιμοποιείται ο **k-means** ως βάση και βελτιστοποιείται από την χρήση ειδικών βαρών (weights) για τον υπολογισμό της απόστασης (τετράγωνο του σχετικού λάθους σε κάθε iteration). Επίσης, η εφαρμογή του γίνεται αναδρομικά και ακολουθείται κάθε φορά μετά τον υπολογισμό των συγκεκριμένων στατιστικών δεικτών από το υπερκυβικό πλέγμα (πράκτορες). Ο αλγόριθμος παίρνει σαν είσοδο κάποιες αρχικές συνθήκες (αριθμός συσταδων, κλπ) και εν συνεχεία συσταδοποιεί τα στοιχεία σε k συστάδες, ανάλογες με αυτές που έχουν οριστεί από την αρχή. Υπολογίζοντας σε κάθε πέρασμα το τετράγωνο της ευκλείδειας απόστασης μεταξύ του κάθε 2-διάστατου δείκτη και του κεντροειδούς (το κοντινότερο σημείο μεταξύ της συστάδας) ο αλγόριθμος ελαχιστοποιεί το τετράγωνο του λάθους της ευκλείδειας απόστασης. Έτσι, οι διάφοροι δείκτες κατηγοριοποιούνται αυτόματα σε εκείνη την συστάδα η μέσα στην οποία ο δείκτης απέχει την πιο κοντινή απόσταση από το κέντρο της.

Στην παρούσα διατριβή, στα παραδείγματα με την ενέργειας, 2-διάστατοι δείκτες χρησιμοποιήθηκαν και επιλέχθηκαν 3 συστάδες και 3 κεντροειδή. Η επιλογή των συστάδων και κεντροειδών είναι βάσει της στρατηγικής που θα χρησιμοποιηθεί για την ανάλυση της δυναμικότητάς τους.

3.2 Αλγόριθμος RKM (recursive K-Means)

Ο αλγόριθμος που δημιουργήθηκε και χρησιμοποιείται σε αυτή την διατριβή, ονομάζεται **Αναδρομικός K-Means (Recursive K-Means, RKM)** και εκτελείται παράλληλα.

Ο αλγόριθμος **RKM** λειτουργεί στα εξής στάδια:

- Ως πρώτο στάδιο επιλέγονται τα υποψήφια clusters (πχ. αριθμός = 3) στα οποία και θα κατηγοριοποιηθούν οι κόμβοι του υπερκύβου και επιλέγονται επίσης και οι δείκτες, βάσει των οποίων θα γίνει η συσταδοποίηση των κόμβων του υπερκύβου. Σε κάθε υπερκυβικό κόμβο μπορεί να αντιστοιχηθεί ένα καταναλωτής, ένας χρήστης, ένας μετρητής ενέργειας, κοκ. Η συσταδοποίηση γίνεται βάσει των δεικτών που αναλύονται και υπολογίζονται από τους υπερκυβικούς πράκτορες και εφαρμόζεται ο αναδραστικός RKM σε κάθε κόμβο, με αποτέλεσμα να έχουμε συσταδοποιήσεις των κόμβων του υπερκύβου σε κάθε εκτέλεση του RKM.

- Ο αλγόριθμος εκτελείται αναδραστικά για κάθε δείκτη και για κάθε κόμβο παράλληλα.
- Ως αποτέλεσμα, έχουμε δυναμικά clusters και προσαρμοστικά κεντροειδή, που έχουν κάποιες ιδιότητες και χαρακτηριστικά. Μέσα στις συστάδες, υπάρχουν πληθυσμοί από υπερκυβικούς κόμβους, που βάσει του RKM εκείνη την χρονική στιγμή είναι μέλη του εκάστοτε cluster. Στο επόμενο πέρασμα του RKM και οι πληθυσμοί αλλά και τα υπόλοιπα στοιχεία / δείκτες αλλάζουν δυναμικά. Αυτή η δυναμικότητα αναλύεται σε πραγματικό χρόνο για συμπερασμό.

Στον αλγόριθμο RKM, το clustering διάνυσμα C ορίζεται από τις συντεταγμένες x, y του κεντροειδούς $\mu(d)$ που ανήκει στην συστάδα (cluster) \mathbf{g}_N , αφού ο αλγόριθμος RKM έχει τελειώσει με τις επαναλήψεις. Η μαθηματική σχέση ορίζεται ως:

$$C_{i,j}^N = [x_{i,j} \quad y_{i,j}]_{\mathbf{g}_N} \quad (3.2.1)$$

Ορίζουμε επίσης την εντροπία e ως το μέσο όρο της στατιστικής διασποράς όλων των ευκλείδιων αποστάσεων, της συστάδας \mathbf{g}_N που εσσωκλείει το κεντροειδές $\mu(d)$:

$$e_{i,j}^N = \left(\frac{1}{n} \sum_{i \in d(n)} \vec{Ed}|\mu|_N \right) \quad (3.2.2)$$

όπου:

$$Ed|\mu|_i = \sqrt{\sum_{n=1}^k (x_{i,n} - \mu(d)_n)^2} \times w_i \quad (3.2.3)$$

αντιπροσωπεύει την i -οστή ευκλείδεια απόσταση του αναδρομικού δείκτη x_i από το πιθανό κεντροειδές $\mu(d)$ της N_{th} συστάδας, πολλαπλασιαζόμενο από ένα προσαρμοστικό βάρος που περιγράφεται από τον Πίνακα \mathbf{w}^T . Ο πίνακας βαρών αντιπροσωπεύει κάποιες δυναμικές μεταβλητές του εκάστοτε υπερκυβικού κόμβου και οι οποίες μοντελοποιούνται ανάλογα. Παράδειγμα, εάν η εφαρμογή γίνει σε ενεργειακά δεδομένα, τότε τα βάρη αντιπροσωπεύουν κάποιες στοχαστικές μεταβλητές του καιρού ή λοιπά σημασιολογικά ή περιγραφικά στοιχεία. Εάν το βάρος είναι $w > 1$ τότε ο κόμβος απομακρύνεται από το κεντροειδές και εάν το βάρος $w < 1$ τότε ο κόμβος (χρήστης) έρχεται πιο κοντά.

Για παράδειγμα, στα ενεργειακά δεδομένα τριών καταναλωτών, έχουμε το διάνυσμα βαρών:

$$w^T = \begin{bmatrix} 2.4 \\ 0.5 \\ 1.8 \end{bmatrix} \text{ με } w^T(1) = 2.4 \text{ που σημαίνει ότι η Ευκλείδεια απόσταση του πρώτου}$$

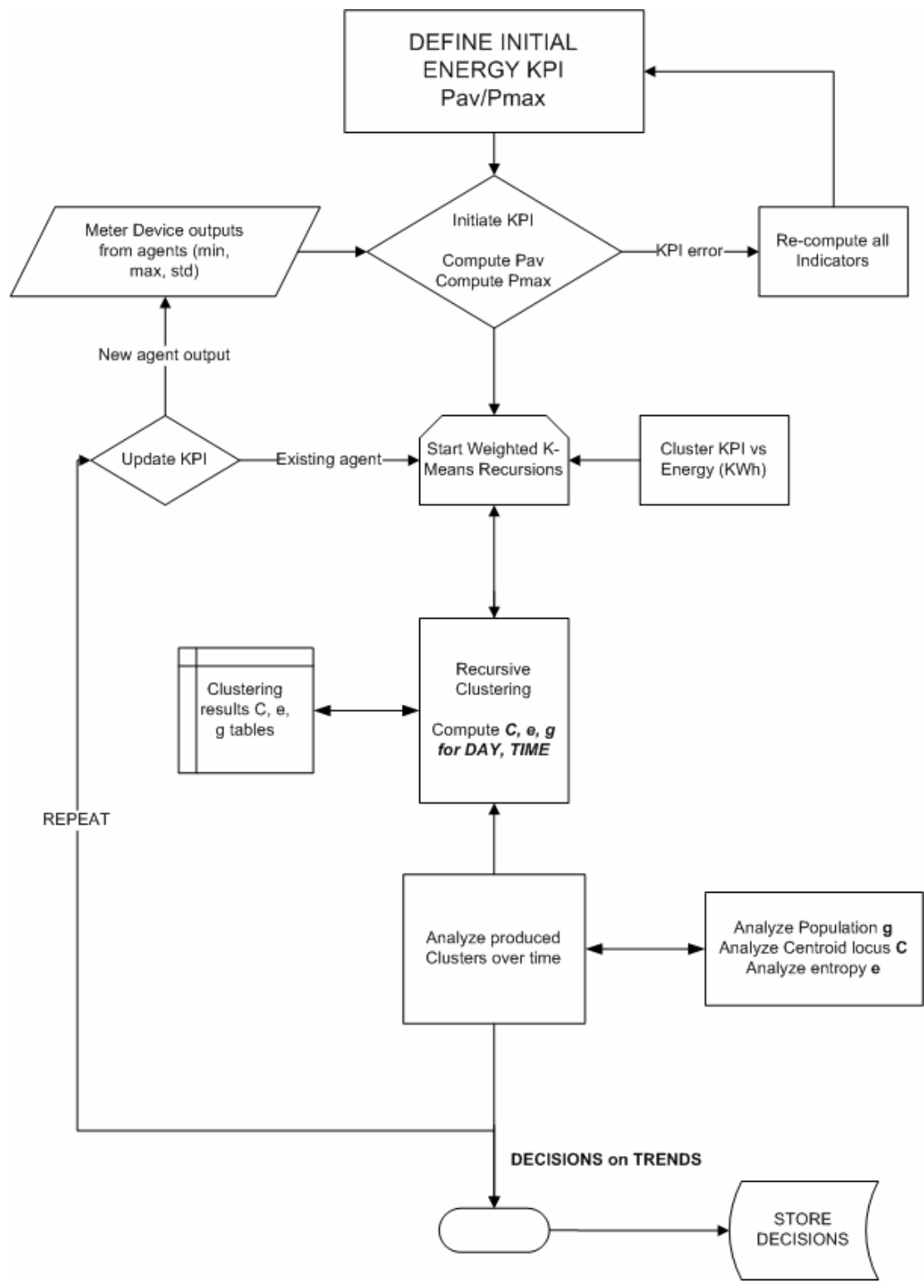
καταναλωτή $(Ed)|\mu|_l$ θα πολλαπλασιαστεί με 2.4 και θα απομακρυνθεί από το κεντροειδές. Αυτό γίνεται για λόγους κανονικοποίησης, καθώς ο συγκεκριμένος καταναλωτής δεν έχει πολύ μεγάλη σημασία για το αποτέλεσμα της συγκεκριμένης ανάλυσης. Οι λόγοι είναι το ότι μένει σε διαφορετικές κλιματικές συνθήκες, ή επειδή καθυστερεί στις πληρωμές. Έτσι ο αλγόριθμος προσφέρει και μια αυτοματοποιημένη κανονικοποίηση, ένα λειτουργικό φιλτράρισμα. Ο 2^{ος} καταναλωτής έχει μεγαλύτερη σημασία στο αποτέλεσμα της συγκεκριμένης ανάλυσης, οπότε ο συντελεστής τον φέρνει πιο κοντά στο συγκεκριμένο κεντροειδές, εξασφαλίζοντας την συμμετοχή του στο cluster.

Έτσι, έχοντας ένα δειγματικό χώρο g , έχουμε έναν συγκεκριμένο υποχώρο από συστάδες g_1 ($N=I$) με διάσταση n όπως:

$$g_{N=1} = \{m_1, m_2 \dots m_n\} \in g \quad (3.2.4)$$

του n -διάστατου χώρου, ή με άλλα λόγια έχουμε n καταναλωτές, όσοι δηλ. και οι υπερκυβικοί κόμβοι στο πλέγμα των πρακτόρων. Αυτοί οι χρήστες (κόμβοι) είναι μέλη του συγκεκριμένου cluster g_N , μετά την εκτέλεση του αλγορίθμου RKM.

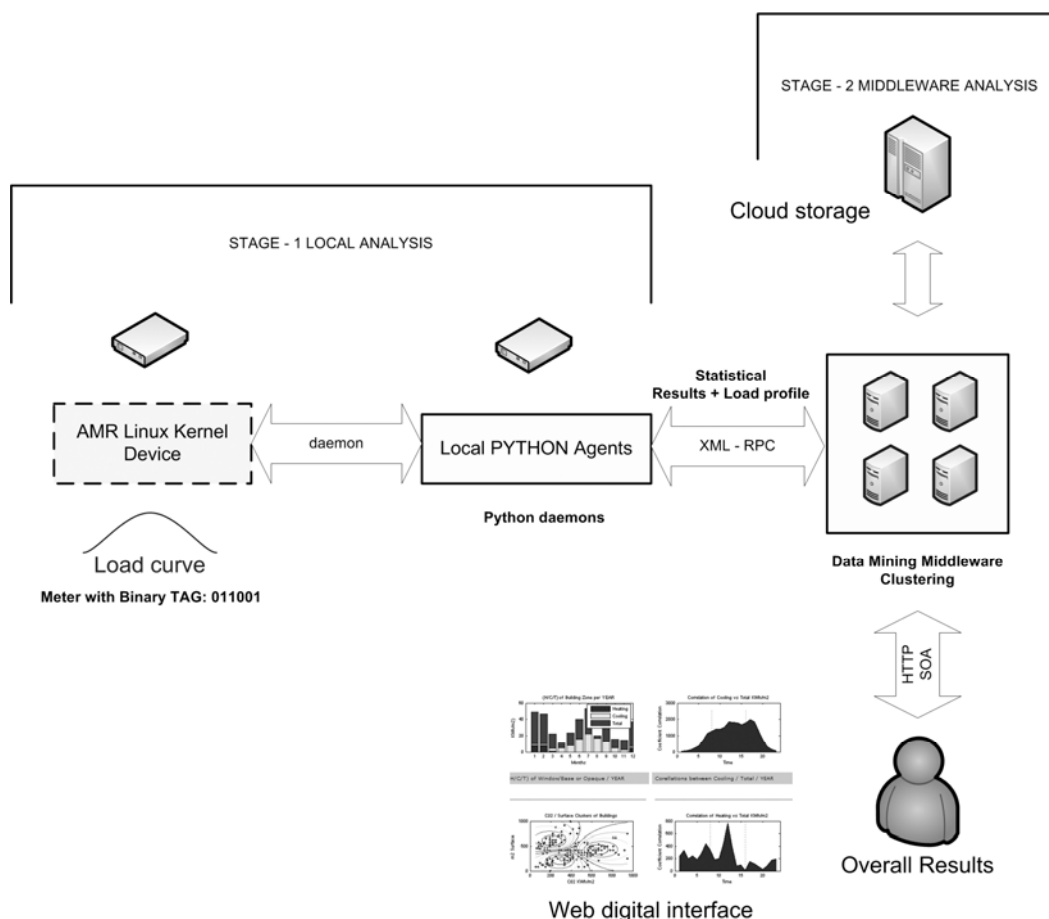
Η λειτουργία του αλγορίθμου, μπορεί να φανεί και στο flow chart που ακολουθεί:



Εικόνα 20. Αλγόριθμος RKM σε εφαρμογή στην ενέργεια, βάσει συγκεκριμένων KPIs

Η συνολική αρχιτεκτονική σε λειτουργικό επίπεδο μπορεί να φανεί στην Εικόνα 17 παρακάτω. Με την χρήση scripting γλώσσας (Python, Perl, Ruby) οι πράκτορες αναλύουν σε επίπεδο firmware τον εκάστοτε υπερκρυβικό κόμβο, που στην συγκεκριμένη περίπτωση

αντιστοιχεί σε έναν έξυπνο ενεργειακό μετρητή. Τα αποτελέσματα της στατιστικής ανάλυσης / κόμβο είναι στην ουσία οι δείκτες που έχουν προ-επιλεγεί και αποτελούν τα input KPIs στον αλγόριθμο RKM.



Εικόνα 21. Λειτουργική αρχιτεκτονική του RKM αλγορίθμου

Η τοπική, ανά υπερκυβικό κόμβο, στατιστική ανάλυση περνάει τα αποτελέσματα στον μερισμικό data mining (data mining middleware clustering) όπου εκεί με βάση τους πράκτορες εκτελείται ο RKM για το συγκεκριμένο χρονικό διάστημα.

Ο αλγόριθμος υπολογίζει το βέλτιστο κεντροειδές, βάσει της ευκλείδειας απόστασης με βάρος w , μεταξύ του συγκεκριμένου δείκτη x_i και του centroid C και καταλήγει σε συστάδες οι οποίες και αναλύονται σε πραγματικό χρόνο. Η διαδικασία επαναλαμβάνεται συνεχώς.

3.3 Ποιοτική Αλγεβρική Ανάλυση Κεντροειδών

Ένα από τα πιο σημαντικά θέματα της παρούσας διατριβής, είναι η ανάλυση των αποτελεσμάτων του αναδρομικού αλγορίθμου RKM. Τα αποτελέσματα του RKM είναι 3 ανύσματα, οι συντεταγμένες των κεντροειδών $C_{i,j}^N$, η εντροπία του cluster $e_{i,j}^N$ και πληθυσμός $P_{i,j}^N$ των συστάδων $g_{N=1} = \{m_1, m_2 \dots m_n\} \in g$ όπου μέσα σε αυτές υπάρχουν συμμετοχές των υπερκυβικών κόμβων οι οποίοι αντιπροσωπεύουν και τον χρήστη.

Η ανάλυση της εντροπίας $e_{i,j}^N$, μας δείχνει πόσο κοντά βρίσκονται τα διάφορα σημεία (κόμβοι) της συστάδας στο κεντροειδές και πόσα είναι αυτά. Είναι ένα μέτρο της πυκνότητας μιας συστάδας.

Η μέτρηση και ανάλυση της στατιστικής μεταβλητότητας (Variance) του μεγέθους του πληθυσμού $P_{i,j}^N$ αντιπροσωπεύει το μέτρο της ποσότητας της εντροπίας $e_{i,j}^N$ της συστάδας N και το κανονικοποιημένο άνυσμα n_{norm} τον αριθμό των κόμβων που είναι μέλη αυτού του cluster.

Το κανονικοποιημένο άνυσμα n_{norm} αντιπροσωπεύει τον επί τοις εκατό % βαθμό πλυθησμού της συγκεκριμένης συστάδας σε σχέση με όλο τον αριθμό των κόμβων που λάβανε μέρος στον RKM αλγόριθμο, και είναι μέρος του δειγματικού χώρου A:

$$P_{i,j}^N = \left| \frac{e_{i,j}^N}{n_{NORM}} \right|_{n \in DIM(A)} \quad \text{where } n_{norm} = n/A \cdot 100 \quad (3.3.1)$$

όπου

$$g_{N=1} = \{m_1, m_2 \dots m_n\} \in g \quad (3.3.2)$$

Συνεπώς, εαν ορίσουμε έναν KRM με 3 συστάδες, ($N=[1,2,3]$), τότε το τελικό διάνυσμα των συστάδων g , υποσύνολο του A είναι:

$$\bigcup g_{N=1}^3 = g_1 \cup g_2 \cup g_3 \in A \quad (3.3.3)$$

Έτσι δημιουργούμε τον Αναδρομικό Πίνακα Συσταδοποίησης RCM (Recursive Clustering Matrix):

$$RCM \equiv g_r^N = \begin{bmatrix} g_{1,1}^N & g_{1,2}^N & g_{1,3}^N \\ g_{2,1}^N & g_{2,2}^N & g_{2,3}^N \\ \vdots & \vdots & \vdots \\ g_{r,1}^N & g_{r,2}^N & g_{r,3}^N \end{bmatrix} \quad (3.3.4)$$

Ο οποίος εμπεριέχει το σύνολο των συστάδων σε κάθε αναδραστική εφαρμογή του RKM αλγόριθμου (recursions) r .

Ορίζουμε τον δείκτη ομοιότητας συστάδας S βάσει του συντελεστή Jaccard [34] με p να είναι το μέγεθος της τομής και το q το μέγεθος της ένωσης δύο διαφορετικών συστάδων / συνόλων g_r και $g_{r,j}$. Συνεπώς, ο αλγεβρικός πίνακας Ομοιότητας $S_{m,n}$ ορίζεται ως:

$$S_{m,n} = \begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,j} & \cdots & S_{1,n} \\ S_{i,1} & S_{i,2} & \cdots & S_{i,j} & \cdots & S_{i,n} \\ S_{m,1} & S_{m,2} & \cdots & S_{m,j} & \cdots & S_{m,n} \end{bmatrix} \quad (3.3.5)$$

Όπου $S_{i,j} = p/q$, για δύο clusters. Οπότε, η συνολική ομοιότητα των δύο συστάδων ορίζεται ως:

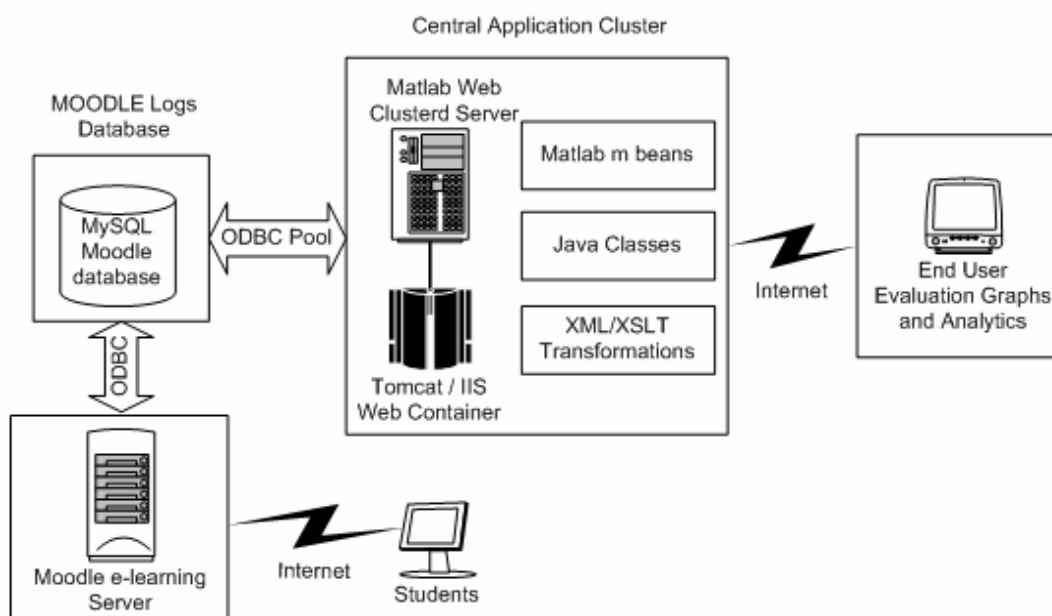
$$Sim(g_{r,i} g_{r,j}) = \frac{\sum_{j=1}^{m,n} S_{m,n}}{\max(m,n)} \quad (3.3.6)$$

Ως αποτέλεσμα λοιπόν του αλγορίθμου RKM ο οποίος εκτελείται συνεχώς και αναδρομικά, μπορούμε σε πραγματικό χρόνο να αναλύσουμε κάποιες σημαντικές παραμέτρους και ποιοτικούς δείκτες των συστάδων που δημιουργούνται και να δούμε πως αυτοί οι δείκτες και συστάδες / κεντροειδή συμπεριφέρονται στο πεδίο του χρόνου, έχοντας χρονοδυναμικά χαρακτηριστικά.

