Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακή Διατριβή
στα Πληροφοριακά και Επικοινωνιακά Συστήματα

«Μηχανισμοί Ανίχνευσης Απάτης σε Δίκτυα VoIP (Fraud Detection)»

Ζαφείριος Γεωργιάδης

Επιβλέπων Καθηγητής
Αναστάσιος Νταγιούκλας

Αύγουστος 2014

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

«Μηχανισμοί Ανίχνευσης Απάτης σε Δίκτυα VoIP (Fraud Detection)»

Ζαφείριος Γεωργιάδης

Επιβλέπων Καθηγητής
Αναστάσιος Νταγιούκλας

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στα Πληροφοριακά Συστήματα

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Αύγουστος 2014

Περίληψη

Ο στόχος της διατριβής αυτής, είναι η ανάπτυξη μηχανισμών, για την ανίχνευση δόλιων κλήσεων (fraudulent calls) σε ένα VoIP δίκτυο, χρησιμοποιώντας CDR δεδομένα των τηλεφωνικών κλήσεων. Για το σκοπό αυτό έχει σχεδιασθεί μια εφαρμογή η οποία δημιουργεί CDR δεδομένα. Η εφαρμογή παράγει αρχεία CDR σύμφωνα με τα στατιστικά στοιχεία του τηλεπικοινωνιακού παρόχου. Για την ανάπτυξη του μηχανισμού μελετήθηκαν κι αξιολογήθηκαν σενάρια απάτης από συνδρομητές VoIP.

Η εισαγωγή της εργασίας αυτής αναπτύσσεται στο πρώτο κεφάλαιο. Το δεύτερο κεφάλαιο αναφέρεται στον όρο VoIP, στα δομικά του στοιχεία και τα πρωτόκολλα τα οποία συνδέονται με την παραπάνω υπηρεσία. Ακολουθούν στο τρίτο κεφάλαιο οι τεχνικές απάτης σε τηλεπικοινωνιακούς παρόχους όπως επίσης και οι τρόποι αντιμετώπισής τους. Οι παραπάνω τεχνικές κατηγοριοποιούνται σε ομάδες. Κλείνοντας το τρίτο κεφάλαιο γίνεται αναφορά για τα πλεονεκτήματα και τα μειονεκτήματα κάθε τεχνικής αντιμετώπισης απάτης σε VoIP.

Το τέταρτο κεφάλαιο είναι αφιερωμένο στην ανάπτυξη μίας εφαρμογής για την ανίχνευση απάτης σε κλήσεις VoIP. Αναλύεται η φάση ανάπτυξης της εφαρμογής όπως και οι δυνατότητες που προσφέρει. Συνεχίζει με το στάδιο της εκπαίδευσης του μηχανισμού ανίχνευσης της απάτης. Από τηλεπικοινωνιακό πάροχο υπηρεσιών VoIP συλλέχθηκαν πακέτα CDRs για διάστημα μίας εβδομάδας. Τα πακέτα αποτέλεσαν τη βάση για τη δημιουργία των προφίλ των χρηστών χωρίς δόλιες τηλεφωνικές κλήσεις. Επίσης χρησιμοποιήθηκε μία γεννήτρια παραγωγής πακέτων CDRs για τη δημιουργία συγκεκριμένων προφίλ νόμιμων και δόλιων κλήσεων. Ο συνδυασμός των πακέτων CDR, από τον τηλεπικοινωνιακό πάροχο και την παραπάνω γεννήτρια, χρησιμοποιήθηκαν για την εκπαίδευση και τον έλεγχο του μηχανισμού ανίχνευσης της απάτης. Κλείνει το τέταρτο κεφάλαιο με την πλήρη περιγραφή των δύο εφαρμογών σε πραγματική λειτουργία.

Το πέμπτο κεφάλαιο παρουσιάζει τα αποτελέσματα από τη λειτουργία των αλγόριθμων ανίχνευσης δόλιων τηλεφωνικών κλήσεων μέσω VoIP υπηρεσιών. Δίνονται οι τρεις αλγόριθμοι που χρησιμοποιεί ο μηχανισμός ανίχνευσης δόλιας τηλεφωνικής κλήσεις και αναλύονται τα αποτελέσματά τους. Οι εφαρμογές που χρησιμοποιούν οι πάροχοι VoIP υπηρεσιών δεν παρέχουν τη δυνατότητα ελέγχου του προορισμού των τηλεφωνικών κλήσεων. Για τον παραπάνω λόγο, η εφαρμογή που υλοποιήθηκε στην παρούσα εργασία, δεν έχει τη δυνατότητα να συγκριθεί με άλλες εφαρμογές που δίνονται στο εμπόριο ή βρίσκονται σε ερευνητικά εργαστήρια.

Η παρούσα εργασία κλείνει με το έκτο κεφάλαιο που είναι ο επίλογος. Στο Παράρτημα Α παρουσιάζονται μέρη από το σύνολο του κώδικα που γράφηκε για την εκτέλεση της εφαρμογής ανίχνευσης απάτης σε κλήσεις VoIP και στο Παράρτημα Β τμήματα από τα αρχεία που χειρίζονται οι δύο εφαρμογές της παρούσας εργασίας.

Summary

The aim of this thesis is the development of mechanisms for detecting fraudulent calls in a VoIP network using CDR data calls. For this purpose designed an application which generates CDR data. The application generates CDR files according to the statistics of the telecommunications provider. For the development of the mechanism studied and evaluated scenarios fraud subscribers VoIP.

The introduction of this work is developed in the first chapter. The second chapter discusses the term VoIP, its structural elements and the protocols associated with this service. We continue in the third chapter with the fraud techniques to telecom providers as well as how to overcome them. These techniques are categorized into groups. Closing the third chapter refers to the advantages and disadvantages of each technique tackle fraud in VoIP.

The fourth chapter is dedicated to developing an application to detect fraudulent calls at a VoIP system. Analyses the development phase of the application as well as potential. The thesis continues with the training stage of the mechanism to detect fraud. By a telecommunications service provider VoIP packets collected CDRs. The packages were the basis for the creation of user profiles without fraudulent VoIP calls. Also is used one packet generator to generate CDRs for specific profiles of legitimate and fraudulent calls. The combination of packages CDR, the telecommunications provider and the above generator were used for training and monitoring mechanism to detect fraud. The fourth chapter closes with the full description of the two applications in actual operation.

The fifth chapter presents the results from the operation of the algorithms detect fraudulent phone calls using VoIP services. Give the three algorithms using the mechanism of detection of fraudulent VoIP calls and analyzed the results. Applications that use the VoIP service providers do not provide the ability to control the destination of the calls. The application which implemented in this thesis is not able to be compared with other applications provided commercially or in research laboratories.

This paper concludes with the sixth chapter is the conclusion. In Annex A presents some parts of the code that was written for the application that detect fraudulent VoIP calls and Annex B sections of the files that handled by the two applications of this work.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής εργασίας, κο Αναστάσιο Νταγιούκλα, Επίκουρο Καθηγητή στη Σχολή Θετικών Επιστημών και Τεχνολογίας του Ελληνικού Ανοικτού Πανεπιστημίου, για την εμπιστοσύνη που μου έδειξε στην εκπόνηση της συγκεκριμένης εργασίας, για την άψογη συνεργασία και την πολύτιμη καθοδήγησή του κατά τη διάρκεια της εκπόνησής της.

Επίσης θα ήθελα να ευχαριστήσω τους καθηγητές του Ανοικτό Πανεπιστήμιο της Κύπρου για το υψηλό επίπεδο σπουδών που παρείχαν και κράτησαν το ενδιαφέρον μου για το παρόν μεταπτυχιακό αμείωτο.

Ένα μεγάλο ευχαριστώ, τέλος, ανήκει στην οικογένειά μου και ιδιαίτερα τη σύζυγό μου, χωρίς την αμέριστη συμπαράσταση και κατανόηση της οποίας, δεν θα ήταν δυνατή η ολοκλήρωση αυτής της εργασίας.

Περιεχόμενα

Αρκτικόλεξα και συντομογραφίες

ANN	Artificial Neural Networks
CDR	Call Detail Record ή Charging Data Record
CPU	Central Processing Unit
DSL	Digital Subscriber Line
IDS	Intrusion Detections Systems
IP	Internet Protocol
IP Telephony	Internet Protocol telephony, Voice over IP Telephony
ISPs	Internet Service Providers
MAC Address	Media Access Control Address
MG	Media Gateway
MGC	Media Gateway Controller
PBX	Private Branch Exchange
PIN	Personal Identification Number
PSTN	Public Switched Telephone Network
RTP	Real Time Transport Protocol
S/MIME	Secure/Multipurpose Internet Mail Extensions,
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SOFM	Self-Organizing Feature Maps
SOM	Self-Organizing Maps (Kohonen Networks)
SS7	Signalling System 7
TFTP	Trivial File Transfer Protocol
UDP	User Datagram Protocol
VLAN	Virtual Local Area Network
VoIP	Voice over IP

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Κεφάλαιο 1
Εισαγωγή

Η IP τηλεφωνία, κοινώς γνωστή ως Voice over IP (VoIP), αναδύεται ως μια βιώσιμη εναλλακτική λύση στα παραδοσιακά τηλεφωνικά συστήματα. Ένας σημαντικός αριθμός εταιρειών τηλεπικοινωνίας έχουν ενσωματώσει τη χρήση της τεχνολογίας Voice-over-IP (VoIP), ενώ άλλες είναι υπό σκέψη να το πράξουν. Η τεχνολογία αυτή, ωστόσο, πάσχει από διάφορα κενά στην ασφάλεια. Τα κλασικά VoIP πρωτόκολλα σηματοδοσίας H.323 και Session Initiation Protocol (SIP) [02], και τα πρωτόκολλα μεταφοράς πολυμέσων: RTP και RTCP θα μπορούσαν να γίνουν στόχος των καλά συνδεδεμένων (well connected) επιθέσεων, όπως Denial of Service (DoS), υποκλοπές, δόλιας χρήσης, κλπ. Έχουν προταθεί για ασφαλή VoIP επικοινωνία οι κρυπτογραφικές λύσεις, π.χ., το πρωτόκολλο ZRTP, αλλά οι εταιρείες χρησιμοποιούν διαφορετικές υλοποιήσεις στα πωληθέντα συστήματά τους. Επιπλέον, οι λύσεις αυτές απαιτούν μια αρχή έκδοσης πιστοποιητικών και σημαντικές γνώσεις στην εγκατάσταση των VoIP τηλεφώνων και VoIP εξυπηρετητών (Servers). Ως εκ τούτου, τα ζητήματα συμβατότητας και οι σημαντικές τεχνικές και οικονομικές διαφορές επηρεάζουν τις μικρές και μεσαίες επιχειρήσεις στη χρήση της κρυπτογραφίας ως λύση [01]. Αν κι έγιναν έρευνες σε VoIP συστήματα, για την ανίχνευση εισβολής σε επίπεδο εφαρμογής, μεταφοράς και δρομολόγησης, στη βιομηχανία, αλλά και στην ακαδημαϊκή κοινότητα, η δυνατότητα εξακρίβωσης μίας εισβολής που εισέρχεται ή υλοποιείται σε υποδομές VoIP βρίσκεται ακόμη σε πρώιμο στάδιο. Αυτό συμβαίνει επειδή τα

περισσότερα συστήματα ανίχνευσης εισβολέων σε VoIP βασίζονται στα υπάρχοντα Τείχη Προστασίας (Firewalls) του δικτύου που αγνοούν κάποια κρίσιμα χαρακτηριστικά των δικτύων VoIP. Για παράδειγμα, ο αλγόριθμος Back propagation στα νευρωνικά δίκτυα για την ανίχνευση εισβολής σε συστήματα VoIP, ο οποίος τροποποιείται από φίλτρα πακέτων SPAM mail/Internet, έχει υιοθετηθεί σε πολλές εθνικού επιπέδου (Κίνα) υποδομές softswitch. Πείραμα εξομοίωσης επιθέσεων από πολλούς γνωστούς παρόχους υπηρεσιών VoIP έχει αποδείξει ότι μια τέτοια δομή εύκολα προκαλεί αρνητικά σφάλματα και μειώνει τις επιδόσεις λόγω των εσωτερικών παγίδων της, όπως τοπικά ελάχιστα και σύγκλισης [03].

Στην παρούσα εργασία μελετήθηκαν και αξιολογήθηκαν σενάρια απάτης από συνδρομητές VoIP. Τα αποτελέσματα οδήγησαν στην ανάπτυξη ενός μηχανισμού clustering για την ανίχνευση απάτης σε κλήσεις αναλύοντας δεδομένα CDR από έναν VoIP τηλεπικοινωνιακό πάροχο .

Η εργασία ξεκίνησε από το αρχείο τριών μηνών με τηλεφωνικές συνδιαλέξεις που μας έδωσε ένας τηλεπικοινωνιακός πάροχος VoIP υπηρεσιών. Το αρχείο περιέχει τα βασικά στοιχεία, για την έναρξη και τη λήξη μίας τηλεφωνικής επικοινωνίας, όπως την καταγράφει το σύστημα VoIP του παρόχου. Δεν δόθηκαν στοιχεία φωνητικά. Τα δεδομένα του παρόχου αποθηκεύονται σε βάση δεδομένων. Για τη μεταφορά τους έγινε εξαγωγή αυτών σε μορφή XML. Τα αρχεία, αποτέλεσαν τη βάση για τη δημιουργία μίας εφαρμογής που θα παράγει τηλεφωνικές κλήσεις. Η εφαρμογή είναι γραμμένη σε γλώσσα προγραμματισμού Python και είναι πλήρως παραμετροποιήσιμη. Με τη χρήση της γεννήτριας τυχαίων κλήσεων, δημιουργήθηκαν εγγραφές τηλεφωνικών συνδιαλέξεων για χρονικό διάστημα δύο μηνών.

Η κύρια εφαρμογή είναι κι αυτή γραμμένη σε γλώσσα Python και χρησιμοποιεί τρεις αλγόριθμους συσταδοποίησης για την εξιχνίαση απάτης σε τηλεπικοινωνιακό πάροχο. Γίνεται χρήση του αλγόριθμου K-Mean,του Αυτό-Οργανωμένου Χάρτη του Kohonen και του Self-Organizing Feature Map.

Η εργασία προσπαθεί να απαντήσει στο ερώτημα εάν είναι δυνατή η ενημέρωση του διαχειριστή συστημάτων VoIP για πιθανή δόλια κλήση. Οι παραπάνω αλγόριθμοι θα συγκριθούν και τα αποτελέσματά τους θα μας δώσουν την απάντηση του παραπάνω ερωτήματος.

Κεφάλαιο 2
VoIP – Αρχιτεκτονική και Πρωτόκολλα

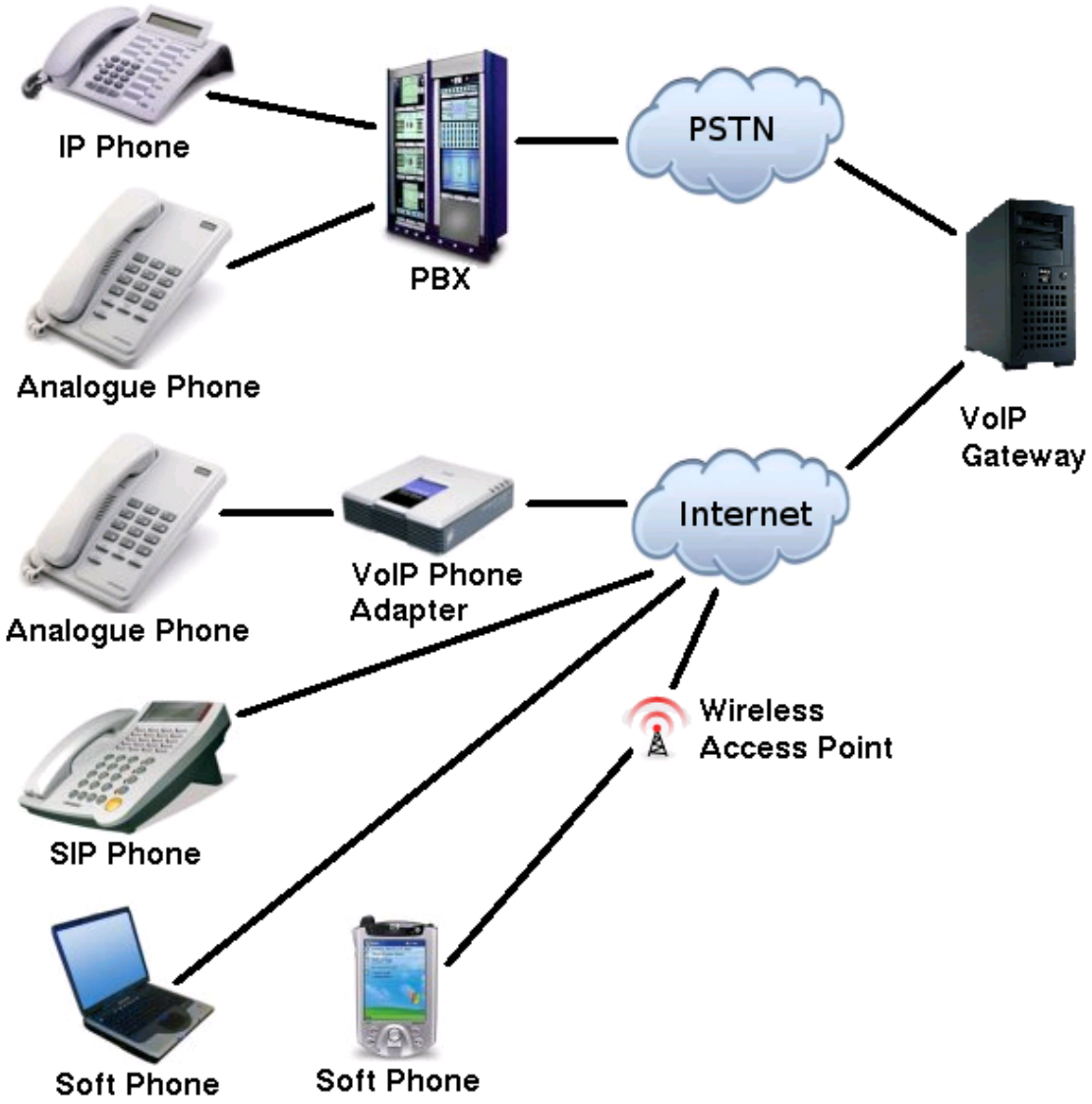
Τα τηλεφωνικά συστήματα έχουν εξελιχθεί κατά τον τελευταίο αιώνα. Η εξέλιξη αυτή αποτελεί μετάβαση από τα αναλογικά στα ψηφιακά συστήματα και από την τεχνολογία μεταγωγής κυκλώματος στα συστήματα μεταγωγής πακέτων. Η ανάπτυξη του Διαδικτύου την τελευταία δεκαετία, μαζί με την υπόσχεση της μείωσης του κόστους για τον πελάτη, έχει οδηγήσει στην ταχεία ανάπτυξη της επικοινωνίας με φωνή μέσω του πρωτοκόλλου Internet (Voice over IP, VoIP) [06]. Αυτή η αύξηση έχει επηρεασθεί ακόμη περισσότερο από την ταχεία διείσδυση της ευρυζωνικότητας σε όλο τον κόσμο. Ως μια εφαρμογή πραγματικού χρόνου που εξυπηρετείται από το Διαδίκτυο (Internet), η υπηρεσία VoIP αντιμετωπίζει πολλές προκλήσεις, όπως η διαθεσιμότητα, η ποιότητα της φωνής, καθώς και η ασφάλεια του δικτύου.

Στο κεφάλαιο αυτό θα παρουσιαστεί η αρχιτεκτονική και τα πρωτόκολλα που απαιτούνται για την υλοποίηση ενός VoIP δικτύου. Θα γίνει αναφορά στους διάφορους λόγους που επηρεάζουν θετικά την υψηλή ποιότητα που προσφέρουν οι κλήσεις VoIP. Στη συνέχεια θα περιγραφούν διάφοροι κωδικοποιητές (codecs) που χρησιμοποιούνται και οι συμβιβασμοί στην τεχνολογία που απαιτούνται μεταξύ της καθυστέρησης (delay) και του εύρους ζώνης (bandwidth). Με αφορμή τους κωδικοποιητές που χρησιμοποιούνται για την εγκαθίδρυση κλήσεων VoIP θα γίνει

αναφορά στα τέσσερα βασικότερα πρωτόκολλα σηματοδοσίας που χρησιμοποιούνται στην υπηρεσία VoIP κι έχουν επικρατήσει στο χώρο των τηλεπικοινωνιών.

2.1 Αρχιτεκτονική

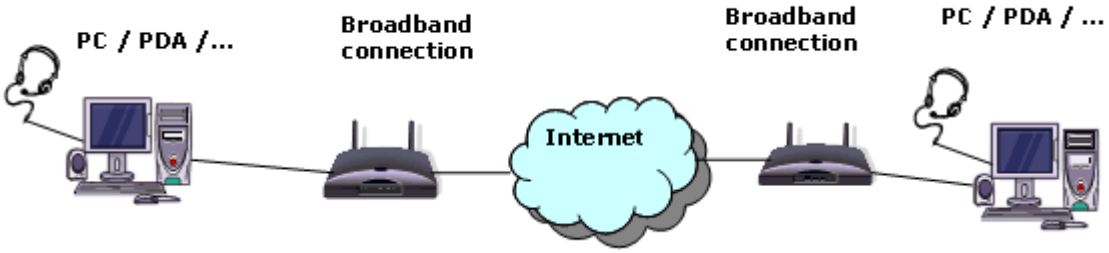
Οι κλήσεις μέσω της υπηρεσίας VoIP μπορούν να υλοποιηθούν μεταξύ ηλεκτρονικών υπολογιστών (PC to PC), τηλεφωνικών συσκευών (Phone to Phone) και τηλεφώνου με ηλεκτρονικό υπολογιστή (phone to PC, PC to phone). Στην παρακάτω εικόνα φαίνεται η βασική της δομή.



Εικόνα 2.1 Η βασική δομή για την παροχή υπηρεσιών VoIP.

2.1.1 Μεταξύ ηλεκτρονικών υπολογιστών

PC to PC call, είναι ένας όρος στη VoIP τεχνολογία, με τον οποίο δηλώνεται ένα επικοινωνιακό σύστημα μεταξύ ηλεκτρονικών υπολογιστικών συστημάτων. Όταν στη δεκαετία του 1990 αναπτύχθηκε η τεχνολογία VoIP, τα πλέον γνωστότερα σε χρήση ηλεκτρονικά υπολογιστικά συστήματα ήταν οι προσωπικοί υπολογιστές (PC). Ίσως αυτό να αποτελεί το λόγο για τον οποίο ο συγκεκριμένος τύπος της τεχνολογίας VoIP πήρε και το όνομα Pc to PC κλήσεις. Στην παρακάτω εικόνα μπορούμε να δούμε τα υλικά που απαιτούνται γι' αυτόν τον τύπο ζεύξης.



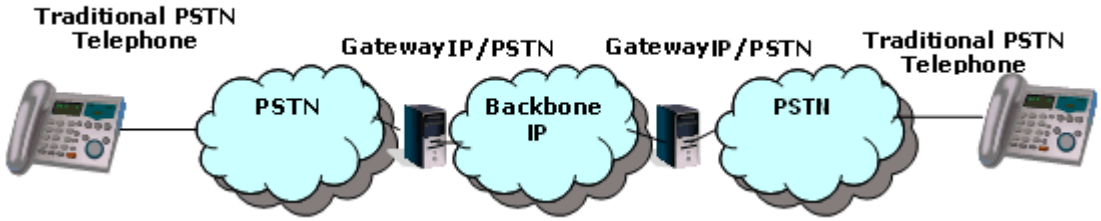
Εικόνα 2.2 Παροχή υπηρεσιών VoIP σε PC to PC.

Με τη χρήση της συγκεκριμένης υπηρεσίας μπορούμε να χρησιμοποιήσουμε τον υπολογιστή μας που έχει συνδεθεί στο Διαδίκτυο για να κάνουμε κλήσεις σε άλλους υπολογιστές που χρησιμοποιούν την ίδια υπηρεσία. Μπορούμε να βρούμε πολλούς παρόχους VoIP στο διαδίκτυο. Μία από τις υπηρεσίες που υποστηρίζει φωνητικές κλήσεις μέσω του Διαδικτύου είναι το Yahoo Messenger. Με τη χρήση του Yahoo Messenger μπορούμε να εγκαθιδρύσουμε μία φωνητική κλήση με άλλους χρήστες που χρησιμοποιούν κι αυτοί την ίδια υπηρεσία. Μπορούμε να επιλέξουμε μια εναλλακτική υπηρεσία που υποστηρίζει φωνητικές κλήσεις μέσω του Διαδικτύου, όπως το Viber, το Google Hangouts , το Skype και πολλά άλλα λογισμικά. Η κλήση VoIP PC to Pc μπορεί να γίνει δωρεάν, κάνοντας χρήση μιας οποιασδήποτε σύνδεσης στο Internet (Dial Up, DSL,Wi-Fi κ.α.) με τον υπολογιστή μας.

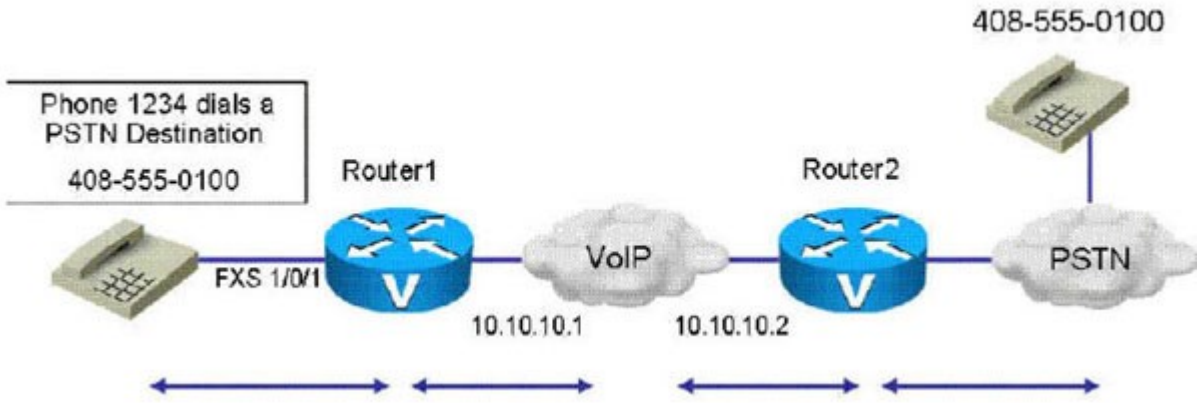
2.1.2 Μεταξύ τηλεφωνικών συσκευών

Η υπηρεσία εκτελείται χρησιμοποιώντας ένα ειδικό τηλέφωνο (IP Phone, SIP Phone) ή συμβατικό τηλέφωνο (Analogue Phone) που συνδέεται με τον προσαρμογέα VoIP (VoIP Phone Adapter). Για να χρησιμοποιήσουμε την υπηρεσία θα πρέπει να συνδεθούμε σε έναν πάροχο VoIP υπηρεσιών (VoIP Provider). Στην Ελλάδα υπάρχουν πολλοί πάροχοι που μπορούν να μας υποστηρίξουν όπως η Κυπριακή εταιρεία Cyta, οι Ελληνικές εταιρείες Viva.gr, OmniNet και άλλες.

Με την υπηρεσία αυτή μπορούμε να πραγματοποιήσουμε κλήσεις οπουδήποτε στον κόσμο, χρησιμοποιώντας τα εργαλεία που μας δίνει ο πάροχος. Δείγμα αυτού του τύπου υπηρεσιών VoIP δίνεται στις επόμενες δύο εικόνες.



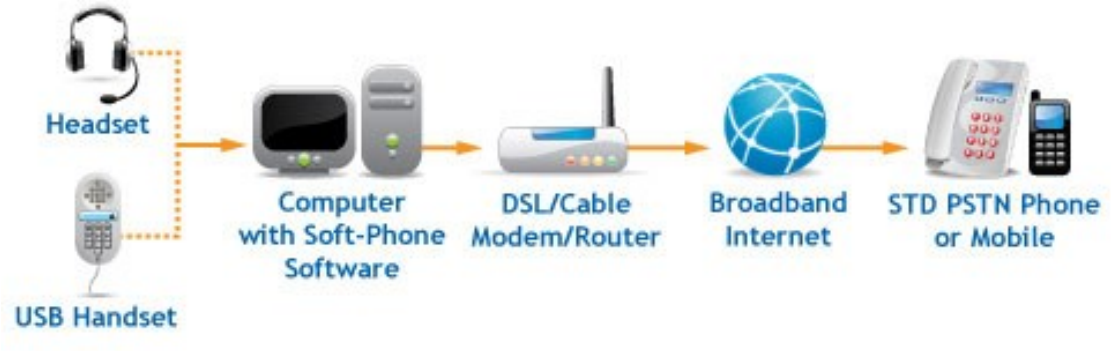
Εικόνα 2.3 Παροχή υπηρεσιών VoIP σε αναλογικές τηλεφωνικές συσκευές



Εικόνα 2.4 Παροχή υπηρεσιών VoIP μεταξύ VoIP τηλεφώνου και PSTN τηλεφώνου

2.1.3 Μεταξύ τηλεφώνου και ηλεκτρονικού υπολογιστή

Όπως και στην παρακάτω εικόνα, αυτός ο τύπος κλήσης μας επιτρέπει την φωνητική ζεύξη ενός ηλεκτρονικού υπολογιστή με μία τηλεφωνική συσκευή, είτε πρόκειται για σταθερό τηλέφωνο (Analogue phone) ή κινητό τηλέφωνο (mobile phone). Ο συγκεκριμένος τύπος απαιτεί τη σύνδεση σε πάροχο φωνητικών υπηρεσιών στο διαδίκτυο. Πάροχοι VoIP που επιτρέπουν αυτού του είδους συνδέσεις υπάρχουν αρκετοί, όπως για παράδειγμα η εταιρεία Microsoft μέσω του τμήματός της Skype (πρόσφατα η εταιρεία Microsoft εξαγόρασε την εταιρεία Skype).

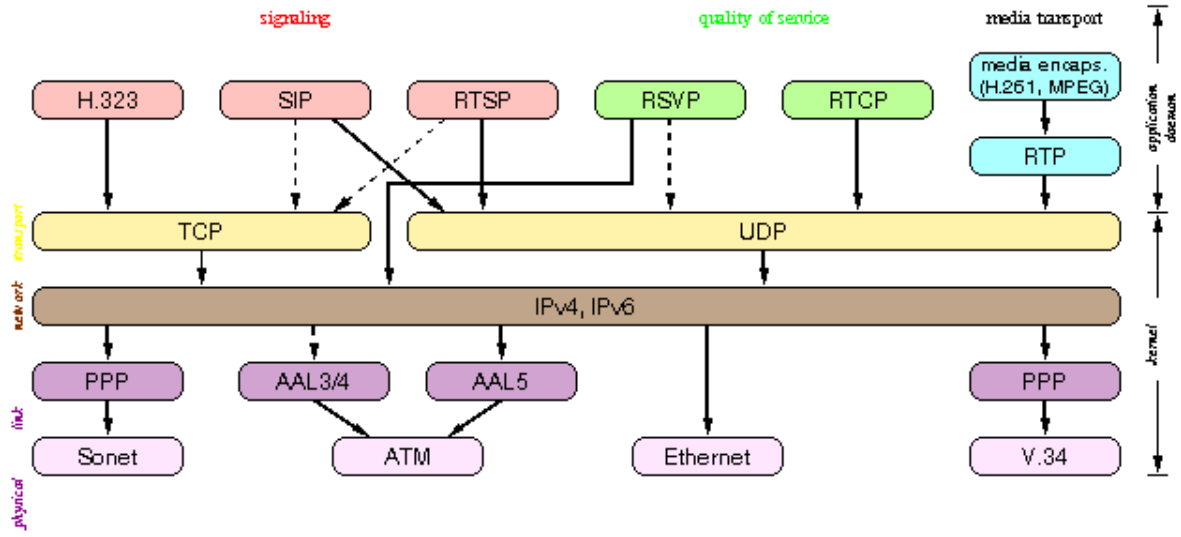


Εικόνα 2.5 Παροχή υπηρεσιών VoIP μεταξύ PC και PSTN τηλεφώνου

Ο συνδυασμός κλήσεων VoIP μεταξύ τηλεφώνων ή μεταξύ τηλεφώνου και ηλεκτρονικού υπολογιστή δεν είναι δωρεάν όπως γίνεται μεταξύ των υπολογιστών. Η υπηρεσία παρέχεται προπληρώνοντας ένα χρηματικό ποσό όπως και στην καρτοκινητή τηλεφωνία.

2.2 Πρωτόκολλα

Από τη δεκαετία του 1990, κυριάρχησε η πλατφόρμα TCP/IP ως αρχιτεκτονική στο διαδίκτυο, από την άλλη η υπηρεσία VoIP χρησιμοποιεί ως πλατφόρμα το RTP/UDP/IP. Η εικόνα 2.6 δίνει την πλήρη αρχιτεκτονική που στηρίζει το διαδίκτυο.



Εικόνα 2.6 Αρχιτεκτονική του διαδικτύου [07]

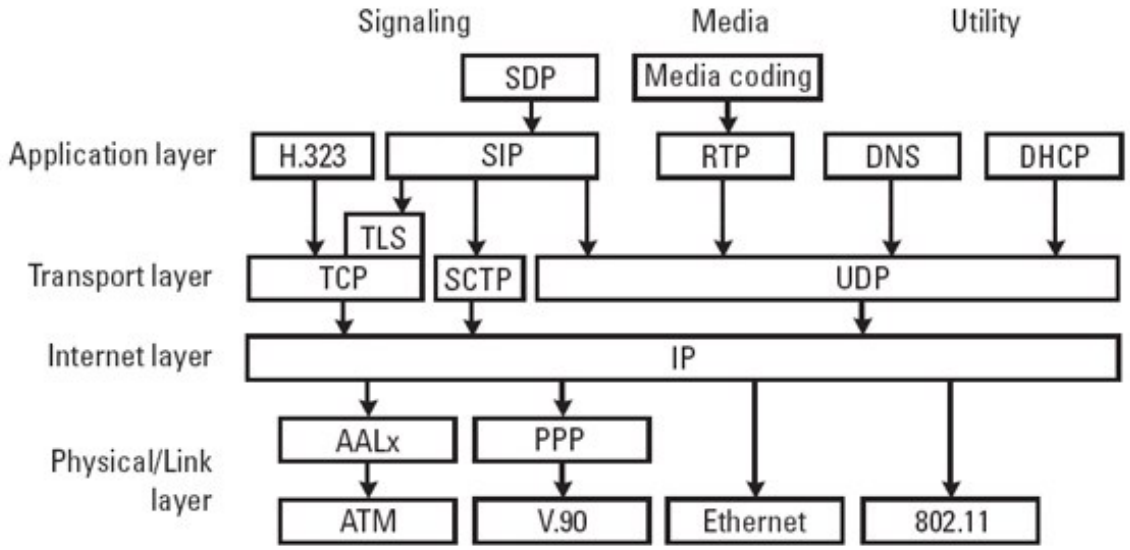
Το Internet Protocol (IP) ασχολείται μόνο με την μη κατευθυνόμενη παράδοση των πακέτων, η οποία βασίζεται σε μια υπηρεσία βέλτιστης προσπάθειας.

Το πρωτόκολλο Real-Time (RTP), που χρησιμοποιείται σε συνδυασμό με το UDP, παρέχει από άκρη σε άκρη (end-to-end) λειτουργίες μεταφοράς πακέτων για εφαρμογές. Μεταδίδει τα δεδομένα σε πραγματικό χρόνο, όπως τον ήχο και το βίντεο, πάνω από unicast και multicast υπηρεσίες δικτύου[08]. Το RTP σε κάθε πακέτο, συμπεριλαμβάνει τους αριθμούς ακολουθίας και τα χρονοσήματα (timestamps), τα οποία είναι χρήσιμα σε εφαρμογές πολυμέσων. Θα πρέπει να τονιστεί ότι το RTP από μόνο του δεν παρέχει κανένα μηχανισμό για να εξασφαλιστεί η έγκαιρη παράδοση των δεδομένων ή άλλες εγγυήσεις ποιότητας υπηρεσιών [09]. Πραγματικά η ενθυλάκωση του RTP μέσα στα πακέτα IP μπορεί να φανεί μόνο στον δέκτη. Δεν υπάρχει δυνατότητα αναγνώρισης του RTP πακέτου μέσα στο IP από τους ενδιάμεσους δρομολογητές.

Το πρωτόκολλο RTCP συγγενικό του RTP χρησιμοποιείται για την δημιουργία στατιστικών σε μια VoIP σύνδεση. Υποστηρίζει τα εξής:

- Ο αποστολέας στέλνει αναφορά σχετικά με τον αριθμό των πακέτων που έχει στείλει σε ένα συγκεκριμένο χρονικό διάστημα.
- Ο παραλήπτης στέλνει αναφορά σχετικά με τον αριθμό των πακέτων που έχει λάβει σε ένα συγκεκριμένο χρονικό διάστημα.
- Ανταλλάσσει πληροφορίες για τις απώλειες και τις καθυστερήσεις μεταξύ των δεκτών.
- Στέλνει πακέτα κατά διαστήματα με βάση τον αριθμό των δεκτών και το διαθέσιμο εύρος ζώνης.

Ωστόσο, μια συνεχής ροή των πακέτων RTP/UDP/IP προσφέρεται στις περισσότερες εφαρμογές VoIP, όπως εμφανίζεται στην επόμενη εικόνα.



Εικόνα 2.7 Αρχιτεκτονική των VoIP πρωτοκόλλων

Στην τεχνολογία VoIP υπάρχουν peer-to-peer πρωτόκολλα ελέγχου σηματοδοσίας όπως οι πλατφόρμες των πρωτοκόλλων H.323 [10] και SIP [02] αλλά και master-slave πρωτόκολλα ελέγχου σηματοδοσίας όπως το Media Gateway Πρωτόκολλο Ελέγχου (MGCP) [11], και Megaco/H.248 [13].

2.2.1 Ποιότητα υπηρεσιών (QoS)

Για να γίνει αποδεκτή η τεχνολογία VoIP οφείλει να προσφέρει παρόμοια ποιότητα με το κλασικό δημόσιο τηλεφωνικό δίκτυο (PSTN). Για την παροχή υπηρεσιών VoIP προϋποθέτονται οι ακόλουθες απαιτήσεις:

- Από άκρη σε άκρη καθυστέρηση (End to End Delay): Είναι η καθυστέρηση μεταξύ της μετάδοσης ενός πακέτου φωνής έως τη λήψη του. Καλείται επίσης η μονόδρομη από άκρο σε άκρο καθυστέρηση. Θα πρέπει να είναι μικρότερη από περίπου 150ms [14], [15] για να διατηρηθεί η καλή ποιότητα στη φωνή και στη συζήτηση μεταξύ των χρηστών. Η συμπίεση της ομιλίας, το πακετάρισμα των δεδομένων, η κωδικοποίηση και η μετάδοση του πακέτου έως τη στιγμή που θα φθάσει στην ουρά του δρομολογητή, είναι τα σημεία που παρατηρείται το μεγαλύτερο τμήμα της καθυστέρησης από άκρο σε άκρο.

- Η διακύμανση της καθυστέρησης ή Jitter: Είναι η παράμετρος που καθορίζει τη χρονική διακύμανση στις αφίξεις των πακέτων μεταξύ τους στο δέκτη. Στις εφαρμογές πραγματικού χρόνου, το Jitter θα πρέπει να είναι μικρότερο από 1 ms [14], [15].

- Ποσοστό Σφάλματος σε Πλαίσιο (Frame error rate, FER): Το ποσοστό σφαλμάτων στη λήψη των πλαισίων υποδηλώνει το ρυθμό των ανολοκλήρωτων ή εσφαλμένων πλαισίων που λαμβάνονται από τον δέκτη. Για φωνητικές κλήσεις, το FER μπορεί να είναι αρκετά υψηλό σε σύγκριση με εφαρμογές όπως της ανάγνωσης ηλεκτρονικών μηνυμάτων (eMail Client) ή περιηγητές στο διαδίκτυο (Web Browser). Πράγματι, η συνομιλία μπορεί να είναι κατανοητή ακόμη και σε ποσοστό σφάλματος 3% [14], [15]. Στις υπηρεσίες VoIP συνήθως το αποδεκτό ποσοστό σφάλματος είναι το 1%.

2.2.2 Πρωτόκολλα σηματοδοσίας στην υπηρεσία VoIP

Η διαχείριση των VoIP ζεύξεων γίνεται με τα πρωτόκολλα σηματοδοσίας. Τέσσερα πρωτόκολλα σηματοδοσίας κυριαρχούν στην υπηρεσία VoIP: Τα πρωτόκολλα αυτά είναι το SIP (Session Initiation Protocol), το H.323, το MGCP-Media Gateway Control Protocol και το MEGACO (MEdia GAteway and COntrol) ή αλλιώς H.248.

2.2.2.1 Session Initiation Protocol (SIP)

Η Voice-over-IP (VoIP) επικοινωνία βασίζεται στο Session Initiation Protocol (SIP) το οποίο κι έχει εγκαθιδρυθεί ως το απόλυτο πρωτόκολλο για τη φωνητική επικοινωνία . Σημειώνει συνεχή ανάπτυξη με την υποστήριξή του από ανοικτές πλατφόρμες βασισμένες στο Internet Protocol [04]

Πέντε χαρακτηριστικά του πρωτοκόλλου SIP του δίνουν τη δυνατότητα να υποστηρίξει τις απαιτήσεις για τη δημιουργία και τον τερματισμό των VoIP κλήσεων [02]:

- Η θέση του χρήστη (User Location): Στη VoIP επικοινωνία που θα επιτευχθεί προηγείται η εύρεση της θέσης του δέκτη.