Εάν τώρα εισάγουμε αυτούς του πίνακες σε μεθοδολογίες προχωρημένες Αριθμητικής Γραμμικής Άλγεβρας και βρίσκοντας τις ιδιοτιμές και τα ιδιο-δυναύσματα τους, τότε μπορεί να έχουμε ακόμα πιο εντυπωσιακά αποτελέσματα.

3.4 Εφαρμογή σε Δεδομένα Μαθησιακού LMS

Από τις πρώτες εφαρμογές του υπερκυβικού αλγορίθμου RKM ήταν σε δεδομένα μαθησιακού πληροφοριακού συστήματος (LMS). Το σύστημα ήταν σε τεχνολογία open source (Moodle) και η προσέγγιση ακολούθησε την τεχνική φιλοσοφία που φαίνεται παρακάτω:

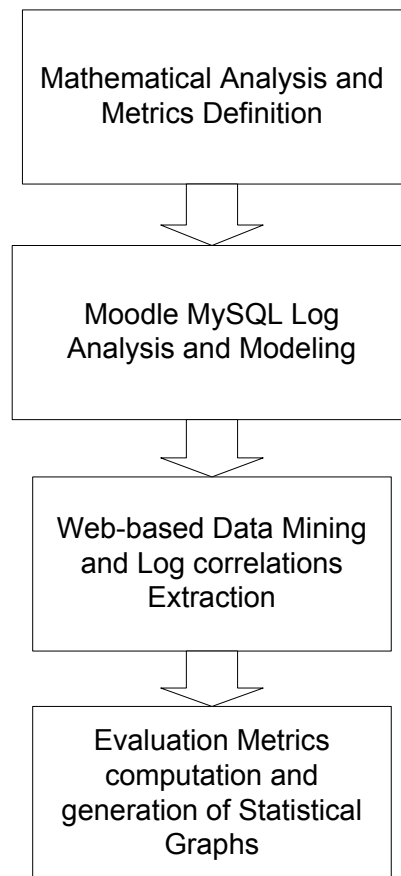


Εικόνα 22. Αρχιτεκτονική ανάλυσης δεδομένων από Moodle LMS

Το σύστημα είχε πρόσβαση στα στατιστικά logs του μαθησιακού συστήματος Moodle, όπου υπήρχε πλήθος μαθητών. Οι μαθητές είχαν πρόσβαση σε μαθήματα τηλεκαίδευσης (e-learning) και μέσω της ανάλυσης του log file της πρόσβασης, μπορούσαν να εξαχθούν σημαντικά συμπεράσματα για την δυναμική τους μαθησιακή συμπεριφορά.

Η βασική παράμετρος και δείκτης (KPI) που ορίστηκε ήταν ο μέσος χρόνος παραμονής κάθε μαθητή σε συγκεκριμένη εκπαιδευτική ενότητα (learning unit) όπου βάσει της ανάλυσης του φαινόταν και το επίπεδο δυσκολίας που έβρισκε στην συγκεκριμένη ενότητα. Στο υπερκυβικό μοντέλο, ο κάθε κόμβος του υπερκύβου αντιστοιχούσε σε έναν μαθητή και ο

υπερκυβικός πράκτορας ανέλυε σε πραγματικό χρόνο τον δείκτη χρονικής σχέσης, από τα data base logs του moodle.



Εικόνα 23. Διαδικασία ανάλυσης μαθησιακών δεδομένων

Οι δείκτες που ορίστηκαν φαίνονται παρακάτω:

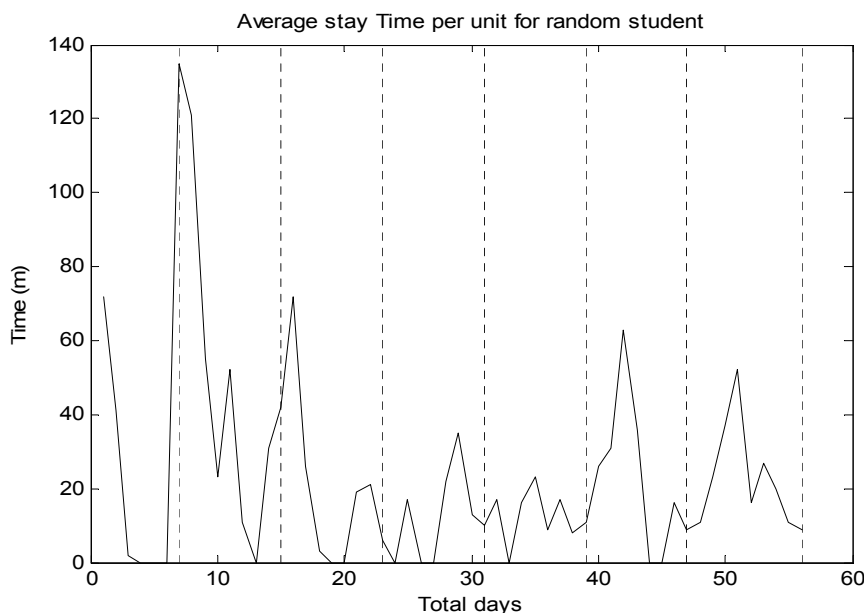
$$\bar{T}_i = \frac{T_{i1} + T_{i2} + \dots + T_{ik}}{k} = \frac{\sum_{j=1}^k T_{ij}}{k} \quad (3.4.1)$$

όπου \bar{T}_i είναι ο μέσος χρόνος ενασχόλησης ενός μαθητή i στην αντίστοιχη μαθησιακή ενότητα j . Επίσης, θέτουμε τον δείκτη B_j , ως:

$$\begin{aligned}
B_j &= \frac{(T_{1j} - \bar{T}_1) + (T_{2j} - \bar{T}_2) + \dots + (T_{nj} - \bar{T}_n)}{n} \\
&= \frac{W_{1j} + W_{2j} + W_{3j} + \dots + W_{nj}}{n} \Rightarrow \\
\Rightarrow B_j &= \frac{\sum_{i=1}^n W_{ij}}{n}
\end{aligned} \tag{3.4.2}$$

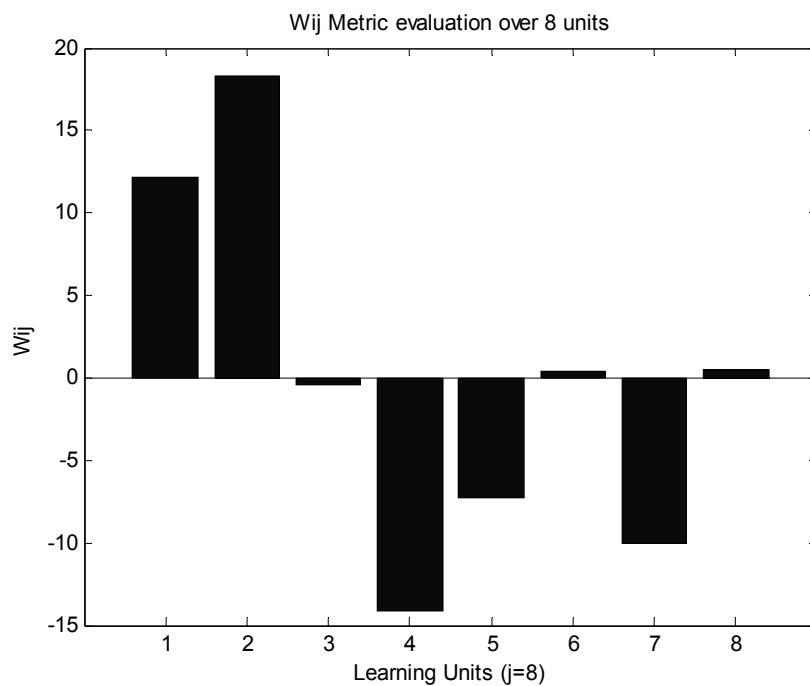
όπου χαρακτηρίζει τον Βαθμό Δυσκολίας B_j , ενός μαθητή στην συγκεκριμένη μαθησιακή ενότητα και στην ουσία είναι ένα μέτρο του μέσου όρου της στατιστικής διασποράς $W_{ij} = T_{ij} - \bar{T}_i$ του χρόνου του συγκεκριμένου μαθητή για το συγκεκριμένο μαθησιακό αντικείμενο, συγκρινόμενο με τον συνολικό μέσο όρο της τάξης.

Θέτοντας λοιπόν το υπερκυβικό μοντέλο για 21 μαθητές (χρησιμοποιήθηκε υπερκύβος με $N=5$ σε 32 διαστάσεις) και έχοντας τους πράκτορες όπου σε κάθε κόμβο (μαθητή που είχε συγκεκριμένα δεδομένα από το log file του moodle) υπολογίζανε Μ.Ο. και Διασπορές, έχουμε τα εξής αποτελέσματα:

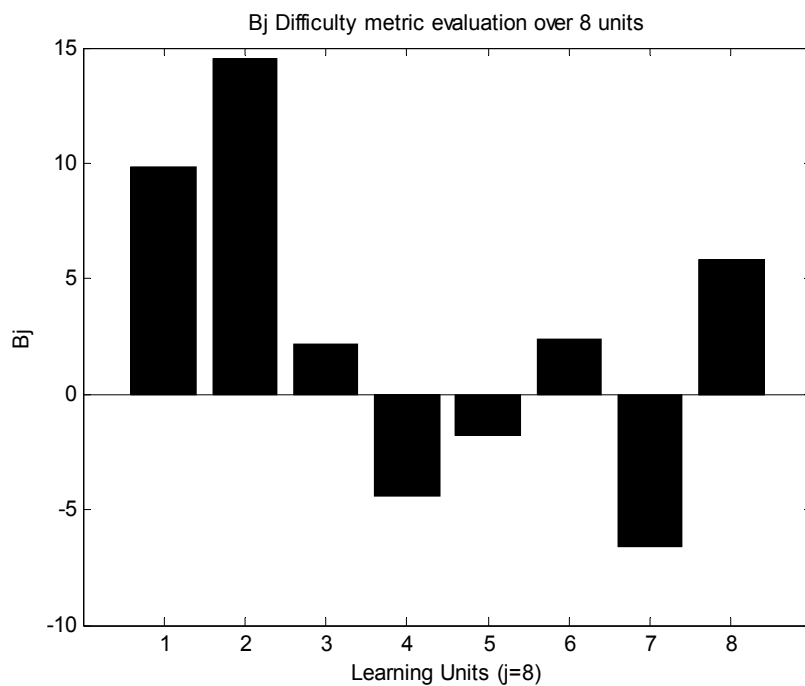


Εικόνα 24. Ανάλυση ΜΟ χρονικής παραμονής σε ενότητα από μαθητή

Παρατηρούμε ότι ο βαθμός δυσκολίας (MO) ανά μαθητή είναι μεταβλητός και δυναμικός ανάλογα με την εκπαιδευτική ενότητα και ο χρόνος παραμονής μεταβάλλεται βάσει και της ενότητας αλλά και του εκπαιδευτικού χρόνου.



Εικόνα 25. Ανάλυση MO χρονικής παραμονής σε ενότητα από μαθητή



Εικόνα 26. Ανάλυση MO χρονικής παραμονής σε ενότητα από μαθητή

Όπως βλέπουμε από τα διαγράμματα, η δυσκολία σε μέσο όρο βγαίνει από την ανάλυση και των 21 μαθητών. Ο κάθε υπερκυβικός πράκτορας ανέλυε σε πραγματικό χρόνο και σε κάθε πέρασμα από τους κόμβους, το κάθε χρονικό timestamp / κόμβο / μαθητή και υπολογιζόταν ο μέσος όρος και του χρόνου αλλά και του συντελεστή δυσκολίας.

Στον παρακάτω πίνακα φαίνονται τα αριθμητικά αποτελέσματα, για 8 εκπαιδευτικές εβδομάδες για έναν (1) τυχαίο μαθητή:

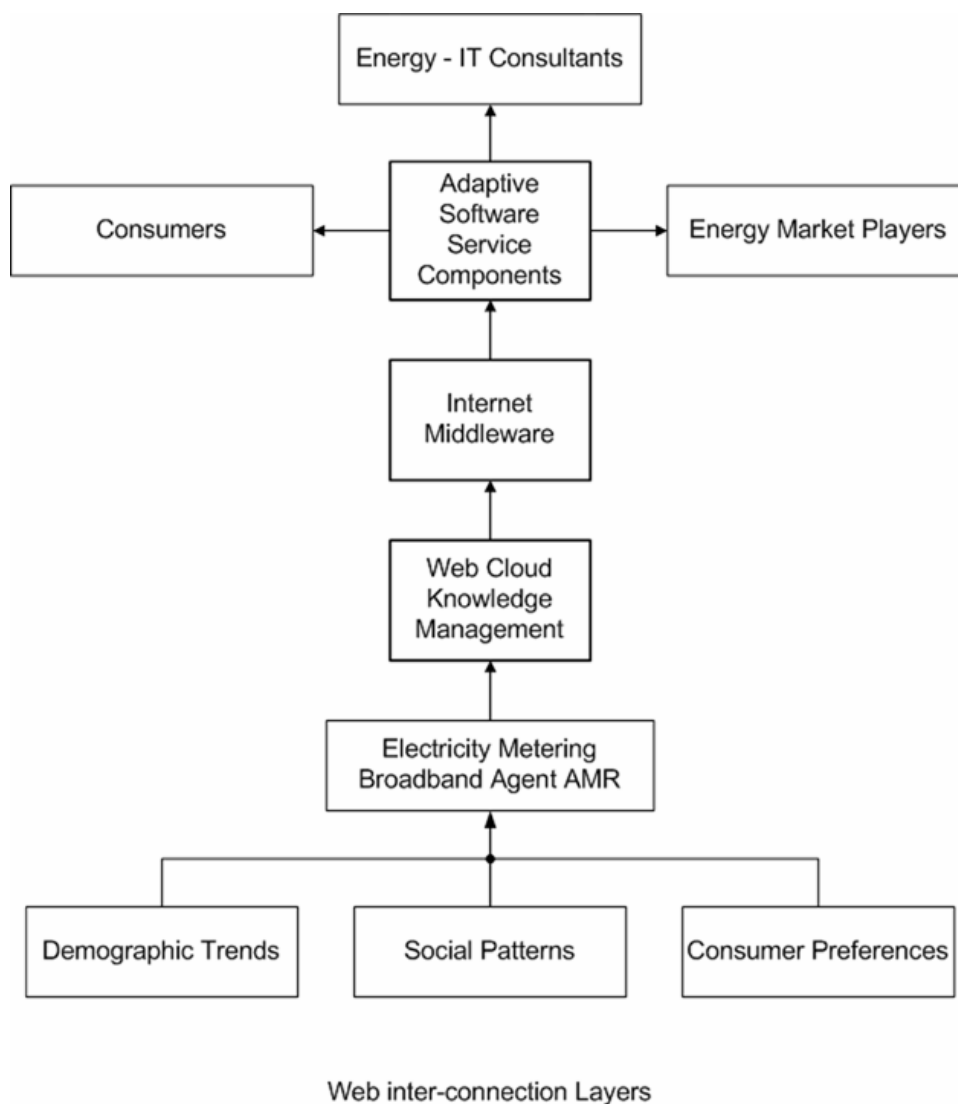
| k=8 units, n=1 (i=1) | | | | | | | | |
|-----------------------|----------------------|------|------|-------|------|------|------|------|
| | Learning Units (k=8) | | | | | | | |
| W_{ij} | 12.2 | 18.3 | -0.4 | -14.1 | -7.3 | 0.4 | -10 | 0.5 |
| \bar{T}_{ij} (m) | 35.7 | 41.8 | 23.1 | 9.4 | 16.2 | 23.9 | 13.5 | 24.5 |
| T_i (m) | 23.5 | | | | | | | |

Πίνακας 4. Ανάλυση δεικτών σε έναν συγκεκριμένο μαθητή από σμήνος πρακτόρων σε υπερκύβο

4. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΛΓΟΡΙΘΜΩΝ ΣΤΗΝ ΕΝΕΡΓΕΙΑ

4.1 Ενεργειακά Πληροφοριακά Συστήματα & Ευφυή Δίκτυα

Τα ενεργειακά πληροφοριακά συστήματα [46] που παρακολουθούν και οργανώνουν την κατανάλωση ενέργειας των πελατών και σχετικά δεδομένα, μέσω του διαδικτύου, εξελίσσονται την τελευταία δεκαετία και μπορούν να θεωρηθούν εξειδικευμένα συστήματα υποστήριξης λήψης αποφάσεων.



Εικόνα 27. Παίκτες που εμπλέκονται σε Ενεργειακά πληροφοριακά συστήματα

Οι δυνατότητες ενός τέτοιου πληροφοριακού συστήματος είναι σημαντικές στη διαχείριση ενέργειας καθώς παρέχει δυνατότητες οργάνωσης ενεργειακών δεδομένων καταναλωτών, αναγνώρισης ανωμαλιών στην κατανάλωση ενέργειας, διαχείρισης κόστους ενέργειας, και αυτοματοποιημένη στρατηγική διαχείριση ζήτησης και εστιασμένο profiling (βλ. παρακάτω) καταναλωτών [47,49,50,51]. Ο πρωταρχικός στόχος των ενεργειακών πληροφοριακών συστημάτων είναι να βοηθήσουν χειριστές εγκαταστάσεων, ιδιοκτήτες και γενικότερα άτομα που παίρνουν τις αποφάσεις, στην σωστή διαχείριση της ενέργειας. Στον τομέα της διαχείρισης κτιρίων αποτελεί ένα σημαντικό εργαλείο καθώς με τη βοήθεια του διαδικτύου καθίσταται μια έγκαιρη και έγκυρη πηγή πληροφοριών για την συμπεριφορά του κτιρίου.

Η πραγματικού χρόνου ενημέρωση των καταναλώσεων επιτρέπει στους χρήστες να αξιολογήσουν την αποδοτικότητα του κτιρίου, που υπό με συμβατικούς τρόπους (π.χ. λογαριασμούς) είναι δύσκολο να παρατηρηθεί. Έτσι με τη συνεχή ενημέρωση των χειριστών δύναται ο άμεσος σχεδιασμός και υλοποίηση της ενεργειακής στρατηγικής άμεσα ή εντός μια ημέρας. Ο χειριστής έχει τη δυνατότητα να παρατηρεί άμεσα ή σχεδόν άμεσα τις επιπτώσεις των επεμβάσεων του στην συμπεριφορά του κτιρίου. Ένα τέτοιο σύστημα έχει να προσφέρει τα μέγιστα στην διαχείριση κτιρίων, σε σχέση με το συμβατικό τρόπο πληροφόρησης των χρηστών που μπορεί ελάχιστα στοιχεία να προσφέρει στην ενημέρωση αποδοτικότητας της εγκατάστασης. Όλες οι πληροφορίες για τις καταναλώσεις που συλλέγονται από το σύστημα αποθηκεύονται σε βάση δεδομένων για να γίνει η επεξεργασία τους και η εξαγωγή χρήσιμων συμπερασμάτων (Data Mining), που αποτελεί και τον πυρήνα λήψης αποφάσεων του συστήματος.

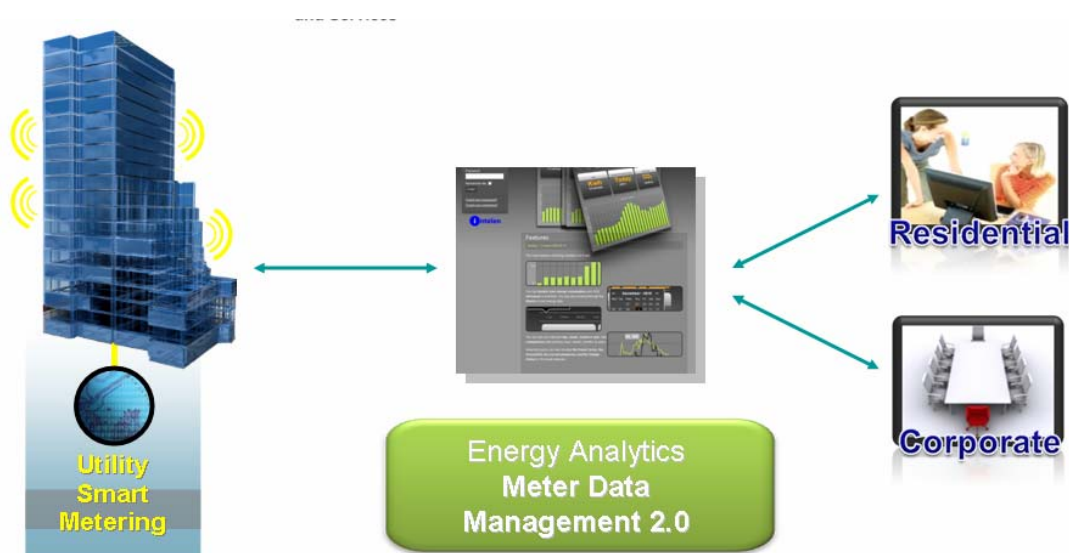
Τα πλεονεκτήματα ενός Ενεργειακού Πληροφοριακού Συστήματος μπορούν να συνοψιστούν ως εξής:

Ενεργειακό πληροφοριακό Σύστημα:

- Συνδυάζει συσσωρευμένες πληροφορίες σε λειτουργικά πληροφορικά συστήματα
- Μειώνει λειτουργικά κόστη
- Μεγιστοποιεί την αποδοτικότητα ελέγχου και λειτουργίας
- Ενεργειακά Προφίλ/Μέτρηση πραγματικού χρόνου
- Προηγμένοι μετρητές δημιουργούν μια ακριβή εικόνα της κατανάλωσης, γνωστή και ο προφίλ φορτίου
- Εγκατάσταση έξυπνων μετρητών σε επιμέρους καταναλώσεις για άμεση πληροφόρηση της χρήσης τους

- Ενημέρωση για την κατανάλωση σε πραγματικό χρόνο(τάξη λεπτών)
- Καταγραφή της χρήσης (πότε , πόσο) και άμεση κοστολόγηση της.

Εκτός από πληροφορίες για τις καταναλώσεις των πελατών, ένα ενεργειακό πληροφοριακό σύστημα μπορεί να διαθέτει και πληροφορίες για το κέλυφος του κτιρίου και τα φορτία του , για τις κλιματολογικές συνθήκες, για τις μεταβολές της τιμής της ενέργειας και γενικά πληροφορίες για προγράμματα διαχείρισης ζήτησης και εξοικονόμησης ενέργειας. Ο συνδυασμός των παραπάνω πληροφοριών δίνει στο σύστημα δυνατότητες για συσχετισμούς και ορθή λήψη αποφάσεων. Η συλλογή των ενεργειακών δεδομένων σε πραγματικό χρόνο, γίνεται από τους έξυπνους μετρητές (smart meters) χρησιμοποιώντας μία ειδική υποδομή, το AMR (Automatic Meter Readings)



Εικόνα 28. Ενεργειακό Πληροφοριακό Σύστημα διαχείρισης Μετρήσεων AMR (MDM)

Η συσκευή του έξυπνου μετρητή, μετράει την ενέργεια που χρησιμοποιείται και στέλνει τις πληροφορίες στο σύστημα και από κει καταλήγουν στον πελάτη, ενημερώνοντας τον για την εκάστοτε κατανάλωση του και το αντίστοιχο κόστος αυτής. Οι έξυπνοι μετρητές έχουν τη δυνατότητα αμφίδρομης επικοινωνίας, συνήθως, δυνατότητα δηλαδή εκτός από την αποστολή δεδομένων, και την λήψη εντολών. Αποτελούν ένα οικονομικό τρόπο για μέτρηση και παρακολούθηση της κατανάλωσης, που επιτρέπει στην καλύτερη ρύθμιση της παραγωγής βασισόμενη σε ημερήσια δεδομένα πραγματικού χρόνου (εξοικονόμηση ενέργειας και χρημάτων-μικρότερες επενδύσεις σε δίκτυα διανομής).