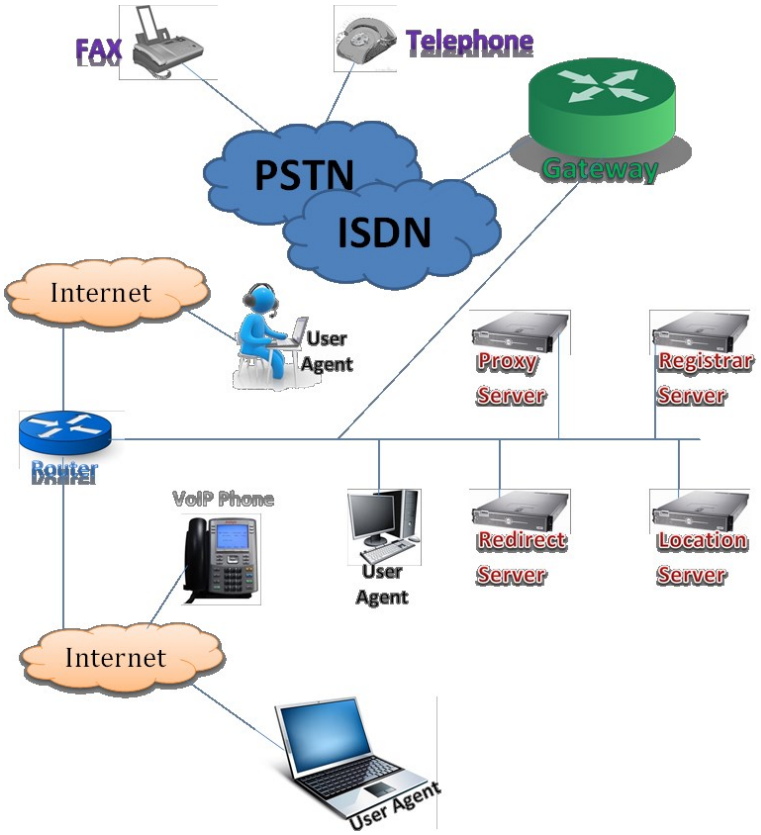
- Δυνατότητες του χρήστη (User Capabilities): Προσδιορίζονται όλες οι παράμετροι των πολυμεσικών συσκευών που θα συμμετέχουν στην επικοινωνία.

- Διαθεσιμότητα του χρήστη (User Availability): Προσδιορίζεται η επιθυμία του δέκτη να απαντήσει στην κλήση VoIP.

- Ρυθμίσεις κλήσεως (Call Setup): Ρύθμιση των τηλεπικοινωνιακών παραμέτρων μεταξύ του πομπού και του δέκτη για την μεταξύ τους επικοινωνία.

- Χειρισμός κλήσεων (Call Handling): Διαχείριση της μεταβίβασης των δεδομένων της κλήσης και τερματισμού αυτής.

Η βασική αρχιτεκτονική του SIP, όπως απεικονίζεται και στην Εικόνα 2.8 βασίζεται σε ένα μοντέλο πελάτη-εξυπηρετητή. Οι κύριες συσκευές για την επίτευξη της επικοινωνίας με το πρωτόκολλο SIP είναι η συσκευή/εφαρμογή του χρήστη που καλεί ή απαντά σε μία κλήση(User Agent), η SIP πύλη (Gateway), ο διακομιστής μεσολάβησης (Proxy Server), ο διακομιστής τοποθεσιών (Location Server), ο διακομιστής ανακατεύθυνσης (Redirector Server) και ο διακομιστής που επικυρώνει τα αιτήματα (Registrar Server):



Εικόνα 2.8 Συστήματα που απαιτούνται στο πρωτόκολλο SIP

- Ο User Agent (UA) είναι μία εφαρμογή πελάτη που στέλνει αιτήματα SIP και ταυτόχρονα μια εφαρμογή εξυπηρετητή που δέχεται τα αιτήματα SIP.

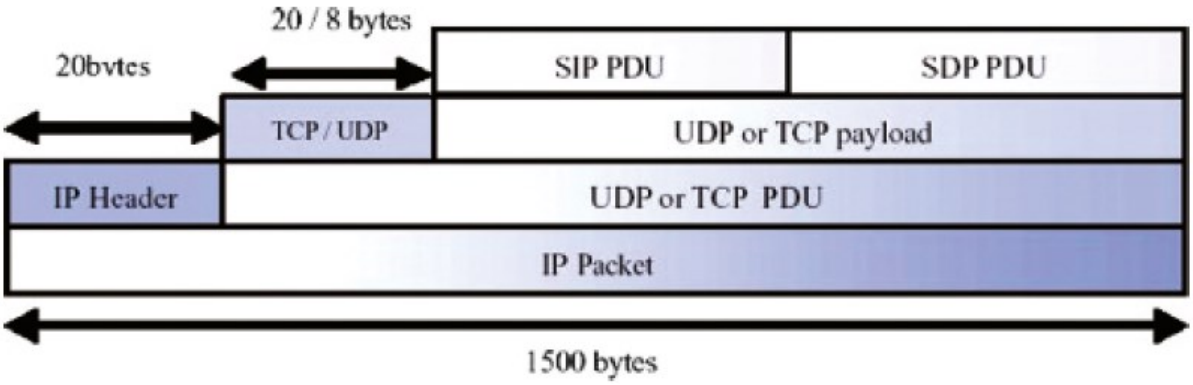
- Η SIP πύλη (SIP Gateway) είναι μια εφαρμογή που διασυνδέει το δίκτυο SIP με ένα άλλο δίκτυο που χρησιμοποιεί διαφορετικό πρωτόκολλο σηματοδοσίας.

- Ο διακομιστής μεσολάβησης (Proxy server) παραλαμβάνει τα αιτήματα, καθορίζει σε ποιον εξυπηρετητή να τα στείλει και στη συνέχεια τα διαβιβάζει.

- Ο διακομιστής ανακατεύθυνσης (Redirector Server) δεν μεταφέρει τα αιτήματά σε άλλους διακομιστές. Ανταποκρίνεται μόνο στα αιτήματα των πελατών και τους ενημερώνει ποιος θα είναι ο επόμενος διακομιστής ώστε να επικοινωνήσουν.

- Ο διακομιστής τοποθεσίας (Location Server) δίνει πληροφορίες στον διακομιστή ανακατεύθυνσης και τον διακομιστή μεσολάβησης, σχετικά με την πιθανή τρέχουσα θέση του δημιουργού της κλήσης .

- Ο διακομιστής επικύρωσης αιτημάτων (Registrar Server) είναι ένας διακομιστής που δέχεται το αίτημα REGISTER, το οποίο περιέχει τις SIP διευθύνσεις και τις αντίστοιχες IP διευθύνσεις. Ο User Agent στέλνει ένα μήνυμα επικύρωσης (registration message) στον διακομιστή Registrar , ο οποίος είναι επιφορτισμένος να το μεταβιβάσει στον διακομιστή τοποθεσίας ως ένα νέο αίτημα. Μόλις μεταβιβασθεί η πληροφορία, ο διακομιστής Registrar στέλνει την κατάλληλη απάντηση πίσω στον User Agent.



Εικόνα 2.9 Το πρωτόκολλο SIP ενθυλακωμένο σε ένα πακέτο IP

Ο κύριος σκοπός του πρωτοκόλλου SIP είναι να εγκαθιστά τις συνεδρίες μεταξύ δύο User Agents. Στην Εικόνα 2.9, το πρωτόκολλο SIP φαίνεται πως λειτουργεί μαζί με το πρωτόκολλο Session Description (SDP) το οποίο είναι υπεύθυνο να περιγράφει τη σύνοδο πριν αυτή ξεκινήσει. Τα μηνύματα SIP μπορούν να ταξιδέψουν ενθυλακωμένα σε UDP ή σε πρωτόκολλο ελέγχου μεταφοράς (TCP):

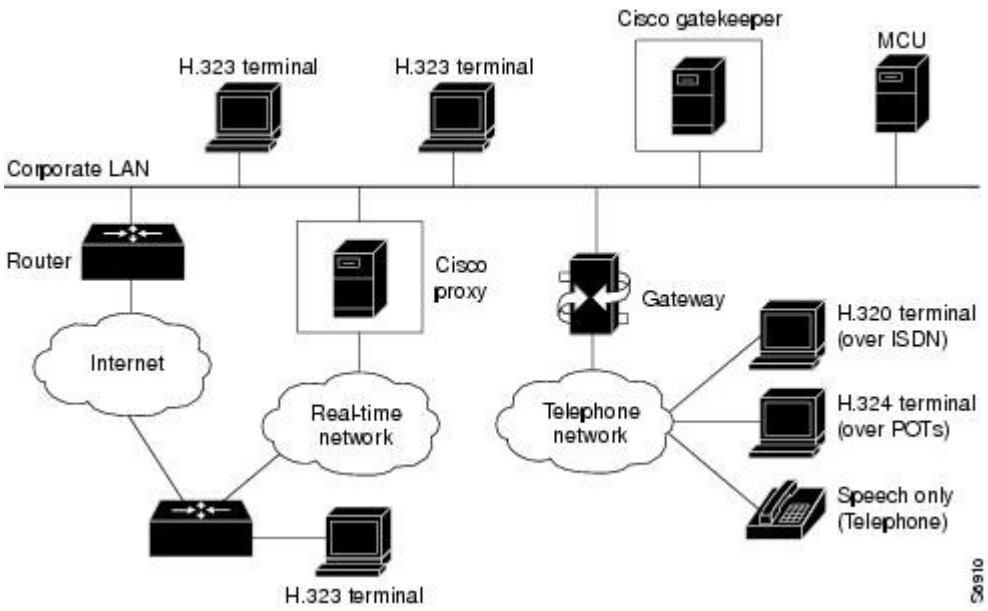
- Όταν το πρωτόκολλο SIP ενθυλακώνεται στο TCP, το επίπεδο μεταφοράς παρέχει αξιοπιστία. Το TCP είναι ένα κατευθυνόμενης σύνδεσης (connection-oriented) πρωτόκολλο και διατηρεί μια εικονική σύνδεση μεταξύ των τερματικών σημείων. Για να διασφαλιστεί ότι τα δεδομένα του έχουν ληφθεί σωστά, το TCP απαιτεί μια αναγνώριση (ACK) από το μηχάνημα προορισμού. Εάν η κατάλληλη ACK που καθορίζεται από τον

αύξοντα αριθμό της δεν έχει ληφθεί εντός ορισμένης χρονικής στιγμής, το πακέτο αναμεταδίδεται. Όταν το δίκτυο είναι κορεσμένο, αυτή η αναμετάδοση οδηγεί σε πολλαπλά πακέτα που στέλνονται. Ωστόσο, το μηχάνημα λήψης χρησιμοποιεί τον αύξοντα αριθμό του πακέτου για να διαπιστωθεί εάν είναι ένα αντίγραφο και το απορρίπτει αν είναι απαραίτητο.

- Όταν το πρωτόκολλο SIP ενθυλακώνεται στο UDP, οι αξιόπιστες διαδικασίες παράδοσης του πακέτου εξασφαλίζονται από το πρωτόκολλο SIP. Το πρωτόκολλο UDP είναι το πλέον διαδεδομένο πρωτόκολλο μεταφορών πακέτων SIP [02].

Το SIP είναι ένα πρωτόκολλο συναλλαγής, υπό την έννοια ότι μια συναλλαγή SIP αποτελείται από μια μόνο αίτηση και τυχόν απαντήσεις σε αυτή την αίτηση. Η δημιουργία μιας συνεδρίας με τη χρήση του πρωτοκόλλου SIP αποτελείται από μια αποστολή INVITE και μια συναλλαγή ACK.

2.2.2.2 H.323



Εικόνα 2.10 Η δομή του πρωτοκόλλου H.323

Για να κατανοήσουμε την αρχιτεκτονική του πρωτοκόλλου H.323, που φαίνεται στην Εικόνα 2.10 θα πρέπει να κατανοήσουμε τα επιμέρους μέρη που το απαρτίζουν:

- Το H.323 τερματικό (Terminal) είναι η συσκευή του τελικού χρήστη και παρέχει υπηρεσίες VoIP.

- Η Η.323 πύλη (Gateway) υλοποιεί τη μετατροπή της μορφής των δεδομένων άλλων δικτύων στη μορφή που απαιτεί το πρωτόκολλο Η.323 κι ελέγχει τη μετατροπή της σηματοδοσίας προς άλλα δίκτυα.

- Η πολλαπλών ελέγχων μονάδα (Multiple Control Unit, MCU) δίνει τη δυνατότητα συνδιάσκεψης πολλών χρηστών. Μεταξύ τριών και περισσοτέρων τερματικών συσκευών ή/και πυλών. Υποστηρίζει τη διαπραγμάτευση των δυνατοτήτων με όλους τους άλλους τερματικούς σταθμούς, προκειμένου να διασφαλιστεί ένα κοινό επίπεδο επικοινωνίας. Η MCU ως δυνατότητα είναι προαιρετική σε ένα δίκτυο Η.323.

- Ο Η.323 φύλακας της πύλης (Gatekeeper) είναι υπεύθυνος για την εγγραφή των χρηστών και παρέχει τη δυνατότητα μετάφρασης των διευθύνσεων κι ελέγχει τις υπηρεσίες των κλήσεων των τερματικών συσκευών του Η.323 . Είναι επίσης υπεύθυνος για τον έλεγχο του εύρους ζώνης που επιτρέπει στις τερματικές συσκευές να αλλάζουν το διαθέσιμο εύρος ζώνης τους στο τοπικό δίκτυο. Ο Η.323 Gatekeeper παρέχει τις ακόλουθες υπηρεσίες:

    - Μετάφραση των Διευθύνσεων: Ο Η.323 Gatekeeper διατηρεί μια βάση δεδομένων για τη μετάφραση των ψευδώνυμων, όπως οι διεθνείς τηλεφωνικοί αριθμοί, με τις διευθύνσεις του δικτύου.

    - Έλεγχο στην είσοδο και την πρόσβαση στις τερματικές συσκευές: ασχολείται με την καταχώρηση των δικαιωμάτων και των προνομίων των τερματικών συσκευών.

    - Διαχείριση του εύρους ζώνης: οι διαχειριστές του δικτύου μπορούν να διαχειριστούν το εύρος ζώνης θέτοντας περιορισμούς στον αριθμό των ταυτόχρονων κλήσεων και αφαιρώντας την άδεια σε ορισμένες τερματικές συσκευές να πραγματοποιούν κλήσεις σε συγκεκριμένες χρονικές περιόδους.

    - Ικανότητα δρομολόγησης: Ο Η.323 Gatekeeper μπορεί να δρομολογήσει όλες τις κλήσεις που ξεκινούν (η τερματική συσκευή είναι ο καλών) ή τερματίζουν (η τερματική συσκευή είναι ο δέκτης μίας κλήσης) στη ζώνη του. Αυτή η δυνατότητα παρέχει πολλά πλεονεκτήματα όπως: (i) τα λογιστικά στοιχεία των κλήσεων που μπορούν να διατηρηθούν για τους σκοπούς της τιμολόγησης και της ασφάλειας. (ii) την επαναδρομολόγηση των κλήσεων στην κατάλληλη πύλη,

ανάλογα με τη διαθεσιμότητα του εύρους ζώνης ή την προώθηση της κλήσης σε άλλη τερματική συσκευή.

Το πρωτόκολλο H.323 χρησιμοποιεί την έννοια των καναλιών για να έχει τη δυνατότητα της ανταλλαγής πληροφοριών μεταξύ των τερματικών συσκευών. Ένα κανάλι είναι μια σύνδεση στο επίπεδο μεταφοράς, το οποίο μπορεί να είναι είτε μονής είτε αμφίδρομης κατεύθυνσης. Όλα τα τερματικά H.323 πρέπει επίσης να υποστηρίζουν το πρωτόκολλο H.245, το οποίο χρησιμοποιείται για να διαπραγματευτεί τη χρήση του καναλιού και τις δυνατότητές του. Δύο άλλα συστατικά που απαιτούνται στη δομή του πρωτοκόλλου H.323 είναι το H.225 για την σηματοδοσία και τη ρύθμιση των κλήσεων και το πρωτόκολλο Εγγραφής/Εισόδου/Κατάστασης (Registration/Admission/Status RAS) που χρησιμοποιείται για την επικοινωνία με τον H.323 Gatekeeper.

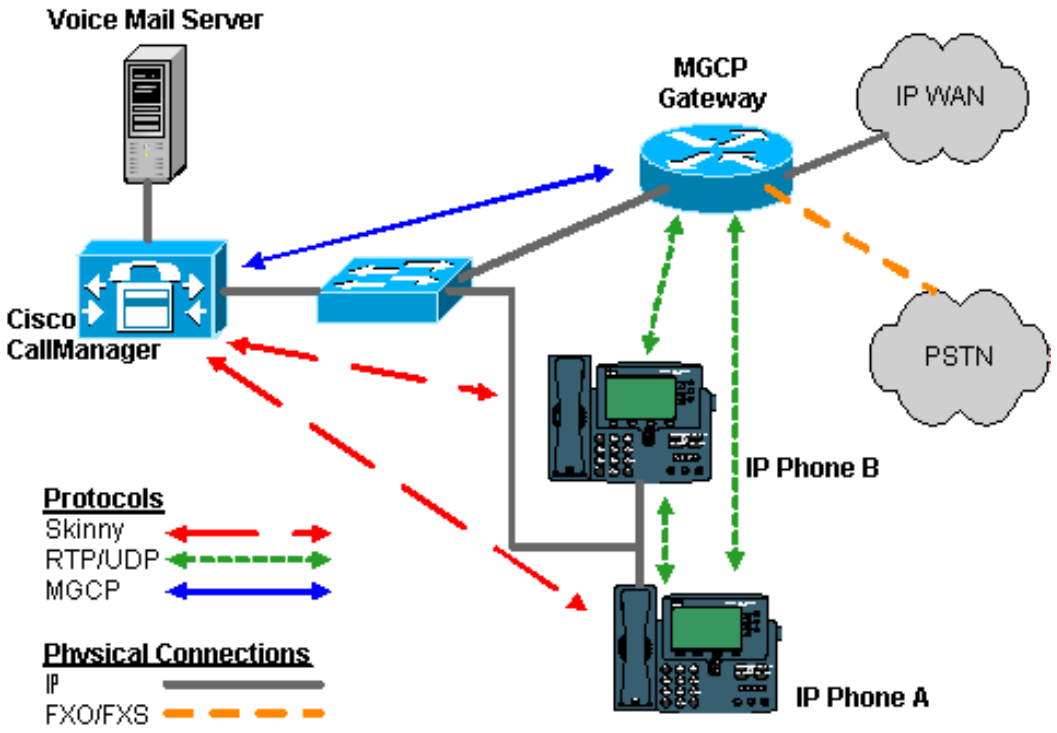
Η διαδικασία για την έναρξη και ρύθμιση μίας κανονικής κλήσης στο πρωτόκολλο H.323 περιλαμβάνει πολλά μηνύματα από τα άλλα πρωτόκολλα με τα οποία συνεργάζεται το H.323, όπως το H.225 [16] και το H.245 [17]. Η έναρξη μίας κλήσης βασιζόμενη στο πρωτόκολλο H.323 είναι μια διαδικασία τριών βημάτων, η οποία περιλαμβάνει την σηματοδοσία της κλήσης, την εγκατάσταση ενός καναλιού για τη σηματοδοσία και τη θέσπιση καναλιών για τα πολυμεσικά δεδομένα. Στην πρώτη φάση της κλήσης σηματοδοσίας, η τερματική συσκευή, που βασίζεται στο πρωτόκολλο H.323 και είναι ο καλών, ζητά άδεια από τον «προαιρετικό» H.323 Gatekeeper για την επικοινωνία της με το δίκτυο. Μόλις η κλήση γίνει αποδεκτή από τον H.323 Gatekeeper, η κλήση σηματοδοσίας προχωρά σύμφωνα με ένα από τα πολλά μοντέλα κλήσης που θα αποφασισθεί μέχρι τον τερματισμό της [18].

2.2.2.3 MGCP

Το πρωτόκολλο ελέγχου πύλης πολυμέσων (Media Gateway Control Protocol, MGCP) χρησιμοποιείται για τον έλεγχο των πυλών τηλεφωνίας (telephony gateways) από εξωτερικές τερματικές συσκευές. Μία τηλεφωνική πύλη είναι ένα κομμάτι του δικτύου που παρέχει τη μετατροπή μεταξύ των ακουστικών σημάτων που διακινούν τα τηλεφωνικά κυκλώματα και των πακέτων δεδομένων που διακινούνται μέσω του διαδικτύου ή μέσω άλλων τύπων δικτύου.

Το πρωτόκολλο MGCP προϋποθέτει μία αρχιτεκτονική ελέγχου των κλήσεων, όπου η νοημοσύνη του ελεγκτή των κλήσεων θα είναι έξω από τις πύλες και θα διεκπεραιώνεται από εξωτερικά στοιχεία ελέγχου των κλήσεων. Το MGCP υποθέτει ότι αυτά τα στοιχεία ελέγχου των κλήσεων ή αλλιώς Call Agents, θα έχουν την ευθύνη του συγχρονισμού μεταξύ τους για να στέλνουν

συνεκτικές εντολές στις πύλες υπό τον έλεγχο τους. Το MGCP είναι ένα master/slave πρωτόκολλο, όπου οι πύλες αναμένεται να εκτελούν τις εντολές που στέλνονται από τα Call Agents. Η λειτουργία του πρωτοκόλλου MGCP φαίνεται στην επόμενη εικόνα.



Εικόνα 2.11 Δομή του MGCP όπως την παρουσιάζει η εταιρεία CISCO [19]

Το MGCP υλοποιεί το περιβάλλον της ελεγκτικής πύλης πολυμέσων ως ένα σύνολο συναλλαγών. Οι συναλλαγές αποτελούνται από μία εντολή και μία υποχρεωτική απάντηση. Υπάρχουν οκτώ τύποι εντολών:

- MGC→MG CRCX, CreateConnection: Δημιουργεί μια σύνδεση μεταξύ των δύο άκρων. Χρησιμοποιεί το SDP για τον καθορισμό των ικανοτήτων λήψεις μεταξύ των τελικών συμμετεχόντων στην VoIP κλήση.

- MGC→MG MDCX, ModifyConnection: Τροποποιεί τις ιδιότητες μιας σύνδεσης που είναι ενεργή. Έχει σχεδόν τις ίδιες παραμέτρους με την εντολή CRCX.

- MGC↔MG DLCX, DeleteConnection: Τερματίζει μία σύνδεση και συλλέγει τα στατιστικά στοιχεία που σχετίζονται με την εκτέλεση της σύνδεσης.

- MGC→MG RQNT, NotificationRequest: Ζητά από την πύλη πολυμέσων να στείλει τα κατάλληλα σήματα σχετικά με την εμφάνιση συγκεκριμένων συμβάντων σε μία τερματική συσκευή (π.χ. σήμα απασχολημένης γραμμής, busy tone).

- MGC←MG NTFY, Notify: Ενημερώνει την ελεγκτική πύλη πολυμέσων όταν συμβαίνουν συγκεκριμένα γεγονότα.

- MGC→MG AUEP, AuditEndpoint: Εξετάζει την κατάσταση μίας τερματικής συσκευής.

- MGC→MG AUCX, AuditConnection: Ανακτά όλες τις παραμέτρους που σχετίζονται με τη σύνδεση.

- MGC←MG RSIP, RestartInProgress: Στέλνει σήμα όταν μία τερματική συσκευή ή μια ομάδα τερματικών συσκευών είναι να συμμετάσχουν στην υπηρεσία ή να εξέλθουν από αυτή.

Δείγματα εκτέλεσης των παραπάνω εντολών μπορούμε να διαβάσουμε στο δικτυακό τόπο της εταιρείας CISCO στην ηλεκτρονική διεύθυνση http://www.cisco.com/c/en/us/support/docs/voice-unified-communications/unified-communications-manager-callmanager/42104-debug-mgcp.html

Οι τέσσερις πρώτες εντολές στέλνονται από τον Call Agent σε μια πύλη. Η εντολή NTFY στέλνεται από την πύλη προς τον Call Agent. Η πύλη μπορεί επίσης να στείλει μία εντολή DLCX. Ο Call Agent μπορεί να στείλει είτε την AUEP είτε την AUCX στην πύλη. Η πύλη μπορεί να στείλει μια εντολή RSIP στον Call Agent.

Όλες οι εντολές που στέλνονται τοποθετούνται πάντα στην αρχή ως κεφαλή, προαιρετικά μπορεί να τους ακολουθεί μια περιγραφή της συνόδου. Όλες οι απαντήσεις τοποθετούνται στην αρχή ως κεφαλή, όπως και στις απεσταλμένες εντολές, προαιρετικά είναι δυνατό να ακολουθεί μια περιγραφή της συνόδου. Οι κεφαλίδες και οι περιγραφές των συνόδων κωδικοποιούνται ως ένα σύνολο από γραμμές κειμένου, που χωρίζονται από τους χαρακτήρες επαναφοράς (carriage return) και γραμμής τροφοδοσίας (line feed) ή, προαιρετικά, από ένα μόνο χαρακτήρα αυτόν της γραμμής τροφοδοσίας. Οι κεφαλίδες διαχωρίζονται από την περιγραφή της συνόδου με μία κενή γραμμή.

Το MGCP χρησιμοποιεί ένα αναγνωριστικό συναλλαγής ώστε να συσχετίσει τις εντολές και τις απαντήσεις. Τα αναγνωριστικά συναλλαγής έχουν τιμές μεταξύ 1 και 999999999. Το πρωτόκολλο MGCP απαγορεύει την επαναχρησιμοποίηση αναγνωριστικού συναλλαγών νωρίτερα από τρία λεπτά μετά την ολοκλήρωση της προηγούμενης εντολής, στην οποία χρησιμοποιήθηκε το αναγνωριστικό.

Η κεφαλίδα της εντολής αποτελείται από:

- Μια γραμμή εντολών, προσδιορίζοντας την απαιτούμενη δράση ή ένα ρήμα, το αναγνωριστικό της συναλλαγής, την τερματική συσκευή προς την οποία ζητείται η δράση, και η έκδοση του πρωτοκόλλου MGCP.

- Ένα σύνολο από γραμμές με παραμέτρους. Κάθε γραμμή αποτελείται από το όνομα της παραμέτρου και ακολουθεί η τιμή της παραμέτρου.

Η γραμμή εντολών αποτελείται από:

- Το όνομα του ρήματος.

- Τον αναγνωριστικό αριθμό της συναλλαγής που συσχετίζει τις εντολές με τις απαντήσεις.

- Το όνομα της τερματικής συσκευής που θα εκτελέσει την εντολή.

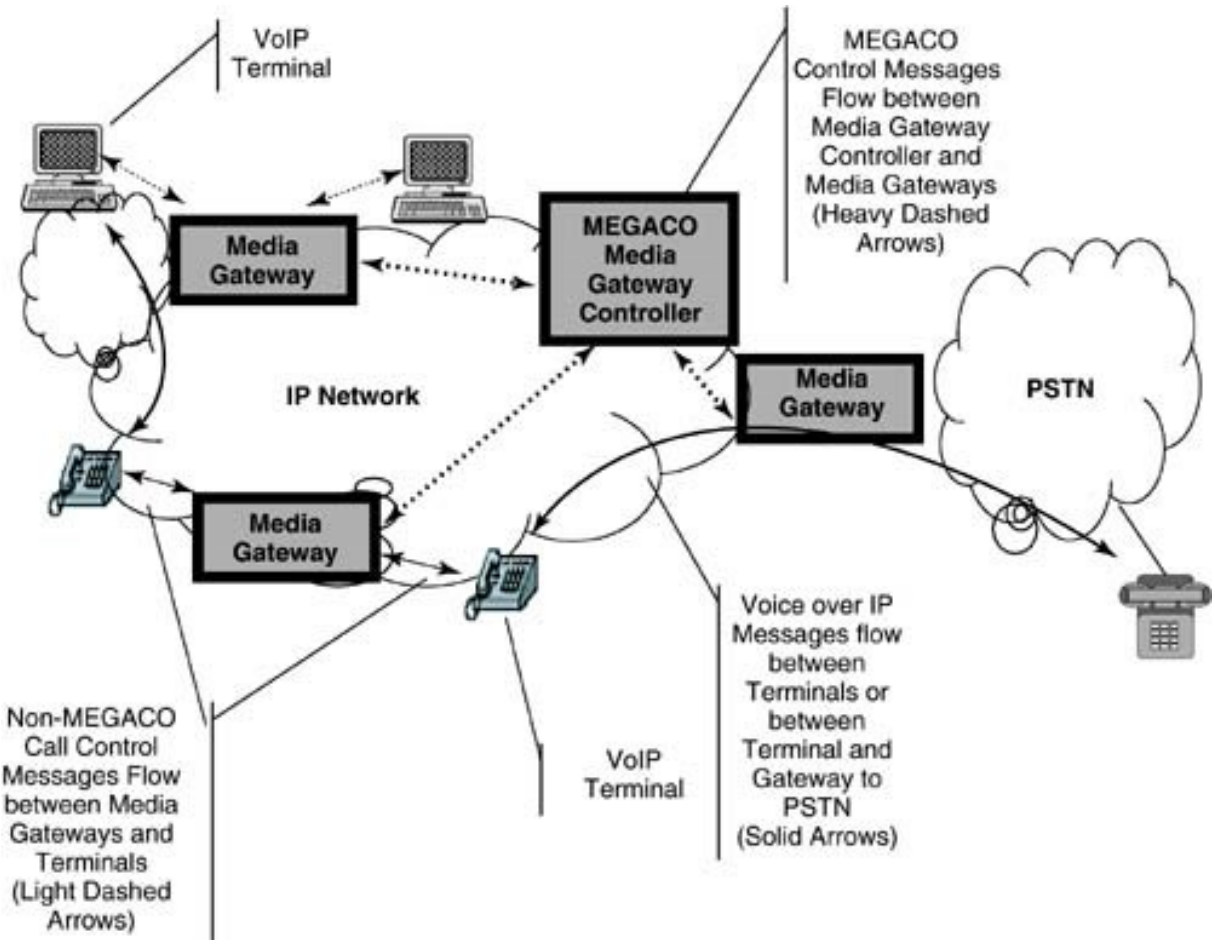
- Την έκδοση του πρωτοκόλλου MGCP.

Τα παραπάνω τέσσερα στοιχεία κωδικοποιούνται ως χαρακτήρες ASCII, που χωρίζονται με τον κενό χαρακτήρα. Συνιστάται να χρησιμοποιείται μόνο ένας κενός χαρακτήρας ASCII.

2.2.2.4 MeGaCo/H.248

Το πρωτόκολλο της πύλης ελέγχου των πολυμέσων (Media Gateway Control, MEGACO) ορίζεται από το πρωτόκολλο H.248 [20]. Η εξήγηση δίνεται από τις εργασίες στο IETF για τις προδιαγραφές του MEGACO. Το MEGACO είναι η κεντρική προσέγγιση στον έλεγχο των υπηρεσιών VoIP. Η προσέγγιση αυτή χρησιμοποιεί μία πύλη ελέγχου των πολυμέσων για τη ρύθμιση και τον έλεγχο των συνόδων μεταξύ των χρηστών. Ο Media Gateway Controller μεσολαβεί στον έλεγχο και τη ρύθμιση του κομιστή με τα άλλα δίκτυα, είτε είναι δίκτυο PSTN είτε κάποιο άλλο δίκτυο VoIP (όπως το H.323). Επίσης ελέγχει τις άλλες πύλες μέσω του δικτύου H.248. Οι άλλες πύλες συντονίζονται και ελέγχονται από τους Media Gateway Controllers που δίνουν οδηγίες στις πύλες για τη δημιουργία μονοπατιών επικοινωνίας μεταξύ τους μέσω του δικτύου πακέτων. Το MEGACO μοιάζει με την κεντρική δομή της σηματοδοσίας και ελέγχου που χρησιμοποιείται στο PSTN. Στο PSTN, το Signalling System 7 (SS7) είναι ένα εξειδικευμένο πρωτόκολλο σηματοδοσίας που καθοδηγεί τα switches των τηλεφώνων στη ρύθμισή τους για να πραγματοποιούν κλήσεις μεταξύ τους. Ο Media Gateway Controller εξυπηρετεί μια παρόμοια λειτουργία για την πύλη ενός δικτύου που βασίζεται στο πρωτόκολλο H.248. Ωστόσο, στο SS7

ελεγκτή PSTN, το δίκτυο SS7 είναι ένα φυσικό διακριτό δίκτυο, ενώ στο περιβάλλον MEGACO ο Media Gateway Controller και οι άλλες πύλες τυπικά υποστηρίζονται πάνω στο ίδιο δίκτυο που βασίζεται στα πακέτα.



Εικόνα 2.12 Η αρχιτεκτονική του πρωτοκόλλου MEGACO

Το MEGACO έχει μια σειρά από πλεονεκτήματα:

- Τα πρωτόκολλα σηματοδοσίας του είναι εύκολα επεκτάσιμα για να ανταπεξέλθουν σε νέα περιβάλλοντα.

- Η ομοιότητα του στην αρχιτεκτονική φιλοσοφία του υπάρχοντος SS7 ελεγκτή PSTN απλοποιεί εσωτερικά τη συνεργασία μεταξύ των περιβαλλόντων που βασίζονται στο MEGACO και του παλαιού τηλεφωνικού δικτύου.

- Είναι ειδικά σχεδιασμένο για να υποστηρίξει κατανεμημένο έλεγχο των σύνθετων υπηρεσιών τηλεφωνίας σε ένα περιβάλλον πακέτων, βοηθώντας έτσι στην ανάπτυξη και εξάπλωση των μεταγωγών τύπου λογισμικού (software switch, softswitch), τα οποία διανέμονται και αντικαθιστούν τους υπάρχοντες τηλεφωνικούς μεταγωγείς.

Όπως το πρωτόκολλο H.248 αναπτύχθηκε σε μεγάλο βαθμό από τους μηχανικούς με τη σκέψη τους περισσότερο προς την τηλεφωνία σε σχέση με το πρωτόκολλο H.323, το πρωτόκολλο MEGACO ασχολείται με θέματα λογιστικής (χρέωσης) και σταθερότητας του δικτύου ως εγγενές στοιχείο της αρχιτεκτονικής.

Η χρήση του πρωτοκόλλου MEGACO όπως και η επικοινωνία του με το υπόλοιπο δίκτυο φαίνεται στην Εικόνα 2.12.

Κεφάλαιο 3
VoIP Επιθέσεις

Τα βασικά πλεονεκτήματα του VoIP είναι η ευελιξία και το χαμηλό του κόστος. Η ευελιξία προέρχεται από τις (κατά κανόνα) ανοικτές αρχιτεκτονικές και τις εφαρμογές λογισμικού στις οποίες βασίζεται, ενώ το χαμηλό κόστος οφείλεται σε νέα επιχειρηματικά μοντέλα, τον εξοπλισμό, τα ενοποιημένα δίκτυα και την πανταχού παρούσα προς τους καταναλωτές ευρυζωνική σύνδεση [05].

Με τα παραπάνω πλεονεκτήματα, η υπηρεσία VoIP έχει δει ταχεία είσοδο στο χώρο των επιχειρήσεων και στην καταναλωτική αγορά. Ένα πλήθος επιχειρήσεων (που συνεχώς αυξάνεται ο αριθμός τους) αντικαθιστούν το αναλογικού τύπου τηλεφωνικό τους κέντρο με εφαρμογές βασιζόμενες στο VoIP, ώστε να εισάγουν νέα χαρακτηριστικά στο χώρο τους και να απομακρύνουν τον περιττό εξοπλισμό. Οι καταναλωτές έχουν αγκαλιάσει μια πληθώρα τεχνολογιών με διαφορετικά χαρακτηριστικά και κόστος, συμπεριλαμβανομένων των P2P κλήσεων (Skype), Internet-to-PSTN γεφύρωσης του δικτύου, καθώς και ασύρματη VoIP. Αυτές οι νέες τεχνολογίες και τα επιχειρησιακά μοντέλα προωθούνται από μια νέα γενιά εταιρειών που θέτουν υπό αμφισβήτηση το στάτους κβο στην παραδοσιακή τηλεφωνία (PSTN) και την προσωπική τηλεπικοινωνία. Ως αποτέλεσμα, ένας αριθμός παρόχων υπηρεσιών αναλογικής τηλεφωνίας έχουν ήδη ολοκληρώσει ή βρίσκονται στο στάδιο της μετάβασης από τα δίκτυα μεταγωγής κυκλώματος (PSTN) σε VoIP υπηρεσίες. Τέλος, όπως ο τομέας του εμπορίου και των

καταναλωτών ακολουθούν την εξέλιξη της τεχνολογίας, το ίδιο κάνουν και οι κυβερνήσεις με τις ένοπλες δυνάμεις λόγω της ανάγκης τους να μειώσουν το κόστος και τη γενική εξάρτηση από τον εξοπλισμό του εμπορίου (Commercial Off-The-Shelf, COTS) για την πλειοψηφία των υπολογιστικών αναγκών τους.

Ο Αμερικανικός Σύλλογος Ελέγχου της Απάτης στις Επικοινωνίες (Communications Fraud Control Association, CFCA) διεξάγει κάθε δύο χρόνια μία παγκόσμια έρευνα στην απώλεια χρημάτων λόγω απάτης στις τηλεπικοινωνίες [27]. Στην τελευταία του έρευνα, ανακοίνωσε τα αποτελέσματα στις 10 Οκτωβρίου 2013, η παγκόσμια απώλεια χρημάτων λόγω της απάτης στις τηλεπικοινωνίες ανέρχεται στο ποσό των $ 46,3 δισεκατομμυρίων δολαρίων, δηλαδή το 2,09% από το σύνολο των εσόδων. Η απώλεια χρημάτων είναι αυξημένη κατά 15% σε σχέση με την προηγούμενη έρευνα της οποίας τα αποτελέσματα δημοσιοποιήθηκαν το 2011, ενώ τα έσοδα των εταιρειών στις τηλεπικοινωνίες έχουν αυξηθεί κατά 3,7%. Το 10,5% των εταιρειών, που συμμετείχαν στην έρευνα, αναφέρουν απώλειες 5% έως 10% των εσόδων τους. Αποκορύφωμα στις απώλειες εσόδων αποτελεί το 8,8% των εταιρειών να δηλώνουν απώλειες πάνω από 10% των εσόδων τους όταν στην έρευνα του 2011 δεν υπήρχε καμία εταιρεία να δηλώνει απώλειες πάνω από το 1/10 των εσόδων της.

3.1 Θέματα Ασφάλειας και Τρωτά σημεία

Για να καταλάβουμε τη διαφορά μεταξύ των όρων απειλή (Threat), επίθεση (Attack) και ευπάθεια ή τρωτό σημείο (Vulnerability), θα πρέπει να χρησιμοποιήσουμε τους σωστούς ορισμούς στα πλαίσια της VoIP ασφάλειας. Συχνά, οι άνθρωποι χρησιμοποιούν τους όρους απειλή και τρωτό σημείο αδιάκριτα, για να καθορίσουν τον κίνδυνο που συνδέεται με έναν πόρο του δικτύου, της υπηρεσίας, ή της αλληλεπίδρασης με τον χρήστη [23]. Πριν συνεχίσουμε στις απειλές και τα τρωτά σημεία σε δίκτυα VoIP, θα πρέπει να ορίσουμε τι λέγεται απειλή κι ευπάθεια.

- Απειλή: Είναι η ύπαρξη ενός κινδύνου, η έκθεση μίας οντότητας σε κίνδυνο ή τέλος η λεκτική ανακοίνωση για την πρόκληση κακού σε μία οντότητα με απώτερο σκοπό τον εξαναγκασμό της οντότητας σε ενέργειες που επιθυμεί ο επιτιθέμενος να εκτελεστούν [24].

- Επίθεση: Είναι το αντίθετο της άμυνας. Η κίνηση εναντίον μίας οντότητας, συνήθως με ορμή, ασκώντας βία επάνω της. Στο χώρο του αθλητισμού θεωρείται η κίνηση παικτών προς την αντίπαλη περιοχή για να πετύχουν θετικό αποτέλεσμα [24].

- Ευπάθεια, Μία σύνθετη λέξη από το πρόθεμα «ευ» και τη λέξη «παθαίνω». Το πρόθεμα «ευ» δηλώνει ότι η σύνθετη λέξη έχει αυτό που εκφράζει το β' συνθετικό δηλαδή το ρήμα «παθαίνω». Παθαίνω σύμφωνα με το ερμηνευτικό λεξικό της νέας Ελληνικής είναι το δυσάρεστο, ανεπιθύμητο, κακό συμβάν σε μία οντότητα [24].

Τα VoIP συστήματα βασίζονται σε κάποιο δίκτυο δεδομένων όπως είναι το διαδίκτυο (Internet). Σε αυτά τα συστήματα εγκλωβίζονται όλων των ειδών οι εφικτές αδυναμίες στην ασφάλεια και την μορφή των επιθέσεων που σχετίζονται με οποιοδήποτε δίκτυο δεδομένων. Για παράδειγμα, σε ένα τηλεφωνικό σύστημα που βασίζεται στο δίκτυο μεταγωγής κυκλώματος, μπορεί να διεξαχθεί μία δραστηριότητα, όπως η υποκλοπή μέσω του καλωδίου, είτε με φυσική πρόσβαση στις τηλεφωνικές γραμμές, είτε με πρόσβαση στο τηλεφωνικό σύστημα της επιχείρησης (Private Branch Exchange, PBX). Στα δίκτυα μεταγωγής πακέτων από όπου παρέχεται η υπηρεσία VoIP, η φωνή μετατρέπεται σε IP πακέτα που θα ταξιδέψουν μέσα από πολλούς ενδιάμεσους δικτυακά προσβάσιμους σταθμούς. Ως εκ τούτου, τα δεδομένα εκτίθενται σε πολύ περισσότερα πιθανά σημεία επίθεσης. Οι εισβολείς θα μπορούσαν να τα υποκλέψουν. Στην πραγματικότητα, όλοι οι κίνδυνοι στην ασφάλεια του IP, όπως οι ιοί υπολογιστών, Denial of Service και οι έμμεσες επιθέσεις από τους ανθρώπους, είναι επίσης τρωτά σημεία για τα συστήματα VoIP.

Ειδικότερα, τα τηλεφωνικά κέντρα βασισμένα στο πρωτόκολλο IP (PC-based IP Phone hosts), που λειτουργούν ως εφαρμογές ηλεκτρονικών υπολογιστών, είναι πιο επιρρεπή σε επιθέσεις. Οι τεχνικές επίθεσης, που εντοπίζονται στους ηλεκτρονικούς υπολογιστές, δημιουργούν τις ευπάθειες των παραπάνω τηλεφωνικών κέντρων. Ορισμένες από τις τεχνικές των επιθέσεων που επιδρούν στην λειτουργία των τηλεφωνικών συστημάτων τύπου λογισμικό αναφέρονται παρακάτω:

- Τα τρωτά σημεία του λειτουργικού συστήματος,

- Τα τρωτά σημεία της εφαρμογής,

- Τα τρωτά σημεία των υπηρεσιών που εκτελούνται στο παρασκήνιο του ηλεκτρονικού υπολογιστή,

- Τα σκουλήκια (worms) που διατρέχουν το διαδίκτυο και εισβάλλουν στους ηλεκτρονικούς υπολογιστές,

- Οι ιοί.

Ένα τηλεφωνικό σύστημα, το οποίο φιλοξενείται σ' έναν ηλεκτρονικό υπολογιστή, είναι συνεχώς ευάλωτο. Οποιαδήποτε επίθεση που θα συμβεί στις περιφερειακές συσκευές αποθήκευσης του ηλεκτρονικού υπολογιστή μπορεί να αποβεί μοιραία και για την εφαρμογή του τηλεφωνικού συστήματος.

Όσα από τα πρωτόκολλα επικοινωνίας της φωνής ανήκουν στο πρωτόκολλο ελέγχου συνόδου (Session Control Protocol, SCP), η διεύθυνση IP και οι πόρτες του TCP/UDP θα περικλείονται στα πακέτα ως πληροφορίες. Σε δίκτυα που χρησιμοποιούν τεχνολογία μεταγλώττισης της δικτυακής διεύθυνσης (Network Address Translation, NAT), η διεύθυνση IP και η πόρτα που περιλαμβάνονται ως πληροφορία στα πακέτα δεν επιτρέπεται να κρυπτογραφηθούν. Οι συσκευές NAT απαιτούν αυτές τις πληροφορίες για να εκτελέσουν τη μετάφραση. Αυτό επιβάλλει μία άλλου τύπου ασφάλεια σε αυτά τα πρωτόκολλα.

Το πρωτόκολλο H.323 ασφαλίζεται με τη χρήση του επιπέδου ασφάλειας στην μεταφορά (Transport Layer Security, TLS). Μία προκαθορισμένη πόρτα στο TCP, η 1300, πρέπει να χρησιμοποιείται για τη δημιουργία του καναλιού που θα συνδέσει την τηλεφωνική κλήση από άκρο σε άκρο. Δε διατίθεται κανένας άλλος μηχανισμός για την ασφάλεια κατά την πρώτη σύνδεση. Αυτή η σταθερότητα και η γνώριμη πόρτα μπορεί να αποτελέσουν απειλή για το πρωτόκολλο.

Για το πρωτόκολλο SIP, η κρυπτογράφηση βασίζεται στη χρήση του Secure/Multipurpose Internet Mail Extensions, (S/MIME) [21]. Μόνο ορισμένες κεφαλίδες στο μήνυμα είναι κρυπτογραφημένες και κρίσιμες κεφαλίδες, όπως τα πεδία του παραλήπτη «To», του αποστολέα «From» και του χαρακτηριστικού της κλήσης «Call-ID», δεν είναι κρυπτογραφημένα.

Μια πολύ γνωστή απειλή στην ασφάλεια του πρωτοκόλλου MGCP είναι το «Uncontrolled barge-in». Εφόσον τα πακέτα φωνής μπορούν να κατευθυνθούν σε μία πύλη μέσω της κατάλληλης πόρτας UDP, ένας επιτιθέμενος θα μπορούσε να είναι σε θέση, ως ενδιάμεσος κρυφός χρήστης, ν' ακούσει τα φωνητικά μηνύματα που ταξιδεύουν ως δεδομένα μεταξύ των τερματικών συσκευών. Αυτού του είδους η απειλή μπορεί να αποφευχθεί εάν θα λειτουργεί σωστά η ασφάλεια. Για την άμβλυνση της απειλής αυτής, μία πύλη θα πρέπει να δέχεται τα δεδομένα μόνο από μία προκαθορισμένη διεύθυνση IP και πόρτα UDP. Το μειονέκτημα είναι η πρόσθετη

εργασία που εισάγεται, και οι διευθύνσεις IP μπορούν να πλαστογραφηθούν. Για την αντιμετώπιση της πλαστογράφησης, το πρωτόκολλο MGCP θα μπορούσε να λαμβάνει την «εξ αποστάσεως περιγραφή της συνόδου» από την πύλη εισόδου και να την προωθεί προς έλεγχο στην πύλη προορισμού. . Ωστόσο, αυτό αυξάνει το χρόνο έναρξης της κλήσης.

Οι υπηρεσίες VoIP υπόκεινται επίσης και στις διαφημιστικές επιθέσεις (spamming), γνωστές ως Spam over Internet Telephony (SPIT) επιθέσεις [22]. Το SPIT αφήνει διαφημιστικά φωνητικά μηνύματα σε στοχευμένα τηλέφωνα IP. Τα φωνητικά μηνύματα είναι τις περισσότερες φορές μεγαλύτερα σε μέγεθος δεδομένων σε σχέση με τα μηνύματα του ηλεκτρονικού ταχυδρομείου (eMail), με συνέπεια το SPIT να φορτώνει το δίκτυο σε υψηλότερο βαθμό από ότι συμβαίνει σε ένα αντίστοιχο τυπικό SPAM email.

Οι επιτιθέμενοι έχουν επίσης προσπαθήσει να αξιοποιήσουν την τεχνολογία του VoIP για να τους γνωστοποιηθούν οι κωδικοί πρόσβασης των θυμάτων και να κλέψουν χρήματα [25]. Αυτή η κατηγορία της επίθεσης είναι παρόμοια με ένα μήνυμα ηλεκτρονικού ταχυδρομείου που βασίζεται στην επίθεση ψαρέματος (phishing), και ως εκ τούτου το όνομά του «vishing» (VoIP phishing). Το θύμα θα λάβει ένα μήνυμα ηλεκτρονικού ταχυδρομείου ή θα ενεργοποιηθεί μία τηλεφωνική κλήση. Στη συνέχεια το θύμα θα κατευθυνθεί σε έναν αριθμό εξυπηρέτησης πελατών. Θα παραπεμφθεί σε ένα τηλεφωνικό μενού επιλογών, στην προσπάθειά τους να κλέψουν αριθμούς λογαριασμών, κωδικούς PIN, και άλλες κρίσιμες πληροφορίες.

Από τα παραπάνω φαίνονται οι αδυναμίες και τα τρωτά σημεία που μπορούν να εντοπιστούν στα συστήματα VoIP. Στο βιβλίο «INTRODUCTION TO VOIP FRAUD» που εκδόθηκε στις 18 Οκτωβρίου 2012 από την Trans Nexus, εταιρεία που παράγει λογισμικό με ειδίκευση στη διαχείριση των VoIP δικτύων, αναφέρει τους παρακάτω τύπους απειλής στα VoIP συστήματα:

- Arbitrage: Δρομολόγηση μίας διεθνούς κλήσης από μία χώρα προς μία άλλη, χρησιμοποιώντας χώρες που έχουν καλύτερες τιμολογήσεις. Παράδειγμα το κόστος μίας κλήσης από τη χώρα Α προς τη χώρα Β μπορεί να κοστίζει 0,28 ευρώ ανά λεπτό (€/min). Η ίδια κλήση εάν δρομολογηθεί από μία άλλη χώρα Γ μπορεί να κοστίσει λιγότερο. Έστω από την Α →Γ το κόστος είναι 0,14 €/min και από τη Γ →Β το κόστος είναι 0,12€/min ευρώ ανά λεπτό. Συνολικό κόστος κλήσης από Α→Γ→Β 0,26 €/min έναντι 0,28 €/min για την απευθείας κλήση Α→Β.

- Buffer Overflow: Ο επιτιθέμενος στέλνει ταυτόχρονα πολλά μηνύματα INVITE ή πακέτα SIP με σκοπό να καταρρεύσει η εφαρμογή του διακομιστή ή να εκτελέσει παράνομο λογισμικό.

- Bypass Fraud: Εταιρείες οι οποίες πουλούν κάρτες τηλεφωνικών συνδιαλέξεων και παρέχουν τη δυνατότητα στον χρήστη να καλεί σε τηλεφωνικό αριθμό του εξωτερικού μέσω ενός τοπικού τηλεφώνου της εταιρείας. Ο πελάτης καλεί αρχικά σε ένα τηλεφωνικό νούμερο που του δίνεται από την εταιρεία και στη συνέχεια μέσω αυτής δρομολογείται η κλήση του προς το διεθνή τηλεφωνικό αριθμό.

- Call Transfer Fraud ή Fraudulent Calls: Ο επιτιθέμενος διεισδύει παράνομα στο τηλεφωνικό κέντρο της εταιρείας (κυρίως όταν το τηλεφωνικό κέντρο είναι τύπου λογισμικό). Χρησιμοποιώντας τις υπηρεσίες της εταιρείας αποκτά πρόσβαση σε όλων των ειδών τις κλήσεις, τοπικές ή διεθνείς. Με τη μέθοδο αυτή ο επιτιθέμενος καταφέρνει να χρησιμοποιεί τα λογισμικά της εταιρείας τα οποία δεν είναι δυνατόν να καταγράψουν τις κλήσεις με συνέπεια να μη χρεώνονται.

- Domestic & International Revenue Share Fraud: Ο επιτιθέμενος στην προσπάθειά του να αποπλανήσει τα θύματα με διάφορα δώρα τους ζητά να τον καλέσουν σε ένα τηλεφωνικό αριθμό. Το θύμα ξεκινώντας μία τέτοια κλήση χρεώνεται με μεγάλα ποσά από τα οποία ένα μέρος το αποκομίζει ο επιτιθέμενος.

- False Answer Supervision: Το θύμα που ξεκινά μία κλήση προς ένα τηλεφωνικό νούμερο που δεν υπάρχει ή είναι απενεργοποιημένη η συσκευή, λαμβάνει ένα ηχητικό μήνυμα για το πρόβλημα που υφίσταται. Το σύστημα, αυτού του είδους την κλήση (κλήση που δεν δέχεται απάντηση από τον καλούμενο), δεν τη χρεώνει. Ωστόσο ο επιτιθέμενος, χρησιμοποιώντας ψεύτικα μηνύματα, δίνει τη δυνατότητα στο σύστημα να χρεώσει την κλήση ως ολοκληρωμένη.

- Location Routing Number (LRN) Fraud: Ο επιτιθέμενος εισάγει ψεύτικες διαδρομές LRN στα πακέτα SIP με σκοπό να αποκομίσει κέρδη από μία διαδρομή τηλεφωνικής κλήσης. Οι εξυπηρετητές VoIP δρομολογούν τα φωνητικά πακέτα από την ψεύτικη φθηνή διαδρομή που δηλώνει ο επιτιθέμενος. Η διαδρομή της τηλεφωνικής κλήσης, στην πραγματικότητα, περνά από ακριβούς σε χρηματικό κόστος ενδιάμεσους σταθμούς

- PBX Hacking: Η απειλή έγκειται στη δυνατότητα του επιτιθέμενου να παραβιάσει το τηλεφωνικό σύστημα για να χρησιμοποιήσει ως παράδειγμα επίθεση τύπου «Domestic &

International Revenue Share Fraud». Η ευπάθεια του συστήματος δίνει τη δυνατότητα σε έναν καλό γνώστη επιθέσεων VoIP να εισβάλλει στα συστήματα. Επισημαίνονται τα τέσσερα βασικότερα προβλήματα ευπάθειας:

  - Ασθενής Αυθεντικοποίηση των χρηστών και των Διαχειριστών του συστήματος.
  - Συστήματα βασιζόμενα μόνο στην ασφάλεια που παρέχουν οι Συνοριακοί Ελεγκτές Περιόδου (Session Border Controllers, SBCs).
  - Ανεπάρκεια διαχωρισμού κι ελέγχου εικονικών δικτύων.
  - Ανεπαρκής χρήση κρυπτογράφησης.

- Phreaking: Η παράνομη πρόσβαση και ο έλεγχος μίας τηλεφωνικής συσκευής. Πολλά παραβιασμένα τηλεφωνικά κέντρα μπορούν να δικτυωθούν μεταξύ τους και να παρέχουν διάφορες υπηρεσίες.

- Roaming Fraud: Η απάτη στην περιαγωγή προκύπτει από υπηρεσίες που προσφέρονται έξω από τη χώρα της εταιρείας κι ο επιτιθέμενος, υποκλέπτοντας στοιχεία, τις χρησιμοποιεί. Η εύρεση της παρανομίας είναι δύσκολη λόγω της μεγάλης καθυστέρησης στην λήψη των αποθηκευμένων κλήσεων.

- SIPVicious: Είναι μία σουίτα τεσσάρων εργαλείων για την επίθεση σε SIP VoIP συστήματα [28]. Ο επιτιθέμενος χρησιμοποιεί Δούρειους Ίππους και εισβάλλει στους εξυπηρετητές VoIP. Χρησιμοποιώντας εντολές και τον έλεγχο των εξυπηρετητών λαμβάνει οδηγίες από ένα .cc domain. Μετά την εγκατάσταση κι εκτέλεση της σουίτας SIPVicious γίνεται αναζήτηση SIP συσκευών στο παραβιασμένο δίκτυο. Εκτελείται βίαια επίθεση στα συστήματα για την εύρεση του κωδικού πρόσβασης που έχει ο διαχειριστής του συστήματος. Δημιουργείται μία βάση από την οποία ο επιτιθέμενος μπορεί να κάνει κλήσεις VoIP. Οι κλήσεις μπορούν να πραγματοποιηθούν προς τηλεφωνικούς αριθμούς με ακριβό κόστος, η κατοχή των οποίων ανήκει στον επιτιθέμενο. Επίσης μπορεί να γίνει τηλεφωνικό διαφημιστικό ψάρεμα (Voice phISHING, VISHING)

- Shell Companies: Περισσότερο ανήκει σε μία οικονομικής φύσης παρανομία στο χώρο του εμπορίου παρά ευπάθεια στο χώρο της πληροφορικής και ιδιαίτερα στις τηλεπικοινωνίες που βασίζονται στο VoIP. Οι «νέες εταιρείες» δεν έχουν καθόλου διαχείριση και στοχεύουν σε πελάτες που δεν διαθέτουν πιστωτικές κάρτες. Η εταιρεία θύμα πουλά σ' ένα «πελάτη» τις VoIP υπηρεσίες της. Ο «πελάτης» από την πλευρά του,

ως επιτιθέμενος, μεταπουλά τις υπηρεσίες της εταιρείας σε περισσότερους δικούς του πελάτες. Οι πελάτες αυτοί έχουν τη δυνατότητα πληρωμής μόνο με μετρητά χρήματα ή με αντικαταβολή. Για να κερδίσει τους πελάτες ο επιτιθέμενος τους δελεάζει με προσφορές κυρίως σε Σαββατοκύριακα και περιόδους διακοπών. Διαστήματα που η εταιρεία ως θύμα δεν μπορεί εύκολα να ελέγξει. Όταν η εταιρεία, θύμα, στείλει το τιμολόγιο, ο επιτιθέμενος ή στέλνει ένα ψεύτικο έγγραφο πληρωμής μέσω ηλεκτρονικού ταχυδρομείου ή συντηρεί τη σύνδεση πληρώνοντας ένα μικρό ποσό στην εταιρεία θύμα. Η πληρωμή ενός μικρού μέρους από το σύνολο του τιμολογίου δίνεται για να μην έχει δικαίωμα η εταιρεία να κόψει την παροχή των υπηρεσιών της.

- Subscription Fraud: Δημιουργείται μία νέα συνδρομή για την οποία δεν υπάρχει η διάθεση της αποπληρωμής της. Αυτού του τύπου η παρανομία συνδέεται τις περισσότερες φορές με άλλα εγκλήματα, όπως η κλοπή της ταυτότητας. Συνήθως ο τύπος αυτός της παρανομίας δεν αναγνωρίζεται, διότι οι εταιρείες το συγκαταλέγουν στη λίστα των κακοπληρωτών. Μόλις η εταιρεία αναγνωρίσει τον «πελάτη» της, ο επιτιθέμενος στην περίπτωση αυτή, ως «καλή την πίστη (bona fide) πελάτη», ξεκινά να δημιουργεί μεγαλύτερα προβλήματα που θα δυσφημίσουν την εταιρεία και θα επηρεάσουν τα κέρδη της.

- Toll Fraud: Παράνομα ο επιτιθέμενος εισχωρεί στο VoIP δίκτυο της εταιρείας και ξεκινά την πώληση χρόνου ομιλίας προς μακρινές αποστάσεις. Ο επιτιθέμενος αποκτά πρόσβαση στο τηλεφωνικό σύστημα της εταιρείας κι ελέγχει τις κλήσεις προς μακρινούς τόπους. Για να κερδίσει χρήματα ο επιτιθέμενος πουλά το χρόνο ομιλίας προς μακρινούς προορισμούς και στη συνέχεια χρησιμοποιεί το παραβιασμένο τηλεφωνικό σύστημα της εταιρείας.

- Unallocated Number Fraud: Τηλεφωνικοί αριθμοί που δεν έχουν αδειοδοτηθεί ακόμη, ο επιτιθέμενος τους προσθέτει στο λογαριασμό του. Στη συνέχεια εξομοιώνει την κίνηση των κλήσεων προς αυτούς τους αριθμούς. Οι αριθμοί αυτοί δρομολογούνται προς μία παραβιασμένη εταιρεία (θύμα) ως οι μοναδικοί αριθμοί που θα μπορούσαν να απαντήσουν στις τηλεφωνικές κλήσεις.

3.2 Μηχανισμοί Αποφυγής VoIP επιθέσεων

3.2.1 Κρυπτογράφηση με IPSEC, TLS ΚΑΙ S / MIME

Η κρυπτογράφηση είναι ένα μέσο για τη διατήρηση της εμπιστοσύνης στο μεταδιδόμενο σήμα. Όπως το SIP και τα υπόλοιπα πρωτόκολλα σηματοδοσίας ανήκουν στο επίπεδο των εφαρμογών (Application Layer) του TCP/IP, οι μηχανισμοί κρυπτογράφησης του IPSec και TLS θα μπορούσαν να χρησιμοποιηθούν σε χαμηλότερο επίπεδο στο TCP/IP, όπως στο επίπεδο της μεταφορά (Transport Layer) ή του δικτύου (Network Layer). Το S/MIME μπορεί να χρησιμοποιηθεί στο κύριο μέρος των μηνυμάτων SIP. Ωστόσο, η κρυπτογράφηση και η αποκρυπτογράφηση είναι απαιτητικές στους πόρους που διαθέτει η Κεντρική Μονάδα Επεξεργασίας (Central Processing Unit, CPU). Επίσης χρειάζεται χρόνο για την επεξεργασία των δεδομένων. Από τα παραπάνω φαίνεται να μειονεκτεί στις επιδόσεις της. Εάν ο χρόνος που προστίθεται στην κλήση VoIP ξεπερνά τα 250 χιλιοστά του δευτερολέπτου, η ποιότητα της κλήσης θα επηρεαστεί αισθητά. Επιπλέον, τα αιτήματα και οι απαντήσεις του πρωτοκόλλου SIP δεν μπορούν να κρυπτογραφηθούν πλήρως για ασφάλεια στα συστήματα από άκρη σε άκρη. Ορισμένα πεδία της επικεφαλίδας, όπως είναι η διεύθυνση IP, ο αριθμός της πόρτας και άλλα, πρέπει να είναι ορατά στους διακομιστές μεσολάβησης για τη σωστή δρομολόγηση των δεδομένων.

Ακόμη και αν χρησιμοποιηθεί η κρυπτογράφηση, η φυσική πρόσβαση ενός εισβολέα δεν μπορεί να αποτραπεί σε εξυπηρετητές VoIP και Πύλες. Ο εισβολέας, εάν καταφέρει και εισχωρήσει στις δικτυακές συσκευές, θα μπορεί να αναλύσει τα πακέτα που βρίσκονται σε κυκλοφορία και να αντλήσει πληροφορίες από τα κρυπτογραφημένα μηνύματα. Για το λόγο αυτό θα πρέπει να είμαστε βέβαιοι για τη συνεχόμενη επάρκεια της ασφάλειας των συστημάτων μας, ώστε να μπορούμε να αποτρέψουμε την ανεπιθύμητη πρόσβαση στις συσκευές του VoIP δικτύου μας [26].

3.2.2 Αυθεντικοποίηση Χρηστών και Συσκευών

Ορισμένοι διακομιστές επεξεργασίας κλήσεων έχουν την δυνατότητα της αυτόματης επικαιροποίησης μίας τηλεφωνικής συσκευής VoIP. Η συσκευή ξεκινά, ως ένα άγνωστο προς το δίκτυο τηλέφωνο, με μια προσωρινή δικτυακή ρύθμιση. Στη συνέχεια της επιτρέπεται η αλληλεπίδραση με το υπόλοιπο δίκτυο. Αυτό το χαρακτηριστικό είναι χρήσιμο για την έγκριση

πολλών IP τηλεφώνων, αλλά κρύβει άλλους κινδύνους. Οι συσκευές των επιτιθέμενων μπορεί να ξεκινήσουν μη εξουσιοδοτημένες υπηρεσίες ή ακόμη να κατευθύνονται προς άλλες συσκευές που συνδέονται με το δίκτυο.

Η ταυτοποίηση της συσκευής χρησιμοποιώντας τη διεύθυνση MAC του τηλεφώνου IP είναι μια λύση σε αυτό το πρόβλημα. Η αυτόματη λειτουργία αναγνώρισης VoIP συσκευών από το διακομιστή επεξεργασίας κλήσεων είναι ορθότερο να απενεργοποιείται. Εάν ένα τηλέφωνο με μια άγνωστη διεύθυνση MAC προσπαθεί να κατεβάσει τις ρυθμίσεις του δικτύου από το διακομιστή επεξεργασίας κλήσεων, η αίτηση θα απορρίπτεται και το δόλιο τηλέφωνο IP, δεν θα εξουσιοδοτείται να χρησιμοποιηθεί εντός του δικτύου. Δεν θα είναι σε θέση να συνδεθεί με το δίκτυο, επειδή ο διακομιστής δεν θα αναγνωρίζει τη διεύθυνση MAC. Ο έλεγχος της ταυτότητας του χρήστη, όπως συμβαίνει με το αναγνωριστικό χρήστη και τον κωδικό πρόσβασης ή έναν προσωπικό αριθμό αναγνώρισης (Personal Identification Number, PIN), είναι ένα αποτελεσματικό μέτρο για την αποφυγή παράνομης κλήσης. Παρέχει ακόμη ένα επίπεδο ασφάλειας στην αναγνώριση ενός άγνωστου χρήστη και δίνει τη δυνατότητα να βεβαιωθεί η ταυτότητα του καλούντος.

3.2.3 Έλεγχος της αλληλεπίδρασης μεταξύ τμημάτων φωνής και δεδομένων

Στα τηλέφωνα που βασίζονται στο IP παρέχεται μία πλατφόρμα για τηλεφωνικές κλήσεις μέσω ενός υπάρχοντος δικτύου δεδομένων IP. Ωστόσο, προκειμένου να διατηρηθεί η ποιότητα της υπηρεσίας (QoS), η επεκτασιμότητα, η δυνατότητα της διαχείρισης και η ασφάλεια, τα πακέτα φωνής και δεδομένων θα πρέπει να διαχωριστούν με τη χρήση διαφορετικών λογικών δικτύων όσο το δυνατόν περισσότερο. Η κατάτμηση των IP πακέτων φωνής από ένα δίκτυο δεδομένων IP ενισχύει σημαντικά την άμυνα στις επιθέσεις VoIP.

Δεδομένου ότι τα τμήματα φωνής και δεδομένων θα πρέπει να διαχωρίζονται, οι τεχνολογίες όπως τα εικονικά δίκτυα LAN (Virtual Local Area Network, VLAN), ο έλεγχος πρόσβασης, και τα τείχη προστασίας με έλεγχο κατάστασης του δικτύου (Stateful Firewalls), μπορούν να προσφέρουν τμηματοποίηση έως το τρίτο επίπεδο που είναι απαραίτητο για να κρατήσει τα τμήματα της φωνής και των δεδομένων διαχωρισμένα για το επίπεδο πρόσβασης. Τα Stateful Firewalls θα πρέπει να τοποθετούνται σε θέσεις του δικτύου όπου επιτρέπεται στα τμήματα να αλληλεπιδρούν. Οι μεταγωγείς του τρίτου επιπέδου (Layer 3) μπορούν επίσης να χρησιμοποιηθούν για τον έλεγχο της πρόσβασης σε τμήματα δεδομένων και φωνής μέσω του ελέγχου πρόσβασης και του φιλτραρίσματος.

Για περισσότερη προστασία στις VoIP υποδομές, οι εταιρείες θα μπορούσε να έχουν ένα κλειστό σύστημα VoIP. Οι χρήστες αυτών των εταιρειών, που διαθέτουν περισσότερα προνόμια, θα έχουν πρόσβαση στις υπηρεσίες VoIP μόνο για ενδοεπικοινωνία. Αυτό θα διαχωρίσει πλήρως τις υπηρεσίες VoIP από το Internet και θα ελαχιστοποιήσει τις απειλές από εξωτερικές επιθέσεις προς το εσωτερικό δίκτυο.

3.2.4 Επιθεώρηση πακέτων

Τα Δικτυακά Συστήματα Πρόληψης Εισχωρήσεων (Network Intrusion Prevention Systems, NIPS) μπορούν να αναλύσουν τα πρωτόκολλα VoIP. Για τα συστήματα VoIP που διαθέτουν υλοποίηση NIPS, οι επιθέσεις με στόχο τις υπηρεσίες VoIP μπορεί να ανιχνευθούν και να αποτραπούν [29][30].

3.2.5 Αντιικό Λογισμικό και προσθήκες στην Ασφάλεια

Οι υπολογιστές που χρησιμοποιούν λογισμικό για τις συνδέσεις VoIP, καθώς και όλοι οι εξυπηρετητές και πύλες VoIP που στηρίζονται σε λογισμικό, θα πρέπει να προστατεύονται. Η προστασία μπορεί να δοθεί από τη χρήση ατομικού τείχους προστασίας σε συνδυασμό με το αντιικό λογισμικό και το λογισμικό για κακόβουλο κώδικα. Η ενημέρωσή τους θα πρέπει να είναι με την πιο πρόσφατη υπογραφή του ιού ή/και τους νεότερους ορισμούς κακόβουλου κώδικα. Αυτό παρέχει μία βασική προστασία των επιθέσεων σε τμήματα δεδομένων που θα μπορούσαν να μεταφερθούν ως τμήματα φωνής. Επιπλέον, οι προσθήκες κώδικα εφαρμογής (patches) για την ασφάλεια των λογισμικών, που εξομοιώνουν τηλεφωνικές συσκευές σε ηλεκτρονικούς υπολογιστές, καθώς και οι δρομολογητές με τις πύλες VoIP, οφείλουν να είναι συνεχώς ενημερωμένες.

Λόγω της επικράτησης των απειλών από ιούς σε υπολογιστές πελάτες, μία VoIP τηλεφωνική συσκευή είναι προτιμότερη από ένα λογισμικό που εξομοιώνει την τηλεφωνική συσκευή (δηλαδή, ένας υπολογιστής που τρέχει το λογισμικό πελάτη VoIP). Ένας εξομοιωτής τηλεφώνου σε ηλεκτρονικό υπολογιστή είναι σε μεγαλύτερο κίνδυνο, διότι το λογισμικό VoIP μπορεί να είναι σε θέση να περνά τα πακέτα δεδομένων μέσα από το τείχος προστασίας. Εάν το λογισμικό δεν είναι σωστά ρυθμισμένο ή υπόκειται σε τρωτά σημεία, μπορεί να επιτρέψει κακόβουλα ή δόλια πακέτα δεδομένων να παρακάμψουν το τείχος προστασίας. Επιπλέον, οι υπολογιστές είναι περισσότερο ευάλωτοι σε ιούς, σκουλήκια και άλλες απειλές που οφείλονται σε ευπάθειες από άλλα λογισμικά του συστήματος, όπως τα προγράμματα περιήγησης στο διαδίκτυο. Ωστόσο, η

VoIP τηλεφωνικές συσκευές μερικές φορές μπορεί να έχουν τα δικά τους προβλήματα ασφάλειας. Ως παράδειγμα, μπορεί να υπάρχει ένα σχεδιαστικό ελάττωμα στο λογισμικό του ίδιου του τηλεφώνου.

3.2.6 Καλές πρακτικές για VoIP Τηλεφωνία σε τελικούς χρήστες

Οι χρήστες των VoIP τηλεφώνων θα πρέπει να γνωρίζουν τις απρόβλεπτες καταστάσεις στις οποίες μπορεί να περιέλθουν κατά την πραγματοποίηση φωνητικών κλήσεων όταν το σύστημα VoIP αποτύχει. Για την προστασία τους σε vishing επιθέσεις, οι χρήστες δεν θα πρέπει να γνωστοποιούν ευαίσθητες πληροφορίες, όπως τα στοιχεία των πιστωτικών καρτών, των τραπεζικών λογαριασμών ή τα στοιχεία σύνδεσης σε αγνώστους κατά τη διάρκεια μιας συνδιάλεξης με VoIP ή μίας απλής συνομιλίας.

Οι χρήστες λογισμικών που εξομοιώνουν IP τηλέφωνα, στους υπολογιστές τους θα πρέπει να προστατεύονται με κάποιο τείχος προστασίας, καθώς και αντιικό λογισμικό ή/και λογισμικό για κακόβουλο κώδικα. Η ενημέρωση των λογισμικών οφείλει να είναι με τις τελευταίες υπογραφές ιών και ορισμών για κακόβουλο κώδικα. Μια συνεπής διαδικασία διαχείρισης των πρόσθετων τμημάτων κώδικα (patches) θα πρέπει να εφαρμόζεται. Οι εταιρείες παραγωγής λογισμικού και οι πελάτες χρήστες οφείλουν να διασφαλίζουν πως όλα τα συστατικά στοιχεία του λογισμικού, συμπεριλαμβανομένων και των λογισμικών εξομοίωσης IP τηλεφώνου, που είναι εγκατεστημένα σε υπολογιστές να επιδιορθώνονται και να ενημερώνονται κατάλληλα. Τα ευαίσθητα δεδομένα δεν θα πρέπει να αποθηκεύονται σε ένα IP τηλέφωνο, καθώς δεν υπάρχουν συστήματα κρυπτογράφησης, για την προστασία των δεδομένων, ενσωματωμένα στις συσκευές αυτές. Άλλα άτομα, όπως οι διαχειριστές συστήματος, ίσως να είναι σε θέση να αποκτήσουν πρόσβαση σε ευαίσθητα δεδομένα ενός IP τηλεφώνου. Γίνεται εφικτό εξ αποστάσεως χρησιμοποιώντας για παράδειγμα το Πρωτόκολλο Μεταφοράς Ασήμαντων Αρχείων (Trivial File Transfer Protocol, TFTP).

3.3 Μηχανισμοί Αντιμετώπισης VoIP επιθέσεων

Οι πιο συνηθισμένες απειλές, που αναπτύσσονται σήμερα, στα VoIP συστήματα έχουν τις ρίζες τους στην πειρατεία των συστημάτων και των υποδομών που τα στηρίζουν. Οι μεγαλύτεροι προμηθευτές τηλεπικοινωνιακών κέντρων IP (Internet Protocol Private Branch Exchange, IP PBX) μπορούν να ασφαλιστούν, αλλά η ασφάλεια πρέπει να εξεταστεί κατά τη διάρκεια της

δημιουργίας των συστημάτων. Είναι σημαντικό να εξετάζονται οι υπάρχουσες συνθήκες ασφάλειας του δικτύου πριν από την προσθήκη VoIP συστημάτων. Ένα τεστ αξιολόγησης της ασφάλειας του συστήματος VoIP και της εισχώρησης σε αυτό θα βοηθήσει τον εντοπισμό των τρωτών σημείων του.

3.3.1 Έλεγχος των αρχείων CDR

Ο απλούστερος τρόπος για τον εντοπισμό της απάτης στα VoIP συστήματα μπορεί να γίνει με την ανάλυση των αρχείων CDRs (Αναλυτικές Εγγραφές των Κλήσεων, Call Detail Records). Οι πάροχοι VoIP υπηρεσιών θα μπορούσαν να θέτουν σε εφαρμογή ένα σύστημα που εντοπίζει τα πιο κοινά συμπτώματα της απάτης και της πειρατείας στα τηλεφωνικά τους κέντρα. Την αύξηση της κυκλοφορίας των φωνητικών πακέτων, τις ασυνήθιστες κλήσεις και άλλες περιπτώσεις. Τα συστήματα αυτά λειτουργούν καλύτερα όταν εκτελούνται σε σχεδόν πραγματικό χρόνο. Ιδιαίτερη προσοχή θα πρέπει να δίνουν τα συστήματα αυτά στις περιόδους της νύχτας, στα Σαββατοκύριακα, στις διακοπές και στις εορτές. Προσθήκες στα συστήματα αυτά θα ήταν η δυνατότητα να προειδοποιούνται οι διαχειριστές των συστημάτων ή ακόμη καλύτερα να μπορούσε το σύστημα ανίχνευσης της απάτης να ενσωματωθεί στο σύστημα δρομολόγησης και να μπλοκάρει προσωρινά ύποπτες κλήσεις.

3.3.2 Call Blocking

Ορισμένοι πάροχοι παίρνουν μια ακόμα πιο επιθετική στάση κατά της απάτης. Εμποδίζουν ολοκληρωτικά τις κλήσεις προς χώρες με πολλές αναφορές σε περιστατικά απάτης ή αποκλείουν κλήσεις σε τηλεφωνικά νούμερα που σχετίζονται με απάτη. Η μέθοδος αυτή αρνείται τη νόμιμη κίνηση των φωνητικών πακέτων και θα μπορούσε επίσης να γίνει μια χρονοβόρα διαδικασία ο συγχρονισμός με τα τελευταία νέα για τις χώρες και τους αριθμούς που σχετίζονται με απάτες σε VoIP δίκτυα.

3.3.3 Δρομολόγηση της κλήσης

Ορισμένοι πάροχοι VoIP υπηρεσιών έχουν επιλέξει να δρομολογούν τις κλήσεις από χώρες με πολλές αναφορές σε περιστατικά απάτης μόνο μέσω των διαχειριστών οι οποίοι εκτελούν διαδικασίες επικύρωσης των κλήσεων πριν την εγκαθίδρυσή τους. Αν και αυτή η μέθοδος μπορεί να περιορίσει την απάτη, είναι επίσης χρονοβόρα και οικονομικά ασύμφορη.

3.3.4 Νομική διαδικασία

Αν και είναι δελεαστικό στους παρόχους VoIP υπηρεσιών να στηρίζονται σε νομική δράση ως λύση στην απάτη στα VoIP συστήματα, δεν αποτελεί την ιδανική επιλογή προς αυτούς. Πρόσφατες μελέτες έχουν δείξει ότι οι εταιρείες επιλέγουν να μην αναφέρουν τις περιπτώσεις της απάτης με την επιβολή του νόμου, λόγω της προφανούς έλλειψης ενδιαφέροντος από μελλοντικούς πελάτες και την ανεπαρκή κατανόηση από τις κρατικές αρχές. Αν και το 88,6% των εταιρειών ανακοινώνουν τουλάχιστον μία προσφυγή στο νόμο ανά έτος, λίγοι απατεώνες τελικά συλλαμβάνονται και τιμωρούνται [27].

Παραδείγματα συλλήψεων από απάτη σε παρόχους VoIP υπηρεσιών δίνονται παρακάτω:

- Τον Μάιο του 2012, δύο άνδρες που έκλεψαν περισσότερα από 4,4 εκατομμύρια δολάρια από τους παρόχους VoIP καταδικάστηκαν σε ποινή φυλάκισης τριών ετών και υποχρεώθηκαν να επιστρέψουν τα χρήματα [31].

- Στην Αυστραλία τρεις απατεώνες, χρησιμοποιώντας τον τύπο απάτης Shell Companies, καταδικάστηκαν σε 40, 30 και 9 χρόνια φυλάκισης σε ομοσπονδιακές φυλακές και αποζημίωση 17,6, 3,3 και 18 εκατομμυρίων δολαρίων στους θιγμένους παρόχους VoIP υπηρεσιών [32].

3.3.5 Ελεγκτής Οριακής Συνεδρίας (SBC)

Οι Ελεγκτές Οριακών Συνεδριών (Session Border Controllers, SBCs), που δουλεύουν πριν από ένα τηλεφωνικό κέντρο IP, μπορεί να αποτελέσουν έναν καλό αποτρεπτικό παράγοντα στην απάτη. Το SBC μπορεί να ανιχνεύσει και να σταματήσει τις προσπάθειες επιτιθέμενου που επιδιώκει να μαντέψει τα πιστοποιητικά ενός χρήστη (όνομα χρήστη και κωδικό πρόσβασης) ή μη εξουσιοδοτημένες προσπάθειες στην κίνηση του δρομολογητή. Μερικοί SBCs μπορούν να αναλύσουν φωνητικά πρότυπα και δυναμικά να δημιουργήσουν πρότυπα από την κίνηση των τηλεπικοινωνιακών μας κλήσεων, στέλνοντας προειδοποιήσεις όταν κάτι παρεκκλίνει από το φυσιολογικό. Ωστόσο ο SBC δεν μπορεί να ανιχνεύσει επιθέσεις τύπου Location Routing Number Fraud και Arbitrage.

Κεφάλαιο 4
Μηχανισμός Ανίχνευσης Απάτης

Στη μάχη της παράνομης πρόσβασης μεταξύ των παρόχων υπηρεσιών VoIP και των ανθρώπων που προσπαθούν να εισχωρήσουν στα συστήματά τους, γίνονται συνεχώς μελέτες κι ανακοινώνονται προτάσεις κι από τις δύο πλευρές. Ο σκοπός στην εργασία αυτή είναι να προτείνει μία νέα εναλλακτική προσέγγιση στην αντιμετώπιση των δόλιων κλήσεων από τους πελάτες των παρόχων VoIP υπηρεσιών.

Η πρόταση της παρούσας εργασίας βασίζεται στη συσταδοποίηση ομάδων με κοινά χαρακτηριστικά ως προς τον προορισμό των κλήσεων, τη διάρκεια που έχουν οι κλήσεις, τη χρονική στιγμή που ξεκινά μία κλήση και τη συχνότητα των κλήσεων ως προς τους προορισμούς. Από τα παραπάνω δεδομένα εξάγονται κι άλλες πληροφορίες οι οποίες συνδράμουν στη συσταδοποίηση.

Η επιλογή της συσταδοποίησης οφείλεται στο βασικό ερώτημα της παρούσας εργασίας. Πώς δηλαδή θα μπορέσει ένας πάροχος VoIP υπηρεσιών να απομονώσει φυσιολογικές κλήσεις και να ενημερωθεί για τις πιθανές δόλιες τηλεφωνικές κλήσεις που περνούν μέσα από τα συστήματά του. Τα πιθανά σενάρια που θέτουμε στην εργασία αυτή, δηλαδή κλήσεις προς προορισμούς που έχουν κόστος στον πάροχο VoIP υπηρεσιών, μας ορίζουν τη χρήση της ομαδοποίησης στις VoIP

κλήσεις. Τα πιθανά σενάρια που θα εξετάσει η παρούσα εργασία βασίζονται στις παρακάτω ομάδες:

- Ομάδα περιοχών με ακριβούς σε χρήματα προορισμούς, όπως είναι οι κλήσεις σε κινητά τηλέφωνα.

- Ομάδα με τη διάρκεια των κλήσεων. Δηλαδή ανάλογα με την ώρα και την ημέρα παρατηρείται εναλλαγή της διάρκειας των τηλεφωνικών κλήσεων. Για παράδειγμα σύντομες σε διάρκεια τηλεφωνικές κλήσεις κατά το διάστημα της εργάσιμης ημέρας και ώρας έναντι των μεγάλης διάρκειας τηλεφωνικών κλήσεων που παρατηρούνται στη διάρκεια του Σαββατοκύριακου.

- Συχνότητα των κλήσεων. Όπως και παραπάνω έτσι κι εδώ, ανάλογα με την ώρα και την ημέρα της κλήσης, παρατηρούνται διακυμάνσεις στο σύνολο των κλήσεων. Πάρα πολλές τηλεφωνικές κλήσεις συναντώνται στη διάρκεια μίας εργάσιμης ημέρας, έναντι των λίγων τηλεφωνικών κλήσεων που δημιουργούνται στη διάρκεια ενός Σαββατοκύριακου.

Τα σενάρια που θα χρησιμοποιηθούν στην εργασία αυτή δίνονται στο επόμενο κεφάλαιο. Στο Κεφάλαιο 5. Όπως φαίνεται από τα παραπάνω, η χρήση της συσταδοποίησης θα βοηθήσει στην αποπεράτωση της εργασίας αυτής, χωρίς να μπορούμε να αποκλείσουμε κι άλλες μεθόδους που μπορούν να ακολουθήσουν άλλες εργασίες σχετικές με το αντικείμενο που εξετάζουμε.

Σύμφωνα με τη Wikipedia ως Συσταδοποίηση (Clustering) μπορούμε να ορίσουμε «*τη διαδικασία εκείνη κατά την οποία ένα σύνολο από «αντικείμενα», διαχωρίζονται σε ένα σύνολο από λογικές ομάδες. Η καταχώρηση αντικειμένων σε ίδια ομάδα μεταφράζεται ως ομοιότητα των αντικειμένων αυτών και αντίστροφα (αντικείμενα που ανήκουν σε διαφορετικές ομάδες είναι ανόμοια). Η ομοιότητα ή μη, μεταξύ των αντικείμενων, ουσιαστικά εξαρτάται από το συγκεκριμένο πρόβλημα και τη μορφή των «αντικειμένων». Στη βιβλιογραφία συναντάται ως ομαδοποίηση και μη επιβλεπόμενη μάθηση. Τα αντικείμενα μπορούν να αναφερθούν και αυτά με διαφορετικούς όρους: πρότυπα, διανύσματα*».