Σύμφωνα με την Κοινοτική Οδηγία, 32006L0032, 2006/32/EK, κεφάλαιο 3, άρθρο 13, παράγραφος 2 και 3 με θέμα «Μετρητές και αναλυτικοί λογαριασμοί για την κατανάλωση ενέργειας», τα κράτη μέλη πρέπει να εξασφαλίζουν ότι, κατά περίπτωση, η χρέωση που πραγματοποιείται από τους διανομείς ενέργειας, τους διαχειριστές συστημάτων διανομής και τις εταιρείες λιανικής πώλησης ενέργειας, πρέπει να βασίζεται στην πραγματική ενεργειακή κατανάλωση, και να παρουσιάζεται με σαφή και κατανοητό τρόπο. Στο λογαριασμό του τελικού καταναλωτή πρέπει να υπάρχουν κατάλληλες πληροφορίες, ώστε να έχει πλήρη εικόνα του τρέχοντος ενεργειακού του κόστους. Η χρέωση με βάση την πραγματική κατανάλωση ενέργειας πρέπει να είναι αρκετά συχνή, ώστε οι καταναλωτές να μπορούν να ρυθμίζουν την ενεργειακή τους κατανάλωση. Επιπλέον, τα κράτη μέλη πρέπει να εξασφαλίζουν ότι, ανάλογα με την περίπτωση, οι διανομείς ενέργειας, οι διαχειριστές συστημάτων διανομής, ή οι εταιρείες λιανικής πώλησης ενέργειας πρέπει να παρέχουν στους τελικούς καταναλωτές, ως μέρος ή μαζί με τους λογαριασμούς τους, τις συμβάσεις τους, τις συναλλαγές τους ή/και τις αποδείξεις των σταθμών διανομής τους, τις ακόλουθες πληροφορίες κατά σαφή και κατανοητό τρόπο:

- Τις τρέχουσες πραγματικές τιμές και την πραγματική κατανάλωση ενέργειας.
- Συγκρίσεις της τρέχουσας κατανάλωσης του τελικού καταναλωτή προς την κατανάλωσή του κατά την ίδια περίοδο του προηγούμενου έτους, κατά προτίμηση υπό μορφή διαγράμματος.
- Συγκρίσεις με κάποιο μέσο κανονικό ή υποδειγματικό χρήστη ενέργειας της ίδιας κατηγορίας, εφόσον τούτο είναι εφικτό και χρήσιμο.
- Διευθύνσεις κ.λπ. οργάνωσεων καταναλωτών, οργανισμών ενέργειας ή παρόμοιων οργάνων, μαζί με διευθύνσεις ιστοσελίδων, από τις οποίες μπορούν να λαμβάνονται πληροφορίες για τα διαθέσιμα μέτρα βελτίωσης της ενεργειακής απόδοσης, συγκρίσεις των διαφόρων κατηγοριών τελικών χρηστών ή/και αντικειμενικές τεχνικές προδιαγραφές για εξοπλισμό που χρησιμοποιεί ενέργεια.

Οι Έξυπνοι μετρητές θα έχουν την δυνατότητα να μετρούν άμεσα την κατανάλωση ηλεκτρικής ισχύος και να μεταδίδουν τις μετρήσεις στις βάσεις δεδομένων στο κέντρο διαχείρισης (Energy Information System). Ο καταναλωτής μπορεί οποιαδήποτε στιγμή να έχει γνώση της πραγματικής κατανάλωσης ηλεκτρικής ενέργειας. Σε συνθήκες απελευθερωμένης αγοράς, οι εταιρίες ηλεκτρικής ενέργειας θα έχουν τη δυνατότητα να επικοινωνούν με τους καταναλωτές μέσω μηνυμάτων πάνω στον Έξυπνο Μετρητή και να προσφέρουν μειωμένες χρεώσεις κιλοβατώρας ή να κάνουν προσφορές ώστε να καταρτίσουν ειδικά προγράμματα χρέωσης με βάση τις ώρες κατανάλωσης της ηλεκτρικής ενέργειας

(Demand Response). Η αύξηση της τιμής της κιλοβατώρας σε περιόδους αιχμής είναι μια μέθοδος που μπορεί να μειώσει την αντίστοιχη ζήτηση με αποτέλεσμα τεράστιο όφελος τόσο για τον παραγωγό όσο και την γενικότερη πολιτική εξοικονόμησης.

Με την αυτόματη αναγνώριση μετρητή, ο διαχειριστής θα είναι σε θέση να γνωρίζει σε πραγματικό χρόνο την κατανάλωση ενέργειας κάθε οικίας, επιχείρησης, βιομηχανίας κτλ., γεγονός που αποτελεί τεράστιο όφελος και εξοικονόμηση οικονομικών και ανθρωπίνων πόρων, αφού μεγάλος αριθμός υπαλλήλων της ΔΕΗ απασχολείται για τη μέτρηση της ηλεκτρικής ενέργειας στους μετρητές ή υποθέτει ένα ποσό κατανάλωσης με βάση στατιστικά δεδομένα και σε επόμενους λογαριασμούς διορθώνει τις αποκλίσεις από την πραγματικότητα. Αυτό το γεγονός μειώνει την αξιοπιστία του παρόχου και προβληματίζει τους πελάτες ως προς το ύψος των λογαριασμών τους.

Επιβάλλεται συνεπώς η εγκατάσταση και η χρήση συστημάτων έξυπνων μετρητών αν θέλουμε να εξασφαλίσουμε καλύτερη διαχείριση της ενέργειας, αλλά και αν θέλουμε να υπακούσουμε στο Κοινοτικό Δίκαιο .

Εκτός από την αυτοματοποίηση της διαδικασίας μέτρησης και υπολογισμού της καταναλισκόμενης ενέργειας το AMR σύστημα παρέχει ένα σύνολο ολοκληρωμένων υπηρεσιών. Κατ' αρχήν μπορεί να απεικονίσει την κατανάλωση ενέργειας σε πραγματικό χρόνο (real – time) καθώς οι μετρήσεις λαμβάνονται σε τακτά χρονικά διαστήματα. Έτσι μπορεί ο πελάτης να ξέρει ακριβώς τι καταναλώνει και τι πληρώνει και επιπλέον μπορεί να δημιουργηθεί ένα ενεργειακό προφίλ του πελάτη (κτιρίου). Το προφίλ αυτό αποτελεί ένα πολύ σημαντικό πιστοποιητικό που του δίνει αγοραστική δύναμη απέναντι σε μια απελευθερωμένη αγορά ενέργειας. Το προφίλ αυτό δείχνει τι καταναλώνει ο πελάτης και ποια χρονική στιγμή, συνεπώς αυτό μπορεί να βοηθήσει σημαντικά στην εξοικονόμηση ενέργειας και χρημάτων, τόσο με εντοπισμό «άχρηστων» φορτίων όσο και από αποφυγή ποινών λόγω υψηλών αιχμών στην κατανάλωση. Επιπλέον μπορεί να βοηθήσει σημαντικά στην ορθή πρόβλεψη φορτίου από την ρυθμιστική αρχή ενέργειας και την αποδοτικότερη ένταξη μονάδων παραγωγής.

Το αυτοματοποιημένο αυτό σύστημα μπορεί να προσφέρει ακόμα δυνατότητες χειρισμού φορτίου, ανίχνευσης σφαλμάτων στο δίκτυο και έγκαιρης ενημέρωσης του συστήματος αλλά και αξιοπιστία στις μετρήσεις. Το AMR είναι ένα σύστημα αυτοματισμού που συλλέγει δεδομένα (μετρήσεις-καταναλώσεις) και τα στέλνει σε μια κεντρική βάση δεδομένων όπου γίνεται η αποθήκευση και η επεξεργασία αυτών των στοιχείων. Η επικοινωνία γίνεται μέσω τηλεπικοινωνιακού διαύλου- ενσύρματου ή ασύρματου- ή μέσω της γραμμής μεταφοράς με

φέροντα κύματα και πραγματοποιείται είτε με μονομερή αποστολή δεδομένων από το σύστημα στο διακομιστή σε τακτά χρονικά διαστήματα, είτε με αποστολή κατόπιν αίτησης του διακομιστή είτε με συνδυασμό των δύο παραπάνω.

Demand Response (Διαχείριση Ζήτησης)

Η Διαχείριση Ζήτησης (Demand side Management - DSM) αφορά τη δράση επηρεασμού της ζήτησης της ενέργειας είτε για να την μειώσει είτε για να την αυξήσει μετατοπίζοντας μεταξύ περιόδων χαμηλής – υψηλής ζήτησης, είτε την ίδια τη διαχείριση των φορτίων από τον παραγωγό ώστε να μειωθεί το κόστος λειτουργία της μονάδας παραγωγής και συνεπώς της μονάδας ενέργειας. Με άλλα λόγια αφορά τη διαχείριση των φορτίων των καταναλωτών ώστε να χρησιμοποιείται όσο πιο αποδοτικά γίνεται η ενέργεια. Η βέλτιστη Διαχείριση Ζήτησης μπορεί να επιτευχθεί αξιοποιώντας συστήματα και τεχνολογίες εξοικονόμηση, με σωστή εποπτεία των εγκαταστάσεων, με διαφορετική τιμολόγηση της KW ανάλογα με την ώρα της ημέρας και με χρήση ηπίων μορφών ενέργειας . Η έννοια αυτή δεν είναι ταυτόσημη με την έννοια της Εξοικονόμησης Ενέργειας, αν και στενά συνδεδεμένη. Η Διαχείριση Ζήτησης αφορά κινήσεις εξοικονόμησης ενέργειας που γίνονται σε καταναλωτές στοχεύοντας στην αλλαγή του προφίλ κατανάλωσης και δεν αφορούν γενικότερα την πολιτική για χρήση συσκευών υψηλής ενεργειακής αποδοτικότητας.

Η Διαχείριση Ζήτησης αφορά την βέλτιστη διαχείριση των φορτίων, ώστε η καμπύλη ζήτησης να είναι όσο το δυνατό ομαλότερη. Αυτό αυτόματα σημαίνει λιγότερο φορτισμένες γεννήτριες και δίκτυα(μικρότερες απώλειες) και συνεπώς εξοικονόμηση ενέργειας και χρημάτων. Το κόστος παραγωγής της ενέργειας δεν αντικατοπτρίζεται άμεσα στον καταναλωτή στις περισσότερες οικονομίες, όμως στην πραγματικότητα όσο αυξάνεται η ανάγκη για ενέργεια τόσο αυξάνεται και το κόστος της μονάδας ενέργειας. Σε υψηλές φορτίσεις του συστήματος εντάσσονται οι λιγότερο αποδοτικές μονάδες για την κάλυψη των αιχμών, με υψηλά λειτουργικά κόστη. Οι μονάδες αυτές, που καλούνται στρεφόμενη εφεδρεία, εντάσσονται για την κάλυψη των επιπλέον αναγκών σε ενέργεια και αυξάνουν δραματικά το κόστος παραγωγής. Συνεπώς ο περιορισμός των αιχμών μειώνει άμεσα και την ανάγκη ένταξη τέτοιων μονάδων. Αν οι καταναλωτές πλήρωναν το πραγματικό κόστος της ενέργειας που καταναλώνουν ανά πάσα στιγμή και όχι μια σταθερή τιμή, τότε οι μεταβολές του κόστους θα αντικατοπτρίζονταν και στη ζήτηση ενέργειας. Αγνοώντας το πραγματικό κόστος της ενέργειας σε ώρες αιχμής οι καταναλωτές δεν έχουν κίνητρο μείωσης της ενέργειας που καταναλώνουν.

Η απόκριση ζήτησης (demand response - DR) αναφέρεται στους μηχανισμούς διαχείρισης της ζήτησης αποκρινόμενη στις συνθήκες προσφοράς, για παράδειγμα η μείωση της κατανάλωσης σε ώρες αιχμής ή ανάλογα με τις τιμές της αγοράς. Τα φορτία των πελατών διαχειρίζονται "έξυπνα" βάση της κοστολόγησης της kW και σε συνεννόηση με τον διαχειριστή. Στην απόκριση ζήτησης σκοπός δεν είναι η εξοικονόμηση αλλά η αποκοπή φορτίων σε κρίσιμες ώρες. Η κινητοποίηση για αποκοπή φορτίων μπορεί να έρχεται είτε από αίτημα της εταιρίας παραγωγής ενέργειας είτε σε απόκριση τιμών της αγοράς. Τα φορτία (φωτισμός, κλιματισμός, μηχανήματα) κόβονται ανάλογα με τη σειρά προτεραιότητας που έχει καθοριστεί. Εναλλακτικό της αποκοπής φορτίων είναι η λειτουργία μονάδων παραγωγής ενέργειας (π.χ. φωτοβολταϊκά, ανεμογεννήτριες, συμπαραγωγή) στην εγκατάσταση, τις κρίσιμες ώρες αιχμής. Ακόμα, δεν αποκλείεται να ζητηθεί η αύξηση της κατανάλωσης από κάποιους πελάτες στις ώρες που η παραγωγή είναι υψηλή αλλά η ζήτηση χαμηλή. Δηλαδή ουσιαστικά έχουμε μια μετατόπιση των αιχμών ζήτησης και υπέρθεση τους ώστε να προκύψει όσο το δυνατόν πιο ομαλή και συμφέρουσα από πλευρά κόστους παραγωγής καμπύλη ζήτησης. Ο όρος συνεπώς αναφέρεται στο σχεδιασμό, υλοποίηση και παρακολούθηση ενεργειών που στοχεύουν στην κινητοποίηση των πελατών να τροποποιήσουν τη ζήτηση τους όσον αφορά τον χρονισμό. Η μείωση αυτή του επιπέδου των αιχμών εξοικονομεί τα υψηλά λειτουργικά κόστη της στρεφόμενης εφεδρείας. Όπως αναφέρει και το Υπουργείο Ενέργειας των ΗΠΑ (DoE): «αφορά την μείωση της ζήτησης σε ώρες αιχμής, ή ακόμα όταν η απειλείται η ευστάθεια του συστήματος».

Οι κινήσεις αυτές αφορούν μόνο κάποιες ώρες της ημέρας όπου η ζήτηση ενέργειας είναι υψηλή ή όταν η εφεδρεία είναι σε χαμηλά επίπεδα. Η κινήσεις αυτές είναι σημαντικές για το διαχειριστή του δικτύου διότι γνωστοποιεί άμεσα της ανάγκες του συστήματος, ενώ είναι ακόμα σημαντικό για το παραγωγό ώστε να μπορεί να καλύψει τη ζήτηση αλλά και για μεγάλους καταναλωτές (π.χ. βιομηχανίες) ώστε να μπορούν να διαχειριστούν καλύτερα τα κόστη λειτουργίας τους. Εάν οι αλλαγές στις τιμές της ενέργειας διαρκέσουν για μεγάλο χρονικό διάστημα ή αναμένεται να γίνει αυτό, τότε μπορεί να προκύψει μια μακρόχρονη στρατηγική μείωσης της κατανάλωσης μέσω της επένδυσης στην ενεργειακή αποδοτικότητα ή μια αλλαγή στη συμπεριφορά του πελάτη.

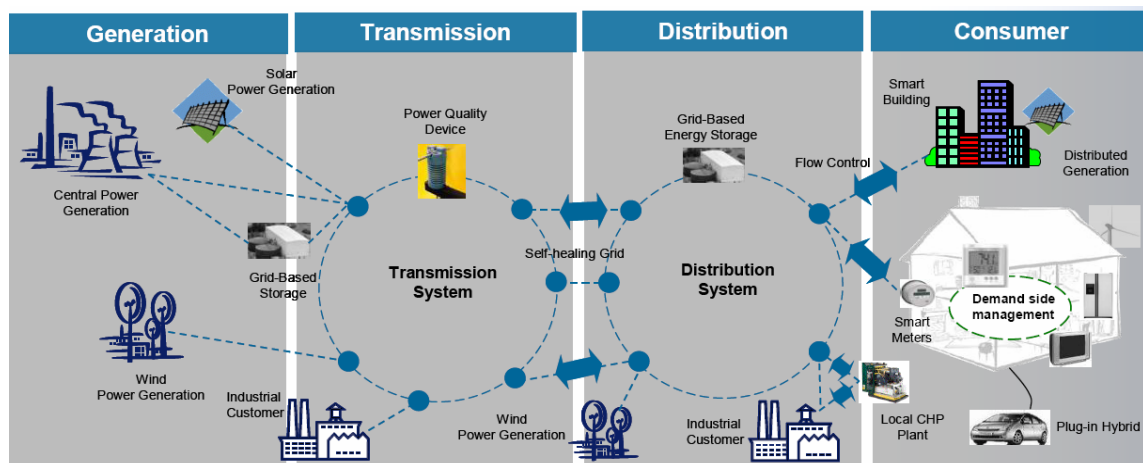
Ευφυή Δίκτυα

Το Ευφυές Δίκτυο (Smart Grid) είναι μια αναβάθμιση του ηλεκτρικού δικτύου, που χρησιμοποιεί προηγμένες τεχνολογίες επικοινωνιών, αυτοματοποιημένου ελέγχου, αυτοματοποιημένες συσκευές μέτρησης και γενικότερα αξιοποιεί την τεχνολογία της

πληροφορίας. Αυτή η ιδέα συνδυάζει βασική υποδομή του ενεργειακού συστήματος, την πληροφορία και τους κανόνες της αγοράς (τιμολογιακή πολιτική) σε μια ολοκληρωμένη διαδικασία με σκοπό την καλύτερη παροχή, έλεγχο και γενικότερα διαχείριση της ενέργειας. Ένα ευφύες δίκτυο επιτρέπει στις συσκευές όλων των επιπέδων να επικοινωνούν με το σύστημα και να έχουν πρόσβαση σε πληροφορίες σε πραγματικό χρόνο ώστε να μπορούν να λειτουργούν όσο το δυνατόν πιο αποδοτικά (συστήματα AMR).

Με τη χρήση έξυπνων συσκευών οι καταναλωτές έχουν τη δυνατότητα να ελέγχουν το φορτίο τους και να εξοικονομούν ενέργεια (πολιτικές DSM). Επιπλέον προηγμένες επικοινωνιακές ικανότητες επιτρέπουν την άμεση ενημέρωση για την τιμολόγηση της ενέργειας, για τα κίνητρα μείωσης ζήτησης και για σήματα άμεσης διακοπής φορτίων (πολιτικές DR).

Επειδή η ζήτηση δεν είναι σταθερή αλλά έχει διακυμάνσεις απαιτούνται οι στρεφόμενες εφεδρείες για να καλύψουν την επιπλέον ζήτηση, όταν χρειαστεί. Αυτός ο τρόπος διαχείρισης έχει υψηλό κόστος, τόσο γιατί το 10% της ενέργειας που είναι διαθέσιμη μπορεί να χρησιμοποιηθεί μόνο για το 1% του χρόνου στο οποίο είναι διαθέσιμο, όσο διότι οι διακοπές παροχής ρεύματος και τα σφάλματα είναι ζημιογόνα για τους χρήστες.



Εικόνα 29. Τμηματοποίηση Ευφύων Δικτύων και υπο-συστημάτων υπηρεσιών

Το ευφύες δίκτυο προσφέρει αλληλεπίδραση μεταξύ φορτίου και παραγωγής σε πραγματικό χρόνο, που επιτρέπει τον καλύτερο υπολογισμό του ισοζυγίου και επιτρέπει στους χειριστές να ανιχνεύουν σφάλματα και να βρίσκουν ταχύτατα εναλλακτική διαδρομή για τη ροή της ενέργειας παρακάμπτοντας το σφάλμα, έτσι αυξάνεται η αξιοπιστία. Επίσης αλλάζοντας ο μηχανισμός κοστολόγησης (υψηλές τιμές ενέργειας τις ώρες αιχμής και χαμηλότερες τις

υπόλοιπες ώρες) γίνεται μετατόπιση φορτίων και μειώνονται οι ανάγκες για εφεδρεία. Θα μπορούσε μάλιστα σε ένα τέτοιο σύστημα η τιμή να μεταβάλλεται συνεχώς ανάλογα με τη ζήτηση. Ενθαρρύνεται επίσης και η χρήση "πράσινης ενέργειας" που μπορεί εύκολα να ενσωματωθεί σε ένα τέτοιο σύστημα καθώς κάθε καταναλωτής μπορεί να γίνει και παραγωγός χρησιμοποιώντας τεχνολογίες ΑΠΕ (φωτοβολταϊκά, ανεμογεννήτριες, μικρά υδροηλεκτρικά, κυψέλες υδρογόνου, συμπαραγωγή) και να προσφέρει την περίσσεια ενέργεια στο δίκτυο ή απλά να καλύπτει μέρος της ζήτησης του. Χρησιμοποιώντας ένα έξυπνο μετρητή ο πελάτης μπορεί άμεσα να γνωρίζει τη ισχύ απορροφά ή προσφέρει στο δίκτυο. Συνοψίζοντας, μπορούμε να πούμε ότι το Ευφύες Δίκτυο υποβοηθούμενο από τεχνολογίες Smart metering, Energy Storage και πολιτικών DR/DSM αλλά και on-line υπηρεσιών πληροφορικής (Energy Information Systems) αποτελεί ένα βέλτιστο σύστημα διαχείρισης της ενέργειας.

Ένα ευφύες δίκτυο δίνει δυνατότητες :

- Ευφυής” συνύπαρξης της κεντρικής και διεσπαρμένης παραγωγής με αποτέλεσμα την μείωση της χρήσης άνθρακα και αποδοτικού χειρισμού της ζήτησης
- Εμπορία ενέργειας και βελτιστοποίηση κόστους μέσω χρονομεταβλητών τιμολογίων και διαφόρων κινήτρων εξαρτώμενων από το μεταβαλλόμενο φορτίο
- Ενεργός συμμετοχή του πελάτη με βάση την επικοινωνία σε δύο κατευθύνσεις και μεγάλη ροή πληροφορίας

Ένα ευφύες δίκτυο προσφέρει :

- Αυξημένη αξιοπιστία
- Αποκεντρωμένη παραγωγή
- Ελαστικότητα στη ζήτηση ενέργειας με τη χρήση ΑΠΕ
- Εξοικονόμηση Ενέργειας - Μείωση Απωλειών

Οι πληροφορίες λοιπόν που βρίσκονται στο ενεργειακό πληροφοριακό σύστημα και ανάλογα με τις υπηρεσίες (AMR, DR, Smart grid management) είναι πάρα πολλές (petabytes) μπορούν με μια έξυπνη εξόρυξη δεδομένων να συνθέσουν το ενεργειακό προφίλ ενός καταναλωτή. Το προφίλ αυτό μας δείχνει τι καταναλώνει ο χρήστης (κτίριο) ανά πάσα στιγμή, αποτελεί την «ενεργειακή ταυτότητα» του, και είναι η πηγή πληροφοριών για τα χαρακτηριστικά του χρήστη, τις συνήθειες του και γενικότερα τη συμπεριφορά του σαν καταναλωτή. Από το προφίλ αυτό μπορούμε να βρούμε αιχμές στην κατανάλωση, μέσες καταναλώσεις και γενικότερα σημαντικές πληροφορίες που μπορούν να συντελέσουν στην ορθότερη διαχείριση της ενέργειας.