Η καθηγήτρια του Πανεπιστημίου Ιωαννίνων Ευαγγελία Πιτουρά, στις διαφάνειες του μαθήματός της, ορίζει τη συσταδοποίηση ως την *«εύρεση συστάδων (ομάδων) αντικειμένων έτσι ώστε τα αντικείμενα σε κάθε συστάδα να είναι όμοια (ή να σχετίζονται) και διαφορετικά (ή μη σχετιζόμενα) από τα αντικείμενα των άλλων συστάδων»*.

Γενικότερα θα μπορούσαμε να ορίσουμε τη συσταδοποίηση ως μία διαδικασία καταμερισμού διαφόρων οντοτήτων, των οποίων τα χαρακτηριστικά δεν είναι όμοια μεταξύ τους, σε ομάδες(Clusters). Ο καταμερισμός τους γίνεται σύμφωνα με την ομοιότητα που παρουσιάζουν κατά προσέγγιση τα χαρακτηριστικά των παραπάνω οντοτήτων.

Δεν αποτελεί κομμάτι μίας επιστήμης η συσταδοποίηση. Μελετάται συνεχώς κι εξελίσσεται, όπως στο χώρο της Εξόρυξης Γνώσης, στη Μηχανική Μάθηση, στις Χωρικές Βάσεις Δεδομένων, στην Στατιστική, στο Marketing και αλλού.

Η συσταδοποίηση σύμφωνα με τους Pang Tan, Michael Steinvach και Vipin Kumar διακρίνεται σε Καλά Διαχωρισμένη, βασισμένη σε Πρότυπο, βασισμένη σε Γράφο, βασισμένη στην Πυκνότητα και σε Συστάδες με Κοινές Ιδιότητες [37]. Για την υλοποίηση της συσταδοποίησης οντοτήτων χρησιμοποιούνται πολλοί αλγόριθμοι. Οι βασικότεροι είναι:

- K-Mean: Ο συγκεκριμένος αλγόριθμος είναι από τους πιο πολύ εφαρμοσμένους και είναι η ρίζα για πολλούς άλλους. Ανήκει στην κατηγορία της επίπεδης συσταδοποίησης διότι παράγει ένα σύνολο συσταδοποιήσεων χωρίς να έχουν καμία ιδιαίτερη δομή-σχέση μεταξύ τους. Ο αλγόριθμος έχει ως στόχο την βελτιστοποίηση μίας συνάρτησης – της συνάρτησης κόστους.

- Ιεραρχικοί: Οι ιεραρχικοί αλγόριθμοι συσταδοποίησης παράγουν μια ιεραρχία εμφωλιασμένων συσταδοποιήσεων. Οι αλγόριθμοι αυτοί διαχωρίζονται στους συσσωρευτικούς και στους διαιρετικούς.

- Αυτό-οργανωμένοι Χάρτες (Self-Organizing Maps, SOM): Οι αυτό-οργανωμένοι χάρτες ή Δίκτυο Kohonen (το όνομα προήλθε από το δημιουργό του αλγόριθμου), είναι μία τεχνική συσταδοποίησης και οπτικοποίησης των δεδομένων που βασίζεται στην επιστήμη των Νευρωνικών Δικτύων. Παρά το ότι ο αυτό-οργανωμένος χάρτης προέρχεται από το χώρο των Νευρωνικών Δικτύων, μπορεί να παρουσιαστεί πιο εύκολα ως μια παραλλαγή της συσταδοποίησης βάση προτύπων. Στόχος του αλγορίθμου είναι η εύρεση ενός συνόλου κέντρων βάρους (Νευρώνες στην ορολογία SOM) και η εκχώρηση κάθε οντότητας, που δίνεται ως είσοδος στο δίκτυο, στον αντίστοιχο Νευρώνα.

Παραλλαγές των παραπάνω αλγορίθμων ή ακόμη εξέλιξη αυτών, υπάρχουν πολλές και συνεχώς παρουσιάζονται νέες. Κάθε αλγόριθμος προσαρμόζεται στις ανάγκες των οντοτήτων που δίνονται στο σύστημα ώστε να εφαρμοσθεί η συσταδοποίηση με όσο το δυνατό καλύτερα αποτελέσματα.

Στην παρούσα εργασία γίνεται χρήση των αλγορίθμων K-Mean, Self-Organizing Maps και Self-Organizing Feature Maps.

Στη συνέχεια δίνεται μία σύντομη βιβλιογραφική ανασκόπηση σχετική με το αντικείμενο που διαπραγματεύεται η εργασία αυτή και θα ακολουθήσει η περιγραφή των βημάτων και των αναγκών που υπήρξαν για το σχεδιασμό της εφαρμογής ανίχνευσης δόλιων κλήσεων.

4.1 Συσταδοποίηση (clustering) και Νευρωνικά Δίκτυα σε VoIP συστήματα

4.1.1 Συσταδοποίηση με K-Mean

Ένας από τους πρωτοπόρους στην ανίχνευση της εισβολής είναι η Dorothy Denning [36]. Στη δημοσίευσή της προτείνει ένα γενικής χρήσης σύστημα ανίχνευσης της εισβολής με βάση:

- την παρακολούθηση και τον έλεγχο των εγγραφών.
- την ανίχνευση των παραβιάσεων στην ασφάλεια από τις αποκλίσεις του φυσιολογικού προφίλ.

Για παράδειγμα, σε μια προσπάθεια διάρρηξης από υποψήφιους εισβολείς, εμφανίζεται ένα υψηλό ποσοστό αποτυχημένων κωδικών πρόσβασης κάποιου χρήστη. Μια επιτυχημένη εισβολή εμφανίζεται ως μια ανώμαλη σύνδεση (π.χ. ο χρόνος, ο τόπος, ο τύπος της σύνδεσης, η συμπεριφορά). Μια διαρροή μπορεί να εμφανιστεί από τη δρομολόγηση των δεδομένων σε μία αχρησιμοποίητη κανονικά συσκευή, όπως για παράδειγμα σε έναν απομακρυσμένο εκτυπωτή σε μια ασυνήθιστη ώρα.

Η Denning στο μοντέλο της προτείνει τη χρήση κανόνων. Αντίθετα οι συγγραφείς του άρθρου «Unsupervised Profiling for Identifying Superimposed Fraud» προτείνουν τη δημιουργία ενός μη ανιχνεύσιμου προφίλ της συμπεριφοράς των πελατών απέναντι στους επιτιθέμενους χρησιμοποιώντας τη συσταδοποίηση. Η προσέγγισή τους, περιλαμβάνει τη συμπεριφορά του καλούντος (για παράδειγμα το χρόνο και τον τόπο), τις αλλαγές στη συμπεριφορά των πελατών και τη συνεχή εμφάνιση νέων τεχνικών απάτης. Με βάση αυτές τις μοναδικές απαιτήσεις, οι συγγραφείς προτείνουν ένα προφίλ με τρία επίπεδα: το προφίλ της κλήσης, το προφίλ κάθε ημέρας και το συνολικό προφίλ. Το πρώτο επίπεδο προτυποποιεί την κατανομή των κλήσεων

στη διάρκεια του χρόνου εντός της ημέρας, τη διάρκεια της κλήσης και τον προορισμό (τοπική κλήση, διεθνής κλήση, κλήση προς κινητό και άλλους τύπους) και στα δύο είδη των VoIP κλήσεων: της φωνής και των δεδομένων. Το δεύτερο επίπεδο προτυποποιεί τα αποτελέσματα ποσοτικών και ποιοτικών προφίλ, όπως είναι ο συνολικός αριθμός των κλήσεων σε μια ημέρα και το ποσοστό των μη μηδενικών κλήσεων αντίστοιχα. Η προτυποποίηση στο τρίτο επίπεδο περιέχει τη μακροπρόθεσμη συμπεριφορά των πελατών. Ένας αλγόριθμος συσταδοποίησης εφαρμόζεται στα καθημερινά προφίλ ώστε να εξαχθούν τα πρότυπα των ημερών και να εμφανιστούν σε διαφορετικές συστάδες. Ο αλγόριθμος συσταδοποίησης βασίζεται στην απόσταση μεταξύ των ποιοτικών προφίλ. Οι συγγραφείς υποστηρίζουν ότι οι γνωστές μορφές της απόστασης (Ευκλείδεια, Hellinger, Mahalonobis, και άλλες) δεν είναι κατάλληλες για την περίπτωσή τους. Προτείνουν μια νέα απόσταση με βάση την αθροιστική κατανομή. Οι συγγραφείς εφαρμόζουν ένα τροποποιημένο K-means αλγόριθμο, αντικαθιστούν την Ευκλείδεια απόσταση με την αθροιστική κατανομή, για να συσταδοποιήσουν τα ποιοτικά χαρακτηριστικά και να υπολογιστεί στη συνέχεια το πρωτότυπο προφίλ για κάθε συστάδα. Η ανίχνευση των καθημερινών ανώμαλων προφίλ γίνεται με ποιοτικό και ποσοτικό έλεγχο στο πλησιέστερο πρωτότυπο προφίλ. Το προφίλ των πελατών είναι μη ανιχνεύσιμο, με την έννοια ότι δεν προκαθορίζεται με πρότυπα χρήσης. Η προσέγγιση αξιολογείται με εξωτερικά στοιχεία (μέσο όρο, τυπική απόκλιση) από μοντέλα που βασίζονται σε κανόνες χρησιμοποιώντας δεδομένα από τον πραγματικό κόσμο και ημι-συνθετικά. Τα πλεονεκτήματα αυτής της προσέγγισης είναι ο δυναμισμός της (δηλαδή τα πρωτότυπα προφίλ μπορούν να ενημερώνονται συνεχώς), η ευαισθησία στην αντιγραφή παράνομων προτύπων και ειδικότερα η ένδειξη καλών επιδόσεων κάτω από ένα χαμηλό θετικό ρυθμό ψεύδους.

4.1.2 Νευρωνικά Δίκτυα με SOM

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ, Artificial Neural Networks, ANN) είναι εμπνευσμένα από τις νευρωνικές συνδέσεις του ανθρώπινου εγκέφαλου και θεωρούνται ένα χρήσιμο εργαλείο στην επιστήμη της εξόρυξης δεδομένων, παράλληλα με τις παραδοσιακές στατιστικές τεχνικές. Τα ΤΝΔ χρησιμοποιούνται σε πολλές εφαρμογές που αντιμετωπίζουν προβλήματα ταξινόμησης είτε συσταδοποίησης. Η εμπειρία μπορεί να εξομοιωθεί σε ένα νευρωνικό δίκτυο με τη σωστή εκπαίδευση. Μετά τη φάση της εκπαίδευσης οι αποφάσεις που θα παίρνει το σύστημα θα μπορούν να έχουν μεγάλη ακρίβεια. Ένα ΤΝΔ αποτελείται από μονάδες που ονομάζονται νευρώνες. Οι νευρώνες οργανώνονται σε επίπεδα. Κάθε ΤΝΔ έχει τουλάχιστον δύο επίπεδα από νευρώνες, το επίπεδο της εισόδου και το επίπεδο της εξόδου. Όλοι οι νευρώνες έχουν τα ατομικά

τους βάρη, που συνδέονται με κάθε εισερχόμενη σύνδεση. Η εκπαίδευση ενός ΤΝΔ σημαίνει την προσαρμογή των τιμών των εν λόγω βαρών, προκειμένου να παράγει σωστές προβλέψεις. Υπάρχουν αρκετοί τύποι Τεχνητών Νευρωνικών Δικτύων. Οι Feed Forward Neural Networks (FFNNs), τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks, RNNs) και οι Αυτό-Οργανωμένοι Χάρτες (Self-Organizing Maps, SOMs) είναι μερικά από τα πιο κοινά είδη τους. Παρά τα πλεονεκτήματά τους, υπάρχουν δύο μειονεκτήματα στα ΤΝΔ:

- Εάν γίνεται σωστή προσαρμογή, στη φάση της εκπαίδευσης, στα προβλήματα που θα κληθεί να λύσει.

- Δεν μπορεί να γίνει καμία ουσιαστική εξήγηση της συμπεριφοράς τους άμεσα, με τη μορφή συγκεκριμένων κανόνων όπως γίνεται στο χώρο των επιχειρήσεων. Ακόμα και η γνώση των συγκεκριμένων τιμών, που έχουν λάβει τα βάρη, δεν έχει νόημα όταν είναι απαραίτητη μια εξήγηση, σχετικά με τη λογική που υπάρχει πίσω από τα αποτελέσματα του ΤΝΔ.

Σημαντική ερευνητική προσπάθεια έχει καταβληθεί για την αντιμετώπιση διαφόρων τύπων απάτης με τη χρήση των τεχνητών νευρωνικών δικτύων. Η έρευνα διεξάγεται κυρίως για απάτες σε ασφαλιστικές επιχειρήσεις, απάτες με πιστωτικές κάρτες, απάτες στα οικονομικά, χωρίς να αποκλείονται και άλλοι χώροι στην καθημερινή ζωή των ανθρώπων. Οι δυνατότητες συσταδοποίησης των Αυτό-Οργανωμένων Χαρτών φαίνεται να κερδίζει τα θετικά σχόλια στην ερευνητική κοινότητα για τα προβλήματα της απάτης. Οι SOMs χρησιμοποιήθηκαν αρχικά για τα προβλήματα σε εικόνες και ήχους, αλλά σύντομα η ικανότητά τους να εντοπίζουν συστάδες στα δεδομένα αναγνωρίστηκε. Διαφέρουν από τα παραδοσιακά ΤΝΔ ως προς την τοπολογία τους και τη μέθοδο της εκπαίδευσής τους. Το επίπεδο της εξόδου αποτελείται από ένα σημαντικά μεγαλύτερο αριθμό νευρώνων, σε σύγκριση με άλλους τύπους των ΤΝΔ, και είναι οργανωμένοι σε ένα πλέγμα με σχήμα συνήθως τετραγώνου. Επιπλέον, στους SOMs ο νευρώνας στο επίπεδο της εισόδου είναι άμεσα συνδεδεμένος με κάθε νευρώνα στο επίπεδο της εξόδου. Οι νευρώνες στο επίπεδο της εξόδου δεν έχουν καμία απευθείας σύνδεση μεταξύ τους. Στη φάση της εκπαίδευσης των SOMs, όταν επιλέγεται ο νικητής νευρώνας στο επίπεδο της εξόδου προσαρμόζονται τα βάρη του. Στη συνέχεια, προσαρμόζονται τα βάρη των γειτονικών νευρώνων. Δεδομένου ότι η εκπαίδευση συνεχίζεται μέχρι την τελευταία εποχή, το μέγεθος της γειτονιάς και των προσαρμογών στα βάρη μειώνεται σε μέγεθος.

Μια αρχιτεκτονική διαχείρισης της απάτης σε δίκτυα επόμενης γενιάς, που χρησιμοποιούν SOMs προτείνεται στην εργασία των Bihina Bella και άλλοι [34]. Χρησιμοποιούν μία διαδικασία

ανίχνευσης με πολλά επίπεδα, όπου αναλύουν τα αρχεία τιμολόγησης, IP Detail Records (IPDRs), για να εμφανίσουν άγνωστα σενάρια απάτης σε δίκτυα νέας γενιάς (New Generation Network, NGN). Οι συγγραφείς υποστηρίζουν ότι «Ο SOM είναι ένα αποτελεσματικό εργαλείο για την ανάλυση των δεδομένων που χρησιμοποιούμε στις υπηρεσίες μας ώστε να ανιχνεύσουμε τα σημάδια απάτης».

Ως συνέπεια, τα συστήματα ανίχνευσης απάτης σε δίκτυα VoIP μπορούν να επωφεληθούν από ένα SOM που χρησιμοποιείται μαζί με ένα σύστημα κανόνων, προκειμένου να ανιχνεύσουν γνωστές και άγνωστες απειλές. Η είσοδος στο εργαλείο SOM θα μπορούσε να είναι σε κάθε κλήση, η διάρκειά της, ο προορισμό της και διάφορες άλλες παράμετροι που περιγράφουν το προφίλ του πελάτη. Ως έξοδο το εργαλείο θα μπορούσε να τοποθετήσει την κλήση σε μία κατηγορία με «πιθανότητα απάτης». Θα πρέπει να αναφέρουμε ότι τα σημεία απάτης που εντοπίζονται από το SOM δεν αρκούν για να διαπιστώσουμε ότι η απάτη έχει συμβεί πραγματικά. Χρειάζεται και περαιτέρω έρευνα από το διαχειριστή του συστήματος.

Στα VoIP δίκτυα δημιουργούνται τεράστιες ποσότητες δεδομένων για την τιμολόγηση και την χρήση τους. Ως εκ τούτου, ένας γρήγορος μηχανισμός ανίχνευσης της απάτης είναι πλέον σημαντικός. Τα ΤΝΔ εγκατεστημένα σε παράλληλες μηχανές μπορούν να επιταχύνουν την παραγωγή κανόνων για την ανίχνευση της απάτης από πελάτες, όπως περιγράφεται στην εργασία των Yufeng Kou και άλλοι [33].

Η εργασία τους [33] υποστηρίζει τη χρήση των νευρωνικών δικτύων σε Συστήματα Ανίχνευσης Εισχώρησης (Intrusion Detections Systems, IDS), τα οποία θα είναι καλύτερα από τη χρήση των δομημένων κανόνων στην ανίχνευση της απάτης σε εμπορικά εργαλεία. Ισχυρίζονται ότι οι τεχνικές αυτές είναι λιγότερο επιτυχείς όταν τα χαρακτηριστικά της επίθεσης διαφέρουν από τους καταχωρημένους κανόνες, και ότι τα τεχνητά νευρωνικά δίκτυα έχουν πολλά πλεονεκτήματα για την επίλυση αυτών των προβλημάτων. Τα συστήματα εύρεσης ανωμαλιών είναι πολύ δύσκολο να κατασκευαστούν, διότι είναι δύσκολο να προσδιοριστούν οι φυσιολογικές και μη φυσιολογικές συμπεριφορές των χρηστών. Αντιθέτως, με τη χρήση των νευρωνικών δικτύων διευκολύνεται η διαδικασία αυτή, επειδή μπορούν να μάθουν να διακρίνουν την κανονική και την ανώμαλη συμπεριφορά ενός συστήματος, λαμβάνοντας τη σωστή εκπαίδευση από τα παραδείγματα που θα δοθούν.

4.2 Εφαρμογή Δημιουργίας Κλήσεων και Αποθήκευσης CDRs

Οι πάροχοι VoIP υπηρεσιών δεν είναι εύκολο να προμηθεύσουν ερευνητικά εργαστήρια με δεδομένα τηλεφωνικών κλήσεων. Δίνεται η δυνατότητα, όταν υπάρχει καλή συνεργασία μεταξύ του παρόχου και του ερευνητή. Η συνεργασία αυτή θεμελιώνεται στην εμπιστοσύνη που υπάρχει στο πρόσωπο του ερευνητή. Ένας νέος τηλεπικοινωνιακός πάροχος στην αναζήτησή του για λογισμικό που θα προστατεύει τον εξοπλισμό μπορεί να βρεθεί σε αδιέξοδο. Το αυξημένο κόστος και οι ελλείψεις που συναντά στις εμπορικές εφαρμογές τον οδήγούν στη συνεργασία με πανεπιστημιακούς φορείς που εξειδικεύονται στις τηλεπικοινωνίες.

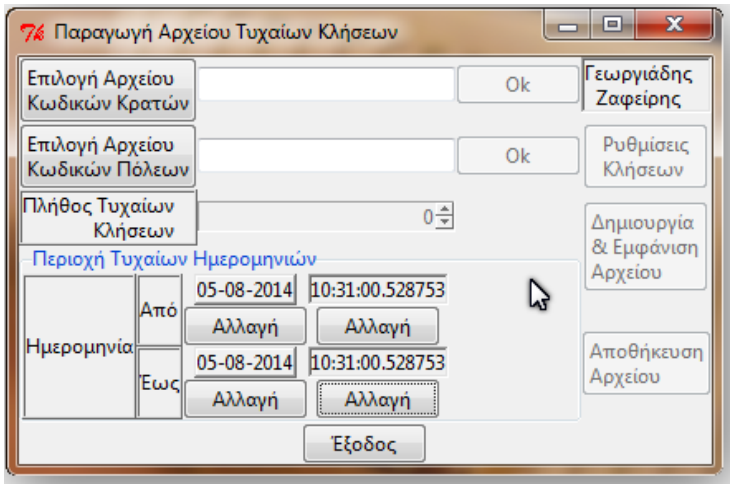
Για την εργασία αυτή μας παραδόθηκε, από πάροχο VoIP υπηρεσιών, ένα δείγμα δύο μηνών με τηλεφωνικές κλήσεις που διεξήχθησαν σε επίπεδο αστικό, υπεραστικό, διεθνές όπως επίσης και προς κινητά τηλέφωνα.

Με βάση το αρχείο αυτό, έγινε μία ανάλυση των δεδομένων από το ερευνητικό εργαστήριο. Τα αποτελέσματα αυτής της έρευνας βοηθούν στο ξεκίνημα της δημιουργίας μίας εφαρμογής, που θα παράγει τυχαίες κλήσεις. Η εφαρμογή θα βασίζεται σε μία γεννήτρια τυχαίων κλήσεων όπως αυτή αναφέρεται στο παράρτημα Α. Οι τυχαίες κλήσεις θα ακολουθούν τους κανόνες που ορίζει το πρότυπο E.164 του International Telecommunication Union [38] όπως και τα εξαγόμενα στοιχεία που δόθηκαν από το ερευνητικό εργαστήριο κι εμφανίζονται στο Παράρτημα Β.

4.2.1 Δημιουργία της Εφαρμογής

Στην Εικόνα 4.1 εμφανίζεται η εφαρμογή που δημιουργεί τυχαίες τηλεφωνικές κλήσεις προς κινητά τηλέφωνα, τοπικούς και διεθνείς προορισμούς. Οι εφαρμογές «Παραγωγή Αρχείου Τυχαίων Κλήσεων» και «Ανίχνευση Δόλιων Κλήσεων» είναι γραμμένες σε γλώσσα περιγραφική (Script), χρησιμοποιήθηκε η Python 3.3, ώστε να είναι εύκολη η επικοινωνία τους με άλλες υποστηρικτικές εφαρμογές του παρόχου VoIP υπηρεσιών. Χρησιμοποιήθηκε το περιβάλλον JetBrains PyCharm Community Edition 3.4 που διανέμεται δωρεάν για μη κερδοσκοπική χρήση και διευκολύνει τον προγραμματιστή με πολλές έξυπνες λειτουργίες που προσφέρει για τη γλώσσα Python. Η ευελιξία που προσφέρει η γλώσσα Python στη δημιουργία εφαρμογών αλλά και το περιβάλλον ανάπτυξης εφαρμογών σε γλώσσα Python της εταιρείας JetBrains, συμπεριλαμβάνουν τη δυνατότητα, ο προγραμματιστής να εκπονεί την εργασία του σε

οποιοδήποτε λειτουργικό σύστημα επιθυμεί, είτε είναι Windows, είτε κλώνος του Linux, είτε Macintosh.

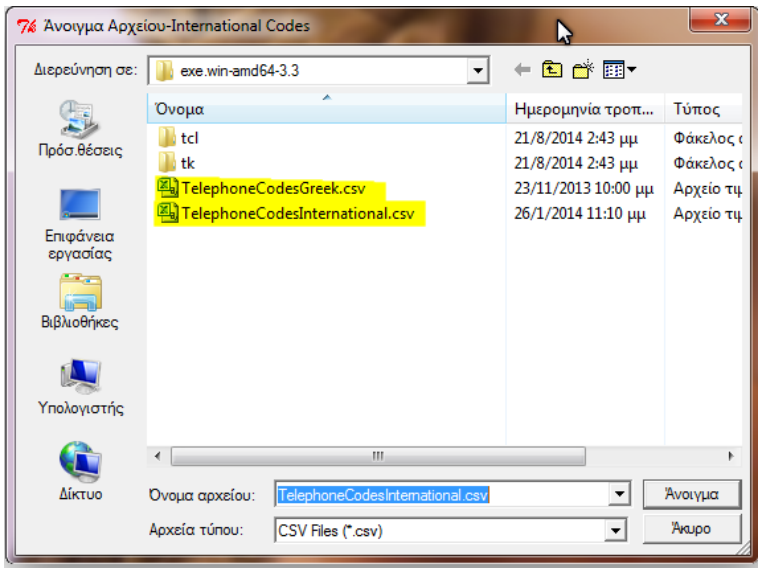


Εικόνα 4.1 Η αρχική οθόνη για την παραγωγή τυχαίων αριθμών κλήσης

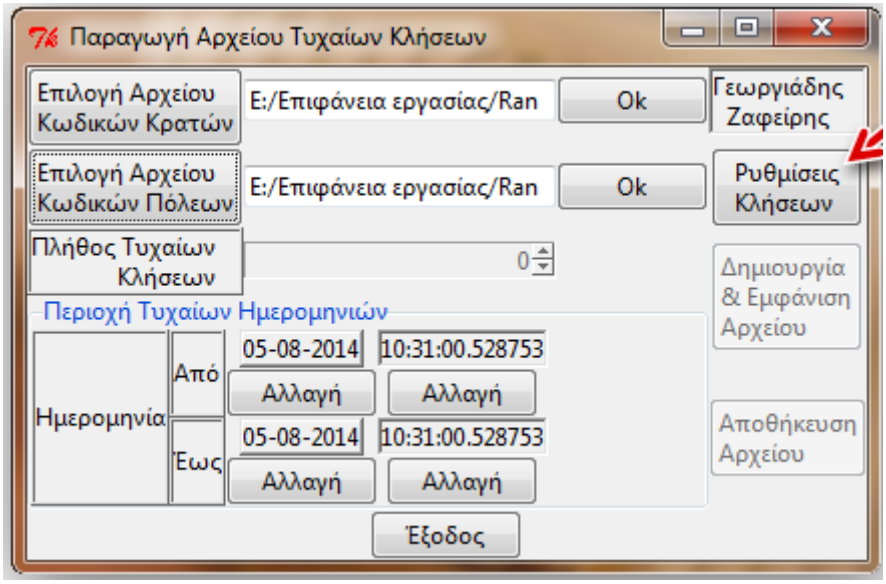
4.2.2 Λειτουργίες της Εφαρμογής

Για να ξεκινήσει η εφαρμογή να παράγει τις κλήσεις θα πρέπει να εισαχθούν οι κανόνες που ορίζουν τη γενική μορφή των τηλεφωνικών αριθμών. Τμήμα από τους κανόνες που περιέχουν τα αρχεία δίνεται στο Παράρτημα Β.

Επιλέγοντας το σωστό αρχείο στο αντίστοιχο πλαίσιο, όπως φαίνονται στην Εικόνα 4.2 και Εικόνα 4.3, για κάθε κανόνα παραγωγής τηλεφωνικών αριθμών ενεργοποιείται το κουμπί για τις ρυθμίσεις των κλήσεων.

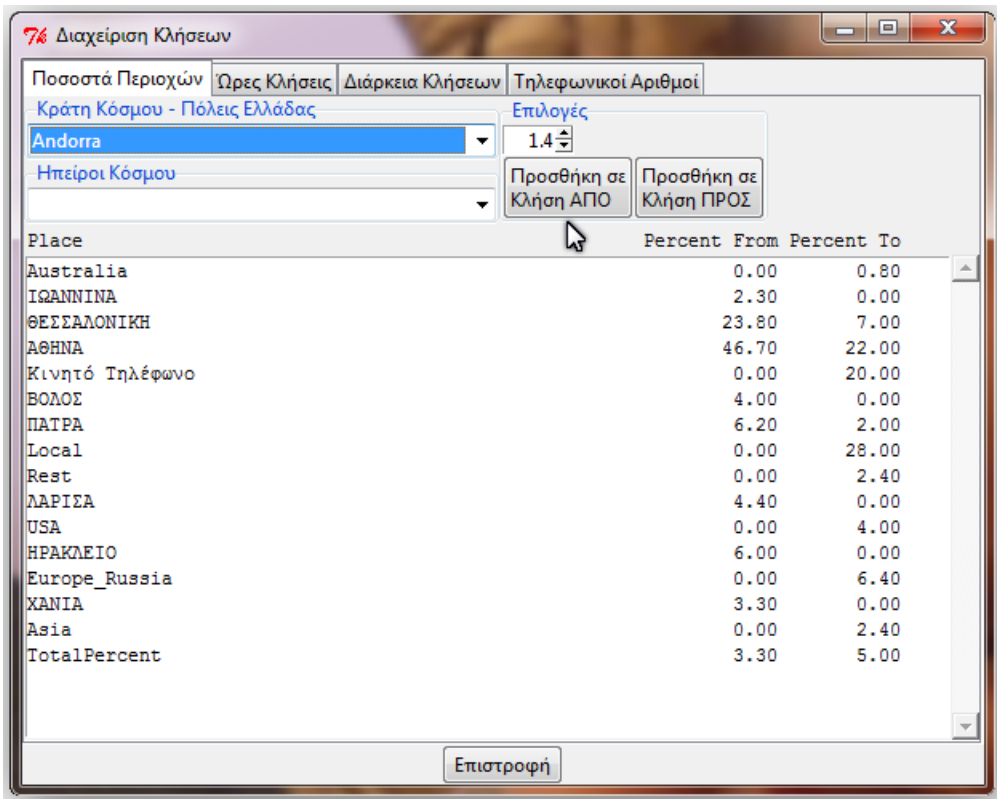


Εικόνα 4.2 Η αρχική οθόνη για την παραγωγή τυχαίων αριθμών κλήσης

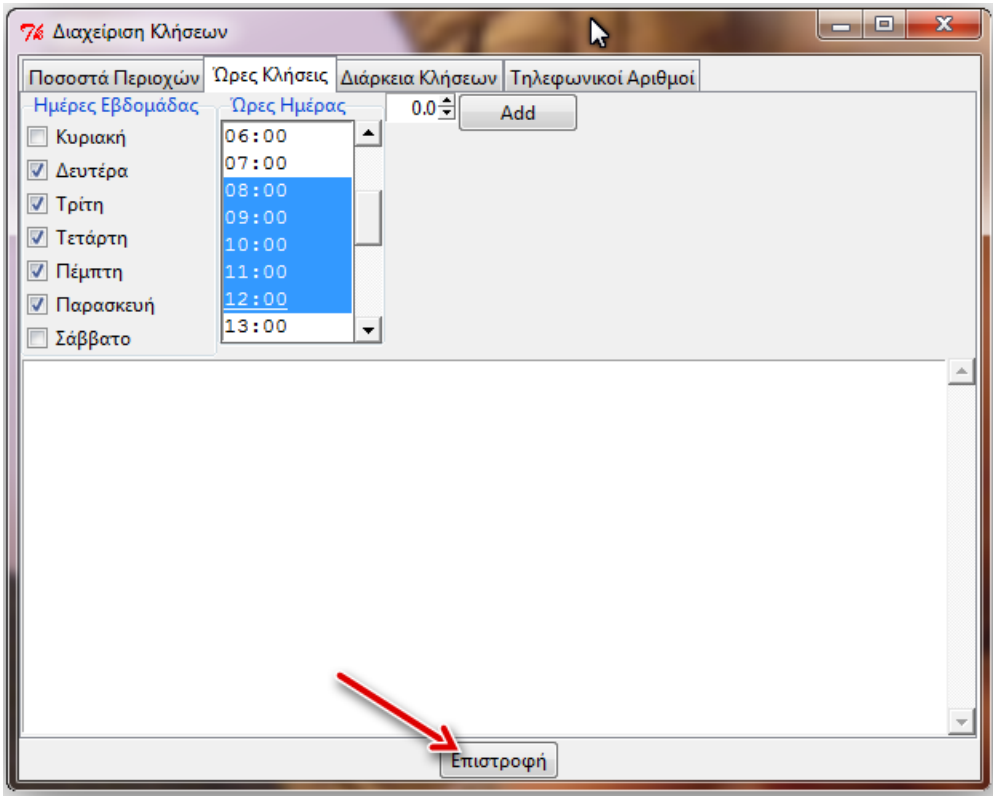


Εικόνα 4.3 Η αρχική οθόνη μετά την επιλογή των αρχείων με τους κανόνες

Στην Εικόνα 4.4 και Εικόνα 4.5, εμφανίζεται μία σειρά από καρτέλες που βοηθούν στον εμπλουτισμό των κανόνων για τη δημιουργία των τυχαίων τηλεφωνικών κλήσεων. Τα ποσοστά που δίνονται αρχικά έχουν εξαχθεί ύστερα από μελέτες που έγιναν από τον πάροχο VoIP υπηρεσιών και το Πανεπιστημιακό Ερευνητικό Εργαστήριο που στηρίζει την παρούσα εργασία.



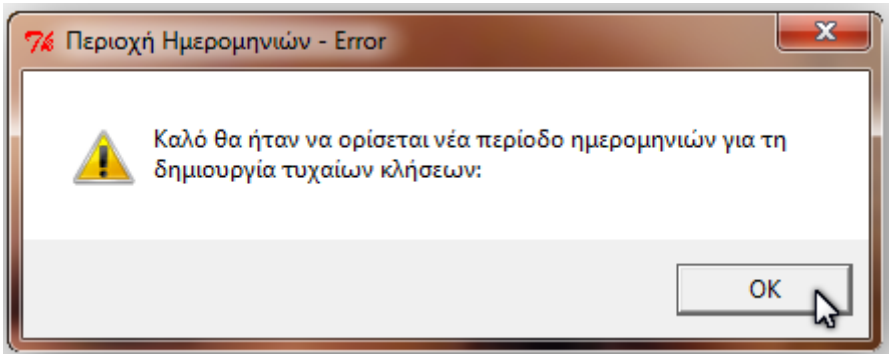
Εικόνα 4.4 Επιλογή ποσοστών τηλεφωνικών κλήσεων Από και Προς διάφορες περιοχές



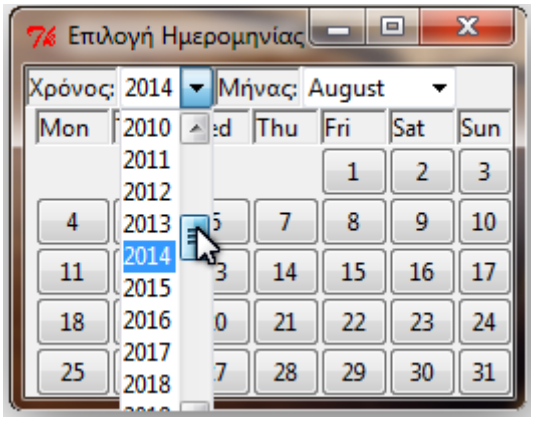
Εικόνα 4.5 Επιλογή ποσοστού τηλεφωνικών κλήσεων στη διάρκεια μίας εβδομάδας

Με την επιστροφή στο κύριο παράθυρο της εφαρμογής ενεργοποιείται και το κουμπί για την έναρξη της δημιουργίας των τηλεφωνικών κλήσεων.

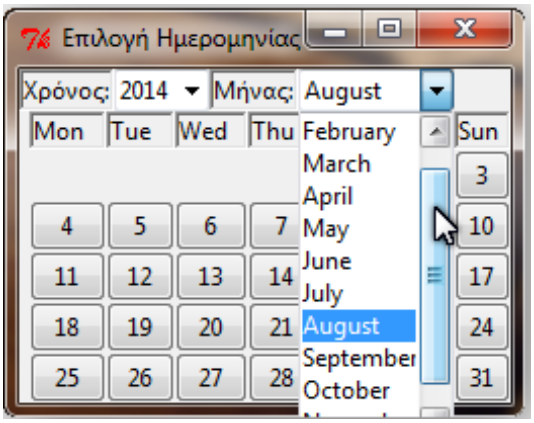
Για οποιοδήποτε σφάλμα συμβεί κατά την εξέλιξη της δημιουργίας των τηλεφωνικών κλήσεων η εφαρμογή μας ενημερώνει με αντίστοιχα μηνύματα που εμφανίζονται σε ξεχωριστά παράθυρα. Ένα προειδοποιητικό παράθυρο εμφανίζεται στην παρακάτω εικόνα. Ο χρήστης της εφαρμογής υπέπεσε στο σφάλμα να μην ενημερώσει την εφαρμογή για το διάστημα που επιθυμεί να παράγει της τυχαίες τηλεφωνικές κλήσεις. Επιστρέφοντας στο κύριο παράθυρο της εφαρμογής μπορούμε να αλλάξουμε τις ημερομηνίες και τις ώρες έναρξης και λήξης της περιόδου των τηλεφωνικών κλήσεων. Στην Εικόνα 4.7, Εικόνα 4.8 και Εικόνα 4.9 φαίνονται τα παράθυρα που ενεργοποιούνται με το αίτημα του χρήστη να αλλάξει την ημερομηνία και την ώρα, είτε πρόκειται για την έναρξη της περιόδου παραγωγής τυχαίων τηλεφωνικών κλήσεων, είτε πρόκειται για τη λήξη τους.



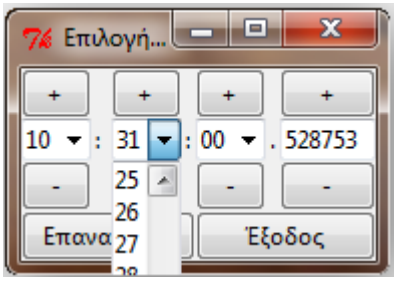
Εικόνα 4.6 Προειδοποιητικό μήνυμα για ένα λογικό σφάλμα του χρήστη της εφαρμογής



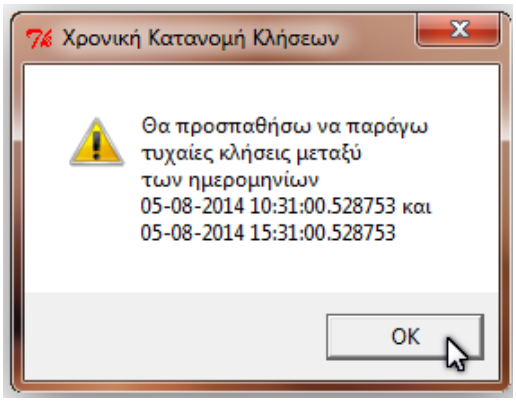
Εικόνα 4.7 Επιλογή έτους



Εικόνα 4.9 Επιλογή μήνα



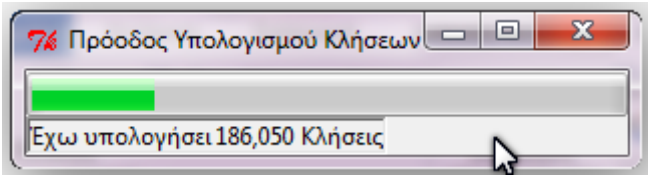
Εικόνα 4.8 Επιλογή ώρας



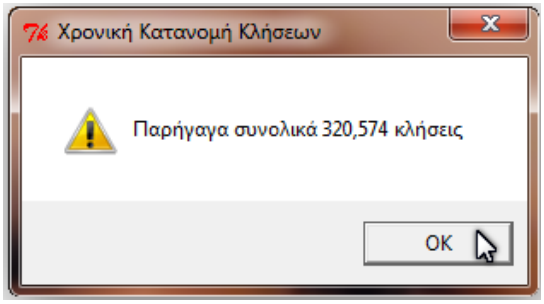
Εικόνα 4.10 Ενημερωτικό μήνυμα για τη διάρκεια των τυχαίων τηλεφωνικών κλήσεων

Στην Εικόνα 4.10 εμφανίζεται ένα προειδοποιητικό μήνυμα που μας ενημερώνει για την ώρα και την ημέρα που θα ξεκινήσει και θα σταματήσει η γεννήτρια να παράγει τυχαίες τηλεφωνικές κλήσεις.

Στην Εικόνα 4.11 εμφανίζεται ένα παράθυρο ενημερωτικό για την εξέλιξη της παραγωγής του συνόλου των τυχαίων τηλεφωνικών κλήσεων. Πληροφορίες περισσότερες δίνονται στο Παράρτημα Α στο αντίστοιχο τμήμα του κώδικα της εφαρμογής.