Για παράδειγμα ο εντοπισμός ετεροχρονισμένων αιχμών στην κατανάλωση διαφόρων χρηστών, και υπέρθεσή τους οδηγεί σε ομαλότερες καμπύλες ζήτησης ενέργειας. Επιπλέον τα προφίλ αυτά μπορούν να ιεραρχικοποιηθούν και να συγκριθούν με πρότυπα προφίλ ορθής ενεργειακής κατανάλωσης ώστε να γίνει εντοπισμός ενεργοβόρων καταναλωτών ή ακόμα και συσταδοποίηση προφίλ ανάλογα με την κατανάλωση ή την περιοχή ώστε να μελετηθεί η επίδραση κατασκευαστικών χαρακτηριστικών του κτιρίου ή η επίδραση του μικροκλίματος μια περιοχής στην κατανάλωση ενέργειας.

4.2 Ενεργειακή Πληροφορική και Ενεργειακή μοντελοποίηση

Το αρχικό ενεργειακό μοντέλο βασίστηκε στην μαθηματική ανάλυση κοστολόγησης του *E. Λεκατσά* η οποία παρουσιάζει και αναλύει διεξοδικά όλα τα σύνθετα σημεία υπολογισμού και κοστολόγησης ηλεκτρικής ενέργειας στην ελεύθερη αγορά. Το μαθηματικό μοντέλο αναλύθηκε σε βάθος και με βάση τις περιγραφικές ολοκληρο-διαφορικές εξισώσεις που περιγράφουν τις διαδικασίες, ενεργειακές ιεραρχικές οντολογίες εξήχθησαν.

Για να επιτευχθεί ο στόχος το πρόβλημα περιγράφεται μαθηματικά με μια Αντικειμενική Συνάρτηση (Objective Function) η οποία με βάση τα ενεργειακά αποτελέσματα που θα εξάγονται από την ενεργειακή Βάση Δεδομένων και σύνθετων on-line αλγορίθμων, θα ελαχιστοποιείται συνεχώς. Η όλη διαδικασία της ελαχιστοποίησης μπορεί να μας δώσει σημαντικές πληροφορίες για την διαδικασία προμήθειας ενέργειας και βέλτιστων ενεργειακών δοσοληψιών, οι οποίες εξαρτώνται χρονικά από τις εκάστοτε ενεργειακές μετρήσεις. Η δυναμικότητα του συστήματος αντιμετωπίζεται δυναμικά από το σύστημα, με βάση την συνεχή διαδικτυακή εξόρυξη μετρήσεων και την συνεχή συσταδοποίηση των αποτελεσμάτων για εξαγωγή απόφασης (Hypercubic Data Mining). Βασική επιδίωξη του web-based συστήματος είναι η εύρεση και στοχαστική πρόβλεψη των ωρών ή ημερών όπου θα υπάρχει κέρδος για τον προμηθευτή, που σημαίνει μικρό διαφορικό κόστος εισαγωγής και η δυνατότητα μεταβλητής τιμολογιακής πολιτικής προς τον καταναλωτή (προσφορές, εκπτώσεις, added-value ενεργειακές υπηρεσίες) που θα διεξάγεται δυναμικά από το σύστημα, με βάση το μεταβλητό ενεργειακό του προφίλ. Ως συνέπεια, η μεταβλητή τιμή της Οριακής Τιμής Συστήματος (ΟΤΣ) συγκριτικά με την μεταβλητή τιμή ενός συμβολαίου και το μεταβλητό διαφορικό κόστος παραγωγής ή προμήθειας θα αποφέρει στον προμηθευτή διαφορικά κέρδη από την πώληση της περίσσειας θετικής ενέργειας στο σύστημα ή στον στοχευμένο ενεργοβόρο καταναλωτή ο οποίος εκείνη την συσχετισμένη χρονική στιγμή θα έχει σημαντικό και κερδοφόρο peak. Μια μέθοδος χρηματοδότησης επενδύσεων εξοικονόμησης ενέργειας στα κτίρια περιλαμβάνει τη χρηματοδότησή τους από

εξειδικευμένες εταιρείες παροχής ενεργειακών υπηρεσιών (Energy Service Companies) [48]. Η λειτουργία τέτοιων εταιρειών δεν έχει βρει αξιόλογες εφαρμογές ακόμη στην Ελλάδα όπως σε άλλες Ευρωπαϊκές χώρες. Οι εταιρείες αυτές χρηματοδοτούν τις επενδύσεις εξοικονόμησης ενέργειας και λαμβάνουν σαν ανταμοιβή για ορισμένα χρόνια τμήμα του οφέλους που προκύπτει από τη μείωση της ενεργειακής κατανάλωσης λόγω της πραγματοποίησης των επενδύσεων αυτών. Η Ευρωπαϊκή Ένωση με την οδηγία 2006/32/EC προάγει τη δημιουργία ESCOs στις χώρες μέλη. Οι ESCOs είναι εταιρίες που σαν στόχο έχουν να βοηθήσουν τους χρήστες ηλεκτρικής ενέργειας αλλά και φυσικού αερίου να μειώσουν το συνολικό κόστος χρήσης της ενέργειας. Αυτές οι εταιρείες υπάρχουν εδώ και τουλάχιστον 50-75 χρόνια αν και αρχικά δεν υπήρχε η συγκεκριμένη ονομασία. Βέβαια οι λειτουργίες που επιτελούσαν ήταν όμοιες με αυτές που εξυπηρετεί μία σύγχρονη ESCO. Οι εταιρείες παροχής ενεργειακών υπηρεσιών είναι εξειδικευμένες εταιρείες σε ενεργειακά θέματα με κατάλληλη τεχνογνωσία και εμπειρία. Μια τέτοια εταιρεία επενδύει για λογαριασμό του πελάτη-συνεργάτη στο χώρο του σε τεχνολογίες εξοικονόμησης ενέργειας και είναι ταυτόχρονα υπεύθυνη για τη λειτουργία τους και βεβαίως με τον τρόπο αυτό μειώνονται οι ενεργειακές δαπάνες του πελάτη.

Η αμοιβή της εταιρείας προκύπτει ακριβώς από την επιτυγχάνουσα μείωση των δαπανών ενέργειας του πελάτη. Για ένα προσυμφωνημένο χρόνο, π.χ. ορισμένα έτη, ο πελάτης πληρώνει στην εταιρεία ένα ποσό που σχετίζεται με την επιτυγχάνουσα μείωση των δαπανών του σε ενέργεια. Έτσι και η εταιρεία παροχής ενεργειακών υπηρεσιών έχει όφελος από τις επενδύσεις που πραγματοποίησε.

Η εταιρεία παροχής ενεργειακών υπηρεσιών δεν είναι απαραίτητο να έχει μεγάλη κεφαλαιουχική υποδομή αλλά μπορεί να χρησιμοποιεί τραπεζικό δανεισμό. Ταυτόχρονα ο πελάτης δε χρειάζεται να επενδύσει κεφάλαια για την εξοικονόμηση ενέργειας που είτε δεν έχει είτε θέλει να τα χρησιμοποιήσει αλλού πιο επωφελώς. Με τον τρόπο αυτό εάν κάποιος φορέας (πελάτης) για οιονδήποτε λόγο δεν μπορεί να πραγματοποιήσει για λογαριασμό του επενδύσεις εξοικονόμησης ενέργειας (π.χ. λόγω έλλειψης κεφαλαίων, λόγω έλλειψης τεχνογνωσίας κ.α.) αναθέτει στην εταιρεία παροχής ενεργειακών υπηρεσιών το έργο αυτό, η οποία και το υλοποιεί για λογαριασμό του.

Σε μια εποχή που το Διαδίκτυο έχει αναπτυχθεί ραγδαία και δημιουργούνται συνεχώς νέες, πρωτοποριακές εφαρμογές γύρω από αυτό, γεννήθηκε η σκέψη της καταμέτρησης ενεργειακών δεδομένων on-line.

Η παραπάνω παραδοχή αποδεικνύεται και μαθηματικά από την παρακάτω εξίσωση Ενεργειακής Ισορροπίας (Λεκατσάς):

$$\sum_{j=1}^n (Q_j - L_{\Delta} - L_E) * SMP + \sum_{i=1}^m (Q_i^a - L_i^a) * SMP = 0 \quad (4.2.1)$$

όπου $\left[Q_i \propto \sum_{i=1}^k W_n^k \right] \propto e_r$ που σημαίνει ότι το φορτίο είναι άμεσα εξαρτώμενο και ανάλογο της πολυδιάστατης Μήτρας W που αντιπροσωπεύει τις ειδικές τιμές και συμφωνίες (n) του συμβολαίου προμήθειας ενέργειας (k) καταναλωτών και επίσης η συνολική σχέση είναι ανάλογη της ζήτησης φορτίου που αντιπροσωπεύεται από την στοχαστική συνάρτηση $e_r(x)_{x_r}$, εξαρτώμενη από διάφορες κλιματολογικές και λοιπές τυχαίες μεταβλητές. Τα ανύσματα L_{Δ} και L_E μοντελοποιούνται από την εξίσωση (4.2.1) και αντιπροσωπεύουν την καταναλισκόμενη ενέργεια (από καταναλωτές) και το $Q_{i,j}$ αντιπροσωπεύει την παραγομένη ή εισαγόμενη ενέργεια από την προμηθευτή.

Το SMP είναι η οριακή τιμή συστήματος (ΟΤΣ). Ένα τυπικό πολυμεταβλητό ενεργειακό διάνυσμα φορτίου περιγράφεται από την εξίσωση:

$$L_D = \mu_m^L + \sum_{i=1}^j w_d^{Li} v_m^{Li} \quad (4.2.2)$$

Το διάνυσμα πρωτεύοντων μεταβλητών v_{Li} και τα συσχετισμένα βάρη w_i ενσωματώνονται έτσι ώστε να υπάρξει η βέλτιστη προσέγγιση ενός συνολικού ενεργειακού διανύσματος φορτίου. Το μ_m συμβολίζει τα ημερήσια ενεργειακά διανύσματα. Ένα μηνιαίο μονοδιάστατο διάνυσμα $[24 \times 1]$ χρησιμοποιείται για να περιγράψει την συμπεριφορά φορτίου, μετασχηματίζοντας την παραπάνω εξίσωση σε: $L_D = \mu_m^L + w_d^L v_m^L$, όπου οι μεταβλητές μ_m και v_m είναι ντετερμινιστικά διανύσματα και το w_d μία ημερήσια στοχαστική διαδικασία. Η προσέγγιση λοιπόν και η πρόβλεψη φορτίων, θα ενσωματωθεί στην αρχική εξίσωση Ενεργειακής Ισορροπίας (4.2.1). Έτσι η διαφορά $Q_i - L_i$ είναι πολύ σημαντική καθώς περιγράφει το «κέρδος» του προμηθευτή, είτε προς το σύστημα (ΟΤΣ) είτε προς τους επιλέγοντες πελάτες του, μέσω της μήτρας συμβολαίων Wn . Η διαφορά κέρδους υπολογίζεται και προβλέπεται από την υπερκυβική συσταδοποίηση των ενεργειακών προφίλ των καταναλωτών. Οι ενεργειακές καμπύλες και δείκτες ανά καταναλωτή Q συσταδοποιούνται γύρω από τα απαραίτητα κεντροειδή, ανάλογα με το ποσοστό προμήθειας της ενέργειας. Αυτή η προσέγγιση επιταχύνει πάρα πολύ τις αλγοριθμικές διαδικασίες και

ενσωματώνει τα κεντροειδή μέσα στο γενικό διάγραμμα . Έτσι, ο προμηθευτής μπορεί, ψάχνοντας μία μεγάλη χρονικά μεταβαλλόμενη ενεργειακή βάση, να βρει συσχετισμούς ενεργοβόρων καταναλωτών και την περιοδικότητά τους από τους υπερκυβικούς δείκτες. Αυτόματα, το σύστημα αναπροσαρμόζει την αντικειμενική συνάρτηση και καθορίζει τις βέλτιστες τιμές συμβολαίων Wh , από όπου και οι καταναλωτές θα εξοικονομήσουν ενέργεια αλλά και ο προμηθευτής θα έχει συγκεντρωτικό κέρδος για μικρό χρονικό διάστημα (peak prediction - correlation). Συνεπώς, μέσω της μεταβαλλόμενης μήτρας W και της στοχαστικής εκτίμησης φορτίου $\overline{e_r(x)}_{x_i}$ μπορούμε να επέμβουμε στην ελεγχόμενη ελαχιστοποίηση. Το σύστημα, με back-end processes και διαδικασίες εξόρυξης, αναλύει τα μετρούμενα ενεργειακά δεδομένα και υπολογίζει τα θετικά ισοζύγια ενέργειας, με βάση τους πελάτες και τα κόστη εισαγωγής του εκάστοτε προμηθευτή. Εν συνεχεία, με την μέθοδο της συσταδοποίησης (clustering) και επιλέγοντας συγκεκριμένο κεντροειδές (centroid) βρίσκει και περιγράφει γραφικά τις βέλτιστες ημέρες και ώρες, όπου ο προμηθευτής αναμένεται να έχει κέρδος. Βέβαια, όλα τα παραπάνω είναι δυναμικά και προσαρμόζονται στην εκάστοτε αγορά και τους κανόνες συνδιαλλαγής.

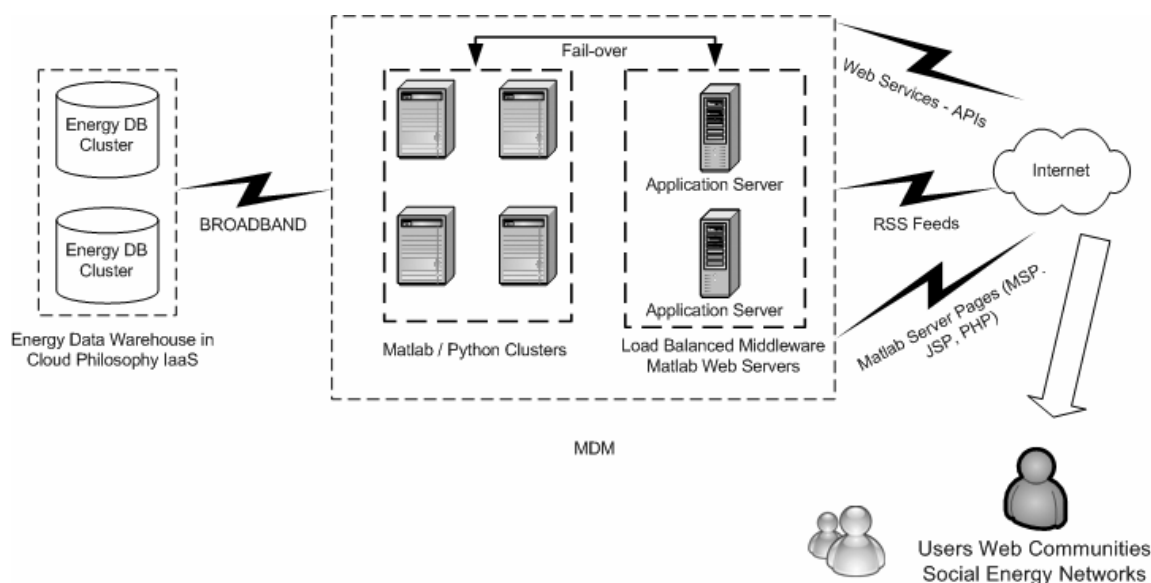
Η παραπάνω προσέγγιση σε συνδυασμό με αλγόριθμους υπερκυβικού clustering, εξάγει συγκριτική απόφαση για το αποτέλεσμα. Μέσω δυναμικών αλγορίθμων, βασισμένων στο Matlab, διάφορες τοπολογικές αποστάσεις (normed distances) μετρούνται από κεντροειδή με βάση διάφορες ενεργειακές ιδιότητες (attributes) οι οποίες βαθμολογούνται σε σχέση με μία βέλτιστη, η οποία χαρακτηρίζεται και ως κεντροειδές. Με αυτό το τρόπο μπορούν να εισαχθούν όσες ιδιότητες θεωρούνται σημαντικές για την εξαγωγή συμπεράσματος (πχ. λεπτομέρειες συμβολαίων, τιμές peaks, ώρες peaks, κλπ)

4.3 Εφαρμογή Υπερκυβικών Αλγορίθμων Γνώσης σε Ενεργειακά Δεδομένα (AMR)

Η μεθοδολογία εφαρμόστηκε σε πραγματικά ενεργειακά δεδομένα, τα οποία προέρχονται από έξυπνους μετρητές (AMR) και εισάγονται σε ένα Ενεργειακό Πληροφοριακό Σύστημα (EIS) το οποίο διαχειρίζεται αυτά τα στοιχεία σε πραγματικό χρόνο (Meter Data Management - MDM) για παροχή υπηρεσιών demand response (DR) αλλά και υπηρεσιών ενεργειακής αποδοτικότητας από εταιρίες ESCO και υπηρεσιών εμπορίας ενέργειας και ενεργειακών συμβολαίων από ένα Utility.

Πρωταρχικός σκοπός είναι η ανάλυση, βάσει του υπερκυβικού πλέγματος και των πρακτόρων, πολλών ενεργειακών δεδομένων στην μονάδα του χρόνου, συσχέτιση των

δεικτών που εξάγονται (KPIs) και χρήση της δυναμικότητάς τους για εξαγωγή συμπερασμάτων, βάσει της υπερκυβικής συσταδοποίησης και του RKM



Εικόνα 30. Τεχνική αρχιτεκτονική για το Ελληνικό Case Study

Η αυτόματη διαδικασία κατάταξης των ενεργειακών καταναλωτών χρησιμοποιεί μία παραλλαγή της Ευκλείδειας απόστασης, η οποία περιλαμβάνει, όπως αναφέρθηκε και πιο πριν, διαφορετικά βάρη για κάθε μεταβλητή ανάλογα με τη σημασία της. Η διαδικασία συσταδοποίησης ακολουθεί διαφορετικά επαναληπτικά στάδια, τον αριθμό των οποίων ορίζουμε από την αρχή έτσι ώστε να αρκούν ώστε να προκύψουν οι διαφορετικές δυναμικές ομάδες από τους υπερκυβικούς κόμβους, οι οποίοι αντιπροσωπεύουν τους καταναλωτές. Στο πρώτο στάδιο της επανάληψης θέτουμε το πλήθος των κλάσεων, ίσον με 3 (λόγω στρατηγικής) καθώς και το πλήθος των καταναλωτών που θα ανήκουν σε κάθε κλάση, χρησιμοποιώντας μία απόσταση ως κατώφλι: όσο η Ευκλείδεια απόσταση με βάρη παραμένει μικρότερη από αυτό το κατώφλι τότε ο καταναλωτής που εξετάζουμε ανήκει στη δεδομένη κλάση και το μόνο που χρειάζεται είναι να ανανεώσουμε το κέντρο βάρους της κλάσης, λαμβάνοντας υπόψη τη συνεισφορά του νέου μέλους της ομάδας. Σε διαφορετική περίπτωση ο υπό εξέταση πελάτης ορίζει νέα κλάση με κέντρο τον ίδιο. Έπειτα κάθε ανακύκλωση υπολογίζεται η Ευκλείδεια απόσταση για κάθε καταναλωτή και κάθε κλάση που έχει δημιουργηθεί ελέγχοντας ταυτόχρονα αν πρέπει όντως να μείνει στην κλάση που ήδη ανήκει ή αν πρέπει να μετακινηθεί σε κάποια άλλη κλάση εξετάζοντας τότε η απόσταση αυτή ελαχιστοποιείται. Ακολούθως ανανεώνονται τα χαρακτηριστικά κάθε κλάσης (κέντρο,

πλήθος μελών). Καθώς μετακινούνται τα μέλη αλλά και τα κεντροειδή των κλάσεων υπολογίζουμε κάποιο μέτρο που υποδηλώνει την ενέργεια του συστήματος.

Το σημαντικό πλεονέκτημα της μεθόδου είναι ότι δεν απαιτεί να προβλέψουμε εκ των προτέρων το πλήθος των απαιτούμενων κλάσεων, αλλά αυτό υπολογίζεται αυτόματα έχοντας επιλέξει κατάλληλα το κατώφλι που υποδηλώνει το εύρος της κλάσης.



Εικόνα 31. το σύστημα Energy Analytics που χρησιμοποιήθηκε για ανάλυση των μετρήσεων

Σε κάθε περίπτωση για να γίνει οποιαδήποτε σύγκριση χρειαζόμαστε αξιόπιστους δείκτες που να περιγράφουν ένα πλαίσιο αναφοράς για την ενεργειακή κατανάλωση και απόδοση, όπως αναφέρθηκε και παραπάνω. Αυτοί οι δείκτες, εφόσον έχουν επιλεγεί σωστά και υπολογισθεί από τους υπερκυβικούς agents, μας βοηθούν να ξεχωρίσουμε τις πραγματικές πληροφορίες από έναν τεράστιο όγκο ανεπεξέργαστων δεδομένων και στη συνέχεια να καθορίσουμε τους

στόχους του συστήματος και να ελέγξουμε αν αυτοί έχουν επιτευχθεί. Αυτού του είδους οι δείκτες χρησιμοποιούνται ευρέως και σε άλλους τομείς εκτός των ενεργειακών project και είναι γνωστοί ως KPIs (Key Performance Indicators). Όπως έχει ήδη περιγραφεί ένα σύγχρονο σύστημα ενεργειακής καταγραφής και διαρκούς παρακολούθησης περιλαμβάνει εξειδικευμένες μετρητικές συσκευές εγκαταστημένες στο χώρο του καταναλωτή, ένα δίκτυο διασύνδεσης για αποστολή των δεδομένων που συλλέγονται και κάποιον server που αναλαμβάνει την αρχειοθέτηση σε βάση δεδομένων αλλά και την real-time επεξεργασία. Τα ίδια ακριβώς στοιχεία συν-αποτελούν ένα οποιοδήποτε πληροφοριακό σύστημα και γι' αυτόν ακριβώς το λόγο η χρήση των ενεργειακών KPIs θεωρείται επιβεβλημένη κατ' αναλογία με τα ήδη χρησιμοποιούμενα KPIs.

Στο συγκεκριμένο παράδειγμα ο εκάστοτε υπερκυβικός πράκτορας θα υπολογίζει τον παρακάτω δείκτη:

$$KPI = \frac{P_{AV_{i,j}}}{P_{MAX_{i,j}}} \quad vs \quad E(KHw) \in [0 \dots 1] \quad (4.2.3)$$

Όπου το διάνυσμα P καθορίζεται ως:

$$P_{AV} = \frac{1}{96} \sum_{i=1}^{96} P_{DATA}^{\rightarrow} \quad (4.2.4)$$

$$P_{MAX} = \max(P_{DATA}^{\rightarrow}) \quad (4.2.5)$$

Το διάνυσμα P_{data} αφορά το άνυσμα των μετρήσεων που προέρχονται από το smart metering σύστημα ανά 15 λεπτά οπότε το σύνολο των ενεργειακών μετρήσεων ανά ημέρα είναι 96.

Εφαρμόζοντας λοιπόν των αναδρομικό υπερκυβικό αλγόριθμο RKM στις ομάδες δεδομένων και παίρνοντας την ευκλείδεια απόσταση Ed με το συγκεκριμένο βάρος (w) για μία συγκεκριμένα ημέρα και ώρα (timestamp), ο αλγόριθμος υπολογίζει αναδραστικά, από τον παραπάνω δείκτη KPI, την εντροπία, τους πληθυσμούς, τον πίνακα κεντροειδών C καθώς και τους πίνακες ομοιότητας (S , Similarity matrix) σε κάθε εφαρμογή του RKM. Το input γίνεται

από τους υπερκυβικούς πράκτορες και το clustering εφαρμόζεται μετά τον υπολογισμό ου εκάστοτε KPI.