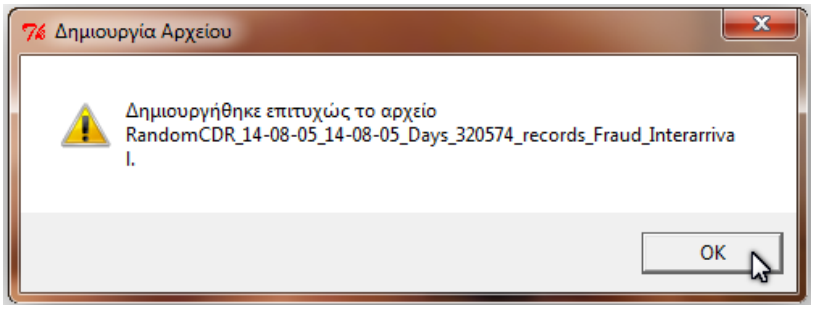


Εικόνα 4.11 Μπάρα προόδου για την πορεία της παραγωγής των τυχαίων τηλεφωνικών κλήσεων

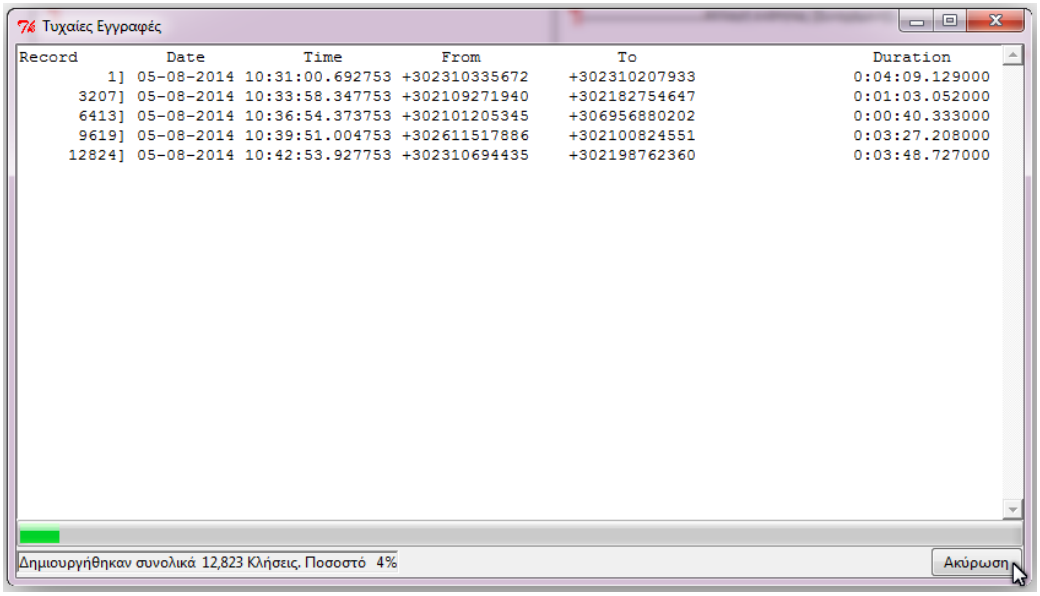


Εικόνα 4.12 Προειδοποιητικό μήνυμα για το σύνολο των τυχαίων τηλεφωνικών κλήσεων

Εφόσον ολοκληρωθεί η διαδικασία παραγωγής των τηλεφωνικών κλήσεων εμφανίζεται το παράθυρο που φαίνεται στην Εικόνα 4.12. Κατόπιν εμφανίζεται νέο παράθυρο, το παρακάτω στην Εικόνα 4.13, που μας ενημερώνει για το όνομα του αρχείου που θα χρησιμοποιηθεί για την αποθήκευση των τυχαίων τηλεφωνικών κλήσεων



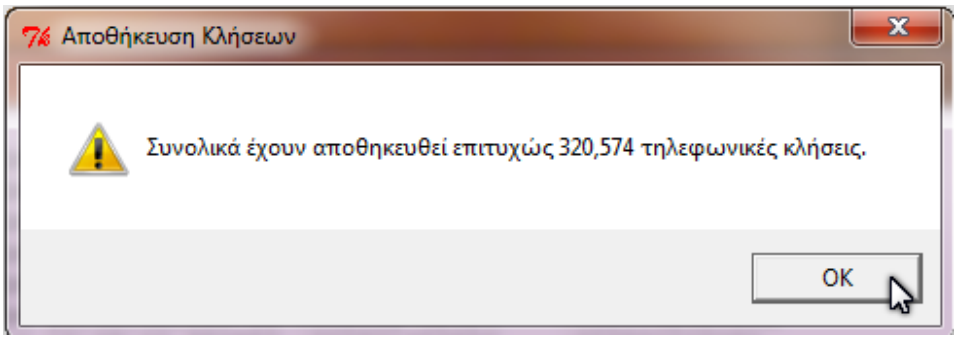
Εικόνα 4.13 Ενημερωτικό μήνυμα για το πού θα περιέχει τις τυχαίες τηλεφωνικές κλήσεις



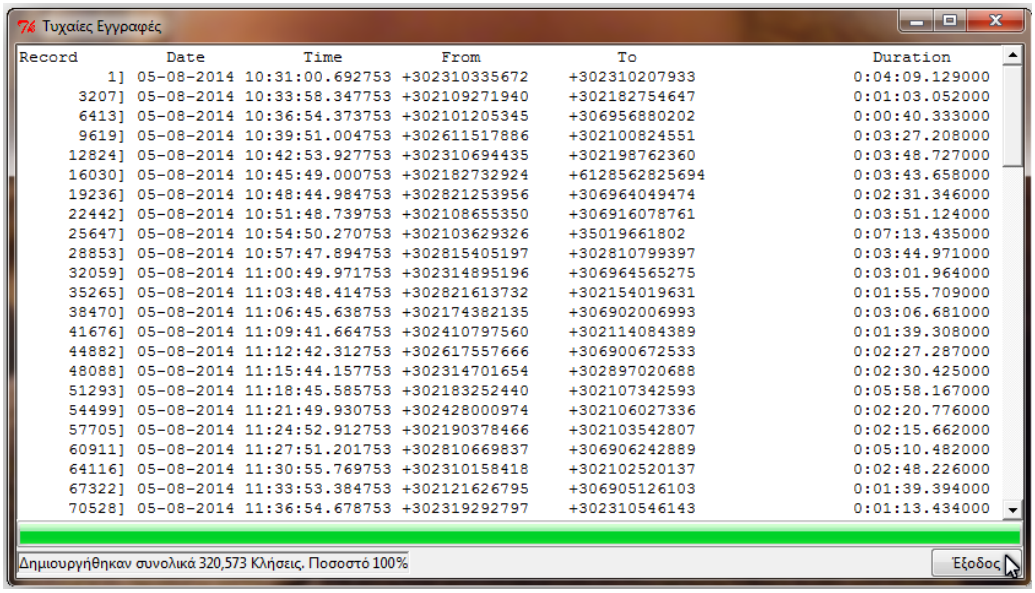
Εικόνα 4.14 Εμφάνιση ενός μικρού μέρους των τυχαίων τηλεφωνικών κλήσεων

Στην Εικόνα 4.14 εμφανίζεται ένα μέρος των τηλεφωνικών κλήσεων που δημιουργήθηκαν (το βασικό μέρος που βλέπει κι ένας πελάτης ενός παρόχου VoIP υπηρεσιών). Στην τρέχουσα φάση της εφαρμογής οι τυχαίες κλήσεις που δημιουργήθηκαν από το προηγούμενο βήμα τώρα λαμβάνουν περισσότερες πληροφορίες όπως είναι ο αριθμός τηλεφώνου που καλεί και καλείται, η πόλη της Ελλάδος από την οποία ξεκινά η κλήση και ο τόπος προορισμού. Δείγμα τέτοιων εγγραφών στα αρχεία CDR δίνονται στο Παράρτημα Β.

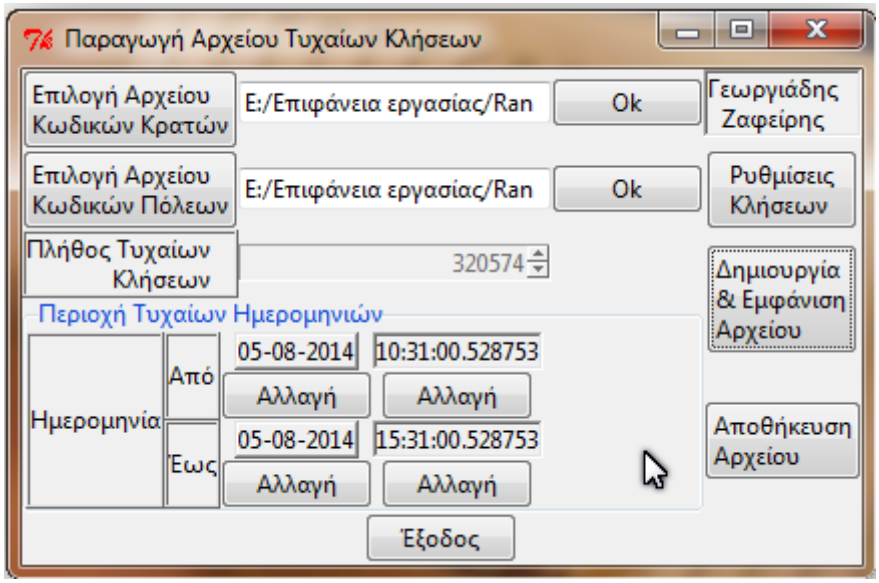
Με την ολοκλήρωση των τυχαίων τηλεφωνικών κλήσεων εμφανίζεται ένα προειδοποιητικό μήνυμα όπως στην παρακάτω εικόνα και μας ενημερώνει για την ομαλή αποθήκευση των δεδομένων στο αρχείο που μας γνωστοποίησε νωρίτερα η εφαρμογή.



Εικόνα 4.15 Ενημερωτικό μήνυμα επιτυχούς αποθήκευσης τηλφωνικών κλήσεων



Εικόνα 4.16 Εμφάνιση ενός μικρού μέρους των τυχαίων τηλεφωνικών κλήσεων



Εικόνα 4.17 Οθόνη κύριας εφαρμογής με το πέρας της δημιουργίας τυχαίων τηλεφωνικών κλήσεων.

Στην Εικόνα 4.16 εμφανίζεται πλέον το κουμπί «Έξοδος» με το οποίο μπορούμε να επιστρέψουμε στο κύριο παράθυρο της εφαρμογής όπως αυτό φαίνεται στην Εικόνα 4.17. Το κουμπί «Αποθήκευση Αρχείου μας δίνει τη δυνατότητα να αποθηκεύσουμε τις τυχαίες τηλεφωνικές κλήσεις και σε ένα άλλο αρχείο με την ίδια (CSV) μορφή ή μία άλλη.

4.3 Εφαρμογή Ανίχνευσης Δόλιων Κλήσεων

Οι εταιρείες που ασχολούνται με εφαρμογές για την ανίχνευση των δόλιων κλήσεων από τους πελάτες ενός παρόχου VoIP υπηρεσιών, όπως αναφέρεται και παραπάνω, διαθέτουν πολλά εργαλεία. Άλλα χρησιμοποιούνται για την πρόληψη, τοποθετώντας εξειδικευμένα τείχη προστασίας, φίλτρα ή άλλους τρόπους, κι άλλα εργαλεία χρησιμοποιούνται για τη διάγνωση των δόλιων κλήσεων.

Τα εργαλεία που χρησιμοποιούνται για την διάγνωση των δόλιων κλήσεων χρησιμοποιούν κυρίως τη μέθοδο της συσταδοποίησης. Η εφαρμογή που παρουσιάζεται παρακάτω χρησιμοποιεί τη μέθοδο της συσταδοποίησης με τη βοήθεια των αλγορίθμων K-Mean, SOM και SOFM.

Στα παρακάτω υπό-τμήματα θα παρουσιασθεί η εφαρμογή που ανιχνεύει δόλιες κλήσεις όταν υπάρξει διαφοροποίηση στο προφίλ των κλήσεων σε ακριβούς προορισμούς.

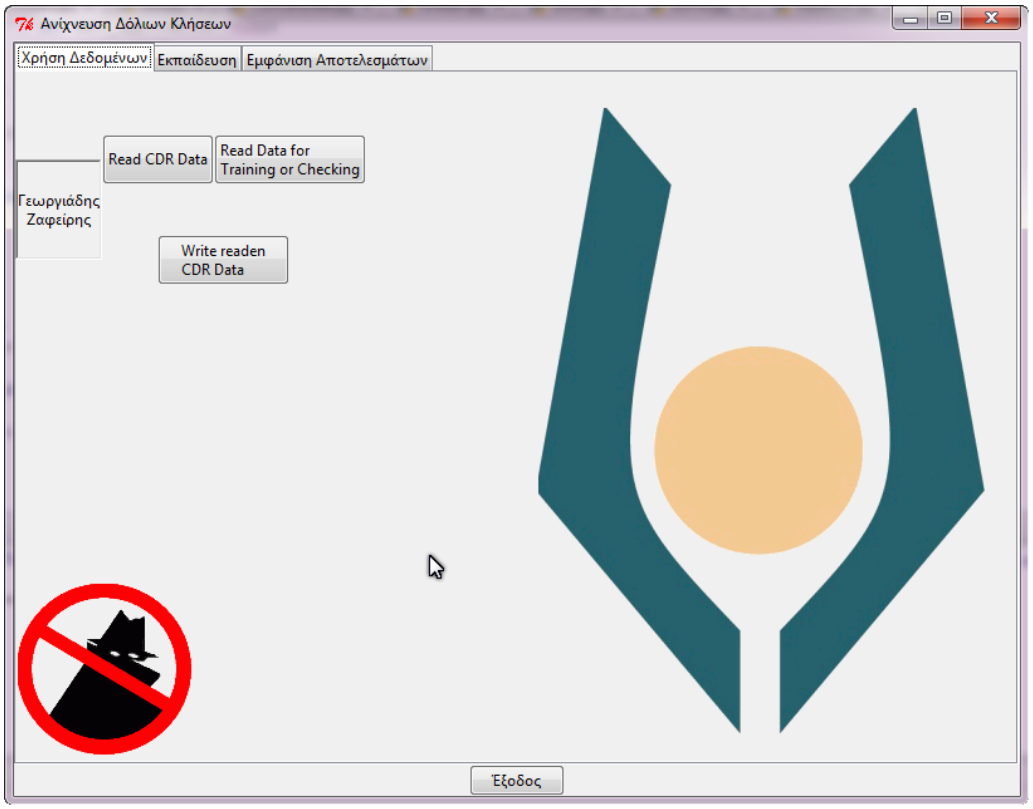
4.3.1 Δημιουργία της Εφαρμογής

Όπως αναφέρθηκε και στο υπό-τμήμα 4.2.1 η εφαρμογή γράφτηκε στο περιβάλλον JetBrains PyCharm Community Edition 3.4 όπως αυτό φαίνεται στην Εικόνα 4.18. Γλώσσα προγραμματισμού κι εδώ χρησιμοποιήθηκε η περιγραφική γλώσσα (script language) Python στην έκδοση 3.3. Στο γραφικό περιβάλλον της εφαρμογής χρησιμοποιήθηκε η βιβλιοθήκη Tkinter και η TTK. Η επιλογή τους έγινε για την καλή ανταπόκριση που έχουν οι δύο αυτές βιβλιοθήκες σε όλα τα λειτουργικά συστήματα που υποστηρίζονται από την περιγραφική γλώσσα Python. Για τη διαχείριση των αρχείων χρησιμοποιείται η βιβλιοθήκη CSV, ώστε να είναι δυνατή η ανάγνωση των δεδομένων σε οποιοδήποτε πρόγραμμα υπολογιστικών φύλλων. Η κωδικοποίηση των γραμμάτων σε «UTF-8» βοηθά στην ομαλή αποθήκευση και ανάκτηση ειδικών χαρακτήρων όπως είναι και το ελληνικό αλφάβητο. Όλες οι παραπάνω βιβλιοθήκες παρέχονται με την περιγραφική γλώσσα Python.

Εικόνα 4.18 Το περιβάλλον προγραμματισμού JetBrains PyCharm Community Edition 3.4

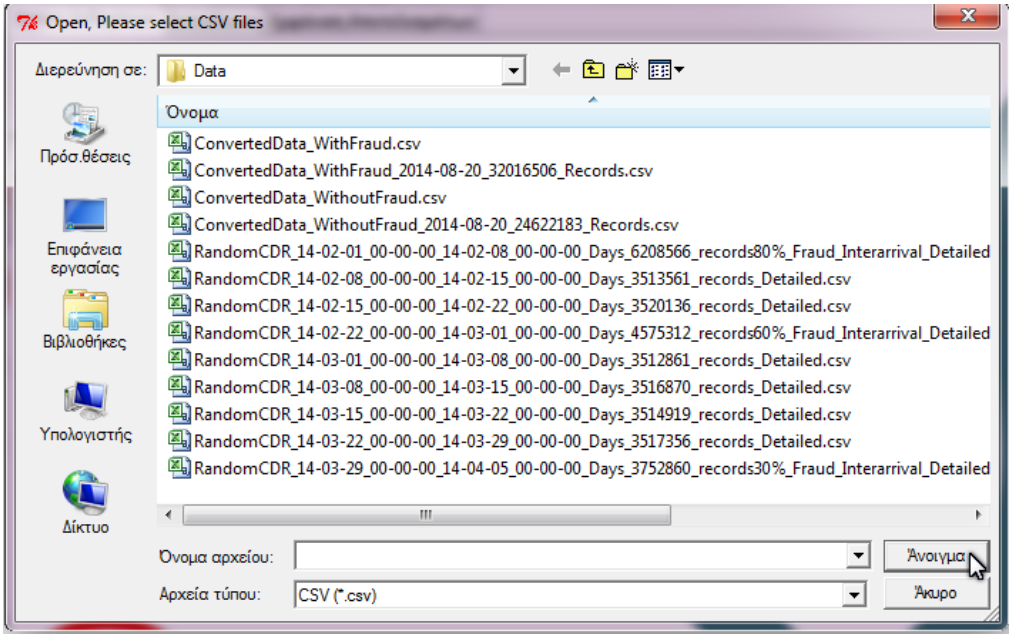
4.3.2 Λειτουργίες της Εφαρμογής

Στην Εικόνα 4.19 εμφανίζεται το αρχικό περιβάλλον της εφαρμογής «Ανίχνευση Δόλιων Κλήσεων». Η εφαρμογή για τις ανάγκες της εργασίας αυτής είναι αποκολλημένη από το σύστημα VoIP. Για την τροφοδοσία της εφαρμογής με δεδομένα CDR χρησιμοποιούνται δύο επιλογές.



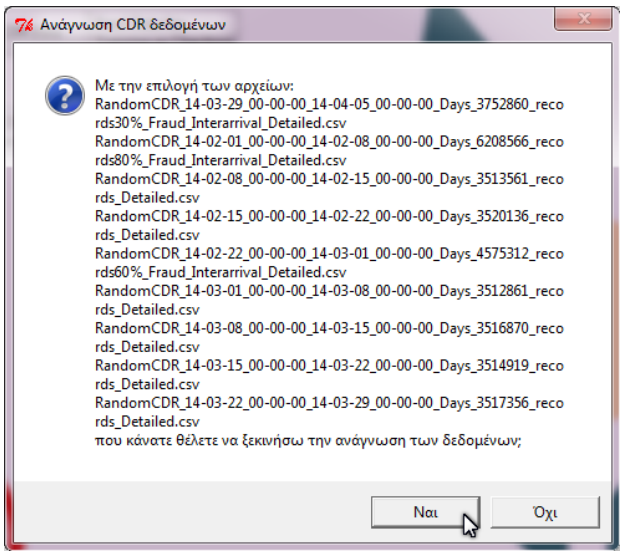
Εικόνα 4.19 Αρχική οθόνη της εφαρμογής «Ανίχνευση Δόλιων Κλήσεων»

Η πρώτη επιλογή γίνεται πιέζοντας το κουμπί «Read CDR Data» για να γίνει η επιλογή των αρχείων με τις τυχαίες τηλεφωνικές κλήσεις, όπως αυτή φαίνεται στην Εικόνα 4.20. Τα αρχεία CDR δομούνται σύμφωνα με τις ανάγκες της εφαρμογής ώστε να επιτευχθεί το ελάχιστο δυνατό μέγεθος των αρχείων, χωρίς να παραλείπονται ουσιώδη δεδομένα όπως είναι η ώρα έναρξης και λήξης της τηλεφωνικής κλήσης, ο αριθμός τηλεφώνου που καλεί και καλείται, κι άλλα δεδομένα.



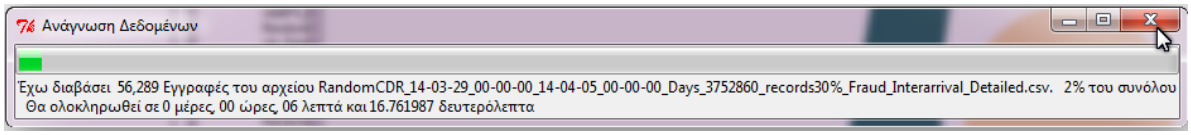
Εικόνα 4.20 Επιλογή αρχείων με CDR δεδομένα

Μετά την επιλογή των αρχείων εμφανίζεται ένα μήνυμα που μας ενημερώνει για την επιλογή που κάναμε. Το παράθυρο του μηνύματος φαίνεται στην παρακάτω εικόνα.



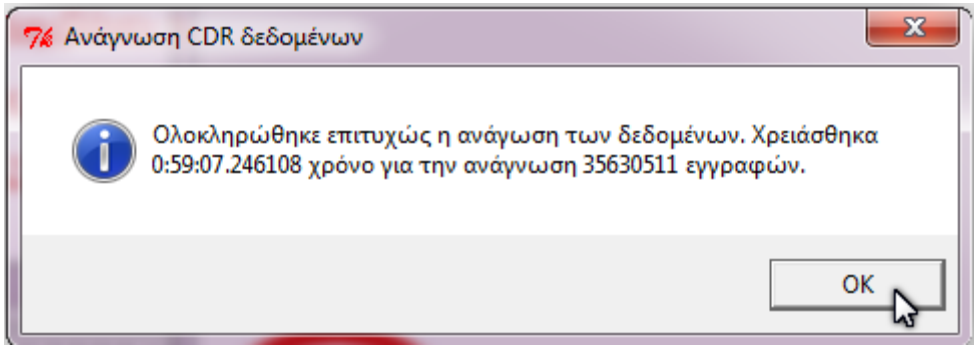
Εικόνα 4.21 Επιβεβαίωση επιλογής αρχείων με CDR δεδομένα

Ένα παράθυρο μας ενημερώνει για την εξέλιξη της ανάγνωσης των αρχείων με τις τηλεφωνικές κλήσεις. Η ενημέρωση περιλαμβάνει το αρχείο που ανοίχθηκε, το ποσοστό ανάγνωσης του αρχείου και τον πιθανό χρόνο τερματισμού της διαδικασίας αυτής. Ένα δείγμα του παραθύρου αυτού δίνεται στην Εικόνα 4.22.

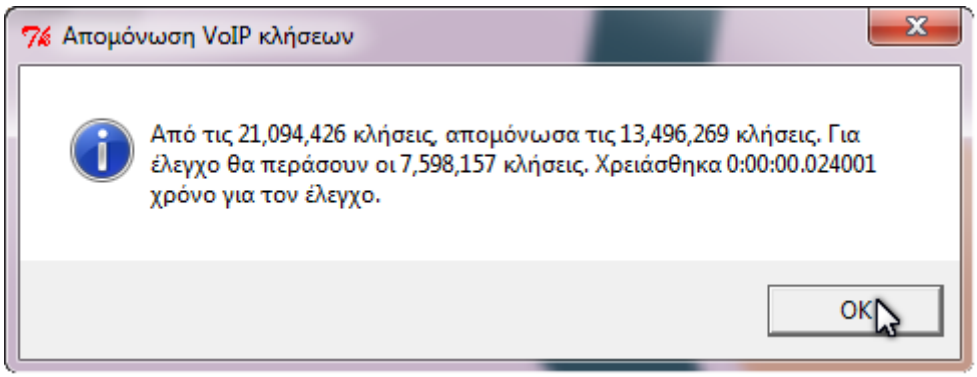


Εικόνα 4.22 Ενημερωτικό παράθυρο για την εξέλιξη της ανάγνωσης αρχείων CDR

Με το πέρας της ανάγνωσης των τηλεφωνικών κλήσεων και τη μετατροπή τους στη μορφή που χρειάζεται το υπόλοιπο μέρος της εφαρμογής, εμφανίζεται ένα ενημερωτικό παράθυρο όπως στην Εικόνα 4.23. Ο όγκος των τυχαίων τηλεφωνικών κλήσεων είναι πολύ μεγάλος για ένα διάστημα μίας εβδομάδας. Μετά την απομάκρυνση των δεδομένων που δεν είναι απαραίτητα για τις ανάγκες της εφαρμογής, όπως αναφέρεται και παραπάνω, το μέγεθος του αρχείου για μία εβδομάδα κυμαίνεται από τα 750 MB έως κι 1.4 GB. Η ανάγκη για τη μείωση του χρόνου της ανάγνωσης των τηλεφωνικών κλήσεων από τα αρχεία CDR μας οδηγεί στη δημιουργία αρχείων που θα περιέχουν συγκεντρωτικά στοιχεία των τυχαίων τηλεφωνικών κλήσεων. Η εφαρμογή για την ταχύτερη αναγνώριση και ανάλυση των τηλεφωνικών κλήσεων απομονώνει όσες δεν έχουν κόστος για τον πάροχο. Στην Εικόνα 4.24 δίνεται ένα ενημερωτικό μήνυμα για την παραπάνω ενέργεια.

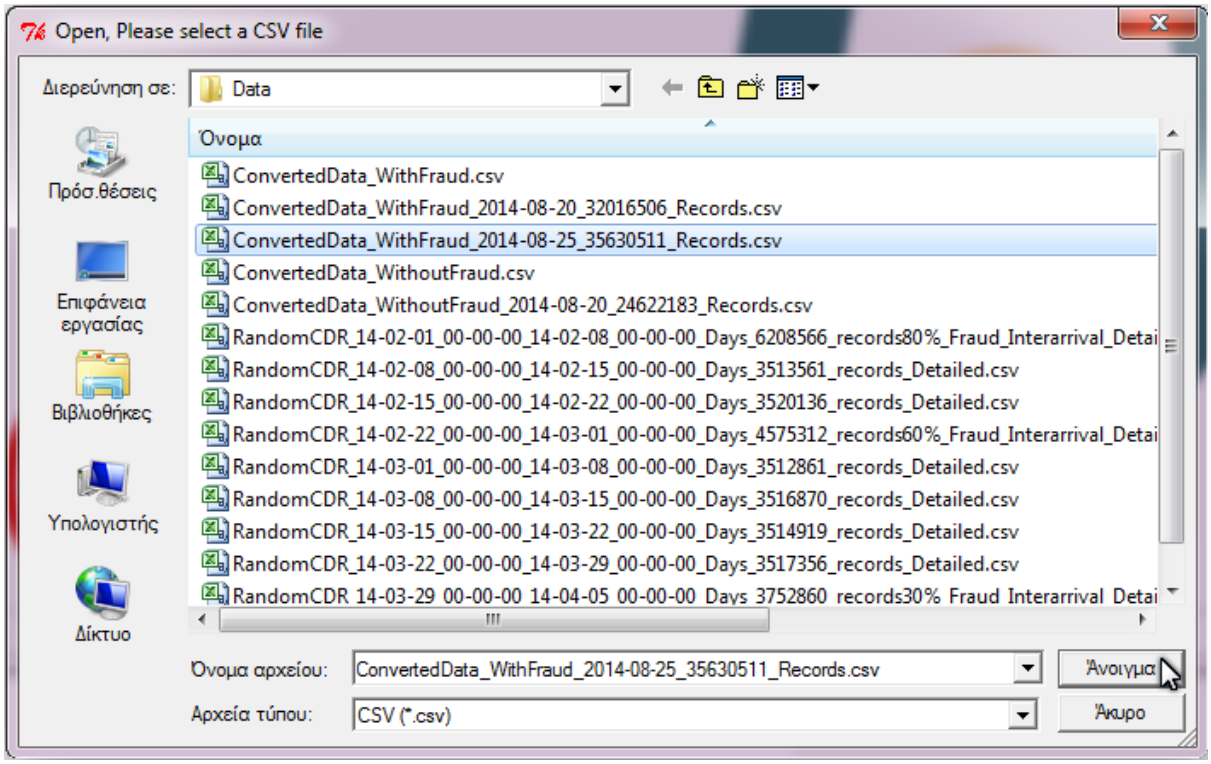


Εικόνα 4.23 Ενημερωτικό παράθυρο για τη λήξη της ανάγνωσης αρχείων CDR



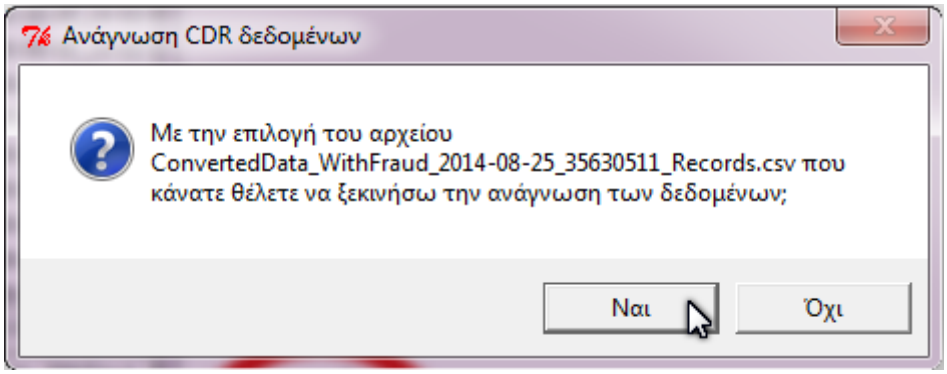
Εικόνα 4.24 Ενημερωτικό παράθυρο για την απομόνωση VoIP κλήσεων.

Η δεύτερη επιλογή «Read Data for Training or Checking» χρησιμοποιείται για το άνοιγμα αρχείου που περιέχει δεδομένα αποθηκευμένα σε μορφή επεξεργάσιμη από την υπόλοιπη εφαρμογή. Τα αρχεία αυτά περιέχουν στοιχεία σχετικά με το πλήθος και τη διάρκεια των κλήσεων προς διάφορες περιοχές. Στην Εικόνα 4.25 εμφανίζεται ένα δείγμα του παραθύρου με το οποίο επιλέγουμε το αρχείο που θα χρησιμοποιηθεί για την ανάλυση και εύρεση πιθανής δόλιας χρήσης της υπηρεσίας VoIP.



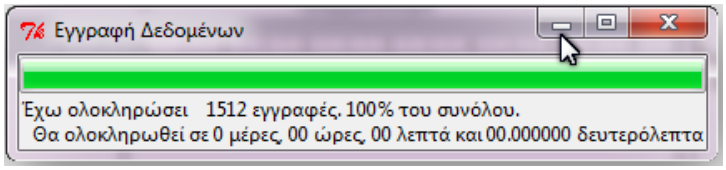
Εικόνα 4.25 Επιλογή αρχείου με δεδομένα για την εφαρογή

Ένα δεύτερο παράθυρο, όπως φαίνεται και στην Εικόνα 4.26 επιβεβαιώνει την επιλογή του αρχείου που θα ανοιχθεί για ανάγνωση.

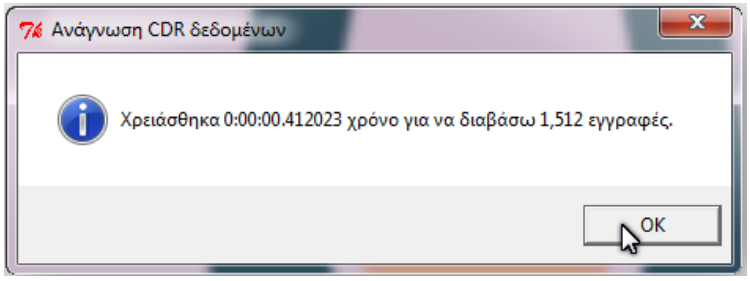


Εικόνα 4.26 Επιβεβαίωση για την επιλογή του αρχείου με δεδομένα για την εφαρμογή

Απαντώντας θετικά στο παραπάνω παράθυρο η εφαρμογή εμφανίζει ένα νέο παράθυρο, όπως φαίνεται και στην Εικόνα 4.27, που μας ενημερώνει για το ποσοστό των δεδομένων που έχει διαβάσει από το αρχείο που ζητήσαμε να ανοίξει, όπως επίσης και το χρόνο που απομένει για την ολοκλήρωση της ανάγνωσης του συνόλου των δεδομένων. Στην Εικόνα 4.28φαίνεται η διαφορά του συνόλου των εγγραφών που έχει διαβάσει η εφαρμογή σε σχέση με τα αρχεία CDR που ανέγνωσε νωρίτερα από την πρώτη επιλογή, όπως αυτά αποτυπώνονται στην Εικόνα 4.23. Οι 1,512 εγγραφές περιέχουν όλα τα απαραίτητα δεδομένα για την εξεύρεση δόλιων κλήσεων και χρειάζονται 0.4 δευτερόλεπτα για την ανάγνωσή τους έναντι των 35,630,511 εγγραφών που περιέχονται στα αρχεία CDR για το σύνολο δύο μηνών και για να αναγνωσθούν χρειάζονται 59 λεπτά. Στο Παράρτημα Β δίνεται ένα δείγμα από τα αρχεία που περιέχουν τα συγκεντρωτικά δεδομένα των αρχείων CDR όπως κι εγγραφές από τα αρχεία CDR.

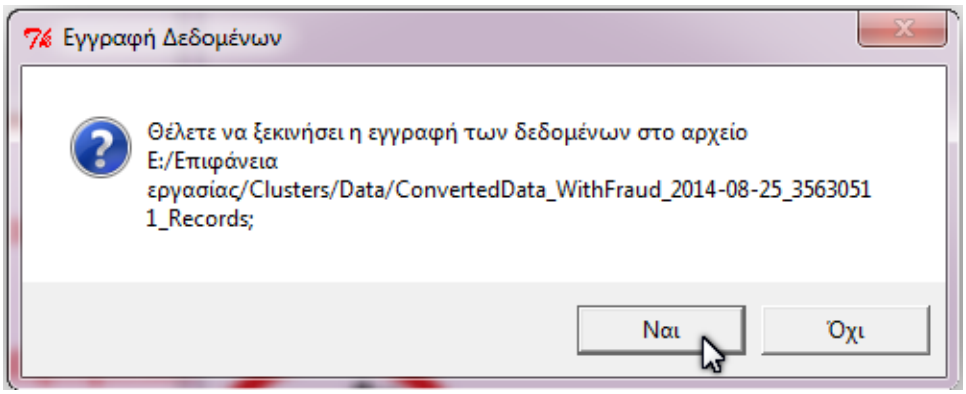


Εικόνα 4.27 Ενημερωτικό παράθυρο για την πρόοδο της ανάγνωσης ενός αρχείου

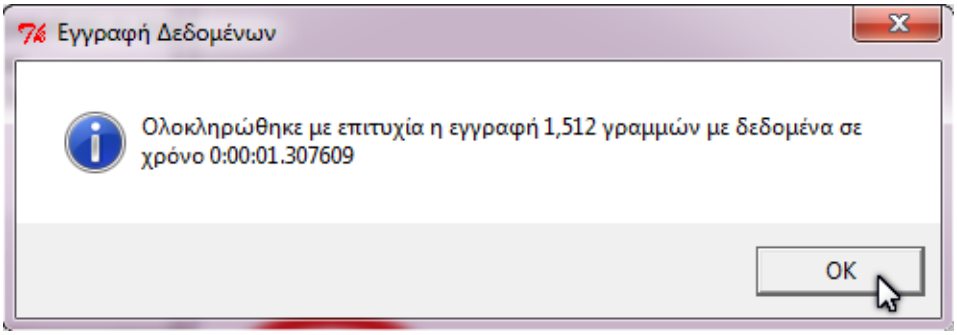


Εικόνα 4.28 Ενημερωτικό μήνυμα για τη διάρκεια και τις εγγραφές που αναγνώσθηκαν

Η τρίτη επιλογή, «Write readen CDR Data», είναι αυτή που αποθηκεύει το σύνολο των αναγνωσμένων τυχαίων τηλεφωνικών κλήσεων σε αρχείο της επιλογής μας. Η ανάγκη της επιλογής αυτής έγκειται στην προαναφερθείσα βελτίωση του χρόνου ανάγνωσης των τυχαίων τηλεφωνικών κλήσεων ώστε να προχωρήσει η εφαρμογή στην ανάλυσή τους κι εν συνεχεία στην ανακοίνωση των αποτελεσμάτων για πιθανή ή μη πιθανή δόλια χρήση των υπηρεσιών του VoIP παρόχου. Στην Εικόνα 4.29 και στην Εικόνα 4.30 εμφανίζονται τα μηνύματα που θα συναντήσει ο χρήστης της παρούσας εφαρμογής.

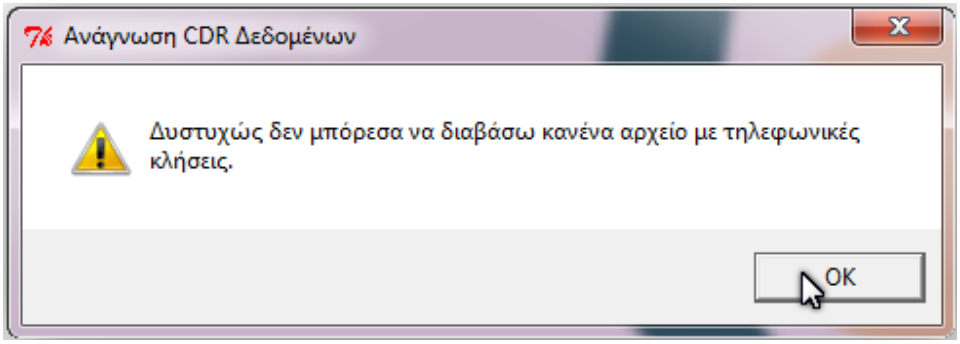


Εικόνα 4.29 Μήνυμα επιβεβαίωσης για την αποθήκευση των δεδομένων της εφαρμογής

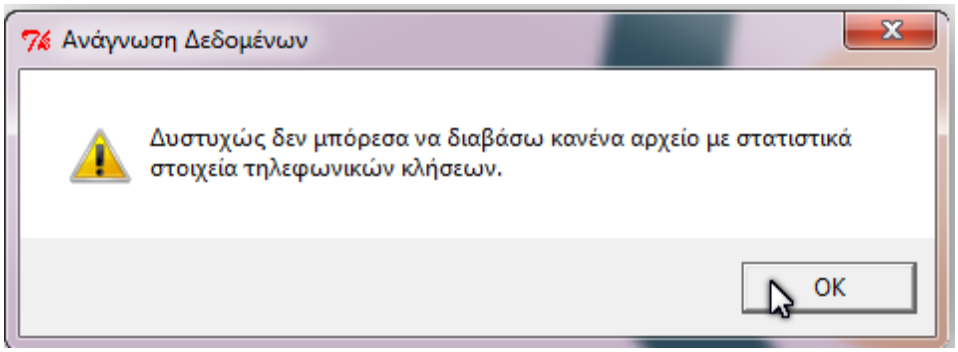


Εικόνα 4.30 Ενημερωτικό μήνυμα για την επιτυχή αποθήκευση των δεδομένων της εφαρμογής.

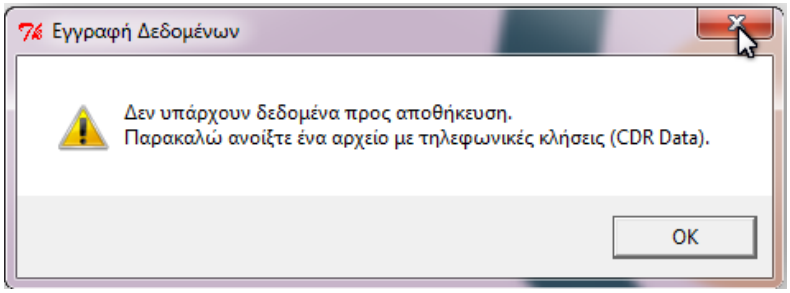
Στο σύνολο της πρώτης οθόνης για οποιοδήποτε πρόβλημα προκύψει εμφανίζονται τα αντίστοιχα μηνύματα στην οθόνη ώστε ο χρήστης της εφαρμογής να είναι έτοιμος να τα αντιμετωπίσει και να συνεχίσει την εργασία του στο κύριο μέρος της εφαρμογής που είναι η εκπαίδευση και ο έλεγχος των αναγνωσμένων αρχείων. Εικόνες των προειδοποιητικών μηνυμάτων εμφανίζονται παρακάτω.



Εικόνα 4.31 Ενημερωτικό μήνυμα για τη μη δυνατή ανάγνωση αρχείων CDR



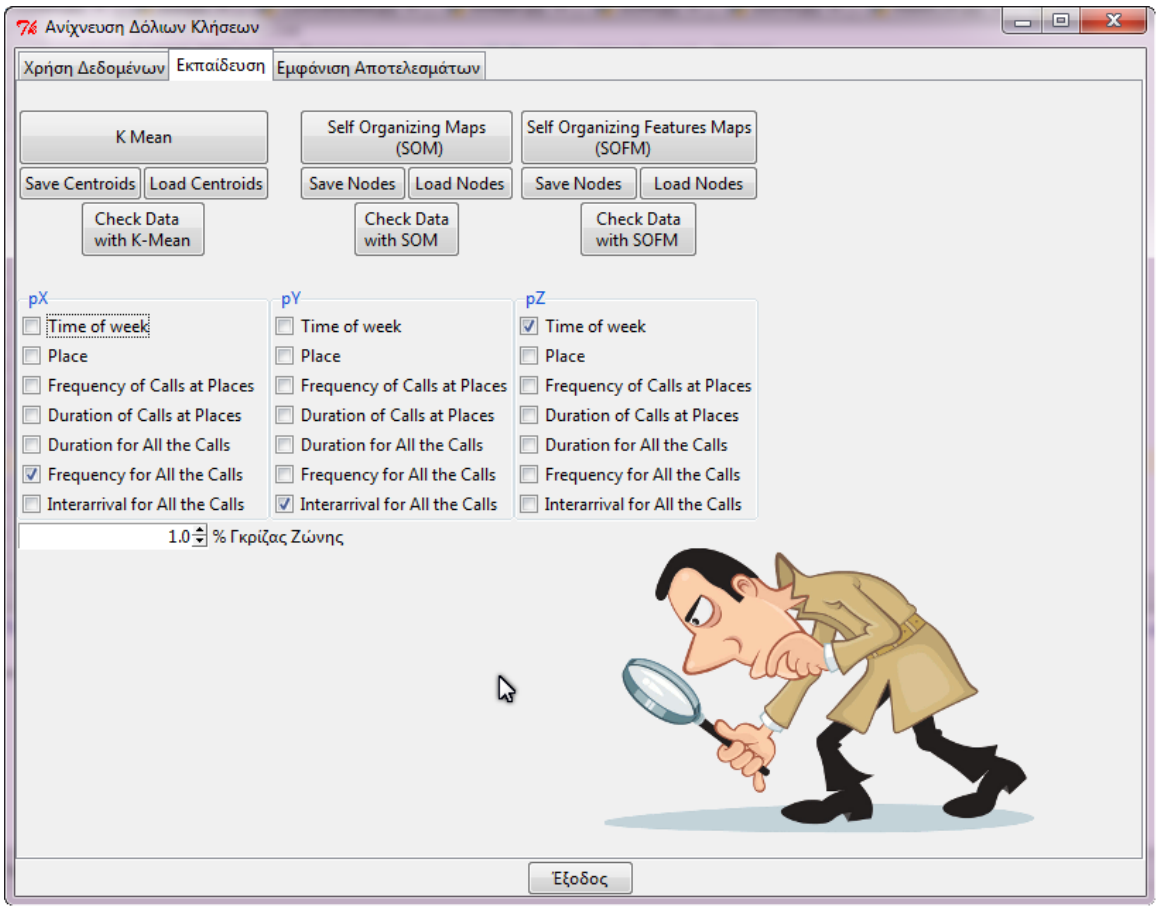
Εικόνα 4.32 Ενημερωτικό μήνυμα για την αποτυχία ανάγνωσης αρχείου με δεδομένα.