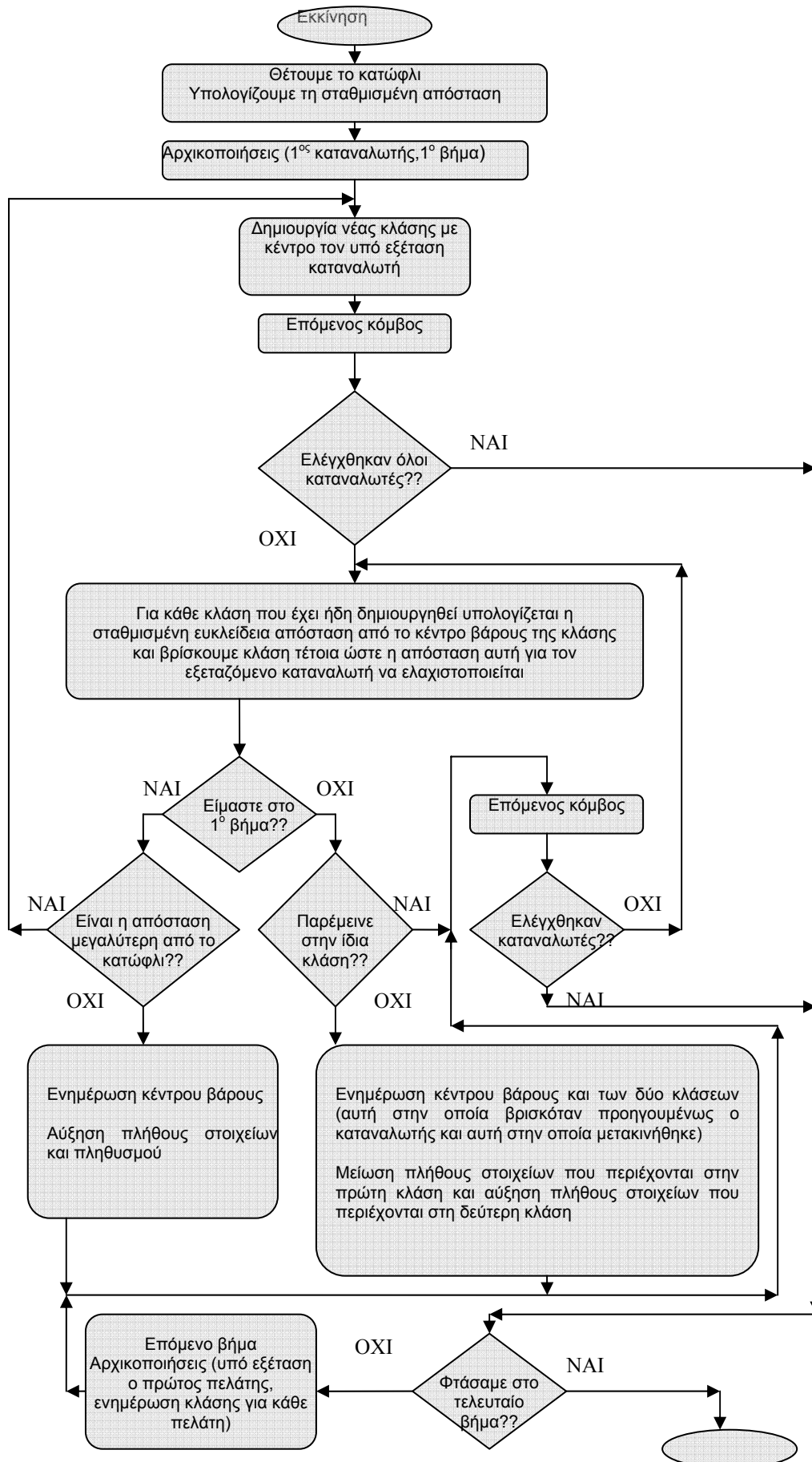
Σε κάθε περίπτωση για να γίνει οποιαδήποτε σύγκριση χρειαζόμαστε αξιόπιστους δείκτες που να περιγράφουν ένα πλαίσιο αναφοράς για την ενεργειακή κατανάλωση και απόδοση ,όπως αναφέρθηκε και παραπάνω. Αυτοί οι δείκτες ,εφόσον έχουν επιλεγεί σωστά , μας βοηθούν να ξεχωρίσουμε τις πραγματικές πληροφορίες από έναν τεράστιο όγκο ανεπεξέργαστων δεδομένων και στη συνέχεια να καθορίσουμε τους στόχους του συστήματος και να ελέγξουμε αν αυτοί έχουν επιτευχθεί.

Όπως έχει ήδη περιγραφεί ένα σύγχρονο σύστημα ενεργειακής καταγραφής και διαρκούς παρακολούθησης περιλαμβάνει εξειδικευμένες μετρητικές συσκευές εγκαταστημένες στο χώρο του καταναλωτή, ένα δίκτυο διασύνδεσης για αποστολή των δεδομένων που συλλέγονται και κάποιον server που αναλαμβάνει την αρχειοθέτηση σε βάση δεδομένων αλλά και την real-time επεξεργασία. Τα ίδια ακριβώς στοιχεία συναποτελούν ένα οποιοδήποτε πληροφοριακό σύστημα και γι' αυτόν ακριβώς το λόγο η χρήση των ενεργειακών KPIs θεωρείται επιβεβλημένη κατ' αναλογία με τα ήδη χρησιμοποιούμενα KPIs. Κάθε ημέρα (24 ώρες) περιλαμβάνει 96 τέταρτα και κάθε εβδομάδα 672 τέταρτα της ώρας. Οπότε για τον υπολογισμό των ημερήσιων δεικτών επεξεργάζονται ταυτόχρονα 96 τιμές.

Μεταξύ αυτών βρίσκουμε το μέγιστο και το ελάχιστο και με βάση το εάν πρόκειται για καθημερινή ημέρα ή για Σαββατοκύριακο τα αποτελέσματα αποθηκεύονται κατάλληλα.

Για την αποθήκευση των δεικτών χρησιμοποιούνται αλγεβρικοί πίνακες. Συγκεκριμένα για τους ημερήσιους δείκτες δημιουργήθηκε ο δισ-διάστατος πίνακας daily_ind με τη μία διάσταση να είναι 4 (όσοι και οι διαφορετικοί ημερήσιοι στατιστικοί δείκτες που υπολογίζονται από τον υπερκυβικό πράκτορα) και την άλλη διάσταση με τόσα στοιχεία όσες είναι και οι ημέρες για τις οποίες έχουμε δεδομένα.

Ο συνολικός αριθμός ημερών προκύπτει ως το αποτέλεσμα της διαίρεσης του πλήθους των δεδομένων προς 96, όσα και τα τέταρτα κάθε 24ώρου.



Ο αλγόριθμος εφαρμόστηκε σε 90 κτίρια και τις συγκεκριμένες μετρήσεις από το Ενεργειακό Πληροφοριακό σύστημα. Οι ειδικοί μετρητές χρησιμοποιούσαν το στρώμα ADSL για την μεταφορά των μετρήσεων στο υπερκυβικό cloud. Ο μετρητής μέτραγε ανά 7 δευτερόλεπτα και aggregation ανά 15 λεπτά τα παρακάτω ενεργειακά στοιχεία:

- Energy consumption (KWh)
- Power factor (cosf)
- Power Demand (KW)
- Voltage and currents (Volts and Amps)

Όπως αναφέρθηκε, κάθε μετρητής αντιπροσωπεύει έναν υπερκυβικό κόμβο και έχει ένα δυαδικό tag, βάσει της ανάλυσης που αναφέραμε για την υπερκυβική δρομολόγηση. Οι πράκτορες, ακολουθώντας τον στοχαστικό αλγόριθμο του L. Valiant, κινούνται συνεχώς μεταξύ των κόμβων, με μεγάλη πιθανότητα να μην υπάρχουν διπλοί πράκτορες στους κόμβους. (πχ. 0110 -> 0111). Με την κίνηση υπολογίζουν τα KPIs τα οποία σε κάθε αναδρομή μπαίνουν σαν input στον αλγόριθμο RKM.

Στη συνέχεια και πριν δημιουργηθεί η τελική δομή cluster , δημιουργείται η δομή cluster_members για να περιγράψει τα μέλη μίας κλάσης, που στην ουσία είναι οι κόμβοι του υπερκύβου, άρα οι καταναλωτές (μετρητές). Κάθε μέλος έχει ένα πεδίο prev_cl τύπου cluster_center, το οποίο κρατά το κέντρο της προηγούμενης κλάσης στην οποία ανήκε το μέλος , σε περίπτωση αλλαγής. Ένα πεδίο valid τύπου int καθορίζει αν το τρέχον μέλος ανήκει πράγματι στην κατηγορία αυτή κι ένα πεδίο dist τύπου float το οποίο κρατάει την απόσταση του μέλους από το κέντρο της κλάσης. Αυτό μας διευκολύνει να αναλύσουμε τις δυναμικές παραμέτρους του αναδραστικού αλγορίθμου, καθώς δεν μας ενδιαφέρει τόσο το στατικό cluster και ο στατικός πληθυσμός αλλά πως αυτά τα ποιοτικά στοιχεία μεταβάλλονται στον χρόνο κα αλλάζουν το βαθμό ομοιότητας (Similarity matrix) των κόμβων

Ένα snapshot της υπερκυβικής διαδικασίας σε 8 τυχαίους υπερκυβικούς κόμβους (πράκτορες που τελείωσαν μία αλγεβρική μετάθεση και εξήγαγαν αποτελέσματα) φαίνεται στον παρακάτω πίνακα:

| Binary Tag | Local Timestamp of the Agent | Run time | User ID | Agent ID | Agent Outputs Results (XML) |
|------------|------------------------------|----------|---------|----------|--|
| 011001 | 2009-07-09-23.14.39 | 0.25 ms | 0254 | A5469 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 011100 | 2009-07-09-23.39.17 | 0.98 ms | 2548 | A5589 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 100011 | 2009-07-09-23.58.59 | 1.25 ms | 1873 | A5422 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 000011 | 2009-07-09-23.59.11 | 0.56 ms | 5598 | A5558 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 010011 | 2009-07-09-23.59.49 | 0.68 ms | 5493 | A9852 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 111000 | 2009-07-09-23.43.52 | 1.58 ms | 5547 | A2221 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 111111 | 2009-07-09-23.43.57 | 0.36 ms | 3659 | A1259 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |
| 000010 | 2009-07-09-23.44.10 | 0.22 ms | 3579 | A9876 | <ul style="list-style-type: none"> ▪ min ▪ max ▪ mean ▪ Standard Deviation |

Πίνακας 5. Αποτελέσματα πρακτόρων σε έναν 6-διάστατο υπερκύβο για υπολογισμό KPIs

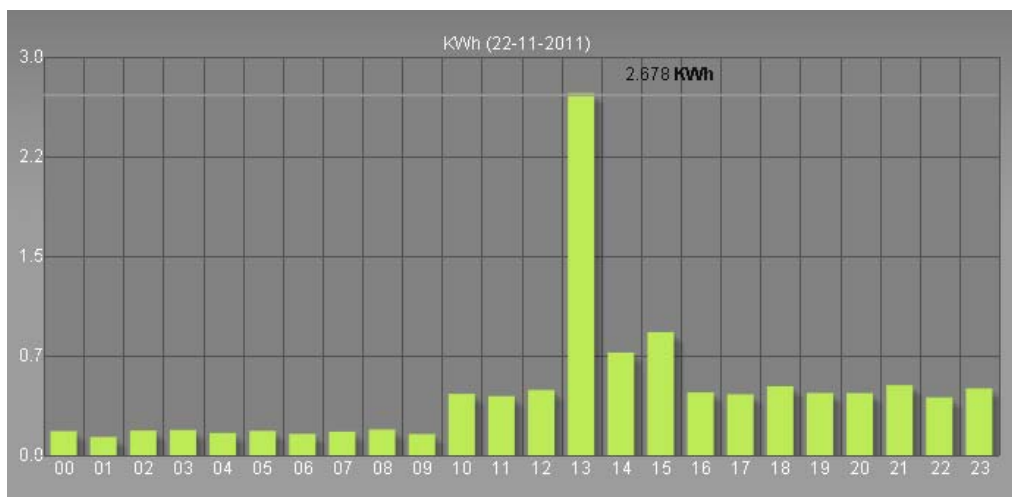
Αφήνοντας τον αλγόριθμο να τρέχει μία φορά την ημέρα, και περνώντας και από τους ενενήντα κόμβους (μετρητές άρα κτίρια), παίρνουμε κάποια συγκεντρωτικά αποτελέσματα από τους υπολογισμούς της εντροπίας e , των πληθυσμών g και των συντεταγμένων των τριών κεντροειδών που είχαμε ορίσει:

| Execution date | Run time (s) | Meter grid (no) | Cluster Populations \mathcal{G}_N | | | $C_{i,j}^N$ | | | $e_{i,j}^N$ | | |
|----------------|--------------|-----------------|-------------------------------------|-------|-------|-------------|--------------|--------------|-------------|-------|-------|
| | | | g_1 | g_2 | g_3 | $C(x,y)^1$ | $C(x,y)^2$ | $C(x,y)^3$ | e^1 | e^2 | e^3 |
| 2009-07-01 | 1.25 s | 90 | 32 | 22 | 36 | (4.2, 0.78) | (5.9, 0.94) | (9.2, 0.95) | 0.67 | 0.84 | 1.02 |
| 2009-07-02 | 1.78 s | 90 | 14 | 29 | 46 | (4.1, 0.76) | (5.9, 0.92) | (9.9, 0.94) | 0.98 | 1.85 | 3.25 |
| 2009-07-03 | 1.35 s | 90 | 21 | 18 | 51 | (5.4, 0.95) | (12.8, 0.81) | (15.1, 0.82) | 0.71 | 2.81 | 2.95 |

| | | | | | | | | | | | |
|------------|--------|----|----|----|----|-------------|--------------|--------------|------|------|------|
| 2009-07-04 | 2.10 s | 90 | 34 | 25 | 31 | (8.1, 0.99) | (11.4, 0.81) | (15.4, 0.83) | 3.10 | 2.98 | 2.15 |
| 2009-07-05 | 1.45 s | 90 | 17 | 24 | 49 | (4.9, 0.99) | (8.1, 0.80) | (12.2, 0.82) | 0.95 | 4.15 | 3.46 |
| 2009-07-06 | 1.81 s | 90 | 29 | 33 | 28 | (7.9, 0.99) | (11.8, 0.99) | (15.1, 0.99) | 1.84 | 1.75 | 1.96 |

Πίνακας 6. Αποτελέσματα του RKM για 90 κτίρια και με 6 iteration ημερήσια, N=3 clusters

Όπως μπορούμε να παρατηρήσουμε από την ανάλυση των δυναμικών δεδομένων, είναι δυνατόν να διακρίνουμε κάποιες ομοιότητες μεταξύ των παραγομένων συστάδων, εάν τα προβάλλουμε στο πεδίο του χρόνου και ανιχνεύσουμε κάποια κρυμμένα trends. Συσχετίζοντας την θέση των κεντροειδών με την ανάλυση της εντροπίας και την δυναμικότητα του πληθυσμού (g members) έχουμε μία αποδοτική εικόνα για το πως κινείται δυναμικά η κατανάλωση και στοιχεία κατανάλωσης των ενεργειακών καταναλωτών (κόμβοι υπερκύβου). Δείκτης KPI (P_{av}/P_{max}) και οι τιμές MIN & MAX υπολογίζονται αυτόματα από τον πράκτορα, μέσω Python agent script.



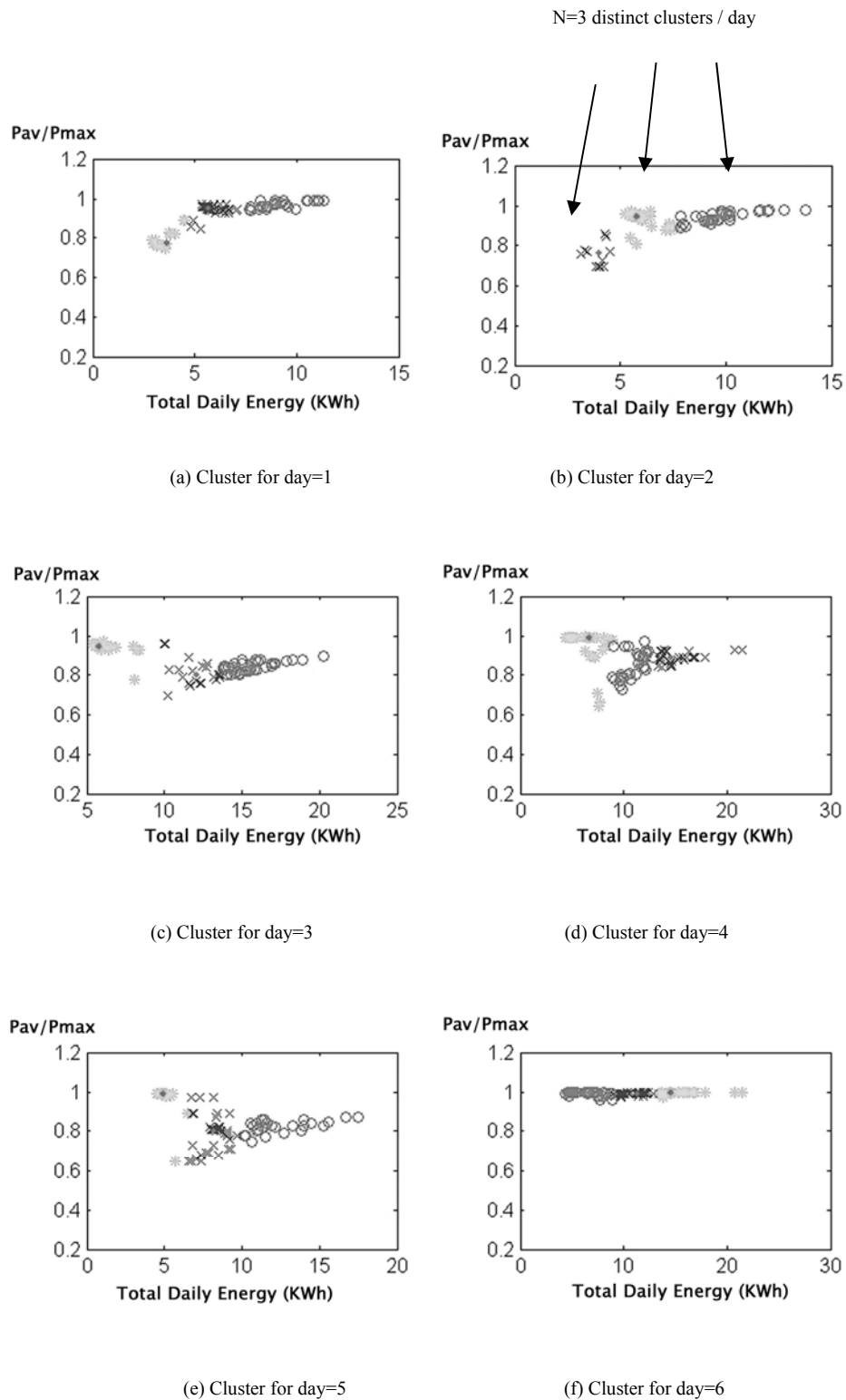
Εικόνα 32. Παράδειγμα αιχμακού φορτίου που ανιχνεύεται με τον δείκτη P_{av}/P_{max}



Εικόνα 33. Μηνιαίες περιοδικότητες που ανιχνεύονται με τον RKM σε συσχετισμένους κόμβους

Έχοντας λοιπόν σαν έξοδο τον υπολογισμό των επί μέρους δεικτών και στατιστικών ανά κόμβο, μπορούμε να παράξουμε τα πρώτα γραφήματα για τις συστάδες που δημιουργούνται, μετά από ημερήσια ανάλυση (υπολογισμός δεικτών σε φορτίο ημέρας / κόμβο) για 6 ημέρες.

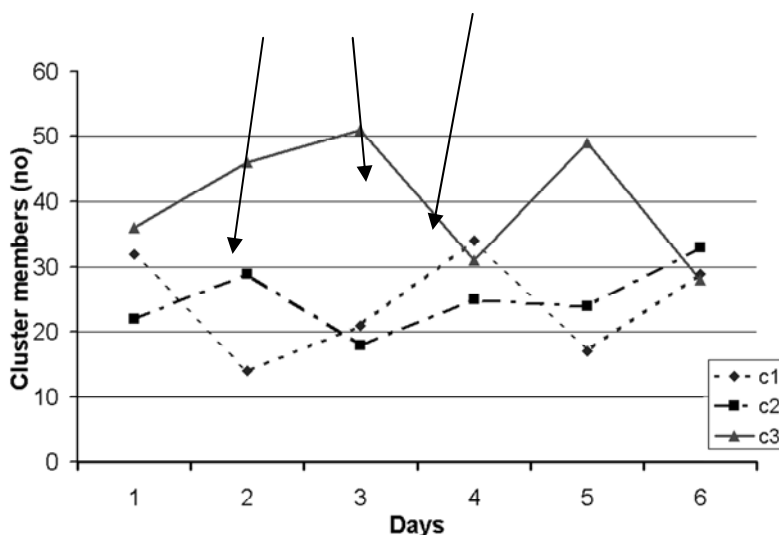
Αυτό φαίνεται παρακάτω:



Εικόνα 34. 3 δυναμικές συστάδες και πως κινούνται σε περίοδο 6 ημερών αναδρομικά

Κάθε συστάδα αλλάζει την μορφή της δυναμικά, καθώς αλλάζουν οι μέρες αλλά και οι ώρες της εφαρμογής του RKM. Μετρώντας τις μεταβλητές των κεντροειδών συντεταγμένων, την εντροπία και την εναλλαγή πληθυσμών, μπορούμε να βγάλουμε πολύ σημαντικά συμπεράσματα για την δυναμική αυτών των ποιοτικών χαρακτηριστικών και να ανιχνεύσουμε κρυμμένα trends. Έχοντας αυτή την ανάλυση και πληροφορία, μια ενεργειακή εταιρία (ESCO ή Utility) μπορεί να σχεδιάζει στρατηγικές και κοινές υπηρεσίες σε συγκεκριμένα group καταναλωτών. Οι συστάδες κινούνται συσχετισμένες στον χρόνο, αναδεικνύοντας κοινά πρότυπα κατανάλωσης. Οι υπερκρυβικοί κόμβοι / καταναλωτές που είναι μέλη ενός συγκεκριμένου cluster (1,2 ή 3) έχουν κοινό δείκτη (P_{av}/P_{max}), mins, max και κάποιες άλλες κοινές στατιστικές παραμέτρους οι οποίες αλλάζουν στο πεδίο του χρόνου. Από τις εικόνες 34 (a) και 24 (b) μπορούμε να παρατηρήσουμε ότι τα 3 κεντροειδή / συστάδες είναι πολύ κοντά η μία στην άλλη αναδεικνύοντας έναν συσχετισμό στα μέλη του, ειδικά τις δύο πρώτες ημέρες μέτρησης της ενεργειακής κατανάλωσης. Εξάγοντας τα μέλη του πληθυσμού από κάθε συστάδα, και αναλύοντας την στατιστική τους διασπορά και variance, έχουμε το γράφημα παρακάτω, που αναλύει την μεταβλητικότητα του πληθυσμού μέσα σε 6 ημέρες:

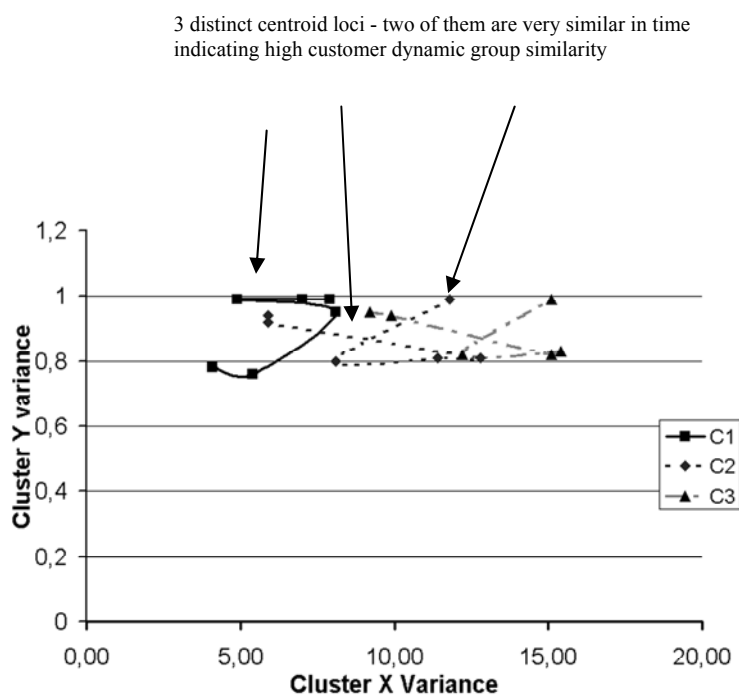
3 distinct time-variant populations over 6 days. The change in patterns among customers is rather big



Εικόνα 35. Δυναμική αλλαγή πληθυσμών P_i

Στο παραπάνω γράφημα (Εικόνα 35), η αλλαγή και μετάπτωση των πληθυσμών φαίνεται καθαρά και μπορεί να φανεί επίσης και μία κρυμμένη περιοδικότητα, στις συστάδες 1 και 2 για τις τελευταίες 4 ημέρες. Αυτό σημαίνει ότι μετά την 3^η ημέρα, η μετάβαση από ένα pattern σε άλλο pattern κατανάλωσης είναι συσχετισμένο και εξάγοντας τα μέλη αυτών των δυναμικών παραμέτρων, είναι δυνατόν η πρόβλεψη κατανάλωσης και δυναμικών αλλαγών σε ομάδες συσχετισμένων καταναλωτών. Η συστάδα 3 έχει τον υψηλότερο πληθυσμό που σημαίνει ότι η πλειοψηφία των καταναλώσεων της 3^{ης} συστάδας μπορούν να θεωρηθούν ως βασικό reference pattern (benchmarking reference). Ειδικά τις μέρες 3 και 5 έχουμε την μέγιστη πληθυσμιακή σύσταση, αναδεικνύοντας υψηλό συσχετισμό.