Εικόνα 4.33 Ενημερωτικό μήνυμα για την αδυναμία αποθήκευσης δεδομένων της εφαρμογής.

4.4 Εκπαίδευση της Εφαρμογής

Στη δεύτερη καρτέλα της εφαρμογής για την ανίχνευση δόλιων κλήσεων εμφανίζονται, όπως δίνεται στην παρακάτω εικόνα, τρεις επιλογές αλγορίθμων και τρεις στήλες με επτά επιλογές έκαστη για την οπτική απεικόνιση των αποτελεσμάτων.

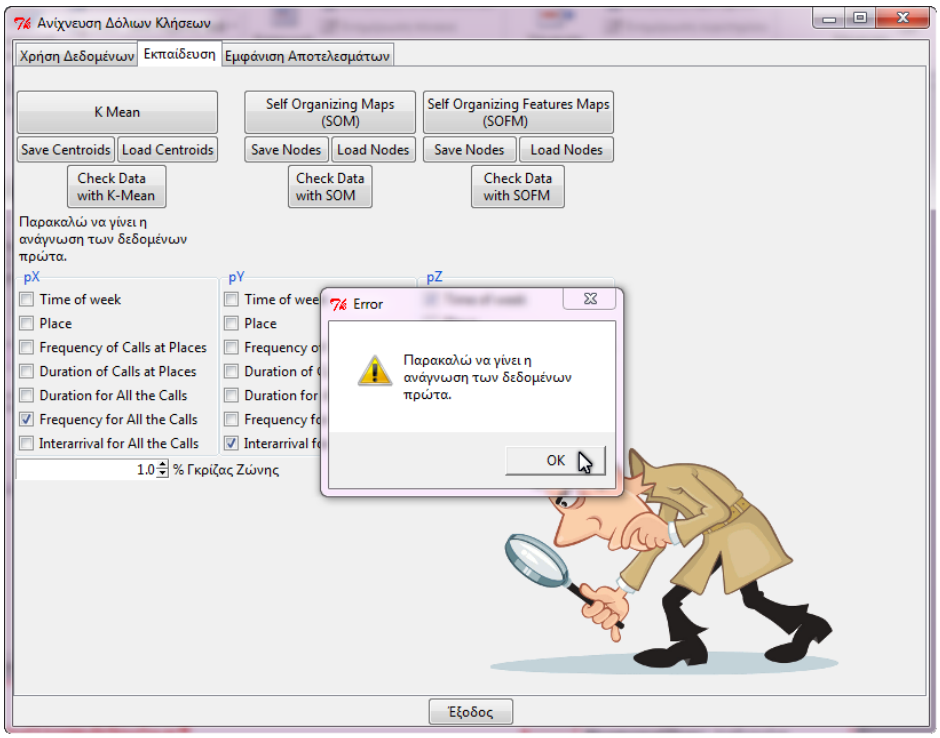


Εικόνα 4.34 Επιλογές για την εκπαίδευση ή τον έλεγχο CDRs κι επιλογές για την εμφάνισή τους.

Στα επόμενα υπό-τμήματα θα περιγραφούν τα βήματα για την εκπαίδευση του συστήματος αλλά και τον έλεγχο για δόλιες τηλεφωνικές κλήσεις.

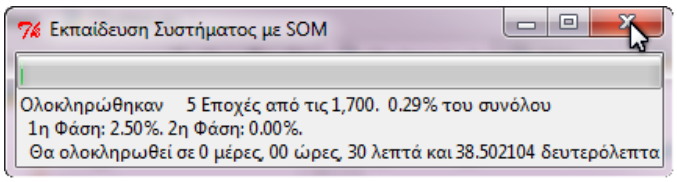
4.4.1 Στάδιο Εκπαίδευσης

Επιλέγοντας έναν από τους τρείς αλγορίθμους (K-Mean, SOM ή SOFM) ξεκινά η διαδικασία της εκπαίδευσης. Εάν ο χρήστης εν αγνοία του δεν έχει ανοίξει κάποιο αρχείο CDR ή αρχείο με δεδομένα των τηλεφωνικών κλήσεων η εφαρμογή τον ενημερώνει σχετικά με αντίστοιχο μήνυμα σε παράθυρο και κάτω από τα κουμπιά του αντίστοιχου αλγορίθμου. Δείγμα αυτής της εσφαλμένης ενέργειας δίνεται από την Εικόνα 4.35.

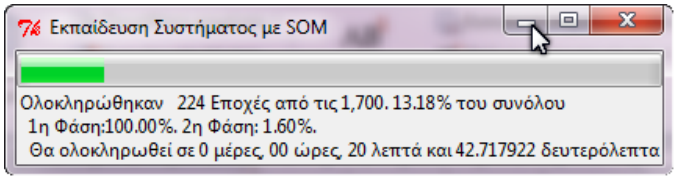


Εικόνα 4.35 Ενημερωτικό μήνυμα για την έλλειψη δεδομένων.

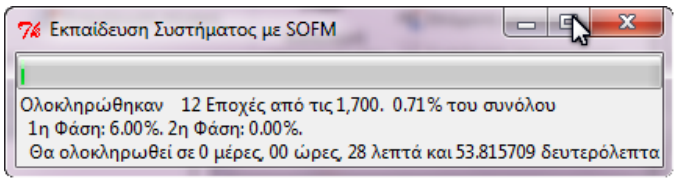
Κατά τη διάρκεια της εκπαίδευσης, η εφαρμογή εμφανίζει παράθυρο για την ενημέρωση του χρήστη σχετικά με την πρόοδο και την πιθανή ολοκλήρωσή της. Η Εικόνα 4.36, η Εικόνα 4.37 και η Εικόνα 4.38 παρουσιάζουν τα ενημερωτικά παράθυρα που θα συναντήσει ο χρήστης της εφαρμογής αυτής.



Εικόνα 4.36 Ενημερωτικό μήνυμα προόδου 1ης φάσης εκπαίδευσης με Αλγόριθμο SOM.

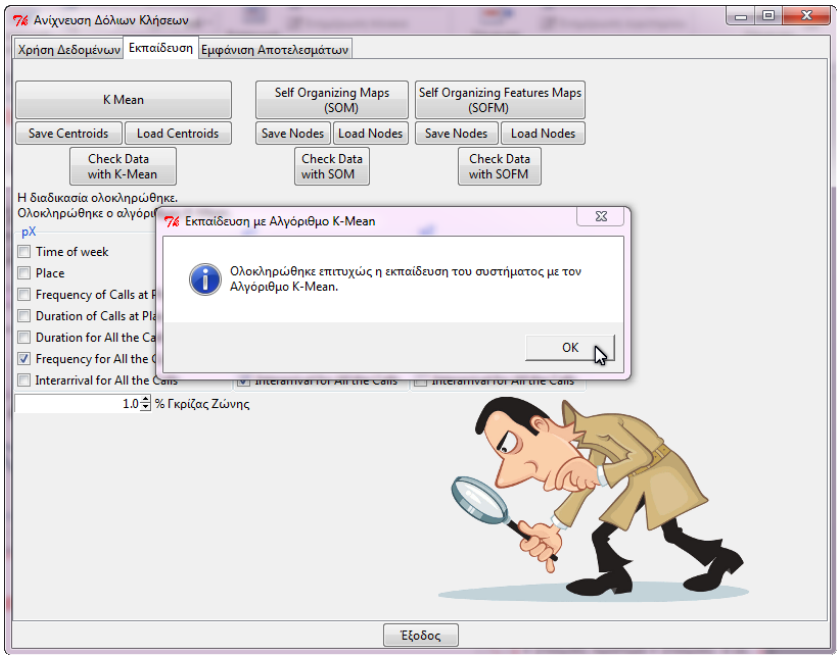


Εικόνα 4.37 Ενημερωτικό μήνυμα προόδου 2ης φάσης εκπαίδευσης με Αλγόριθμο SOM.



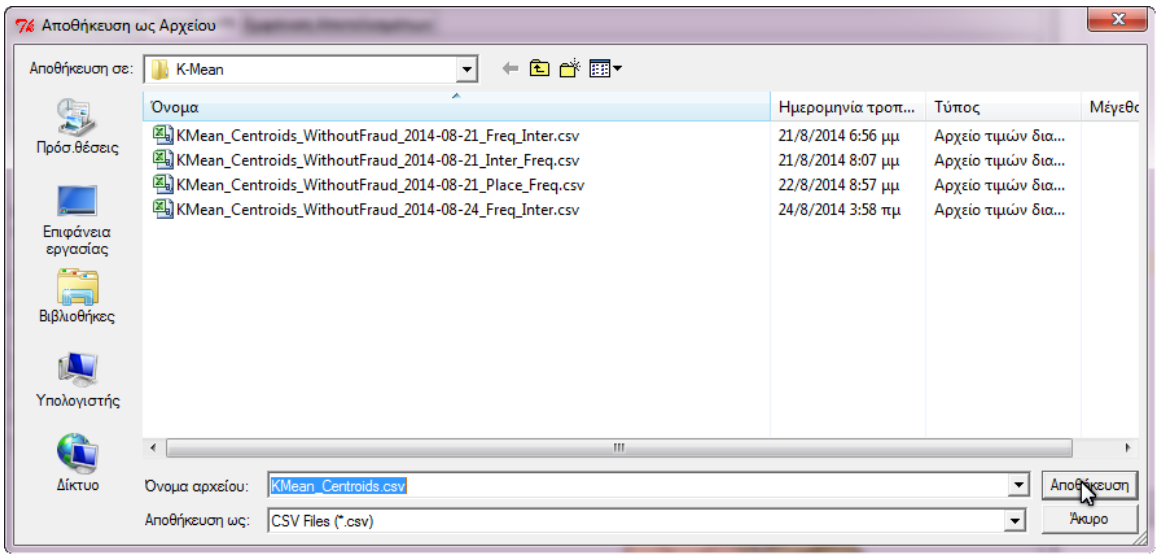
Εικόνα 4.38 Ενημερωτικό μήνυμα προόδου 1ης φάσης εκπαίδευσης με Αλγόριθμο SOFM

Η εφαρμογή ολοκληρώνει τη διαδικασία της εκπαίδευσης με αντίστοιχο μήνυμα. Η Εικόνα 4.39 παρουσιάζει ένα ενημερωτικό παράθυρο σχετικό με την ολοκλήρωση της εκπαίδευσης του συστήματος με τη χρήση του Αλγόριθμου K-Mean.



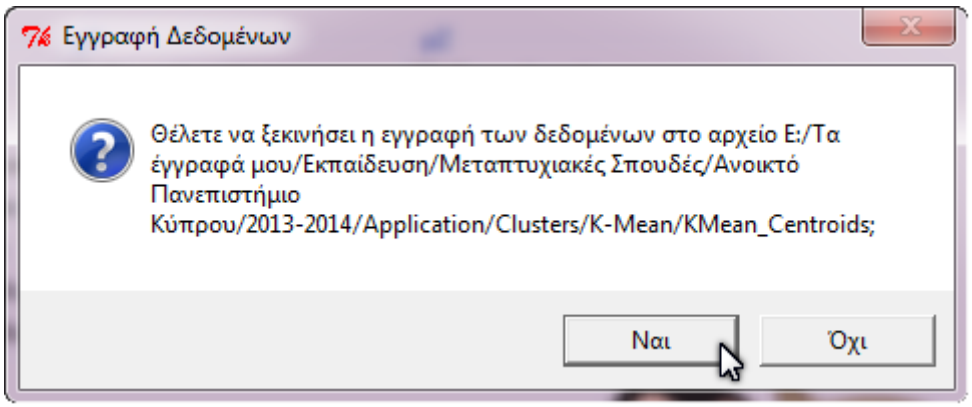
Εικόνα 4.39 Ολοκλήρωση της εκπαίδευσης με τον Αλγόριθμο K-Mean.

Με το πέρας της εκπαίδευσης του συστήματος με όποιον Αλγόριθμο επιθυμεί ο χρήστης της εφαρμογής του δίνεται η δυνατότητα να αποθηκεύσει τα αποτελέσματα του εκπαιδευμένου συστήματος σε αρχείο της επιλογής του. Η Εικόνα 4.40 παρουσιάζει το παράθυρο που εμφανίζεται στο χρήστη όταν ζητήσει την αποθήκευση των δεδομένων ενός εκπαιδευμένου συστήματος με το αντίστοιχο κουμπί «Save (Centroids;ή Nodes ή Nodes)» για τους αντίστοιχους Αλγορίθμους, K-Mean, SOM και SOFM.

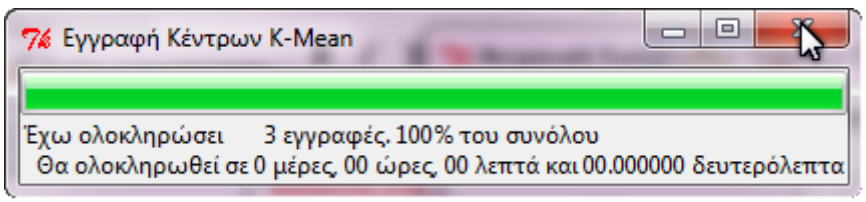


Εικόνα 4.40 Αποθήκευση δεδομένων εκπαιδευμένου συστήματος με τον Αλγόριθμο K-Mean.

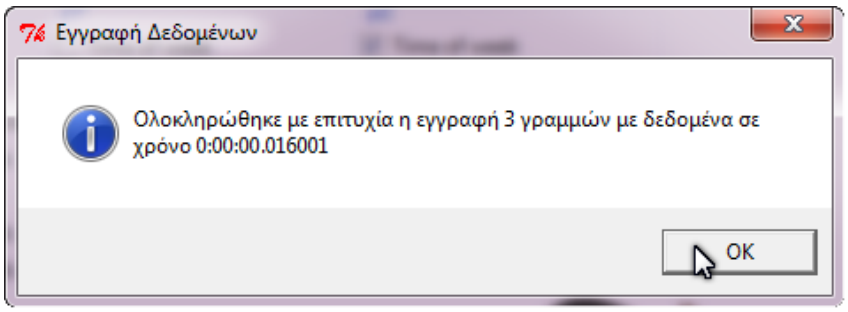
Τα δεδομένα που αποθηκεύονται στον Ηλεκτρονικό Υπολογιστή του χρήστη της εφαρμογής είναι κωδικοποιημένα και αποθηκευμένα σε δυαδική μορφή για να αποφευχθεί η πιθανή αλλοίωσή τους από παράνομους εισβολείς στο σύστημα VoIP.



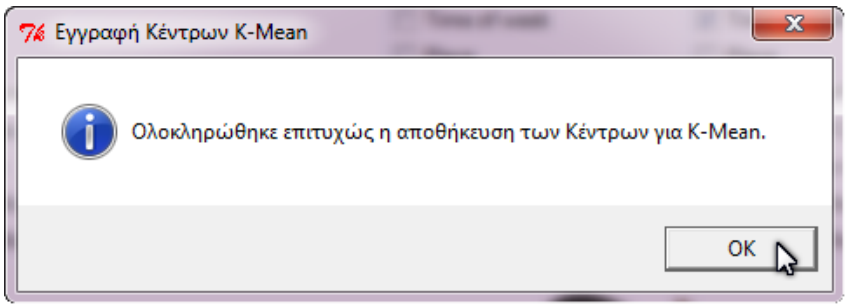
Εικόνα 4.41 Αποθήκευση δεδομένων εκπαιδευμένου συστήματος με τον Αλγόριθμο K-Mean.



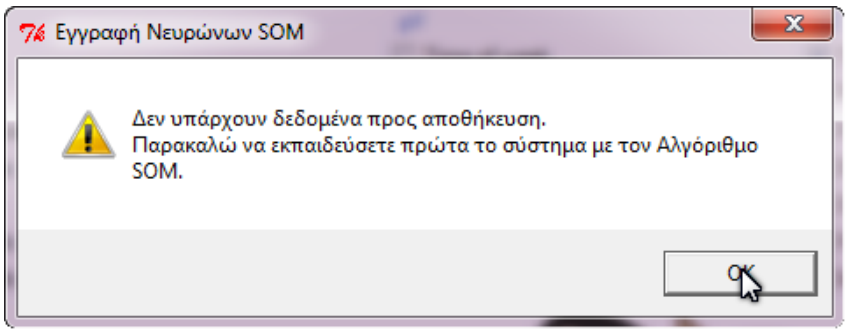
Εικόνα 4.42 Ενημερωτικό παράθυρο προόδου αποθήκευσης των Centroids του K-Mean.



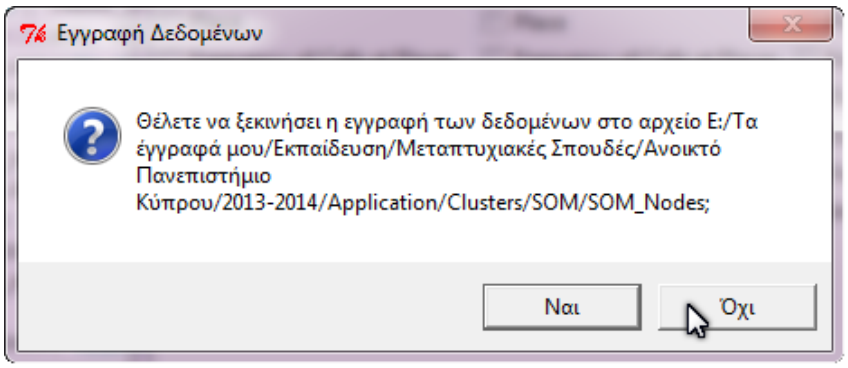
Εικόνα 4.43 Ενημερωτικό μήνυμα επιτυχούς ολοκλήρωσης αποθήκευσης με πληροφορίες για τα Centroids του K-Mean.



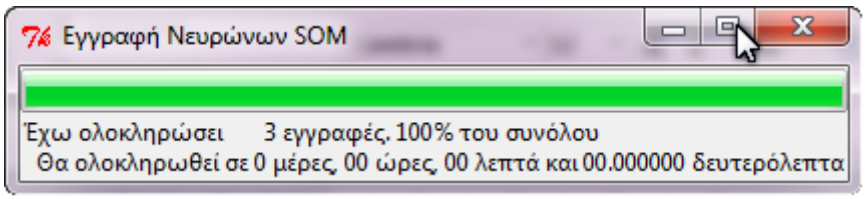
Εικόνα 4.44 Ενημερωτικό μήνυμα για την επιτυχή αποθήκευση των Centroids του K-Mean.



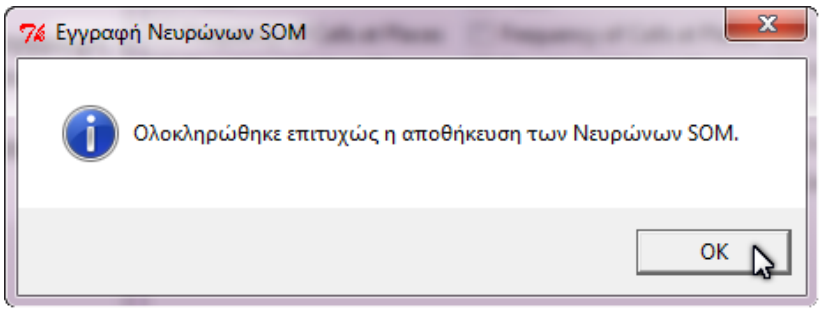
Εικόνα 4.45 Ενημερωτικό μήνυμα για την αποτυχία αποθήκευσης των Νευρώνων του SOM.



Εικόνα 4.46 Αποθήκευση δεδομένων εκπαιδευμένου συστήματος με τον Αλγόριθμο SOM.



Εικόνα 4.47 Ενημερωτικό παράθυρο προόδου αποθήκευσης των Νευρώνων του SOM.



Εικόνα 4.48 Ενημερωτικό μήνυμα για την επιτυχή αποθήκευση των Νευρώνων του SOM.

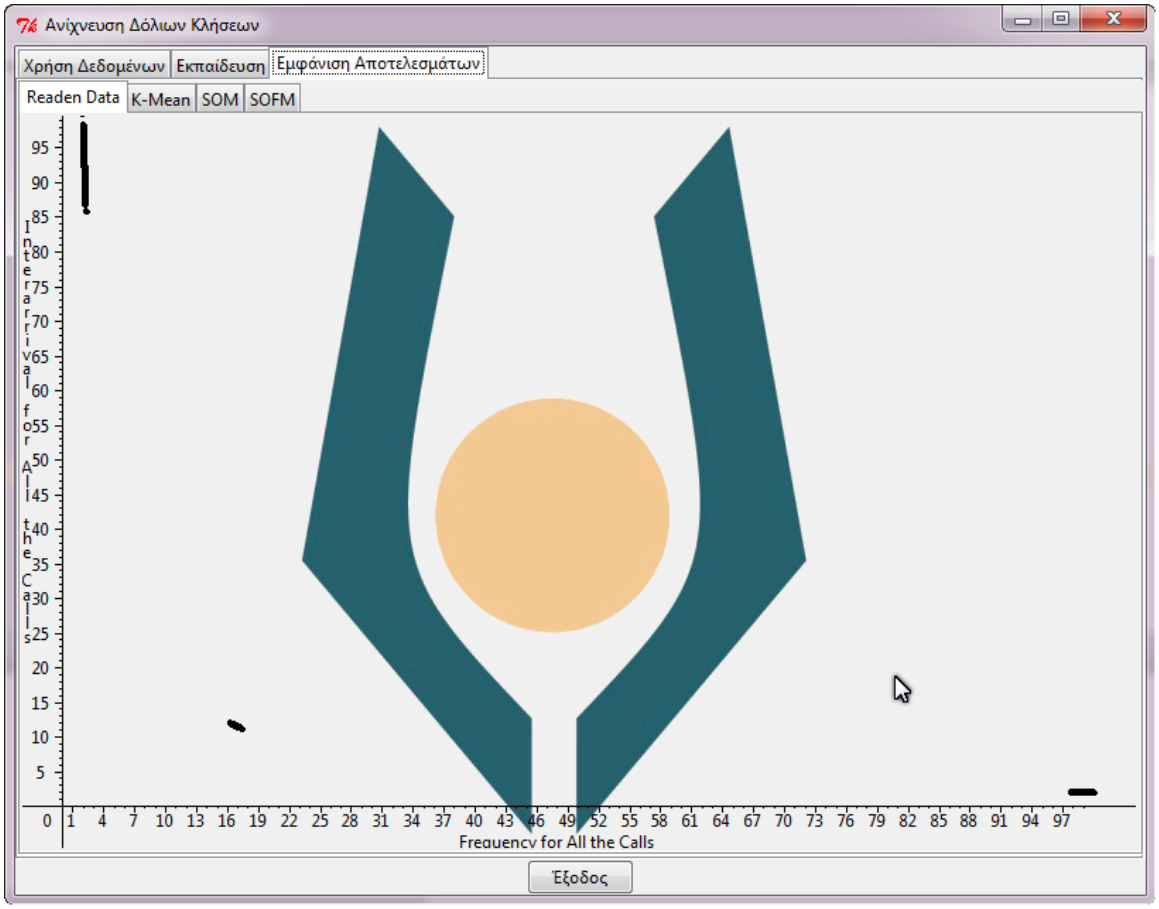
Στις παραπάνω εικόνες εμφανίζονται τα ενημερωτικά μηνύματα για την πρόοδο της εγγραφής των δεδομένων στον ηλεκτρονικό υπολογιστή όπως επίσης και η επιτυχής ή εσφαλμένη εγγραφή αυτών.

4.4.2 Στάδιο Ελέγχου

Με την ολοκλήρωση της εκπαίδευσης του συστήματος από τον εκάστοτε Αλγόριθμο, K-Mean, SOM και SOFM, ο διαχειριστής της εφαρμογής μπορεί να προχωρήσει στον έλεγχο των δεδομένων που έχει εισάγει στην εφαρμογή (σε κανονική λειτουργία, δηλαδή όταν η εφαρμογή είναι συνδεδεμένη με το σύστημα VoIP του παρόχου, ανακτά τα δεδομένα των CDR απευθείας από το σύστημα καταγραφής των τηλεφωνικών κλήσεων). Τα αποτελέσματα με το πέρας της εκπαίδευσης του συστήματος εμφανίζονται στην Τρίτη καρτέλα με όνομα «Εμφάνιση Αποτελεσμάτων» της εφαρμογής όπως φαίνονται στις επόμενες τέσσερις εικόνες.

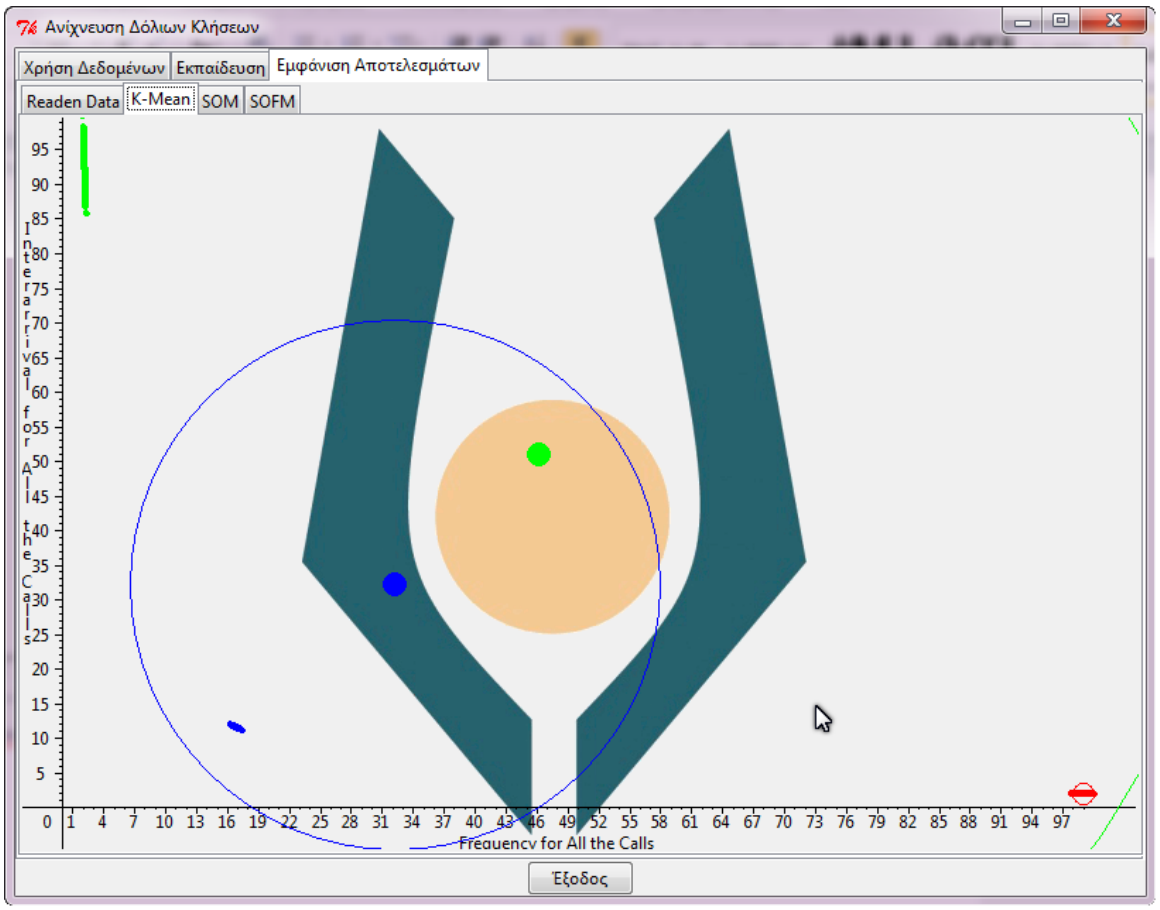
Στις επόμενες τέσσερις εικόνες εμφανίζονται τα δεδομένα των VoIP κλήσεων. Στον οριζόντιο και κάθετο άξονα εμφανίζονται οι ετικέτες από τις ομάδες των τιμών που επέλεξε ο χρήστης της εφαρμογής να εμφανίζει. Κάθε εγγραφή που εισάγεται στην εφαρμογή προς έλεγχο περιέχει επτά τιμές, όπως αυτές φαίνονται στην Εικόνα 4.34 στις στήλες pX, pY και pZ. Η τιμή που εμφανίζεται στο πλαίσιο της «Γκρίζας Ζώνης» στην Εικόνα 4.34, χρησιμοποιείται ως ποσοστό για τον προσδιορισμό των εγγραφών που ο διαχειριστής του συστήματος VoIP υπηρεσιών θα

πρέπει να δώσει ιδιαίτερη προσοχή καθώς αυτές βρίσκονται οριακά στις δόλιες ή φυσιολογικές κλήσεις και θα χρειαστούν περαιτέρω έλεγχο από τον ίδιο.



Εικόνα 4.49 Εμφάνιση VoIP κλήσεων όπως αναγνώσθηκαν από τα αρχεία.

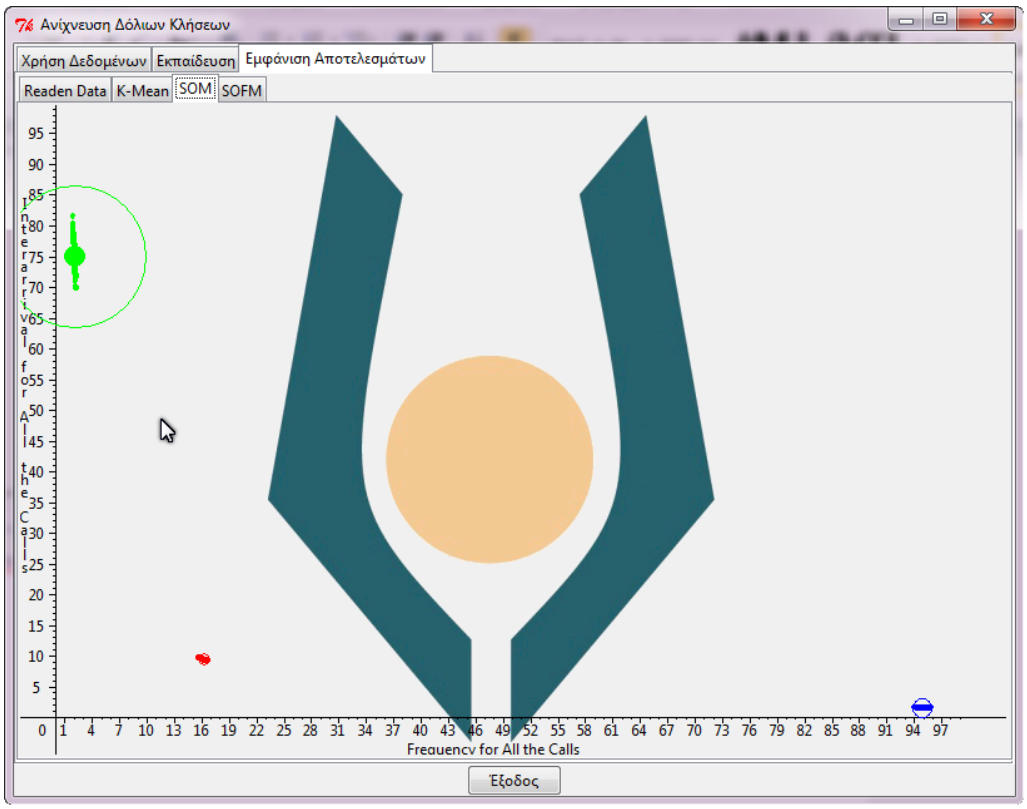
Στην Εικόνα 4.49 εμφανίζονται οι εγγραφές των κλήσεων που χειρίστηκε η εφαρμογή μετά την είσοδό τους από τα αρχεία που επέλεξε ο χρήστης της εφαρμογής αυτής. Η κλίμακα που χρησιμοποιείται στους δύο άξονες είναι κοινή για όλες τις καρτέλες στην «Εμφάνιση Αποτελεσμάτων». Στα δεδομένα, που εμφανίζονται ως διδιάστατα σημεία, έχει γίνει αναλογικά με τις μέγιστες τιμές των αξόνων μετατροπή των αρχικών τους τιμών.



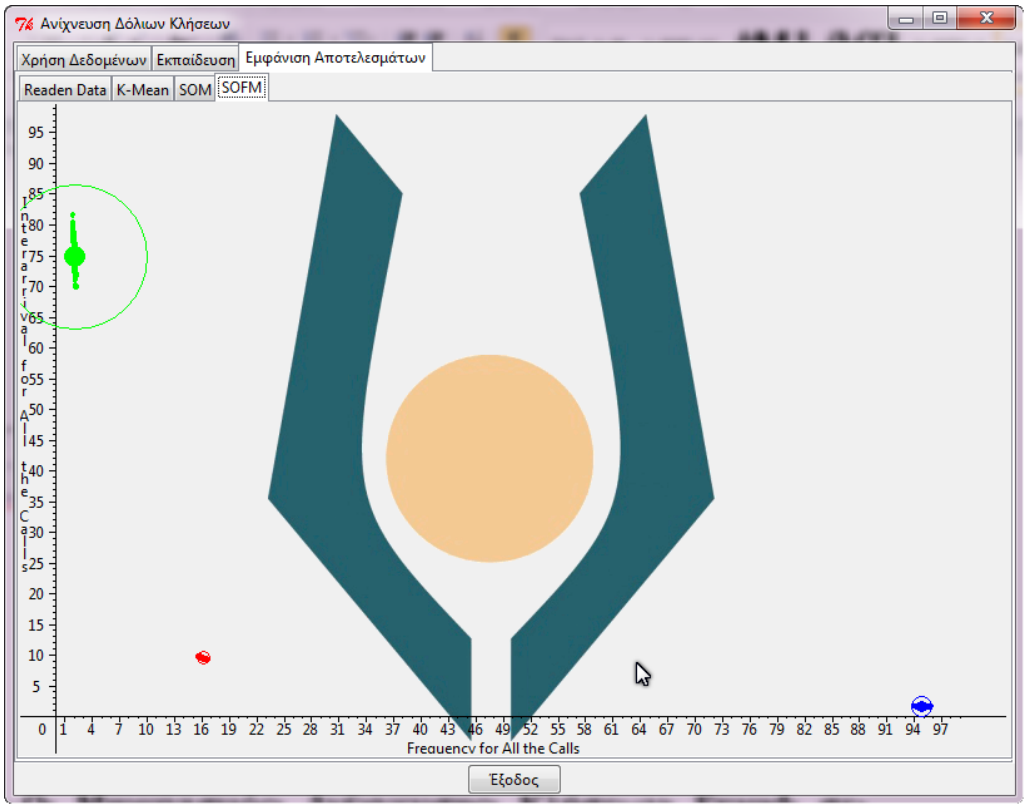
Εικόνα 4.50 Εμφάνιση VoIP κλήσεων μετά από την εκπαίδευση με τον Αλγόριθμο K-Mean.

Στον Αλγόριθμο K-Mean, όπως φαίνεται στην παραπάνω εικόνα, μπορούμε να παρατηρήσουμε πως τα κέντρα (centroids) που έχει δημιουργήσει ο ίδιος δεν βρίσκονται πλησίον των σημείων πέραν του ενός κόκκινου κέντρου. Στην παραπάνω εικόνα, που είναι και η φάση της εκπαίδευσης, μπορούμε να παρατηρήσουμε πώς οι κύκλοι που σχηματίζονται με τα αντίστοιχα χρώματα των κέντρων περικλείουν όλα τα δεδομένα που έχουν εισαχθεί στην εφαρμογή.

Στις επόμενες δύο εικόνες εμφανίζονται τα αποτελέσματα των Αλγορίθμων SOM και SOFM. Μπορούμε να παρατηρήσουμε πως οι νευρώνες βρίσκονται πολύ κοντά στην ομάδα τους και οι κύκλοι που σχηματίζονται από τις ακραίες τιμές περικλείουν όλα τα εισαγόμενα δεδομένα όπως και στον Αλγόριθμο K-Mean.

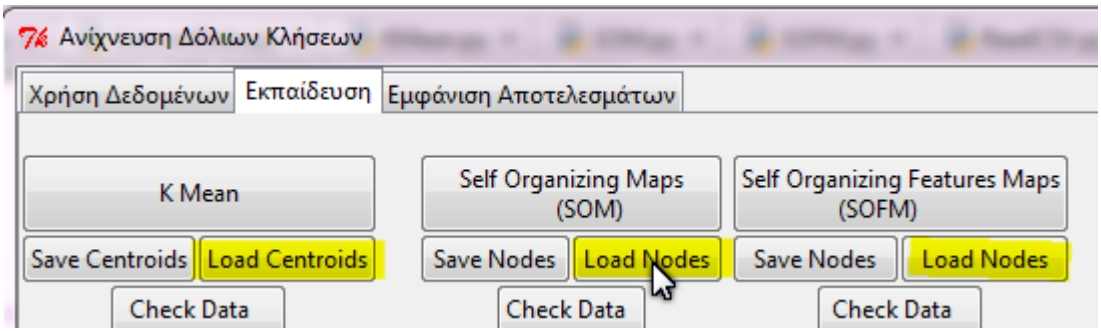


Εικόνα 4.51 Εμφάνιση VoIP κλήσεων μετά από την εκπαίδευση με τον Αλγόριθμο SOM.

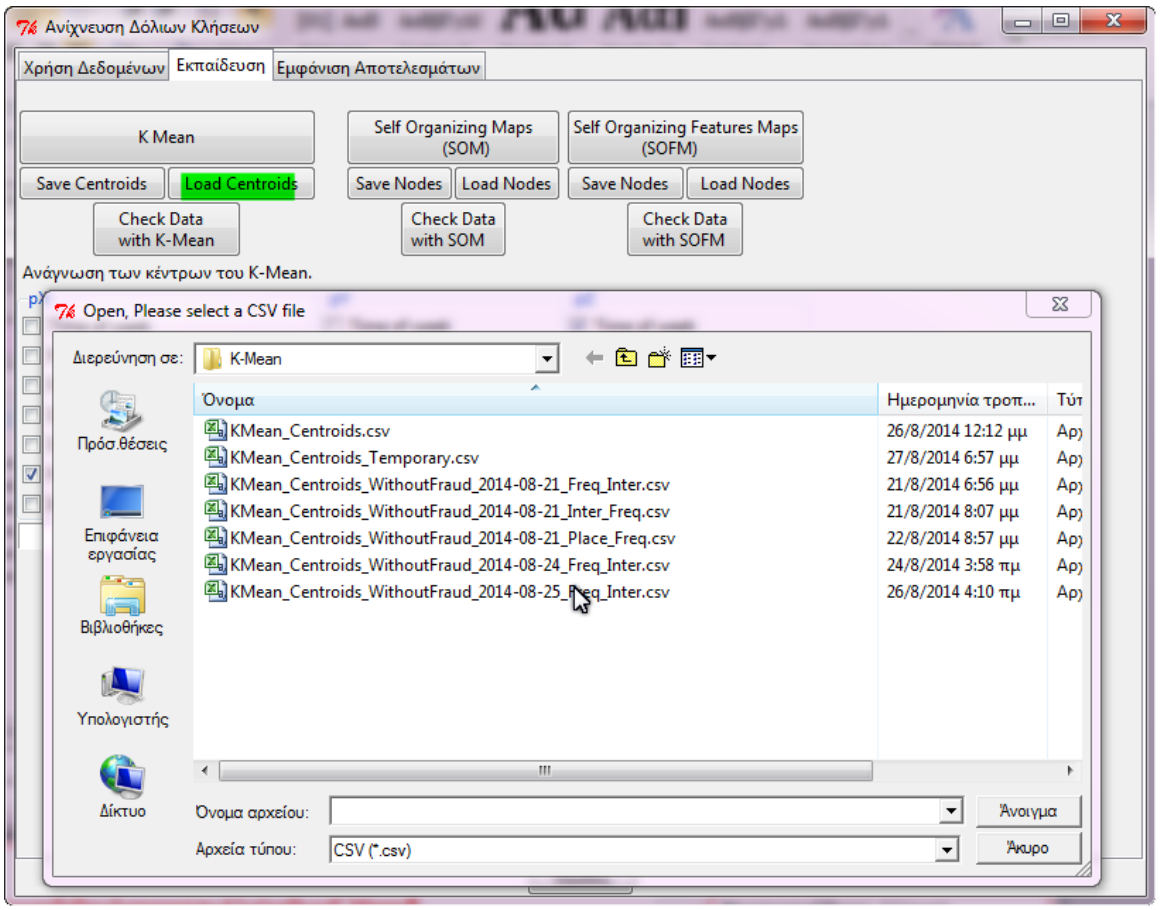


Εικόνα 4.52 Εμφάνιση VoIP κλήσεων μετά από την εκπαίδευση με τον Αλγόριθμο SOFM.

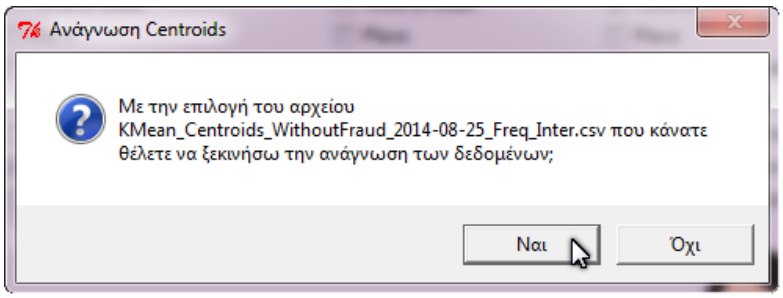
Μετά τη φάση της εκπαίδευσης ο χρήστης της εφαρμογής μπορεί να ανασύρει τα δεδομένα των τριών αλγορίθμων πατώντας στο αντίστοιχο κουμπί «Load» όπως φαίνεται στην παρακάτω εικόνα.



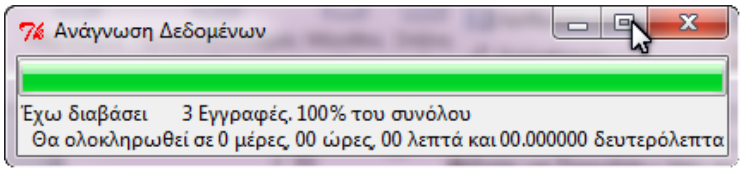
Εικόνα 4.53 Κουμπιά επιλογής αρχείου με δεδομένα των Αλγορίθμων.



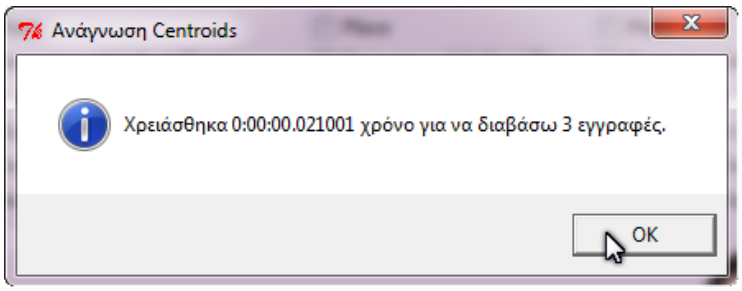
Εικόνα 4.54 Παράθυρο επιλογής αρχείου με δεδομένα για τον Αλγόριθμο K-Mean.



Εικόνα 4.55 Επιβεβαίωση επιλογής αρχείου με δεδομένα για τον Αλγόριθμο K-Mean.



Εικόνα 4.56 Πρόοδος ανάγνωσης αρχείου με δεδομένα για τον Αλγόριθμο K-Mean.

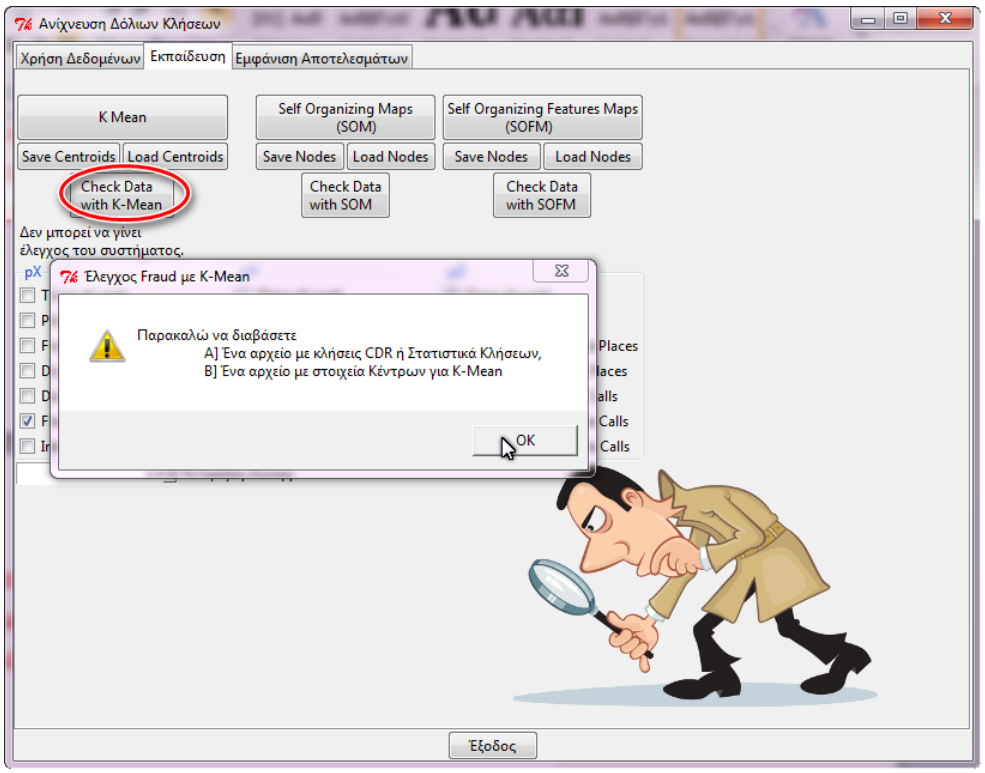


Εικόνα 4.57 Ενημερωτικό μήνυμα επιτυχούς ανάγνωσης αρχείου με δεδομένα για τον Αλγόριθμο K-Mean.

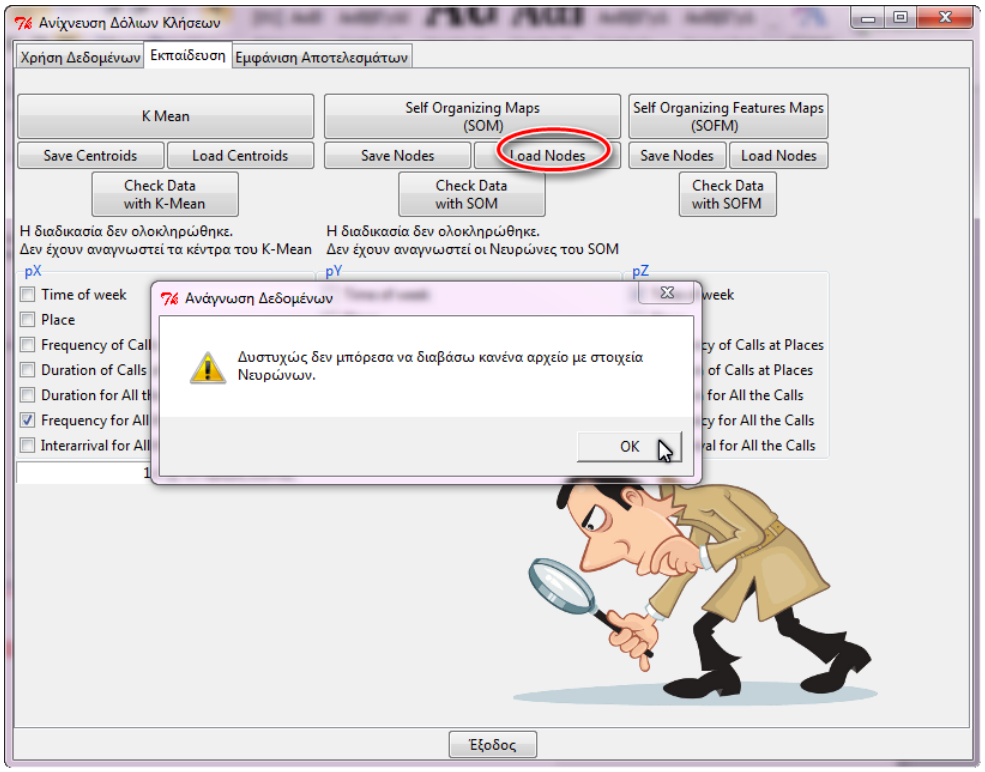
Η Εικόνα 4.54, η Εικόνα 4.55, η Εικόνα 4.56 και η Εικόνα 4.57 εμφανίζουν τα διαδοχικά παράθυρα που θα αντικρίσει ο χρήστης της εφαρμογής για οποιοδήποτε Αλγόριθμο επιλέξει να ανασύρει τα δεδομένα του.

Όταν η ανάγνωση των δεδομένων του εκάστοτε Αλγορίθμου που επέλεξε ο χρήστης της εφαρμογής ολοκληρωθεί επιτυχώς κι έχει η εφαρμογή δεδομένα προς έλεγχο τότε και μόνο μπορεί να προχωρήσει ο χρήστης στην επιλογή του κουμπιού «Check Data» για τον έλεγχο των δεδομένων όπως φαίνεται στην επόμενη εικόνα.

Για οποιοδήποτε πρόβλημα προκύψει στην ανάγνωση των δεδομένων των αλγορίθμων ή την εμφάνιση των αποτελεσμάτων η εφαρμογή εμφανίζει μηνύματα σε παράθυρα. Οι επόμενες εικόνες παρουσιάζουν τα ενημερωτικά μηνύματα.



Εικόνα 4.58 Ενημερωτικό μήνυμα για την έλλειψη δεδομένων στον Αλγόριθμο K-Mean.

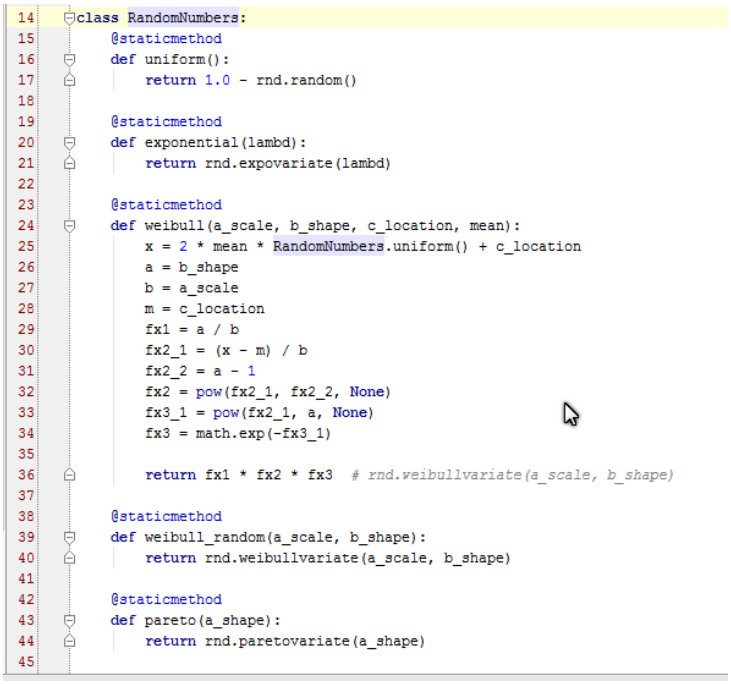


Εικόνα 4.59 Ενημερωτικό μήνυμα για την έλλειψη δεδομένων στον Αλγόριθμο K-Mean.

Στο κεφάλαιο που ακολουθεί θα παρουσιασθούν τα αποτελέσματα της εφαρμογής «Ανίχνευση Δόλιων Κλήσεων» χρησιμοποιώντας διαφορετικά σενάρια VoIP κλήσεων.

Κεφάλαιο 5
Αποτελέσματα της Εφαρμογής «Ανίχνευση Δόλιων Κλήσεων»

Η εκπαίδευση της εφαρμογής έγινε με δεδομένα από τα αρχεία που παρήγαγε η εφαρμογή «Παραγωγή Αρχείου Τυχαίων Κλήσεων». Η γεννήτρια τυχαίων τηλεφωνικών κλήσεων λαμβάνει υπόψη της του κανόνες που τις θέτονται και στη συνέχεια χρησιμοποιεί τη βιβλιοθήκη Random της περιγραφικής γλώσσας Python για να παράγει σε τυχαίες χρονικές στιγμές τηλεφωνικές κλήσεις. Στην παρακάτω εικόνα δίνεται η τάξη (class) RandomNumbers του αρχείου RandomInstances.py. Η τάξη αυτή καλείται για να παράγει τυχαίους αριθμούς σύμφωνα με τις κατανομές Weibull και Exponential. Οι κατανομές Weibull και Exponential χρησιμοποιούνται αντίστοιχα για την αναμονή μεταξύ των κλήσεων και τη διάρκεια κάθε κλήσης. Η επιλογή των δύο αυτών κατανομών έγινε για την πιστότερη προσομοίωση των κανονικών VoIP κλήσεων και την αποφυγή μίας ομοιόμορφης κατανομής που δίνει η γεννήτρια τυχαίων αριθμών Random και Uniform.



```python
class RandomNumbers:
    @staticmethod
    def uniform():
        return 1.0 - rnd.random()

    @staticmethod
    def exponential(lambd):
        return rnd.expovariate(lambd)

    @staticmethod
    def weibull(a_scale, b_shape, c_location, mean):
        x = 2 * mean * RandomNumbers.uniform() + c_location
        a = b_shape
        b = a_scale
        m = c_location
        fx1 = a / b
        fx2_1 = (x - m) / b
        fx2_2 = a - 1
        fx2 = pow(fx2_1, fx2_2, None)
        fx3_1 = pow(fx2_1, a, None)
        fx3 = math.exp(-fx3_1)

        return fx1 * fx2 * fx3  # rnd.weibullvariate(a_scale, b_shape)

    @staticmethod
    def weibull_random(a_scale, b_shape):
        return rnd.weibullvariate(a_scale, b_shape)

    @staticmethod
    def pareto(a_shape):
        return rnd.paretovariate(a_shape)

```

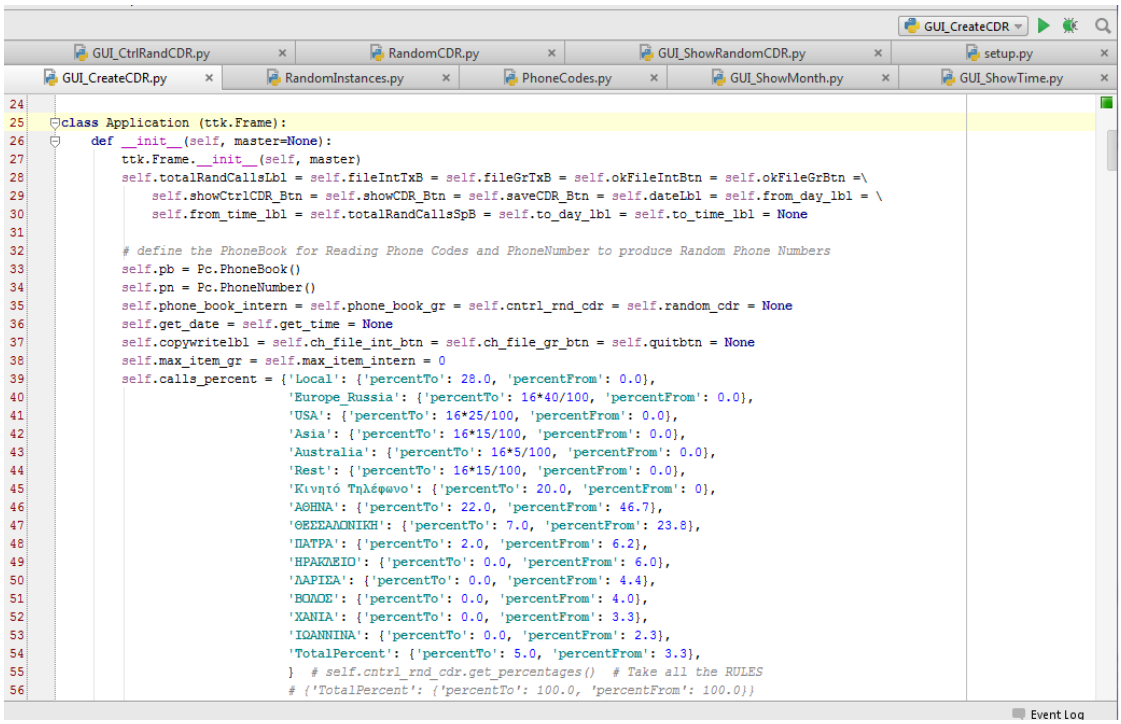
Εικόνα 5.1 Τμήμα κώδικα για την παραγωγή τυχαίων αριθμών

Τα σενάρια για τη δημιουργία των προφίλ των τυχαίων VoIP κλήσεων ως προς τη διάρκεια κάθε κλήσης και το χρόνο αναμονής μεταξύ των κλήσεων δίνεται παρακάτω ως τμήμα από το σύνολο του κώδικα. Η εφαρμογή καλεί τη διαδικασία «calculate» με παράμετρο το παράθυρο που δείχνει την πορεία της δημιουργίας των τυχαίων χρονικών στιγμών. Η επόμενη εικόνα εμφανίζει ένα τμήμα της διαδικασίας «calculate»

```python
    def calculate(self, pr_bar_window=None):
        is_valid = True
        start_of_time = dt.now()
        valid = 'Valid Call'
        fraud = 'Fraud Call'
        tmp_hour = -1  # self.start_date.hour
        #ia=call_inter_arrival, dur=call_duration, a=a, b=b, c=c
        # a = duration / math.gamma(1 + (1 / b)

        var = {
            # NORMAL Check if it is time for work (8.00a.m. to 4.00 p.m
            0: {'ia': 0.05648, 'dur': 149.92, 'a': 1.5533, 'b': 1.6615, 'c': 17.213, 'label': 'Job Time',
                'weekday': {0, 1, 2, 3, 4}, 'hour': {8, 9, 10, 11, 12, 13, 14, 15},  # 9, 10, taken for fraud 60%
                'type': valid},
            # NORMAL Use the above for the Rest time for each day except the previous rules
            1: {'ia': 0.32848, 'dur': 100.83, 'a': 0.83048, 'b': 1.2517, 'c': 18, 'label': 'Rest Day after 4 p.m.',
                'weekday': {0, 1, 2, 3, 4}, 'hour': {16, 17, 18, 19, 20, 21, 22, 23, 0, 1, 2, 3, 4, 5, 6, 7},
                'type': valid},  # 20, 21, 22, taken for fraud 30%
            # NORMAL Check if it is Weekend (Saturday or Sunday)
            2: {'ia': 2.59, 'dur': 104.12, 'a': 0.39613, 'b': 0.57025, 'c': 18, 'label': 'Weekend',
                'weekday': {5, 6}, 'hour': {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,
                                            15, 16, 17, 18, 19, 20, 21, 22, 23},  # 18, 19, 20, taken for fraud 30%
                'type': valid},

# ***************************************************************************************************************************
            # Start Fraud Scenario with more calls
# ***************************************************************************************************************************
#             # FRAUD 10% Start Check if it is Weekend (Saturday or Sunday) and time is 11.00 a.m. to 1.00 p.m.)
#             3: {'ia': 2.33, 'dur': 114.53, 'a': 0.39613, 'b': 0.57025, 'c': 18, 'label': 'Weekend Fraud 10%',
#                 'weekday': {5, 6}, 'hour': {11, 12},
#                 'type': fraud},
#             # FRAUD 20% Start Check if it is time for work (9.00a.m. to 11.00a.m.)
```

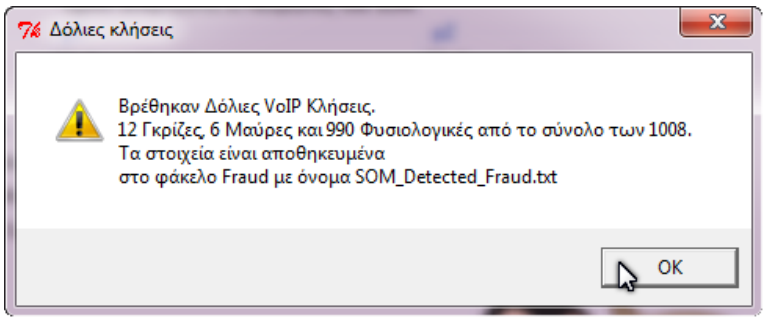
Εικόνα 5.2 Τμήμα κώδικα για την παραγωγή τυχαίων αριθμών

Στο δομητή (Constructor) της κύριας τάξης «Application», που δημιουργεί την εφαρμογή «Παραγωγή Αρχείου Τυχαίων Κλήσεων», δίνονται ως όρισμα στη μεταβλητή «self.calls_percent», τύπου «Λεξικό (Dictionary)», τα ποσοστά που έχουν εξαχθεί από το ερευνητικό εργαστήριο. Τα ποσοστά έχουν εξαχθεί ύστερα από ανάλυση του αρχείου με τις VoIP κλήσεις που δόθηκε από συνεργαζόμενο πάροχο VoIP υπηρεσιών. Τα ποσοστά αυτά μπορεί ο χρήστης της εφαρμογής να τα αλλάξει όπως φαίνεται και στην Εικόνα 4.4. Η Εικόνα 5.3 δίνει το τμήμα του δομητή της κύριας τάξης της εφαρμογής «Παραγωγή Αρχείου Τυχαίων Κλήσεων».



Εικόνα 5.3 Τμήμα κώδικα για την καταχώρηση των ποσοστών κλήσεων προς και από περιοχές της Ελλάδος και του κόσμου.

Οι παραγόμενες τηλεφωνικές κλήσεις αποθηκεύονται σε αρχείο που έχει ορίσει ο χρήστης της εφαρμογής όπως αναφέρεται και στα βήματα στο υπό-τμήμα 4.2.2 του προηγούμενου κεφαλαίου.



Εικόνα 5.4 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων

5.1 Δεδομένα Παρόχου VoIP Υπηρεσιών

Η μορφή των δεδομένων του Παρόχου VoIP Υπηρεσιών που δόθηκαν στο ερευνητικό εργαστήριο, ακολουθεί τη δομή που ορίζει η εταιρεία CISCO [39] σε παλαιότερη έκδοση, με λίγες διαφοροποιήσεις, σύμφωνα με τις ανάγκες τις επιχείρησής του.

Από το παραπάνω αρχείο κρατήθηκαν μόνο οι στήλες που χρειάζονται στην παρούσα εφαρμογή όπως αυτές φαίνονται στο Παράρτημα Β.

5.2 Λειτουργία της Εφαρμογής με Πραγματικά Δεδομένα

Η εφαρμογή «Ανίχνευση Δόλιων Κλήσεων» έχει εκπαιδευτεί με αρχεία που περιέχουν εγγραφές τηλεφωνικών κλήσεων από μία (01) εβδομάδα έως και δύο μήνες συνεχόμενους (συνολικά εννέα εβδομάδες). Κάθε αρχείο μίας εβδομάδας περιέχει εκατόν εξήντα οκτώ εγγραφές, όσες είναι και οι ώρες της εβδομάδας. Η ομαδοποίηση των VoIP κλήσεων σε παράθυρο μίας ώρας οφείλεται στην επιλογή μας να μην έχουμε μεγάλο πλήθος εγγραφών για ανάγνωση αλλά ούτε να επιτρέπουμε μεγάλες χρονικές στιγμές χωρίς έλεγχο για τον πάροχο VoIP υπηρεσιών.

Για τον έλεγχο της αποτελεσματικότητας της εφαρμογής χρησιμοποιήθηκαν τρία σενάρια:

- Αύξηση της διάρκειας των κλήσεων και μείωση του διαστήματος αναμονής μεταξύ των κλήσεων κατά 30%, στη διάρκεια ενός Σαββατοκύριακου και στη διάρκεια της καθημερινής ημέρας, πέρα του ωραρίου εργασίας.

- Μείωση της διάρκειας των κλήσεων και του διαστήματος αναμονής μεταξύ των κλήσεων κατά 60%, σε ώρες εργασίας (9.00 π.μ. με 11.00 π.μ).

- Μείωση της διάρκειας των κλήσεων και του διαστήματος αναμονής μεταξύ των κλήσεων κατά 80%, σε ώρες εργασίας (9.00 π.μ. με 11.00 π.μ).

Κάθε εγγραφή συγκεντρώνει τα παρακάτω πεδία, όπως αυτά δίνονται και στο Παράρτημα Β.

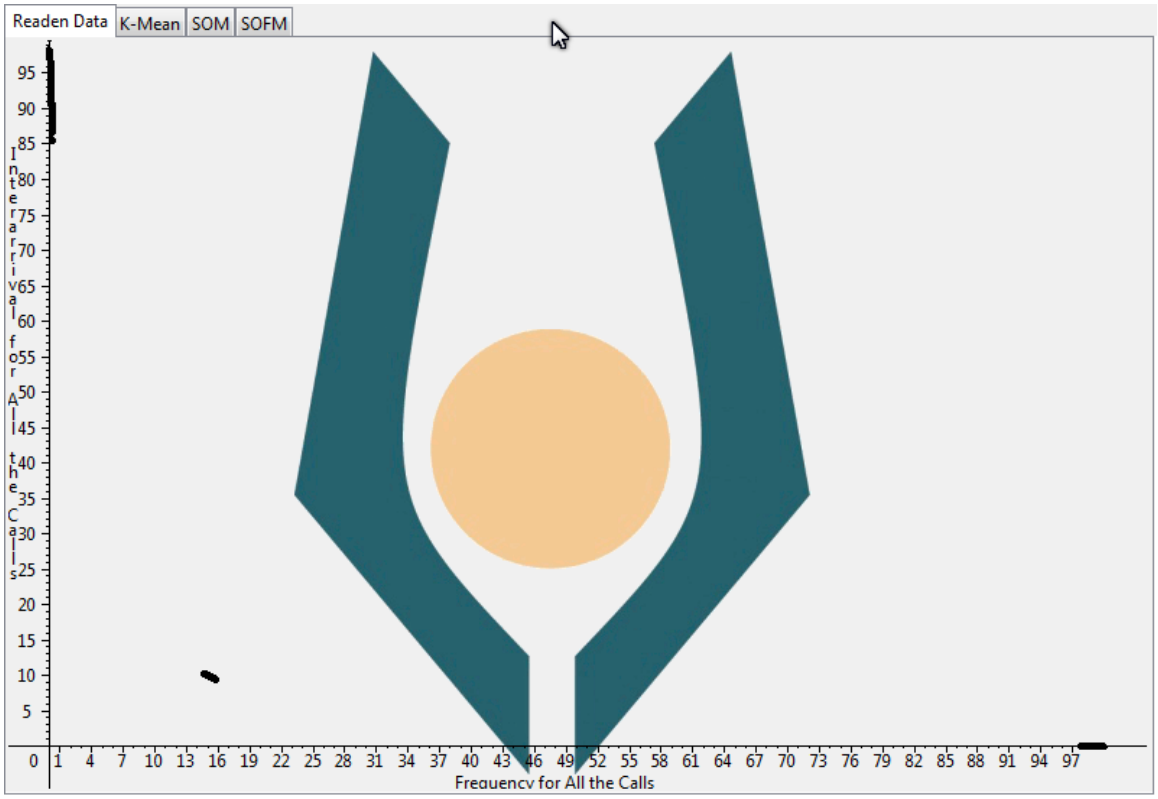
- Όταν είναι αρχείο με εγγραφές τηλεφωνικών κλήσεων: 'StartTime', 'EndTime', 'Duration', 'CallerTelephone', 'CallerCountryCode', 'CallerCode', 'CallerMinNationalNumber', 'CallerMaxNationalNumber', 'CallerZone', 'CallerCountry', 'CallerCounty', 'CallerCity',

'CallingTelephone', 'CallingCountryCode', 'CallingCode', 'CallingMinNationalNumber', 'CallingMaxNationalNumber', 'CallingZone', 'CallingCountry', 'CallingCounty', 'CallingCity', 'Info'

- Για αρχείο με δεδομένα τηλεφωνικών κλήσεων: 'Week', 'Day', 'Hour', 'FromAthensFrequency', 'FromAthensDuration', 'FromAthensCallPercent', 'FromThessalonikiFrequency', 'FromThessalonikiDuration', 'FromThessalonikiCallPercent', 'FromPatraFrequency', 'FromPatraDuration', 'FromPatraCallPercent', 'FromIraklioFrequency', 'FromIraklioDuration', 'FromIraklioCallPercent', 'FromLarisaFrequency', 'FromLarisaDuration', 'FromLarisaCallPercent', 'FromVolosFrequency', 'FromVolosDuration', 'FromVolosCallPercent', 'FromXaniaFrequency', 'FromXaniaDuration', 'FromXaniaCallPercent', 'FromIoanninaFrequency', 'FromIoanninaDuration', 'FromIoanninaCallPercent', 'FromTotalPercentFrequency', 'FromTotalPercentDuration', 'FromTotalPercentCallPercent', 'ToEurope_RussiaFrequency', 'ToEurope_RussiaDuration', 'ToEurope_RussiaCallPercent', 'ToUSAFrequency', 'ToUSADuration', 'ToUSACallPercent', 'ToAsiaFrequency', 'ToAsiaDuration', 'ToAsiaCallPercent', 'ToAustraliaFrequency', 'ToAustraliaDuration', 'ToAustraliaCallPercent', 'ToRestFrequency', 'ToRestDuration', 'ToRestCallPercent', 'ToLocalFrequency', 'ToLocalDuration', 'ToLocalCallPercent', 'ToMobileFrequency', 'ToMobileDuration', 'ToMobileCallPercent', 'ToAthensFrequency', 'ToAthensDuration', 'ToAthensCallPercent', 'ToThessalonikiFrequency', 'ToThessalonikiDuration', 'ToThessalonikiCallPercent', 'ToPatraFrequency', 'ToPatraDuration', 'ToPatraCallPercent', 'ToTotalPercentFrequency', 'ToTotalPercentDuration', 'ToTotalPercentCallPercent', 'InterArrival', 'Duration', 'TotalCalls', 'TotalFraud', 'TotalValid'

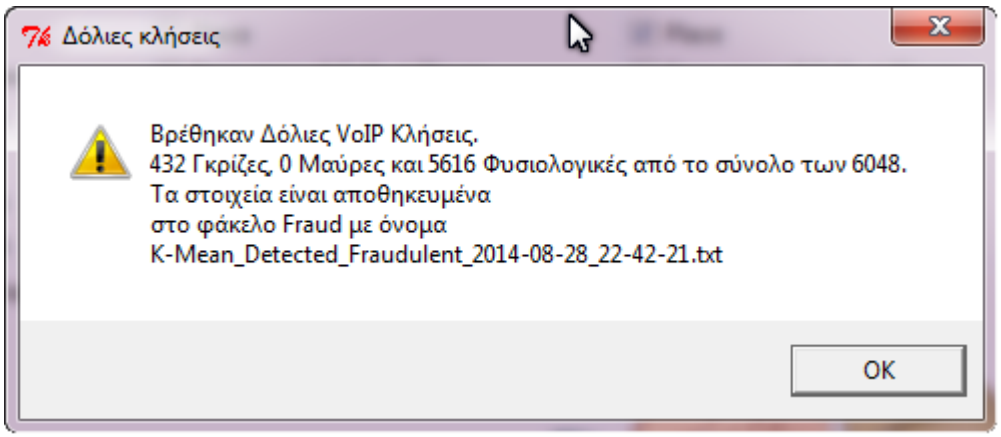
Με την ανάγνωση των δεδομένων η εφαρμογή εκπαιδεύεται με τον ορισμό αρχικά 8 συστάδων.

Στην παρακάτω εικόνα εμφανίζονται τα δεδομένα χρησιμοποιώντας τη συχνότητα των κλήσεων (Call Frequency) στον άξονα των X και την χρονική καθυστέρηση (Interarrival) μεταξύ των κλήσεων στον άξονα των Y.



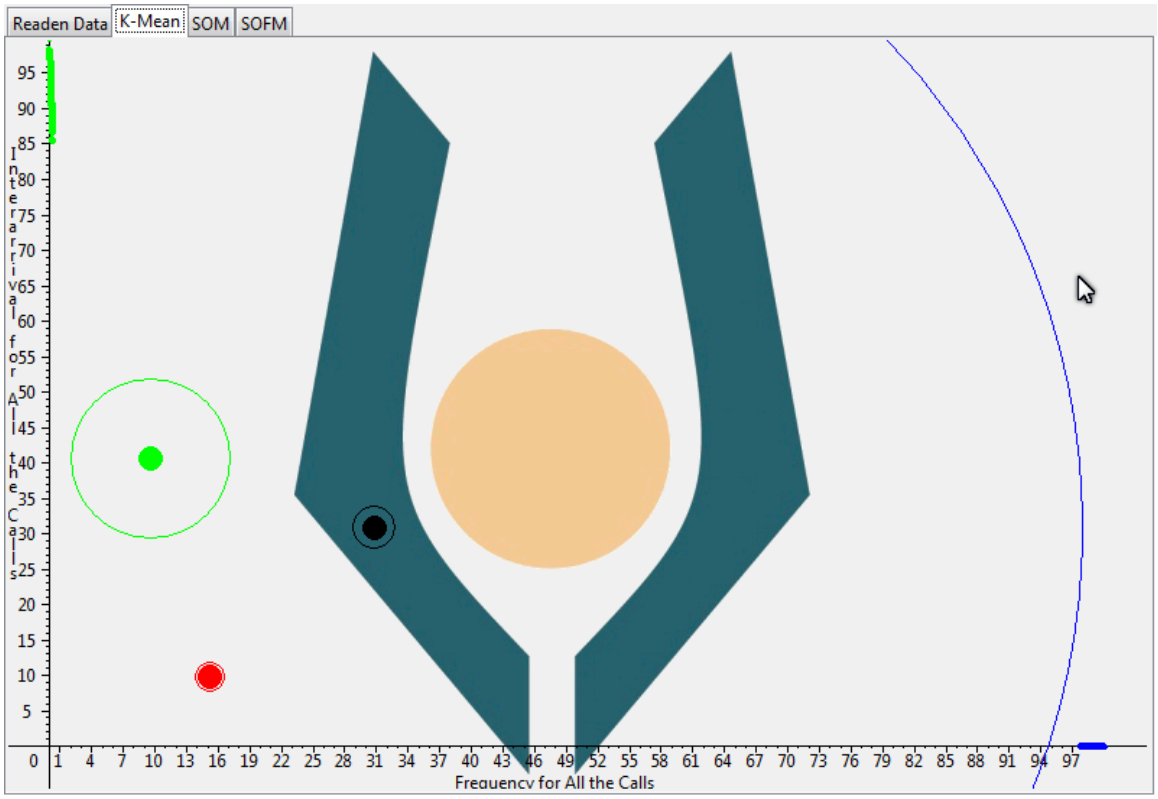
Εικόνα 5.5 Οπτική απεικόνιση των δεδομένων που θα επεξεργαστεί η εφαρμογή. Φυσιολογικές VoIP κλήσεις.

Στον έλεγχο του συστήματος ο Αλγόριθμος K-Mean αναγνωρίζει στην Γκρίζα ζώνη 432 εγγραφές κλήσεων που είναι λανθασμένες διότι γνωρίζουμε πως με αυτές εκπαιδεύτηκε το σύστημα της εφαρμογής. Η τιμή αυτή μπορεί να γίνει αποδεκτή διότι, ορίσαμε για το 1% των εγγραφών που βρίσκονται πλησίον των ορίων κάθε συστάδας το σύστημα να μας ενημερώνει γι' αυτές ώστε ο διαχειριστής να τις ελέγχει σε δεύτερη φάση με άλλα εργαλεία. Η Εικόνα 5.6 εμφανίζει τα αποτελέσματα μετά από έλεγχο που ζητήσαμε από την εφαρμογή



Εικόνα 5.6 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

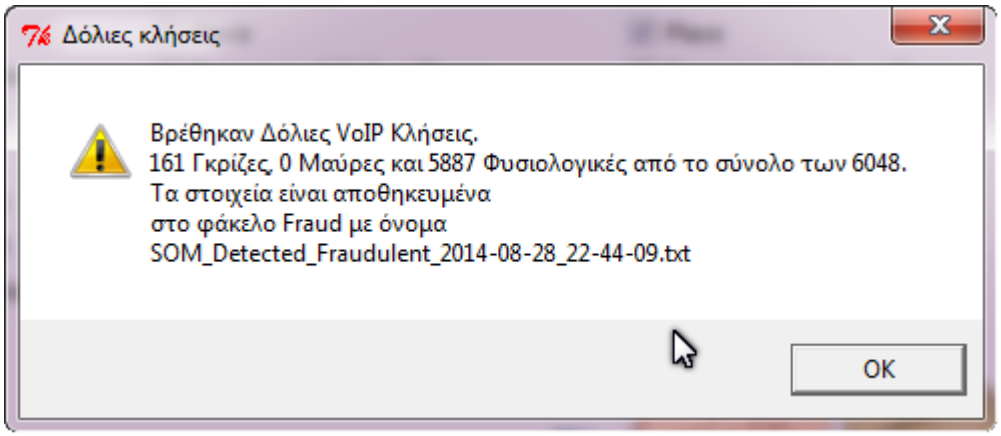
Παρακάτω εμφανίζεται και οπτικά η συσταδοποίηση που έγινε από την εφαρμογή.



Εικόνα 5.7 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο K-Mean.

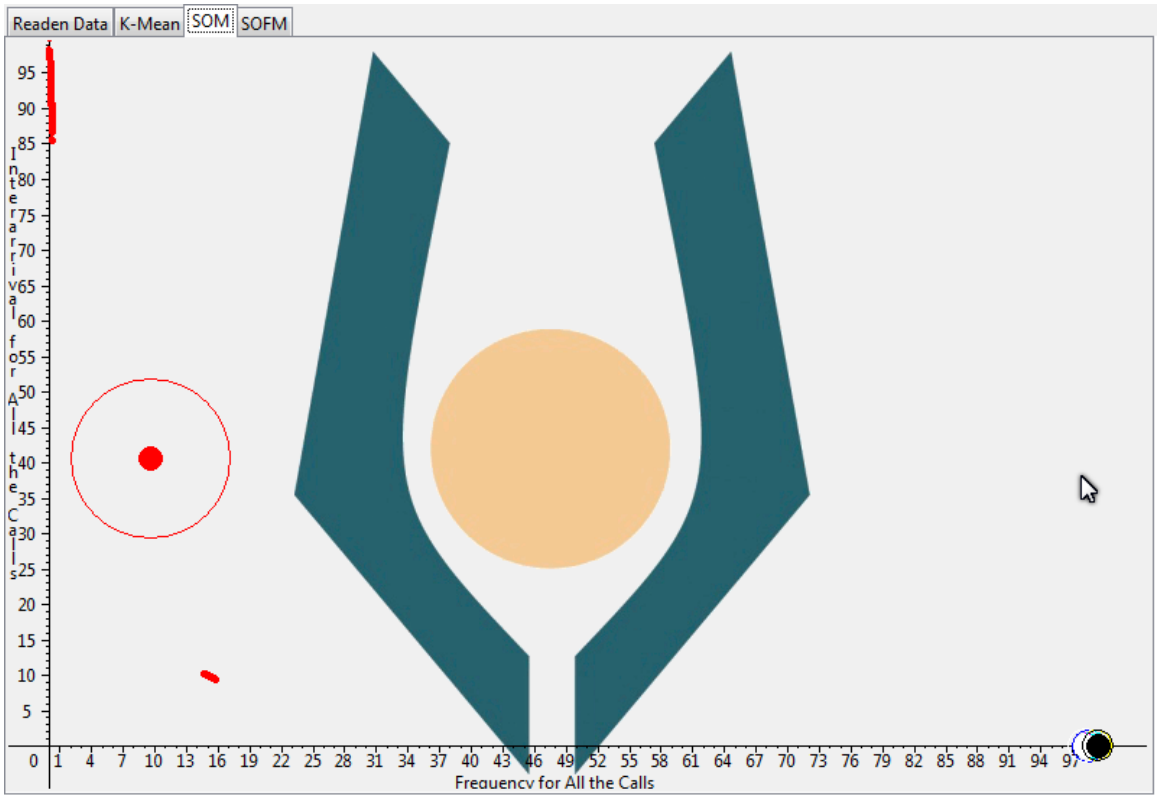
Μεγάλη διαφορά στη Γκρίζα ζώνη από τον αλγόριθμο K-Mean συναντάμε στους αλγορίθμους SOM και SOFM.

Ο αλγόριθμος SOM αναγνώρισε 161 εγγραφές κλήσεων. Κι αυτές ανήκουν στις λάθος ανακοινώσεις, αλλά για τον ίδιο λόγο που δόθηκε στον αλγόριθμο K-Mean, κι αυτές είναι αποδεκτές διότι βρίσκονται πλησίον των άκρων.



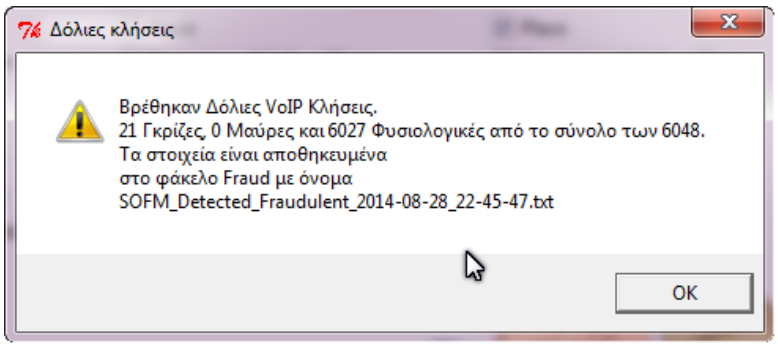
Εικόνα 5.8 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

Στην Εικόνα 5.9 εμφανίζεται και οπτικά η συσταδοποίηση που έγινε από την εφαρμογή.



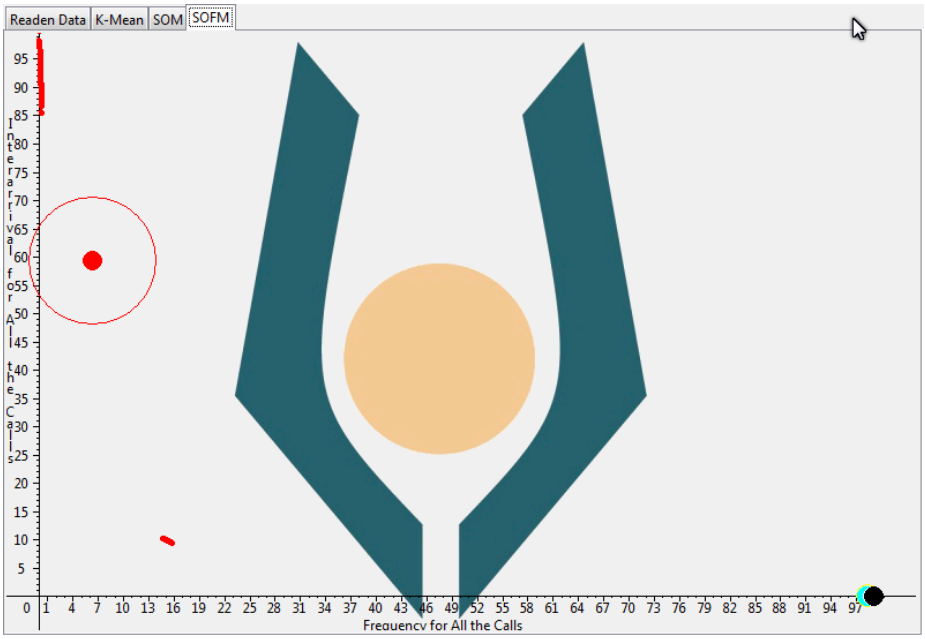
Εικόνα 5.9 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOM.