Συγκρίνοντας τις μεταβολές του πληθυσμού και εάν τώρα συσχετίσουμε και τις καρτεσιανές x,y θέσεις των κεντροειδών στον χρόνο, παρατηρούμε ένα πολύ ενδιαφέρον γράφημα ενός Τόπου των θέσεων των κεντροειδών, όπου αναλύεται δυναμικά η μεταβλητότητα (Variance):



Εικόνα 36. Γεωμετρικός τόπος κεντροειδών (X,Y Variance)

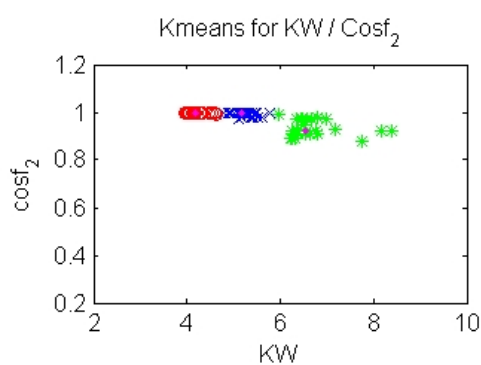
μπορούμε να παρατηρήσουμε ότι τα κεντροειδή 2 και 3 έχουν την ίδια καρτεσιανή μεταβλητότητα που σημαίνει ότι τα μέλη (υπερκυβικοί κόμβοι άρα καταναλωτές / κτίρια) που είναι μέλη αυτών των συστάδων έχουν συσχετισμένους δείκτες (*KWh*, *Pmax*, *Pav*) αλλά όχι απαραίτητα κοινό πληθυσμιακό δείκτη (*P*). Τώρα, εάν συσχετίσουμε και παραβάσουμε τα

σχήματα 35 και 36, και εξάγουμε όλα τα κοινά μέλη στο κοινό πεδίο του χρόνου, τότε αυτοί οι καταναλωτές είναι πολύ συσχετισμένοι και μπορούμε να πούμε τα εξής:

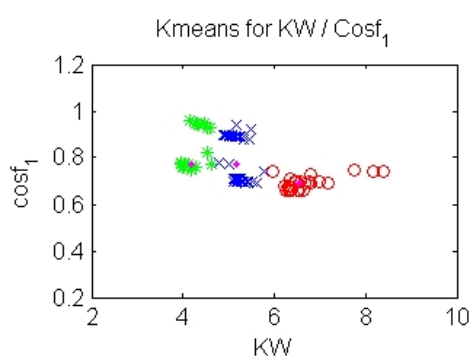
- Στις πρώτες 2 ημέρες, υπάρχει μία ομοιότητα στις καμπύλες φορτίου (Similarity matrix S)
- Τα μέλη του 3^{ου} cluster θεωρούνται σαν reference μοντέλο βάσης για συγκρίσεις με τους λοιπούς καταναλωτές
- Το 1^ο και το 3^ο cluster αλλάζουν τους πληθυσμούς τους συσχετισμένα, μετά της 3^{ης} ημέρα μέτρησης
- Τα μέλη του 2^{ου} και 3^{ου} cluster έχουν κοινούς στατιστικούς δείκτες (*KWh*, *Pmax*)
- Τα μέλη του 1^{ου} cluster έχουν έναν μέγιστο πληθυσμό την 4^η ημέρα

Εάν επαναλάβουμε τον RKM για άλλο ζεύγος δεικτών, ο οποίος υπολογίζονται από τους υπερκυβικούς πράκτορες (*Kw*, *cosf1* & *cosf2*), για ανάλυση 8 ημερών (μία εβδομάδα) έχουμε τα παρακάτω γραφήματα.

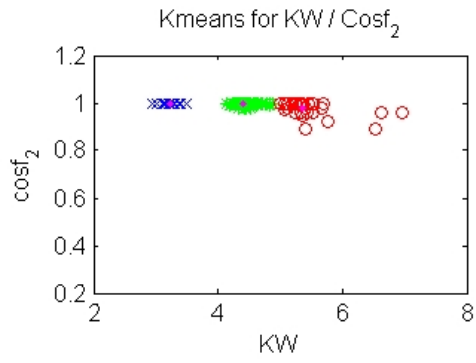
Ανάλυση βάσει του RKM της κατανάλωσης μεγάλων κτιρίων για μία εβδομάδα (10-17/7/2011):



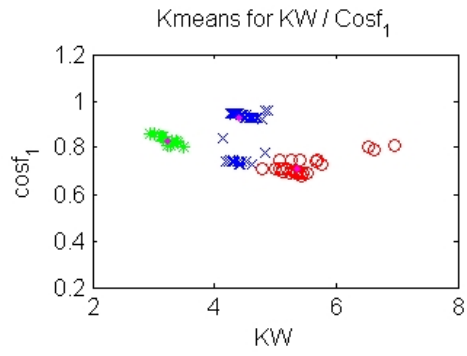
(a) Cluster for day=1



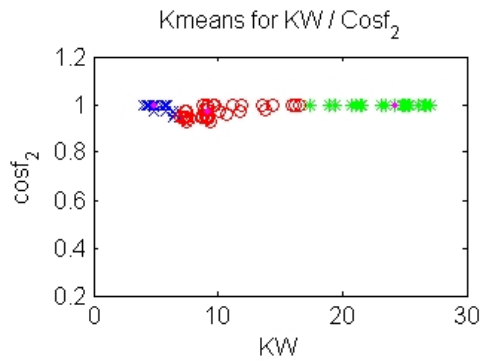
(b) Cluster for day=1



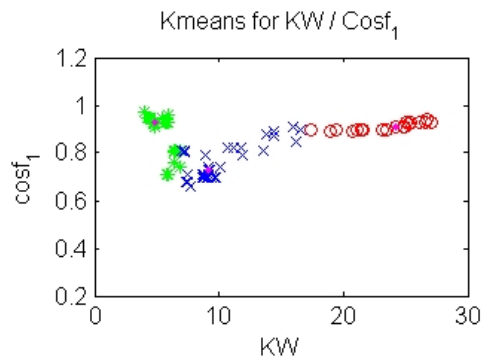
(c) Cluster for day=2



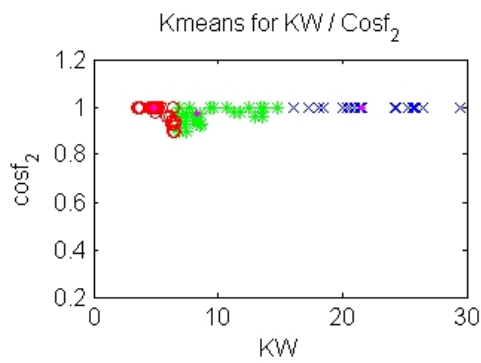
(d) Cluster for day=2



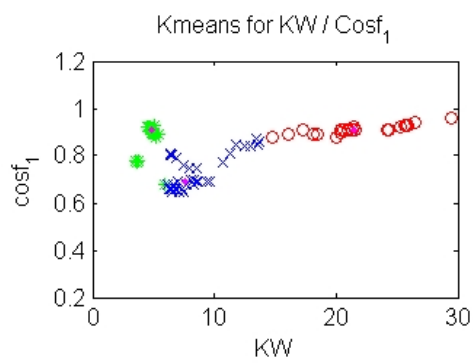
(e) Cluster for day=3



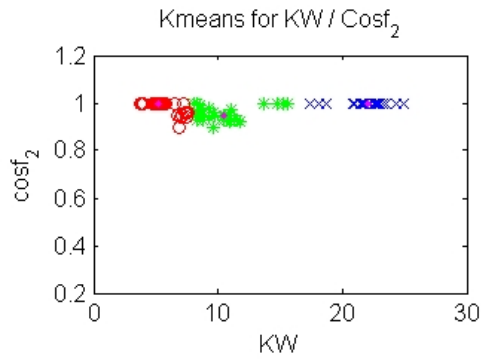
(f) Cluster for day=3



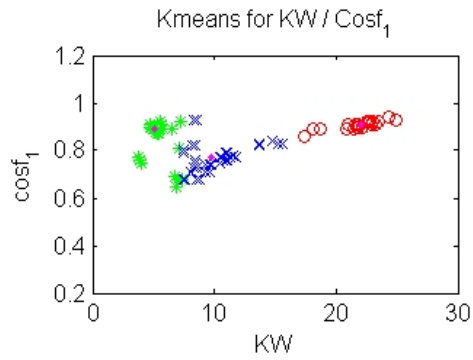
(g) Cluster for day=4



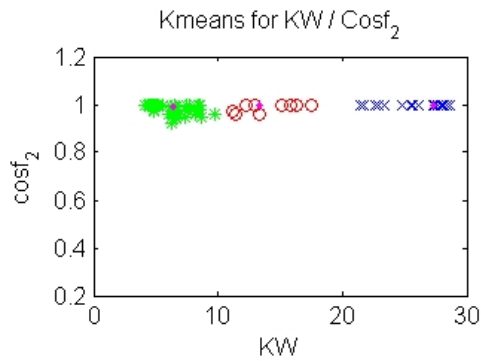
(h) Cluster for day=4



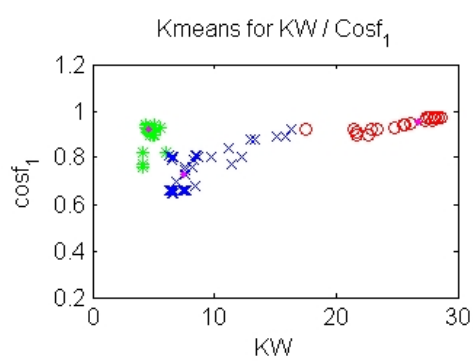
(i) Cluster for day=5



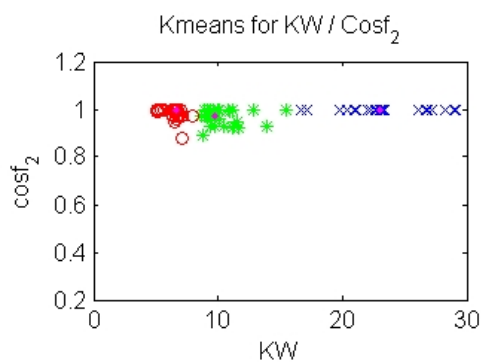
(j) Cluster for day=5



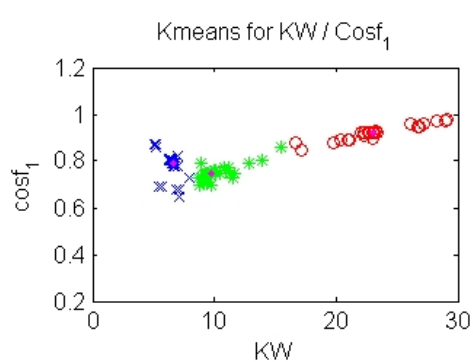
(k) Cluster for day=6



(l) Cluster for day=6



(m) Cluster for day=7



(n) Cluster for day=7

Εικόνα 37. Αναδραστικά clusters για 50 κτίρια και 7 ημέρες ανάλυση

Και πάλι από την παραπάνω αναδραστική διαδικασία, ανιχνεύουμε κοινά patterns και trends δεικτών, ανάμεσα στις συστάδες και την ανάλυση των δεικτών C, e, g και Similarity S. Για παράδειγμα στις ημέρες 1 και 2 οι συστάδες εμφανίζουν διαφοροποίηση από τις άλλες ημέρες καθώς είναι Σαββατοκύριακο και μπορούμε να δούμε τις διαφορές και τα διαφορετικά μέλη των συστάδων, αλλά και την διαφορετική θέση των κεντροειδών. Η εντροπία e είναι γενικά σταθερή, αναδεικνύοντας σταθερά trends χωρίς μεγάλες αυξομειώσεις, κάτι που σημαίνει ότι η γενική συμπεριφορά είναι ντετερμινιστική. Και αυτό είναι γεγονός, καθώς τα κτίρια ήταν μεγάλα εμπορικά καταστήματα με σταθερή συμπεριφορά, μικρή εντροπία, σταθερούς πληθυσμούς και σταθερά κεντροειδή.

Σε μία επιπλέον ανάλυση συσχέτισεων και των αιχμών, μπορούμε να έχουμε συσχετίσεις, βάσει της διαδικασίας της αυτοσυσχέτισης.

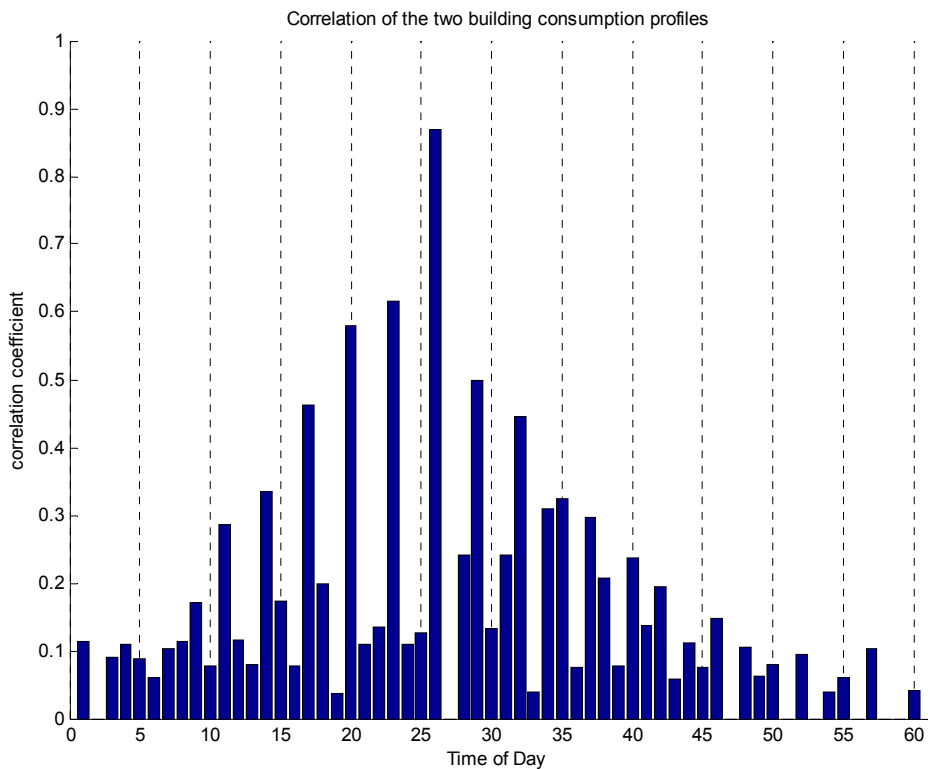
Αρχικά επιλέγουμε ένα «παράθυρο» παρατήρησης από τις 07:30 το πρωί έως τις 22:30 το βράδυ για να μελετήσουμε τη συσχέτιση, καθώς δεν μας ενδιαφέρει η κατανάλωση τις νυχτερινές ώρες. Παρ' όλα αυτά το παράθυρο αυτό μπορεί να καθοριστεί ανάλογα με τον σκοπό της μελέτης, θα μπορούσε για παράδειγμα να περιοριστεί μόνο τις μεσημεριανές ώρες όπου υπάρχει μεγάλη φόρτιση του δικτύου. Για την εύρεση σχέσης ανάμεσα στα δύο προφίλ θα βρούμε την ετεροσυσχέτιση τους όπως ορίζεται από την θεωρία στατιστικής. Αν x το διάνυσμα που περιγράφει τις μετρήσεις της μίας ημέρας και y το διάνυσμα που περιγράφει τις καταναλώσει τις επόμενης ημέρας τότε η ετεροσυσχέτιση (correlation r) τους δίνεται από τον τύπο:

$$r_{x,y}(m) = \sum x(n) * y(n - m) \text{ με το } m \in (-\infty, +\infty) \quad (4.2.6)$$

Επειδή εμείς έχουμε πεπερασμένο αριθμό στοιχείων, το m είναι πεπερασμένο και δείχνει πόσο θα «ολισθήσει» το ένα διάνυσμα πάνω στο άλλο. Εμείς επιλέξαμε το $m=30$ (βασιζόμενοι στο παράθυρο παρατήρησης για συγκεκριμένες ώρες την ημέρα).

Ο υπολογισμός της συσχέτισης των καταναλώσεων έγινε στο MATLAB με χρήση της ρουτίνας `xcorr`, ενώ έγινε χρήση της παραμέτρου 'coef' (συντελεστής ετεροσυσχέτισης) ώστε να κανονικοποιηθούν οι τιμές στο διάστημα [0,1] (η τιμή ένα αντικαθιστά την τιμή που έχει η αυτοσυσχέτιση).

Ακολουθεί το διάγραμμα (Εικόνα 38) των συντελεστών ετεροσυσχέτισης:



Εικόνα 38. Συσχέτιση 2 καταναλώσεων στη μονάδα της ημέρας, βάσει υπερκυβικών πρακτόρων

Στην κεντρική τιμή παρατηρούμε ότι η συσχέτιση δεν είναι μεγάλη όμως μέγιστη για τιμή 26 δηλαδή για $m = 31-26 = 5$. Δηλαδή παρουσιάζεται μέγιστη συσχέτιση μεταξύ των προφίλ 75 λεπτά νωρίτερα από το σημείο 0, όσο και η ολίσθηση που κάναμε. Παρατηρούμε ακόμα ότι υπάρχουν και άλλες τιμές του συντελεστή με σημαντική συσχέτιση, όπως στα σημεία 20, 23, 29. Σημαντική συσχέτιση έχουμε όταν η τιμή του συντελεστή είναι πάνω από 0.5. Μια ερμηνεία μπορεί να είναι ότι οι καταναλώσεις παρουσιάζουν αιχμές οι οποίες αν υπερτεθούν δίνουν σημαντικές αιχμές.

Τι σημαίνουν τα παραπάνω για μια πολιτική ορθής διαχείρισης της ενέργειας; Σημαίνει ότι αν θέλουμε να ολισθήσουμε τις αιχμές του ενός καταναλωτή θα πρέπει να προσέξουμε ώστε να μην συμπέσουν σε σημεία με σημαντική συσχέτιση. Αν επεκτείνουμε τον παραπάνω υπολογισμό για περισσότερα κτίρια τότε μπορούμε να αναζητήσουμε την βέλτιστη ενεργειακή πολιτική που μπορούμε να εφαρμόσουμε σε ένα σύνολο κτιρίων(π.χ. μιας περιοχής) ώστε να έχουμε την ελάχιστη δυνατή υπέρθεση αιχμών. Βέβαια για περισσότερα

κτίρια δεν αρκεί μόνο ο υπολογισμός των συντελεστών συσχέτισης αλλά απαιτείται και μια ιεραρχικοποίηση των συσχετίσεων.

Εκτός αυτού μπορούμε να υπολογίσουμε και τους συντελεστές και για άλλες τιμές του m και να βαθμολογήσουμε την κάθε συσχέτισης μεταξύ ενός προφίλ και των υπολοίπων, βάση των τιμών των συντελεστών και των τιμών του m . Φυσικά όσο πιο ταυτοχρονισμένες είναι οι αιχμές τόσο υψηλότερη θα είναι η βαθμολογία. Για παράδειγμα θα μπορούσαμε να υπολογίζουμε μια νόρμα με όλους του συντελεστές συσχέτισης δύο προφίλ, ή τουλάχιστον όσους έχουν αξιόλογη τιμή, όπως η παρακάτω:

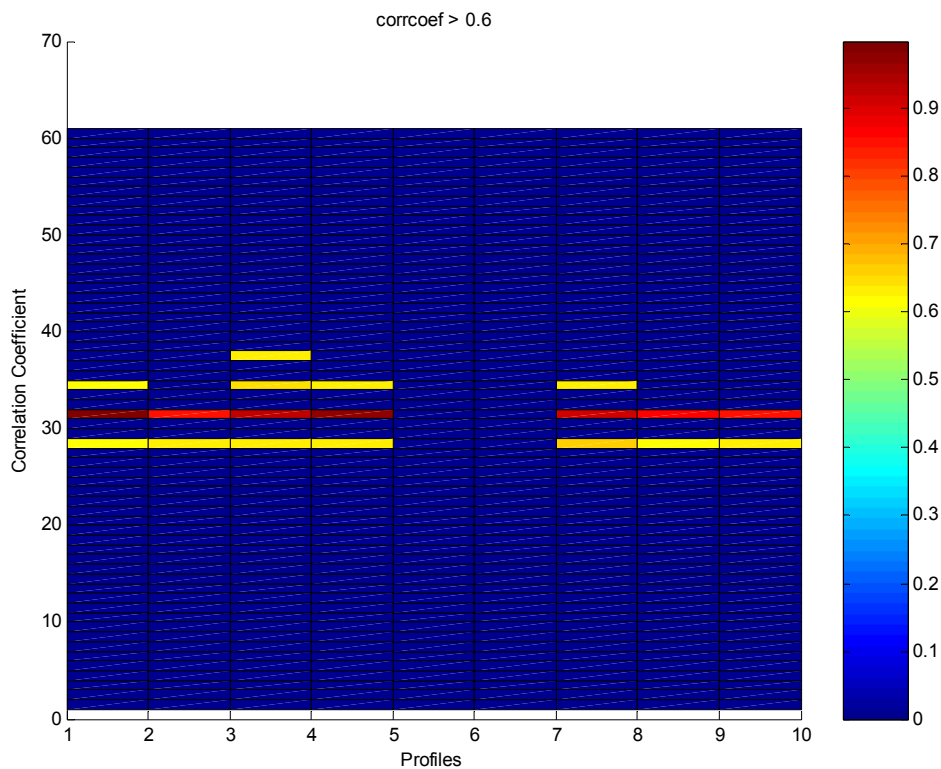
$$\|x_1, x\| = \left\{ \sum [coef(x_1, x_2)^k] \right\}^{1/k} \quad \forall k \text{ ακέραιο} \quad (4.2.7)$$

Υψηλές τιμές της παραπάνω νόρμα σημαίνουν, υψηλή συσχέτιση.

Έτσι θα δημιουργηθεί ένας πίνακας που περιέχει μόνο τις χρονικές «μετατοπίσεις» ή καθυστερήσεις στις οποίες υπάρχει υψηλή συσχέτιση. Ο πίνακας αυτό περιέχει σημαντικές πληροφορίες για τον συγχρονισμό αλλά και τη διαδοχή των αιχμών στα προφίλ των καταναλωτών.

Εντοπίζοντας γραμμές με μεγάλο αριθμό στοιχείων μπορούμε να πούμε ότι τα προφίλ είναι ταυτοχρονισμένα ενώ εντοπίζοντας διαγώνιους μπορούμε να βρούμε προφίλ που εμφανίζουν αιχμή με χρονική καθυστέρηση.

Είναι σαφώς ποιο οικονομικό για ένα σύστημα να υπερθέσει προφίλ καταναλωτών που βρίσκονται πάνω σε διαγώνιους και να αποφύγει όσο το δυνατόν περισσότερο να συμπέσουν προφίλ που βρίσκονται στην ίδια γραμμή.



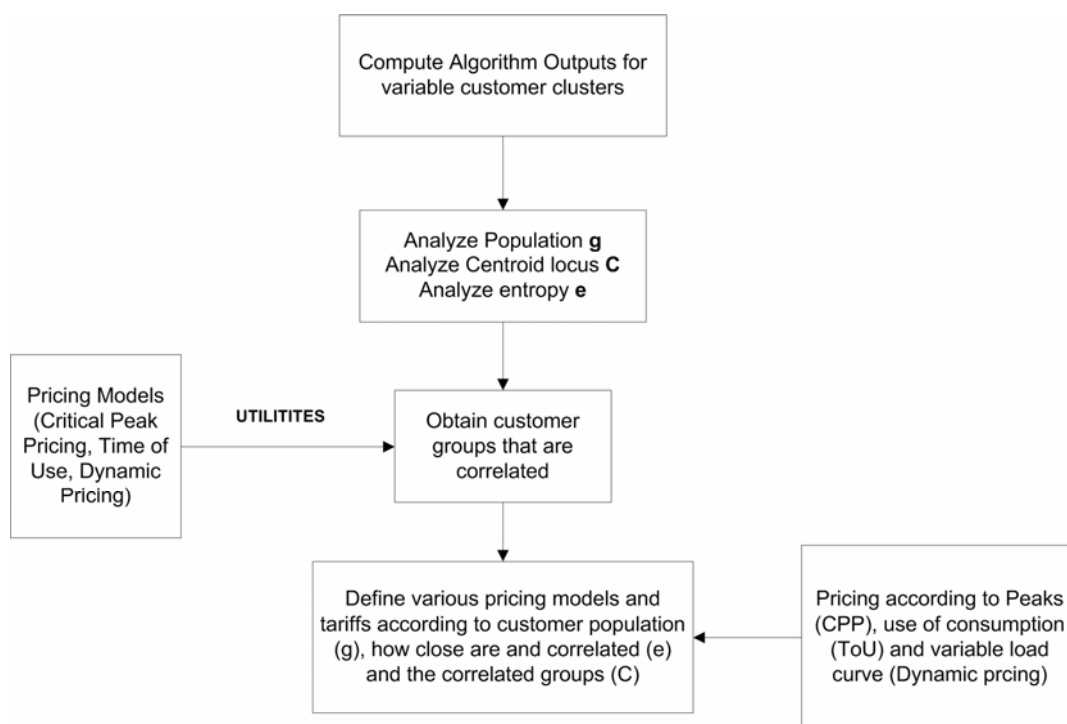
Εικόνα 39. Συντελεστές συσχέτισης μεγαλύτεροι από 0,6 μεταξύ πρώτου προφίλ και υπολοίπων

Στην παραπάνω Εικόνα 39, μπορούμε να παρατηρήσουμε τους συντελεστές συσχέτισης που είναι μεγαλύτεροι του 0.4 και τις νοητές οριζόντιες ευθείες και διαγώνιους που χρησιμοποιούνται. Η κόκκινη γραμμή στο 32 δηλαδή στο $m = 0$ δείχνει υψηλό ταυτοχρονισμό (αναμενόμενο αφού τα προφίλ προέρχονται από το ίδιο κτίριο), ενώ η κίτρινη δείχνει ότι υπάρχει και υψηλή συσχέτιση για καθυστέρηση 45 λεπτών. Δηλαδή, στη συγκεκριμένη περίπτωση, υπάρχουν υψηλές αιχμές ζήτησης με χρονική διαφορά 45 λεπτών.

Οι παραπάνω διαδικασίες, υλοποιούνται από τους αναδραστικούς πράκτορες σε κάθε κόμβο, έχοντας και την πληροφορία από τους δυαδικούς κόμβους που έχουν περάσει, και αποτελούν input στον RKM αλγόριθμο.

Αναλύοντας λοιπόν τα στοιχεία, μπορούμε να τα χρησιμοποιήσουμε για ανάλυση στρατηγικής στην ενέργεια και σχεδιασμό υπηρεσιο-κεντρικής φιλοσοφίας με συγκεκριμένες τάσεις και δομές.

Στο παρακάτω σχήμα, φαίνεται αναλυτικά η χρησιμότητα των αποτελεσμάτων και των ποιοτικών δεικτών που πηγάζουν από τον αλγόριθμο RKM, για σχεδίαση υπηρεσιών στον τομέα της ενέργειας και των ενεργειακών υπηρεσιών.



Εικόνα 40. Διάγραμμα συμπερασμού βάσει των ποιοτικών χαρακτηριστικών g , C , e και S

Το αποτέλεσμα και η χρήση των αποτελεσμάτων, φαίνεται και στον παρακάτω πίνακα, όπου αναλύεται ξεχωριστά η χρήση του κάθε δείκτη και το impact σε μία ανάλογη ενεργειακή υπηρεσία.

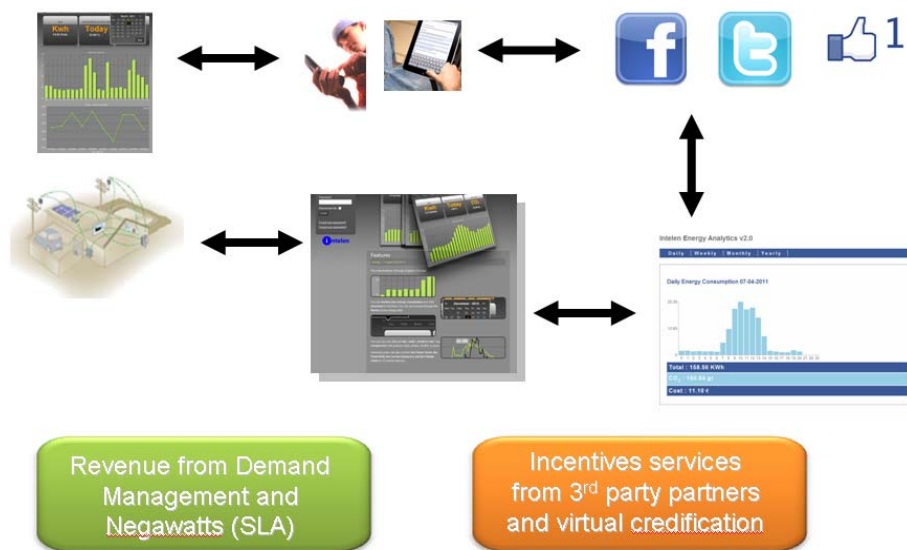
| Output Metric | Description | Pricing model mapping | Impact |
|----------------------|---|---|---|
| Population g | Number of customers at the specific time that are part of the variable cluster | Time of Use pricing (ToU) targeted to similar target groups / populations | Need to know the exact population for the specific timeslot |
| Entropy e | Density of customers that are close to the specific centroid at a specific time | Critical Peak Pricing (CPP) | Need to know this value in order to identify higher correlated peaks and higher correlated KPIs |

| | | | |
|--------------------------|--|----------------------|--|
| Centroid Locus C | Correlated centroids that have specific customer population and specific entropy value | Dynamic Pricing (DP) | Need to know the locus in order to apply batch pricing models to specific customers that change with respect to time (different tariff per hour or mins) |
| Combination of the above | Full decision making regarding the correlations, the trends and the customer variable groups | All pricing models | The combination gives us variable correlated clusters that can be mapped to variable pricing models |

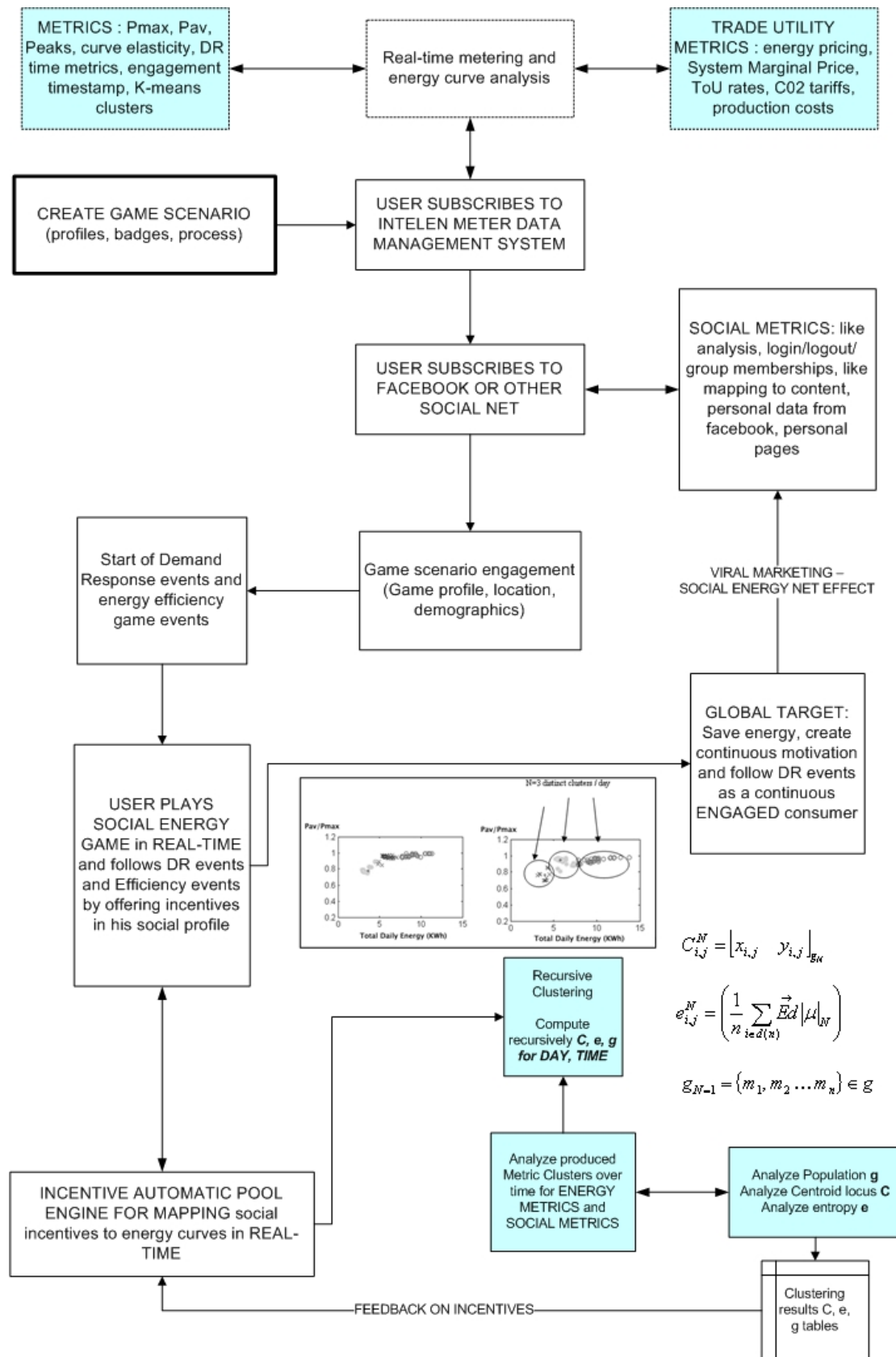
Πίνακας 7. Χρήση δεικτών από RKM για Ενεργειακή Στρατηγική και υπηρεσίες

4.4 Ενεργειακά Κοινωνικά Δίκτυα και Εφαρμογή του RKM

Η χρήση των κοινωνικών δικτύων στον τομέα της ενέργειας είναι μία νέα καινοτόμος προσέγγιση, η οποία ασφαλώς εσωκλείει μεγάλες αναλύσεις δεδομένων σε πραγματικό χρόνο, καθώς η συσχέτιση του ενεργειακού και του κοινωνικού προφίλ είναι κάτι το εξαιρετικά ενδιαφέρον αλλά και πολύπλοκο. Η εφαρμογή τέτοιων προσεγγίσεων, θα προσφέρουν μία νέα προοπτική και αντίληψη στην συσχέτιση των παραπάνω οντοτήτων.



Εικόνα 41. Ενεργειακά Κοινωνικά Δίκτυα με συσχέτιση Ενεργειακού και κοινωνικού προφίλ



Εικόνα 42. Ενεργειακά Κοινωνικά Δίκτυα και ο συνδυασμός τους με τον RKM

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 Συμπεράσματα Διατριβής

Η παρούσα διατριβή προσέγγισε το θέμα της εφαρμογής υπερκυβικής τοπολογίας και ανάλυσης σε επίπεδο πρακτόρων, καταναμημένης πληροφορίας και γνώσης σε πραγματικό χρόνο. Οι αλγόριθμοι εφαρμόστηκαν σε πρώτη φάση σε δεδομένα μαθησιακού περιεχομένου και σε δεύτερη και πιο εξελιγμένη φάση σε ενεργειακά δεδομένα πραγματικού χρόνου, από ενεργειακό πληροφοριακό σύστημα και πλέγμα έξυπνων μετρητών.

Η εξομοίωση του μαθηματικού μοντέλου και των αλγορίθμων δρομολόγησης, ανέδειξαν την στοχαστική μέθοδο με ενδιάμεση τυχαία κατανομή των πρακτόρων, ως πολύ πιο αποδοτική μέθοδο και με ανώτερο φραγμένο όριο πολυπλοκότητας (Borodin-Horcroft). Η μέθοδος έδωσε πολύ ικανοποιητικά αποτελέσματα κατά την διάρκεια εξομοιώσεων και απέδειξε ότι μπορεί να διαχειριστεί δεδομένα με έντονο χρονοδυναμικό προφίλ.

Η εφαρμογή του αλγορίθμου και των μεθοδολογιών στον τομέα της ενέργειας είχε εντυπωσιακά αποτελέσματα, καθώς η ανάλυση και η μαθηματική μοντελοποίηση δίνει ταχύτητα και αποδοτικότητα στην ανάλυση των δεδομένων καθώς και την δυνατότητα αύξησης των δεδομένων χωρίς να υπάρξει πρόβλημα υπολογιστικής πολυπλοκότητας. Η φύση και μαθηματική τοπολογία του υπερκύβου, επιτρέπει την εύκολη και αποτελεσματική διαχείριση των μετρήσεων και δεδομένων, ανοίγοντας έτσι νέες προοπτικές και ερευνητικές πόρτες στην εξόρυξη γνώσης σε δεδομένα πραγματικού χρόνου (stream mining) καθώς και σε νέες προσεγγίσεις αξιολόγησης και ανάλυσης δυναμικών δεικτών με σκοπό της διαχείριση γνώσης και τον συμπερασμό. Η ισομορφία της υπερκυβικής δομής επιτρέπει την δυναμική μοντελοποίηση και εφαρμογή του υπερκύβου σε οποιοδήποτε πρόβλημα της σύγχρονης διαχείρισης γνώσης και εξόρυξης, καθώς οι πράκτορες ακολουθούν ακόμα και τυχαίες αλγεβρικές μεταθέσεις, χωρίς να υπάρχει πρόβλημα κατεύθυνσης και πολυπλοκότητας του υπερκυβικού κλειστού γράφου.

Εάν παραθέσουμε αναφορικά τα συμπεράσματα της Διατριβής, μπορούμε να πούμε ότι:

- Αναλύθηκε και συγκρίθηκε η ντετερμινιστική και στοχαστική προσέγγιση της δρομολόγησης πρακτόρων στο υπερκυβικό πλέγμα, με σημαντικά αποτελέσματα όσον αφορά την αποδοτικότητα και ταχύτητα του αλγορίθμου

- Δημιουργήθηκε ο αλγόριθμος RKM (Recursive K-Means) ο οποίος βάσει της υπερκυβικής τοπολογίας εκτελεί έναν weighted-αναδραστικό K-means αλγόριθμο από κάθε υπερκυβικό πράκτορα. Ο συγκεκριμένος αλγόριθμος είναι πολλά υποσχόμενος και δημοσιεύθηκε με επιτυχία σε Διεθνές περιοδικό με κριτές
- Η παραπάνω μεθοδολογία εφαρμόστηκε με επιτυχία σε ενεργειακά δεδομένα πραγματικού χρόνου και τα αποτελέσματα χρησιμοποιήθηκαν στον χώρο των ευφυών δικτύων και στις υπηρεσίες που υπάρχουν στον χώρο της ενέργειας.
- Η υπερκυβική πολυδιάστατη διαχείριση της ενεργειακής πληροφορίας αποτελεί μία από τις πολλές καινοτομίες της Διατριβής, η οποία οδηγεί την παροχή ειδικευμένων καταναλωμένων διαδικτυακών ενεργειακών προϊόντων καθώς και αποτελεσματικό behavioral control ενεργειακών καταναλωτών, για παροχή εξατομικευμένων ενεργειακών και λοιπών υπηρεσιών. Οι συσχετίσεις που μπορούν να γίνουν, με βάση τα δυναμικά ενεργειακά προφίλ των καταναλωτών και οδηγούν σε πολύ αποτελεσματικές τεχνικές διαχείρισης και αποκρυπτογράφησης ενός σύνθετου ενεργειακού προφίλ.

5.2 Μελλοντικές Προεκτάσεις στην Έρευνα και Καινοτομία

Βάσει των αποτελεσμάτων της συγκεκριμένης Διατριβής, ανοίγονται νέοι δρόμοι εφαρμοσμένης έρευνας σε ανάλυση στον χώρο των «Μεγάλων Δεδομένων» (big data), στην φιλοσοφία Internet of Things και σε προχωρημένα συστήματα ενεργειακής διαχείρισης, συστημάτων Meter data management βασισμένο στο cloud computing και τέλος σε αλγοριθμικές μεθοδολογίες καταναμημένης διαχείρισης γνώσης.

Ειδικά στο τελευταίο, με το σύστημα και την μεθοδολογία που παρουσιάστηκε, καταργείται η ανάγκη για χρησιμοποίηση Βάσεων Δεδομένων, καθώς το υπερκυβικό πλέγμα και οι πράκτορες αναλύουν και αποθηκεύουν τις δυναμικές παραμέτρους και τα χαρακτηριστικά της ανάλυσης δεικτών (KPIs). Μελλοντικές προεκτάσεις υπάρχουν επίσης και στα κοινωνικά δίκτυα, δεδομένης της πολυπλοκότητας και του ρυθμού παραγωγής δεδομένων και πληροφορίας στο Internet. Η συσχέτισή τους με το υπερκυβικό πλέγμα, βάσει κάποιων τοπικών πρακτόρων (screen savers) μπορούν να δώσουν ένα πολύ προχωρημένο σύστημα ανάλυσης πληροφορίας σε πραγματικό χρόνο και σύστημα συσχέτισής της με άλλους εξωγενείς παράγοντες (σενσορες, internet of things, κλπ)

ΣΕ ΔΙΕΘΝΗ ΕΠΙΣΤΗΜΟΝΙΚΑ ΠΕΡΙΟΔΙΚΑ ΜΕ ΚΡΙΤΕΣ

1. **Nikolopoulos, V**; Mpardis, G; Giannoukos, I; Lykourantzou, I; Loumos, V (2011). Web-based Decision Support System Methodology for Smart Provision of Adaptive Digital Energy Services over Cloud Technologies, IET Software Journal, **Issue Date:** Oct. 2011 **Volume:** 5 Issue:5 **On page(s):** 454 - 465
2. Pazalos K., Loukis E., **Nikolopoulos, V.**, (2010). A structured methodology for assessing and improving e-services in digital cities, Telematics & Informatics Journal, Elsevier doi:10.1016/j.tele.2010.05.002
3. Journal of the American Society for Information Science and Technology, 60(2), pp. 372-380.
4. Mpardis, G., **Nikolopoulos, V.**, Kotsilieris, T., Lykourantzoy, I., Giannoukos, I., Loumos, V., (2009). Optimization and Re-Engineering of internal Greek Banking Processes using BPMN Modelling and adaptive KPIs Analysis, International Journal of Cooperative Information Systems (IJCIS), In Press, Corrected Proof, ISSN: 0218-8430.
5. Lykourantzou, I., Giannoukos, I., Mpardis, G., **Nikolopoulos, V.**, Loumos, V., (2009). Early and Dynamic Student Achievement Prediction in E-Learning Courses Using Neural Networks, Journal of the American Society for Information Science and Technology, 60(2), pp. 372-380.
6. Lykourantzou, I., Giannoukos, I., **Nikolopoulos, V.**, Mpardis, G., Loumos, V., Dropout prediction in e-learning courses through the combination of machine learning techniques, Computers & Education, In Press, Corrected Proof, Available online 31 May 2009, ISSN 0360-1315, DOI: 10.1016/j.compedu.2009.05.010, May/2009.

ΣΕ ΠΡΑΚΤΙΚΑ ΣΥΝΕΔΡΙΩΝ ΜΕ ΚΡΙΤΕΣ

1. **Nikolopoulos, V**, Loumos, V, A Web-based Information System for Optimal Energy Sources Management through Ontologies and Hypercubic Clustering, Energy International conference, 2006
2. **Nikolopoulos, V**, Loumos, V, A complete Ontological Model for effective web-based Energy Management through a Hypercubic Semantic Grid, 2nd Panhellenic Conference of PSDMH, 2007
3. **Nikolopoulos, V**, Loumos, V, Web-based energy information system for optimal bi-directional behavioural control of various energy customers using ADSL hypercubic clustering and internet services (emir system), PSM Conference 2007

4. **Nikolopoulos, V.**, Lykourantzou, I., Mpardis, G., Loumos, V., Kayafas, E., A web-based statistical evaluation method for e-learning courses using student logs analysis, ICIE 2nd International Conference on Interdisciplinarity in Education, Athens, Greece, 2006.
5. **Nikolopoulos V**, Giorgos Mpardis, Vassili Loumos, Eleftherios Kayafas, “Dynamic simulation of a student e-learning performance profile through web-based data-log mining“Eurosim 2007 Ljubljana, Slovenia, September 9 - 13, 2007
6. **Nikolopoulos V**, Ioanna Lykourantzou, Giorgos Mpardis, Vassili Loumos, Eleftherios Kayafas, “A Web-based Interoperable Method for accessing e-learning Remote Repositories using directed XML-based Remote Procedure Calls”, WSEAS TRANSACTIONS on COMMUNICATIONS
7. **Nikolopoulos V**, Giorgos Mpardis, Vassili Loumos, “Διαχείριση Ενεργειακών Δυναμικών Προφίλ με χρήση Διαδικτυακών Προσαρμοστικών Αλγορίθμων και Google Maps, Φοιτητικό Συνέδριο Διοικητικής Επιστήμης και Τεχνολογίας Συνεδριακό κέντρο ΔΑΪΣ
8. “Dynamic Reliability Assessment of Multiple Choice Tests in distributed E-Learning Environments: A Case Study” Ioanna Lykourantzou, Giorgos Mpardis, **Nikolopoulos V**, Vassili Loumos, IFIP Conference on Artificial Intelligence Applications & Innovations
9. “Δυναμική Μοντελοποίηση Ηλεκτρονικών Ενεργειακών / Περιβαλλοντικών Υπηρεσιών, με βάση το BPMN και Προσαρμοστικό Πληροφοριακό Σύστημα “, Giorgos Mpardis, **Nikolopoulos V**, Vassili Loumos, Φοιτητικό Συνέδριο Διοικητικής Επιστήμης και Τεχνολογίας Συνεδριακό κέντρο ΔΑΪΣ
10. Giannoukos, I., Lykourantzou, I., Mpardis, G., **Nikolopoulos, V.**, Loumos, V. and Kayafas, E., 2008, An intelligent mechanism for adaptive peer user modeling in web-based environments, 3rd International Workshop on Semantic Media Adaptation and Personalization, SMAP 2008, Prague, Czech Republic, December 15-16, 2008.
11. Giannoukos, I., Lykourantzou, I., Mpardis, G., **Nikolopoulos, V.**, Loumos, V., and Kayafas, E. 2008. Collaborative e-learning environments enhanced by wiki technologies. In Proceedings of the 1st ACM international Conference on Pervasive Technologies Related to Assistive Environments, PETRA '08 , Athens, Greece, July 16 - 18, 2008 , vol. 282. ACM, New York, NY, pp. 1-5.
12. Lykourantzou, I., Mpardis, G., **Nikolopoulos, V.**, Loumos V., E-Learning Reliability Assessment Of Multiple Choice Test Data: Method Comparison And Reliability Quality Estimation, AEA2007, Stockholm, Sweden, November 8-10, 2007.

ΣΕ ΚΕΦΑΛΑΙΑ ΒΙΒΛΙΩΝ

1. Mpardis, G., Karadimas, N., **Nikolopoulos, V.**, Kotsilieris, T., Loumos, V., Bank Loan Processes using BPMN Modelling, Springer Lecture Notes in Electrical Engineering, In Press, Corrected Proof, Available online 31 March 2010, ISSN1876-1100.

2. Likourentzou, I., Mpardis, G., **Nikolopoulos, V.**, Loumos, V., Dynamic Reliability Assessment of Multiple Choice Tests in distributed E-Learning Environments: A Case Study, In: Artificial Intelligence and Innovations 2007: from Theory to Applications, IFIP International Federation for Information Processing, Springer Boston, ISSN 1571-5736 (Print) 1861-2288 (Online), Vol. 247/2007, ISBN 978-0-387-74160-4, pp. 73-80.

ΣΥΓΓΡΑΦΗ ΒΙΒΛΙΩΝ

1. Μελέτες - Εφαρμογές & Υλοποίηση Δικτύων Η/Υ Γ.Μπάρδης, Β.Νικολόπουλος, Ι.Μπράττος (Β.Γκιούρδας 2007) ISBN -13: 978-960-387-577-2

ΕΘΝΙΚΕΣ ΠΑΤΕΝΤΕΣ (ΟΒΙ)

1. ΔΙΑΔΙΚΤΥΑΚΗ ΕΝΕΡΓΕΙΑΚΗ ΜΗΧΑΝΗ ΑΝΑΖΗΤΗΣΗΣ ΚΑΙ ΜΕΘΟΔΟΣ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ ΓΙΑ ΒΕΛΤΙΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΤΙΜΟΛΟΓΙΑΚΗ ΕΚΤΙΜΗΣΗ ΕΝΕΡΓΕΙΑΚΩΝ ΠΟΡΩΝ (Αρ. Εθνικού ΔΕ 1006221), **Β. Νικολόπουλος**
2. ΔΙΑΔΙΚΤΥΑΚΗ ΜΕΘΟΔΟΣ ΚΑΙ ΣΥΣΤΗΜΑ ΕΝΕΡΓΕΙΑΚΩΝ ΔΟΣΟΛΗΨΙΩΝ ΚΑΙ ΑΓΟΡΑΠΩΛΗΣΙΑΣ ΕΝΕΡΓΕΙΑΚΩΝ ΦΟΡΤΙΩΝ ΓΙΑ ΕΠΙΛΕΓΟΝΤΕΣ ΚΑΤΑΝΑΛΩΤΕΣ, ΜΕΣΩ ΑΥΤΟΜΑΤΗΣ ΕΝΕΡΓΕΙΑΚΗΣ ΔΙΑΧΕΙΡΙΣΗΣ, ΕΝΕΡΓΕΙΑΚΩΝ ΚΟΥΠΟΝΙΩΝ ΚΑΙ ΠΙΣΤΩΤΙΚΩΝ ΚΑΡΤΩΝ (Αρ. Εθνικού ΔΕ 1006308), **Β. Νικολόπουλος**

7. REFERENCES

1. M. Klein, L.B. Methlie, Knowledge-Based Decision Support Systems with Applications in Business, John Wiley & Sons, Chichester, 1995.
2. M.S. Scott Morton, Management decision systems; computerbased support for decision making, Graduate School of Business Administration, Harvard University, 1971.
3. S. Ba, Kalakota, A.B. Whinston, Executable documents DSS, Proceedings of 3rd International Conference on DSS, Hong Kong, 22–23 June, 1995.
4. P. Barton, The George Washington University Data-Driven Decision Support Project, URL DSSResources.COM (2003).
5. P. Bharati, A. Chaudhury, An empirical investigation of decision-making satisfaction in Web-based decision support systems, *Decision Support Systems* 37 (2) (2004) 187–197.
6. H.K. Bhargava, R. Krishnan, The World Wide Web: opportunities for operations research and management science, *INFORMS Journal on Computing* 10 (1998 (Fall)) 359–383.
7. H.K. Bhargava, D.J. Power, Decision support systems and Web technologies: a status report, Proceedings of the 2001 Americas Conference on Information Systems, Boston, MA, August 3–5, 2001.
8. H.K. Bhargava, A. King, D. McQuay, DecisionNet: an architecture for modeling and decision support over the World Wide Web, Proceedings of 3rd International Conference on DSS, Hong Kong, 22–23 June, 1995.
9. H.K. Bhargava, R. Krishnan, D. Kaplan, On generalized access to a WWW-based network of decision support services, Proceedings of 3rd International Conference on DSS, Hong Kong, 22–23 June, 1995.
10. H.K. Bhargava, C.L. Herrick, S. Sridhar, Beyond spreadsheets: software for building decision support systems, *IEEE Computer* 32 (1999 (March)) 31–39.
11. M. Goul, A. Philippakis, M. Kiang, D. Fernandes, B. Otondo, Towards a client/server open-DSS protocol suite for automating DSS deployment on the World Wide Web, Proceedings of 3rd International Conference on DSS, Hong Kong, 22–23 June, 1995.
12. P.D. Coddington, K.A. Hawick, H.A. James, Web-based access to distributed high-performance geographic information systems for decision support, Proceedings of the 32nd Hawaii International Conference on System Sciences, January 1999.
13. M. Cohen, C.B. Kelly, A.L. Medaglia, Decision support with Web-enabled software, *Interfaces* 31 (2001) 109–129.
14. D. Delen, R. Sharda, P. Kumar, Movie forecast guru: a Web-based DSS for hollywood managers, *Decision Support Systems* (this issue), doi:10.1016/j.dss.2005.07.005.
15. C. Holsapple, A. Whinston, Exploring the next generation of decision support, *Decision Support Systems* 14 (3) (1995 (July)) 185–186.
16. Insightful Staff, Merck Deploys Web-Based Drug Discovery Visualization System, URL DSSResources.COM (February 2002).

17. P. Stephenson and M.Paun, "Consumer advantages from half-hourly metering and load profiles in the UK competitive electricity market". Proc. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies 2000, City University, London, 4-7 April 2000, pp.35-40.
18. B. Iyer, G. Shankaranarayanan, M.L. Lenard, Model management decision environment: a Web service prototype for spreadsheet models, Decision Support Systems 40 (2) (2005 (August)) 283–304.
19. Stein, Jeff Ross, Aditi Raychoudhury, Charles Eley. 2000. "The Jury is (Halfway) In: New Building Performance Contracting Results", Proceedings of the 2000 ACEEE Summer Study on Energy Efficiency in Buildings
20. S.V.Allera and A.G.Horsburgh, "Load profiling for the energy trading and settlements in the UK electricity markets", Proc. DistribuTECH Europe DA/DSM Conference, London, 27-29 October 1998.
21. Chen CS, Hwang JC, Tzeng YM, Huang CW, Cho MY. Determination of customer load characteristic by load survey system at Taipower. IEEE Trans Power Deliv 1996;11(3):1430–6.
22. C. S. Chen, J. C. Hwang, and C. W. Huang, "Application of load survey systems to proper tariff design," Power Systems, IEEE Transactions on, vol. 12, pp. 1746-1751, 1997.
23. Yezioro, A.; Dong, B.; Leite, F. (2008) "An Applied Artificial Intelligence Approach towards Assessing Building Performance Simulation Tools". In: Energy and Buildings, Volume 40, Issue 4, p. 612-620
24. Akkermans, H, C. Warmer, K. Kok, J. Kester, R. Kamphuis and P. Carlsson (2004): "ELEKTRA - Play the DER electronic market power game. An interactive experience with advanced Information and Communication Technologies for DER management" Contribution to IRED 2004, Integration of Renewable and Distributed Energy Resources, 1-3 Dec 2004, Brussels.
25. Darby, S. (2006): "The Effectiveness of Feedback on Energy Consumption", A Review for DEFRA of the Literature on Metering, Billing and direct Displays, Environmental Change Institute, University of Oxford.
26. J.K. Kok, C.J. Warmer, I.G. Kamphuis, "The PowerMatcher: Multi-agent Control of Electricity Demand and Supply", InM. Pechoucek and S.G. Thompson, "Agents in Industry: The Best from the AAMAS 2005 Industry Track, IEEE Intelligent Systems, March/April 2006.
27. D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "A methodology to classify distribution load profiles," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.
28. M.Emoult and F.Meslier, "Analysis and forecast of electrical energy demand". RCE, No.4, 1982.
29. M. Kitayama, R. Matsubara, and Y. Izui, "Data mining for customer load profile analysis," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.
30. F. J. Duarte, F. Rodrigues, V. Figueiredo, Z. Vale, and M. Cordeiro, "Data mining techniques applied to electric energy consumers characterization," presented at Artificial Intelligence and Soft Computing, 14-16 July 2003, Banff, Alta., Canada, 2003.

31. M. Kitayama, R. Matsubara, and Y. Izui, "Application of data mining to customer profile analysis in the power electric industry," presented at Proceedings of Winter Meeting of the Power Engineering Society, 27-31 Jan. 2002, New York, NY, USA, 2002.
32. B. D. Pitt and D. S. Kitschen, "Application of data mining techniques to load profiling," presented at Power Industry Computer Applications, 1999. PICA '99. Proceedings of the 21st 1999 IEEE International Conference, 1999.
33. J. R. Filho, E. M. Gontijo, A. C. Delaiba, E. Mazina, J. E. Cabral, and J. O. P. Pinto, "Fraud identification in electricity company customers using decision tree," presented at Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004.
34. R. Jiang, H. Tagaris, A. Lachsz, and M. Jeffrey, "Wavelet based feature extraction and multiple classifiers for electricity fraud detection," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.
35. R. Mano, R. Cespedes, and D. Maia, "Protecting revenue in the distribution industry: a new approach with the revenue assurance and audit process," presented at Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES, 2004.
36. Mary Ann Piette, Satkartar Kinney. October 1999. "Performance Assessment and Adoption Processes of an Information Monitoring and Diagnostic System Prototype", Final Report for CEC/CIEE Project: Diagnostics for Building Commissioning and Operation.
37. Nasraoui O. , Cerwinske J., Rojas C., and Gonzalez F., "Collaborative Filtering in Dynamic Usage Environments", in Proc. of CIKM 2006 - Conference on Information and Knowledge Management, Arlington VA , Nov. 2006
38. Scholz, Martin and Klinkenberg, Ralf: "Boosting Classifiers for Drifting Concepts". In Intelligent Data Analysis (IDA), Special Issue on Knowledge Discovery from Data Streams, Vol. 11, No. 1, pages 3–28, March 2007.
39. M.J.A. Berry and G. Linoff, "Data Mining Techniques for Marketing, Sales, and Customer Support," John Wiley & Sons, New York, USA, 1999.
40. Valiant L. G., Brebner G., "Universal Schemes for Parallel Communications", Proc. 13th ACM Symp. on Theory of Computing, pp. 263-277, 1981
41. Valiant L. G., "A scheme for fast Parallel Communications", Siam Journal Comp., Vol. 11, No. 2, pp. 350-361, May 1982
42. Lai, J. Z. C., Huang, T.-J., & Liaw, Y.-C. (2009). A fast k-means clustering algorithm using cluster center displacement. Pattern Recognition, In Press, Corrected Proof.
43. MacQueen, J. (1967). Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297
44. MATLAB Web Server. Natick, MA: The Mathworks Inc., 1999.
45. **E.M.I.R. Project** - Energy Management & Intelligent Reporting <http://www.plugme.eu>
46. Nikolopoulos, V, Loumos, V, "Web-based energy information system for optimal bi-directional behavioural control of various energy customers using ADSL hypercubic clustering and internet

- services (emir system)", Power Systems Modelling Conference 2007, ([PSM 2007](#)), June 2007, Athens GREECE, pp. 45-46
47. D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "An approach to customers daily load profile determination," presented at 2002 IEEE Power Engineering Society Summer Meeting, Jul 21 25 2002, Chicago, IL, United States, 2002.
 48. Karen Herter, Seth Wayland, "Residential response to critical-peak pricing of electricity: California evidence" *Energy*, Volume 35, Issue 4, April 2010, Pages 1561-1567, Elsevier
 49. J.K. Lin, S.K. Tso, H.K. Ho, C.M. Mak, K.M. Yung, Y.K. Ho "Study of climatic effects on peak load and regional similarity of load profiles following disturbances based on data mining" *International Journal of Electrical Power & Energy Systems*, Volume 28, Issue 3, March 2006, Pages 177-185, Elsevier
 50. José A. Dominguez-Navarro, José L. Bernal-Agustín, Rodolfo Dufo-López, "Data mining methodology for disaggregation of load demand" *Electric Power Systems Research*, Volume 79, Issue 10, October 2009, Pages 1393-1399
 51. Melody Stokes, Mark Rylatt, Kevin Lomas, "A simple model of domestic lighting demand" *Energy and Buildings*, Volume 36, Issue 2, February 2004, Pages 103-116
 52. Babcock B., Datar M., Motwani R., O'Callaghan L.: Maintaining Variance and k-Medians over Data Stream Windows. To appear in Proceedings of the 22nd Symposium on Principles of Database Systems (PODS 2003) (2003).
 53. Burl M., Fowlkes C., Roden J., Stechert A., and Mukhtar S. Diamond Eye: A distributed architecture for image data mining. In SPIE DMKD, Orlando, April (1999).
 54. Charikar M., O'Callaghan L., and Panigrahy R.: Better streaming algorithms for clustering problems. In Proc. of 35th ACM Symposium on Theory of Computing (STOC) (2003).
 55. O'Callaghan L., Mishra N., Meyerson A., Guha S., and Motwani R.: Streamingdata algorithms for high-quality clustering. Proceedings of IEEE International Conference on Data Engineering, March (2002).
 56. Cormode G., Muthukrishnan S.: What's hot and what's not: tracking most frequent items dynamically. PODS 2003. (2003) 296-306
 57. Datar M., Gionis A., Indyk P., Motwani R.: Maintaining Stream Statistics over Sliding Windows (Extended Abstract). In Proceedings of 13th Annual ACM SIAM Symposium on Discrete Algorithms (SODA 2002) (2002).
 58. Domingos P. and Hulten G., A General Method for Scaling Up Machine Learning Algorithms and its Application to Clustering. Proceedings of the Eighteenth International Conference on Machine Learning, 106--113, Williamstown, MA, Morgan Kaufmann. (2001)
 59. Domingos P. and Hulten G. Mining High-Speed Data Streams. In Proceedings of the Association for Computing Machinery Sixth International Conference on Knowledge Discovery and Data Mining, (2000) 71—80.
 60. Gaber, M.M., Krishnaswamy, S. and Zaslavsky, A. (2004). Cost-Efficient Mining Techniques for Data Streams. In Proc. Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), Dunedin, New Zealand. CRPIT, 32. Purvis, M., Ed. ACS.

61. Gaber, M. M., Krishnaswamy, S., and Zaslavsky, A., Adaptive Mining Techniques for Data Streams Using Algorithm Output Granularity, The Australasian Data Mining Workshop (AusDM 2003), Held in conjunction with the 2003 Congress on Evolutionary Computation (CEC 2003), December, Canberra, Australia, Springer Verlag, Lecture Notes in Computer Science (LNCS).
62. Gaber, M. M., Zaslavsky, A., and Krishnaswamy, S., A Cost-Efficient Model for Ubiquitous Data Stream Mining, Accepted for publication in the Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004), Perugia Italy, July 4-9.
63. Gaber, M. M., Zaslavsky, A., and Krishnaswamy, S., (2004), Towards an Adaptive Approach for Mining Data Streams in Resource Constrained Environments, Accepted for publication in the Proceedings of Sixth International Conference on Data Warehousing and Knowledge Discovery - Industry Track (DaWak 2004), Zaragoza, Spain, 30 August - 3 September, Lecture Notes in Computer Science (LNCS), Springer Verlag.
64. Ganti V., Gehrke J., Ramakrishnan R.: Mining Data Streams under Block Evolution. SIGKDD Explorations 3(2): (2002) 1-10.
65. Garofalakis M., Gehrke J., Rastogi R.: Querying and mining data streams: you only get one look a tutorial. SIGMOD Conference 2002: 635. (2002).
66. Giannella C., Han J., Pei J., Yan X., and Yu P.S.: Mining Frequent Patterns in Data Streams at Multiple Time Granularities. In Kargupta H., Joshi A., Sivakumar K., and Yesha Y. (eds.), Next Generation Data Mining, AAAI/MIT (2003).
67. Guha S., Mishra N., Motwani R., and O'Callaghan L.: Clustering data streams. In Proceedings of the Annual Symposium on Foundations of Computer Science. IEEE, November (2000).
68. Anderberg, M.R. (1973). Cluster Analysis for Applications, Academic Press.
69. Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop, AI Magazine, **11** (5), 68-70.
70. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining, AAAI Press.
71. Gibson, D., Kleinberg, J. and Raghavan, P. (1998). Clustering Categorical Data: An Approach Based on Dynamical Systems, In Proceedings of the 24th International Conference on Very Large Data Bases, 311-322, Morgan Kaufmann.
72. Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values, Data Mining and Knowledge Discovery, **2** (3), 283-304.
73. Johnson, R.A. and Wichern, D.W. (1998). Applied Multivariate Statistical Analysis, Prentice Hall.
74. Ganti, V., Gehrke, J. and Ramakrishnan, R. (1999). CACTUS: Clustering Categorical Data Using Summaries, In Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 73-83, ACM Press.
75. Guha, S., Rastogi, R. and Shim K. (2000). ROCK: A Robust Clustering Algorithm for Categorical Attributes. Information Systems, **25**, 345-366.

76. Han, J. and Kamber, M. (2001). *Data Mining: Concepts and Techniques*, Morgan Kaufman Publishers.
77. Hand, D., Mannila, H. and Smyth, P. (2001). *Principles of Data Mining*, MIT Press, Cambridge.
78. Marques de Sá, J. P. (2001). *Pattern Recognition: Concepts, Methods and Applications*, Springer.
79. Andritsos, P. (2004). *Scalable Clustering of Categorical Data and Applications*, PhD Thesis, University of Toronto, Department of Computer Science.
80. Maimon, O. and Rokah, L. (2005). *Data Mining and Knowledge Discovery Handbook*, Springer.
81. Tan, Steinbach and Kumar (2005). *Introduction to Data Mining*, Addison Wesley.
82. Ferreira Alfonso, "Issues in parallel computing with Hypercube multiprocessors", CNRS - LIP, ENS Lyon, DIMACS Technical Report 94-49, October 1994
83. Baudon Olivier, Fertin Guillaume, Havel Ivan, "Routing Permutations and 2-1 Routing Requests in the Hypercube", LABRI - Univ. Bordeaux I, Czech Math. Institute, Preprint submitted to Discrete Applied Maths, , 18 June 2000
84. Aiello A. William, Leighton F. T., Maggs M. Bruce, Newman Mark, "Fast Algorithms for Bit-Serial Routing on a Hypercube", Bell Labs, MIT Mathematics Dept., NEC Research Institute, Technical Report, 1989
85. Nesson Ted, Johnsson S. Lennart, "ROMM Routing: A Class of efficient Minimal Routing Algorithms", TR-21-94, Parallel Computing Res. Group, Harvard University, Technical Report to appear in Proceedings of the 1994 Parallel Computer Routing and Communications Workshop, May 16-18 1994
86. Sprague P., Tamaki Hisao, "Routings for Involutions of a Hypercube", Dept. Computer Science, Univ. of Alabama USA, Dept. of Computer Science, Ibaraki Univ. Japan, Technical Report, 1994
87. Saad Y., Schultz M. H., "Topological Properties of hypercubes", IEEE Trans. Computers 37 867-872, 1988
88. Cormack, R. M. A review of classification. *Journal of the Royal Statistical Society (Series A)*, 1971, 134, 321-367.
89. Gordon A.D. , *Classification* , Chapman & Hall/CRC , Boca Raton, FL, 1999
90. Liu G.L. *Introduction to Combinatorial Mathematics*, McGraw Hill ,New York 1968
91. Hartigan, J.A., Wong, M.A. 1979. A K-Means Clustering Algorithm. *Applied Statistics* 28:100–108
92. Kohonen, T., "Self-Organisation and Associative Memory", 3rd Ed., Springer-Verlag, Berlin, 1989.