Ο αλγόριθμος SOFM αναγνώρισε 21 εγγραφές στη Γκρίζα ζώνη. Όπως αναφέρεται και παραπάνω η τιμή αυτή είναι αποδεκτή αν και αποτελεί λάθος σήμανση.



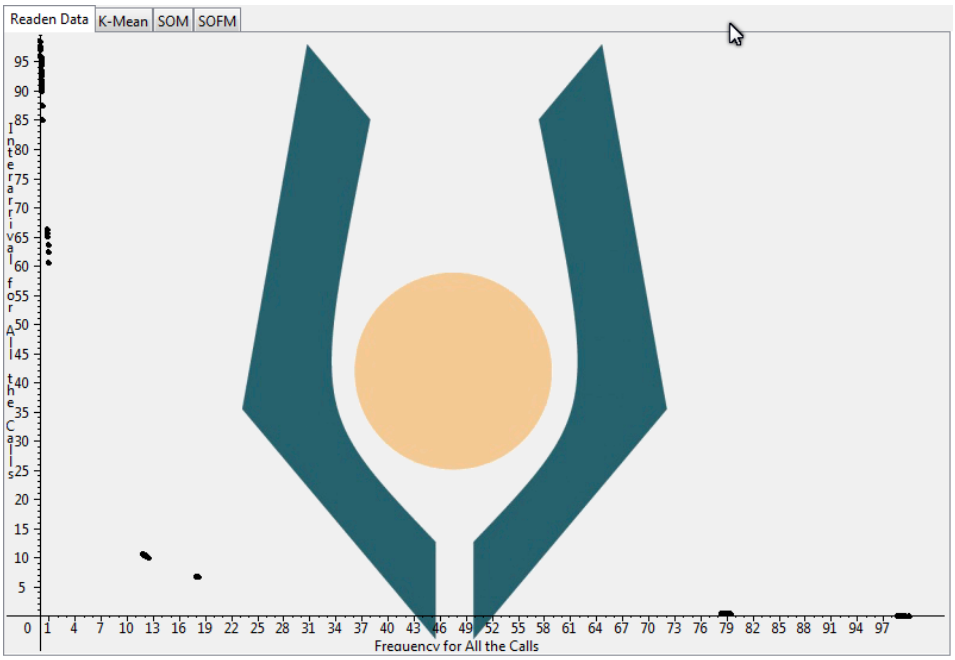
Εικόνα 5.10 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

Στην επόμενη εικόνα εμφανίζεται οπτικά η συσταδοποίηση που έγινε από την εφαρμογή.

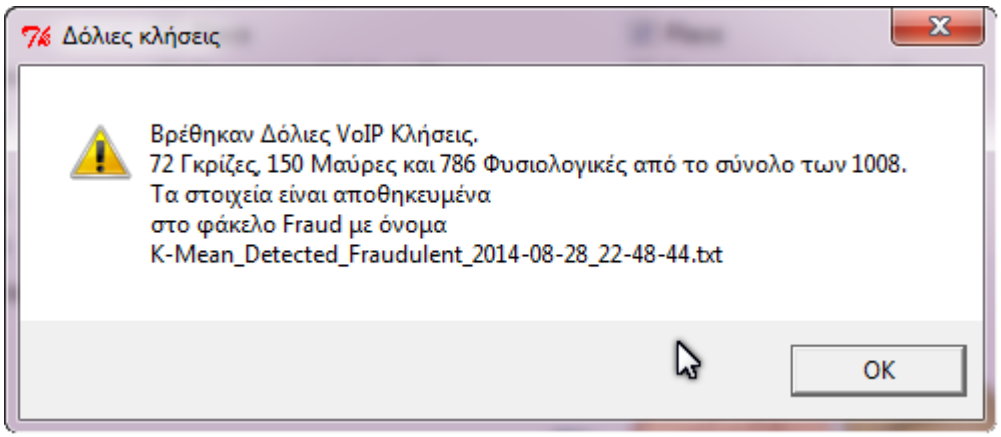


Εικόνα 5.11 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOFM.

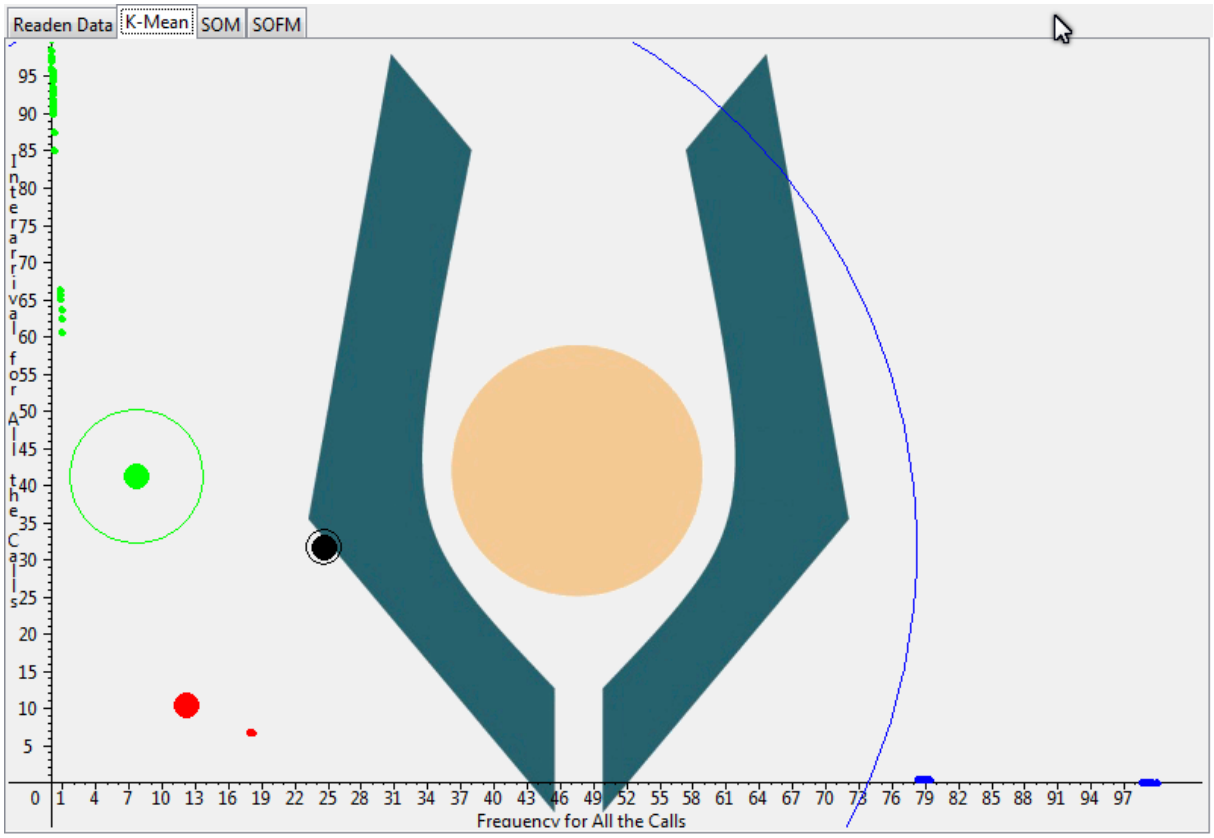
Στην περίπτωση που εισάγουμε δόλιες κλήσεις σε ποσοστό 30% στη διάρκεια της ημέρας και του Σαββατοκύριακου σύμφωνα με τα σενάρια που αναφέραμε, οι αλγόριθμοι εμφανίζουν τα παρακάτω αποτελέσματα όπως δίνονται από την Εικόνα 5.13, την Εικόνα 5.15 και την Εικόνα 5.17. Η Εικόνα 5.12, η Εικόνα 5.14, η Εικόνα 5.16 και η Εικόνα 5.18 παρουσιάζουν την οπτική εικόνα που θα έχει ο χρήστης της εφαρμογής όταν θα ζητήσει τον έλεγχο των δεδομένων που έχει εισάγει.



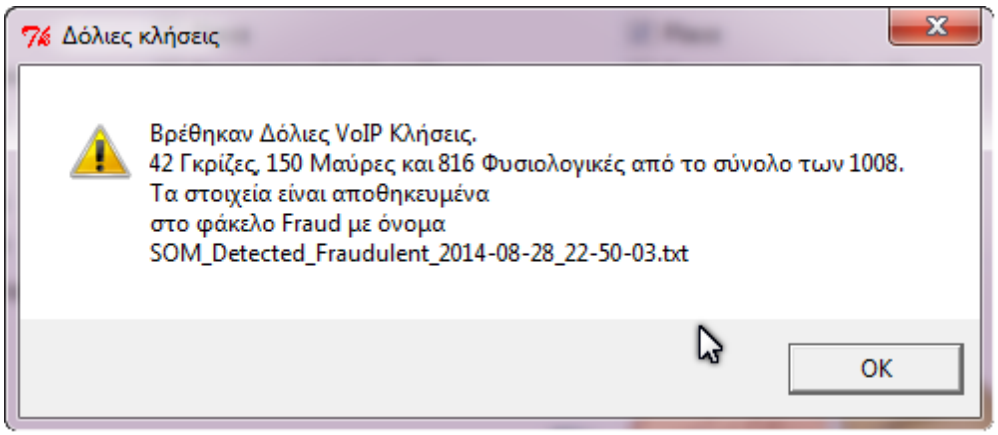
Εικόνα 5.12 Οπτική απεικόνιση των δεδομένων που θα επεξεργαστεί η εφαρμογή. 30% δόλιες VoIP κλήσεις.



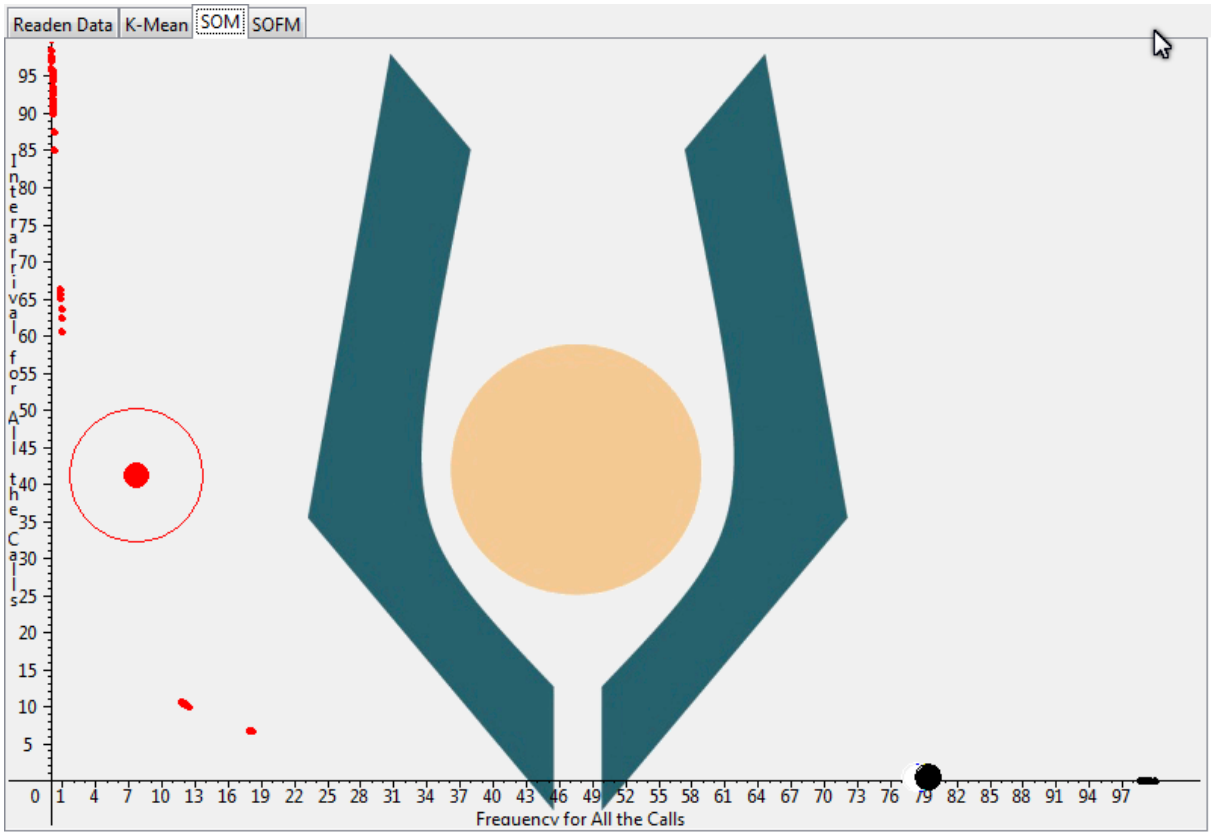
Εικόνα 5.13 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.



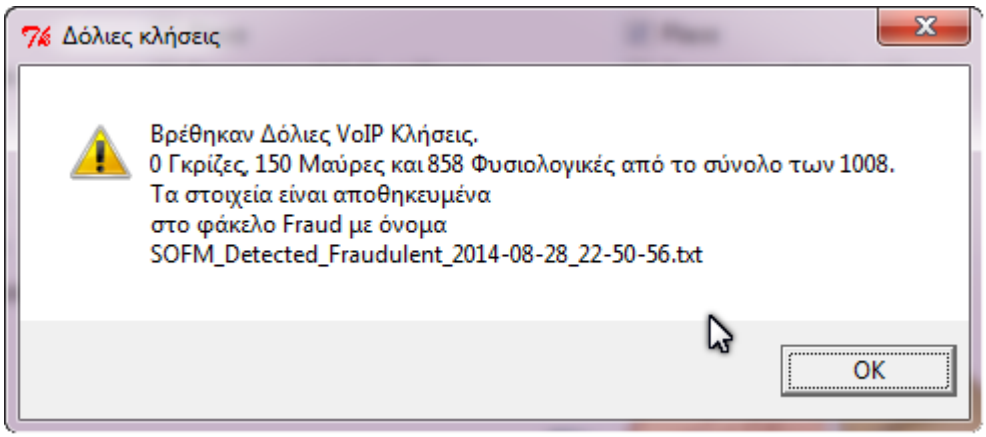
Εικόνα 5.14 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο K-Mean. 30% δόλιες κλήσεις



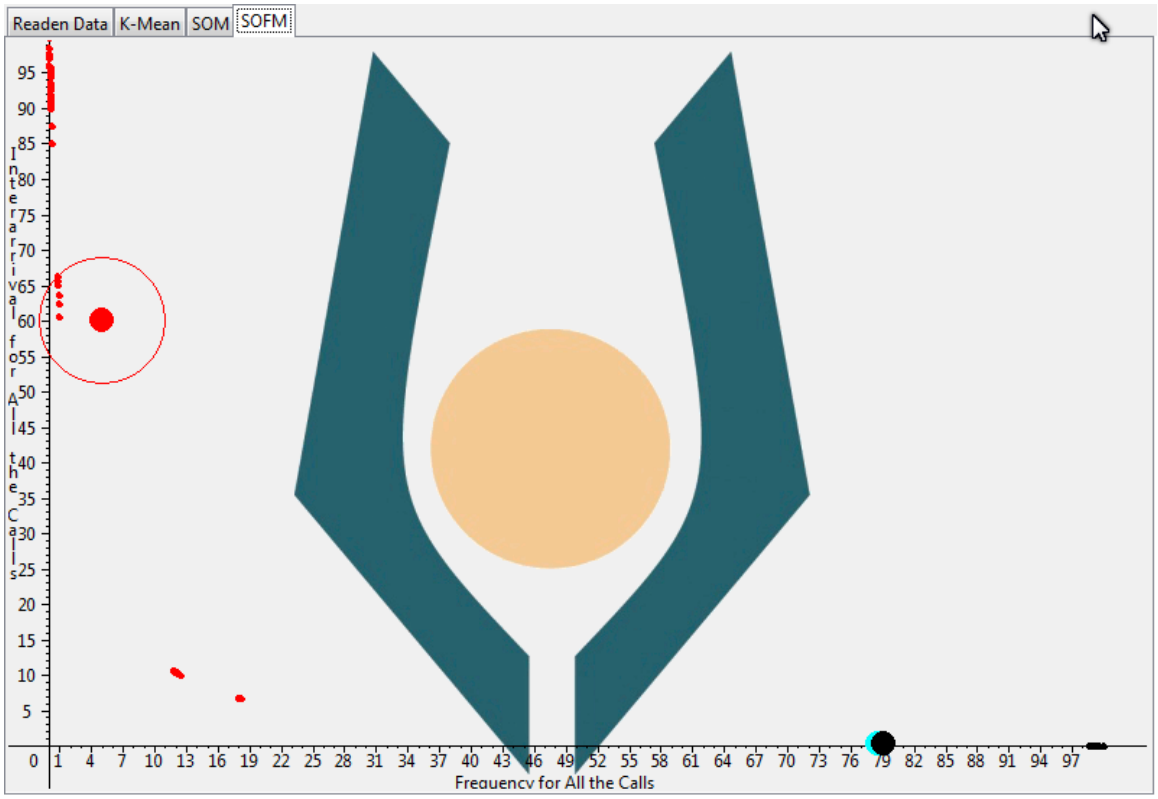
Εικόνα 5.15 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.



Εικόνα 5.16 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOM. 30% δόλιες κλήσεις

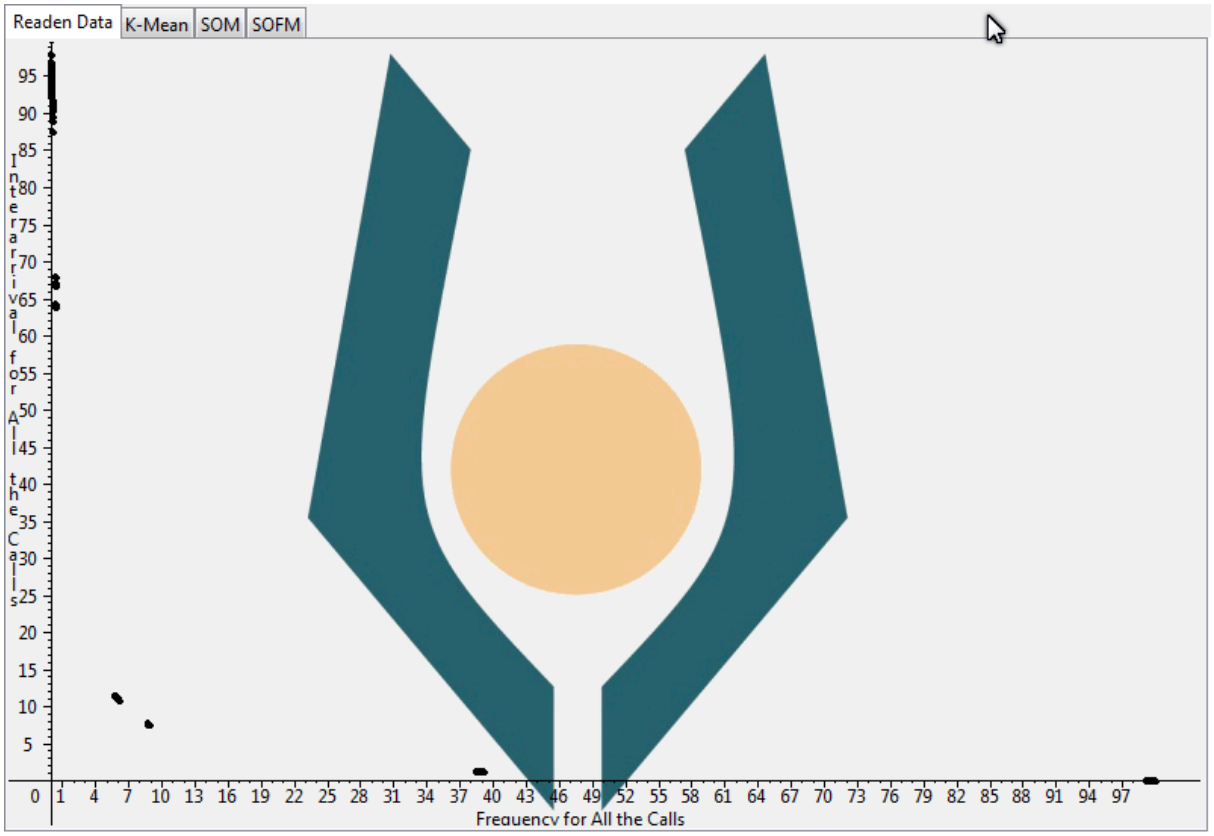


Εικόνα 5.17 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

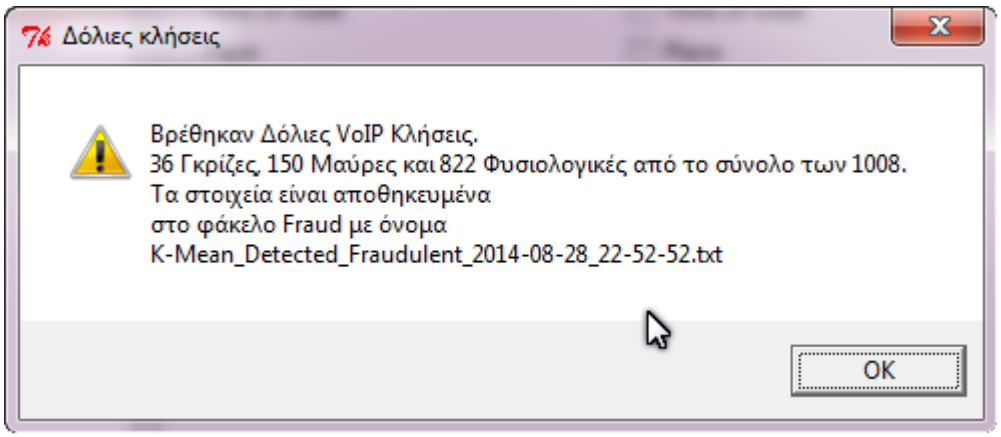


Εικόνα 5.18 Οπτική απεικόνιση τις συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOFM. 30% δόλιες κλήσεις

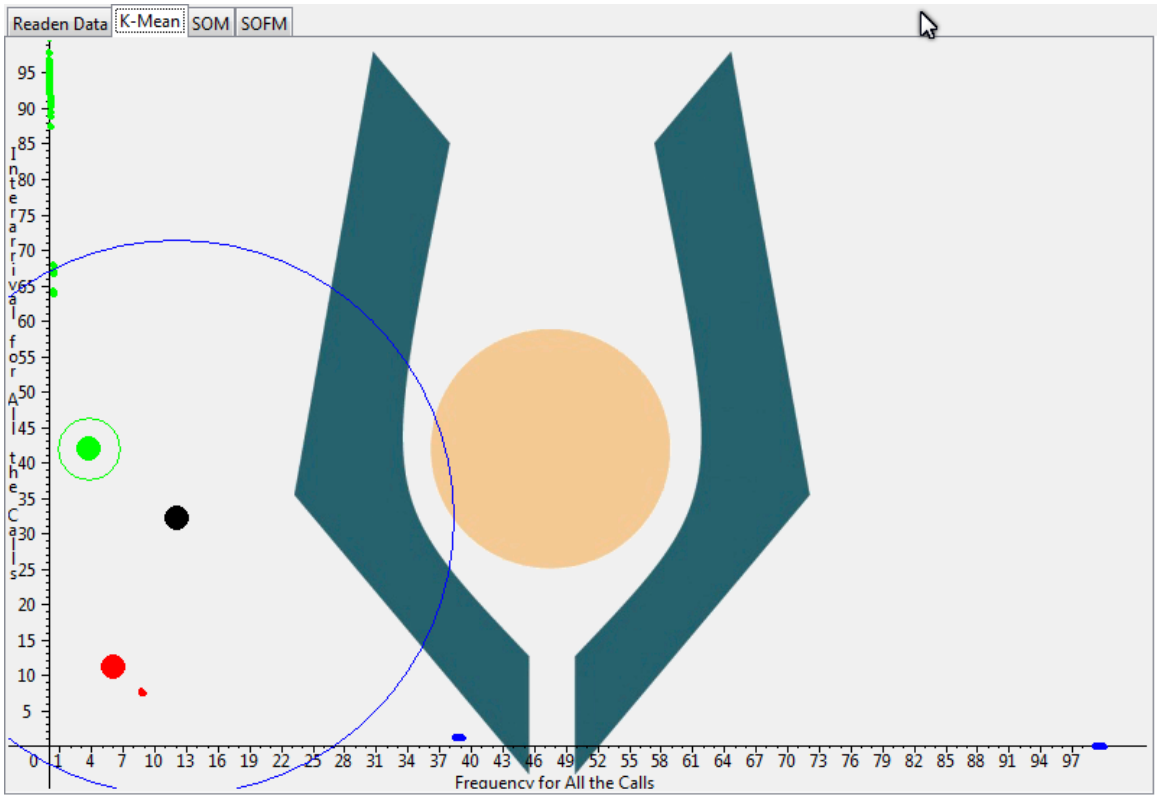
Στην περίπτωση που εισάγουμε δόλιες κλήσεις σε ποσοστό 60% στη διάρκεια της ημέρας και κυρίως τις ώρες εργασίας 9.00 π.μ. με 11.00 π.μ., όπως αναφέρεται παραπάνω, στα σενάρια που θα χρησιμοποιήσουμε για την αξιολόγηση του μηχανισμού ανίχνευσης δόλιων κλήσεων, μπορούμε να δούμε τα αποτελέσματα στις επόμενες εικόνες.



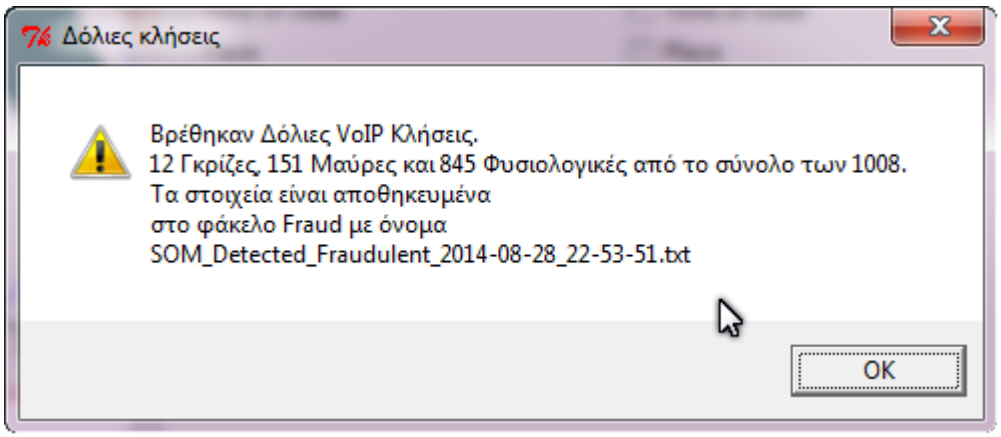
Εικόνα 5.19 Οπτική απεικόνιση των δεδομένων που θα επεξεργαστεί η εφαρμογή. 60% δόλιες VoIP κλήσεις.



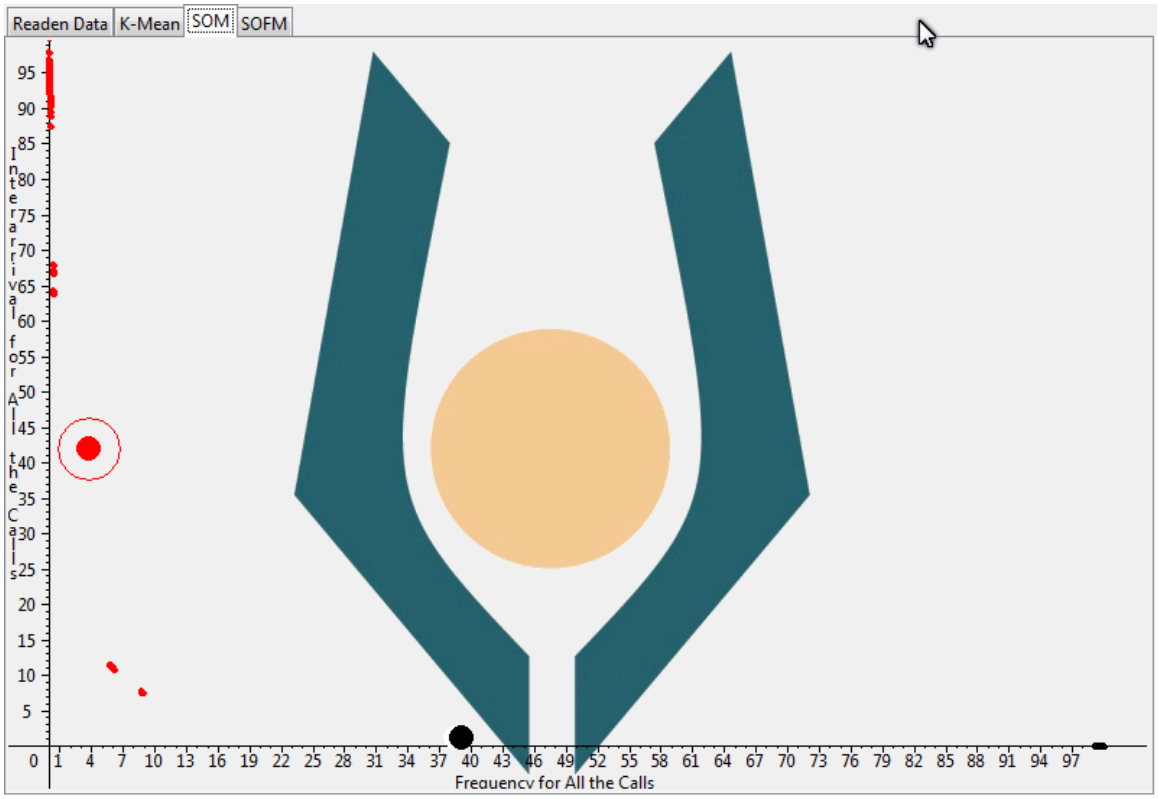
Εικόνα 5.20 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.



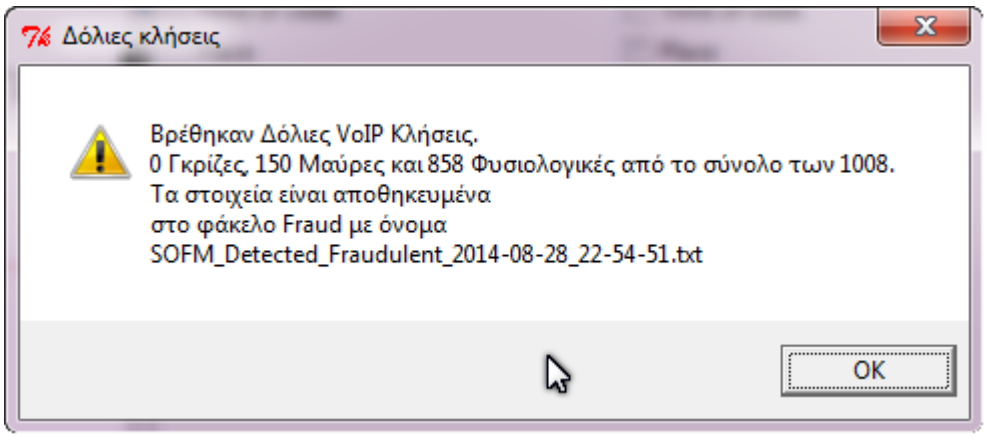
Εικόνα 5.21 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο K-Mean. 60% δόλιες κλήσεις



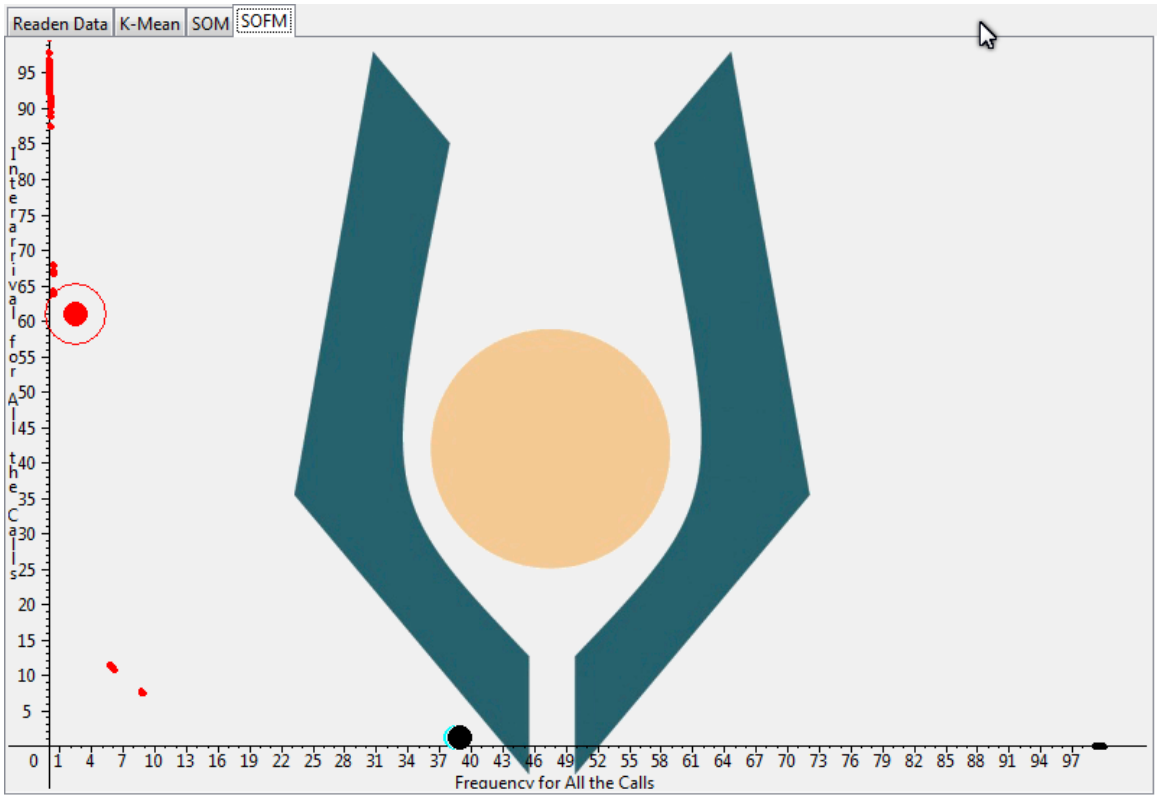
Εικόνα 5.22 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.



Εικόνα 5.23 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOM. 60% δόλιες κλήσεις

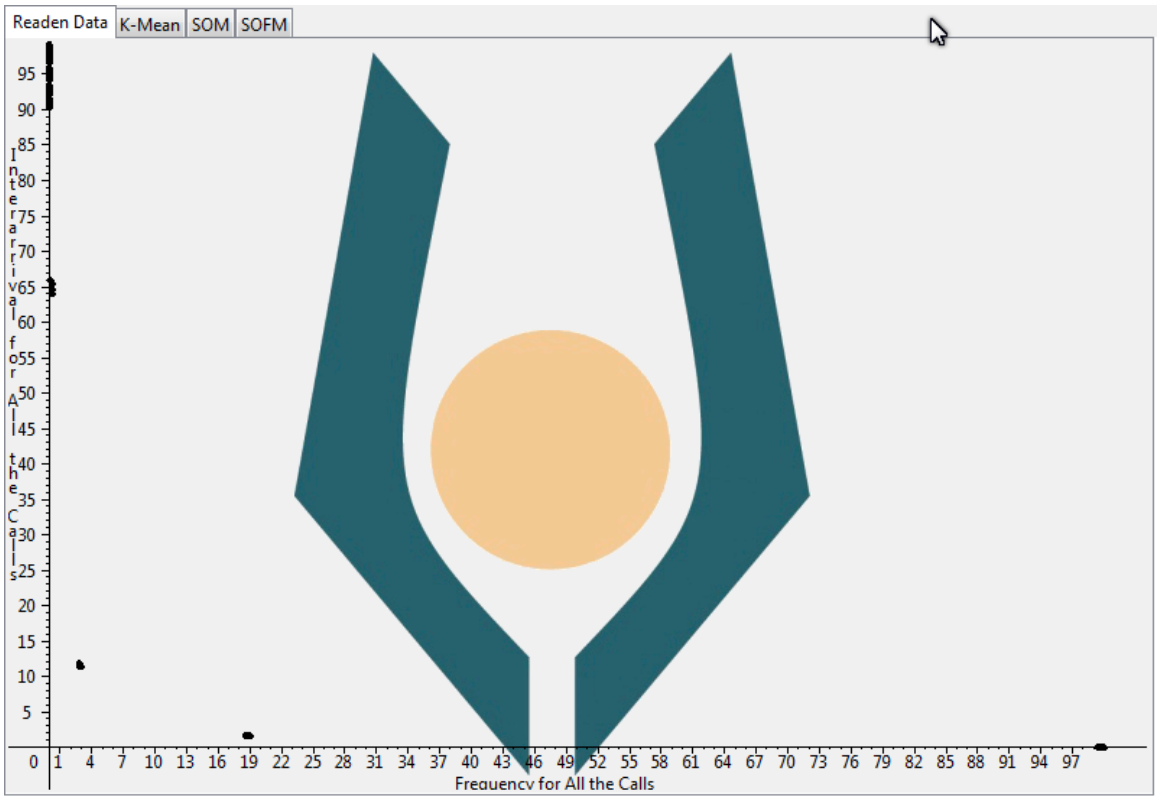


Εικόνα 5.24 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

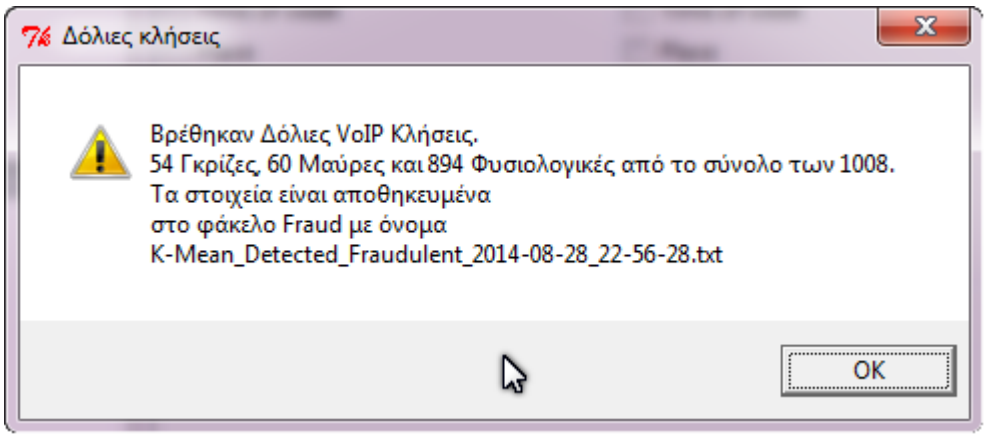


Εικόνα 5.25 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOFM. 60% δόλιες κλήσεις

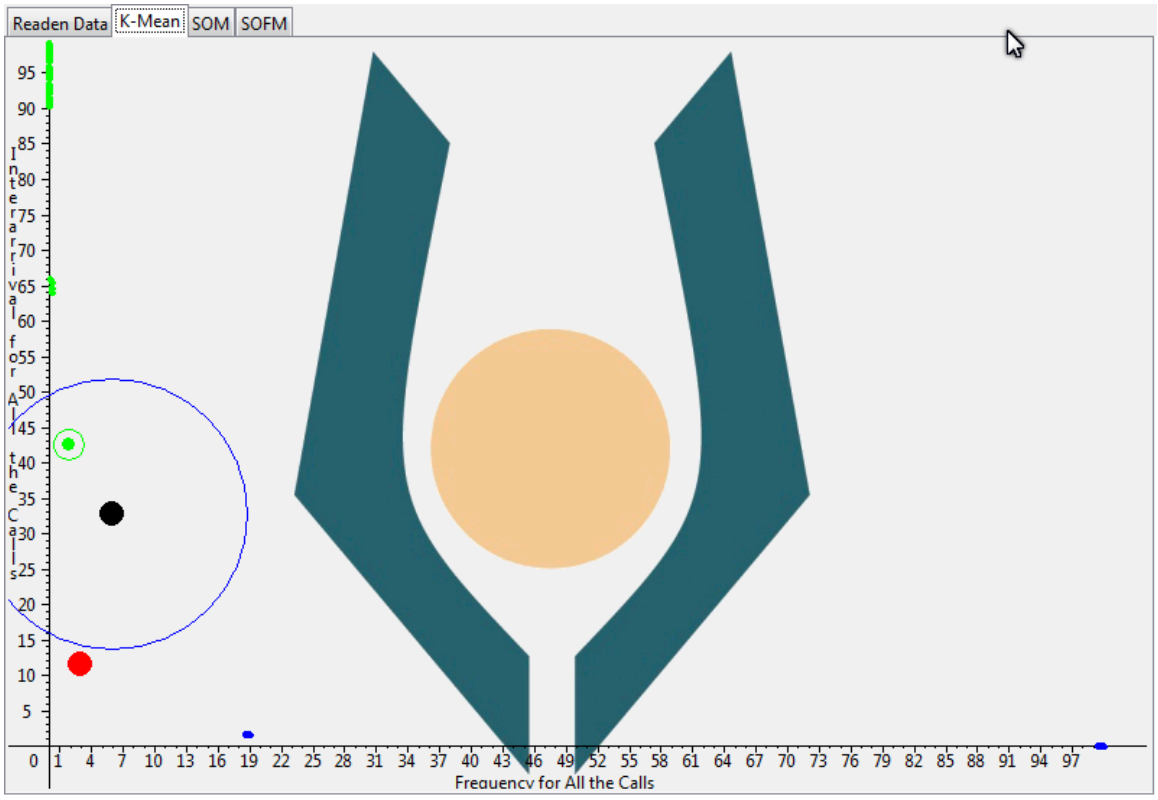
Στην περίπτωση που οι δόλιες κλήσεις ανέβουν σε ποσοστό 80% στη διάρκεια της ημέρας και κυρίως τις ώρες εργασίας 9.00 π.μ. με 11.00 π.μ., τα αποτελέσματα μετά τον έλεγχο του μηχανισμού που χρησιμοποιεί η εφαρμογή δίνονται από τις παρακάτω εικόνες:



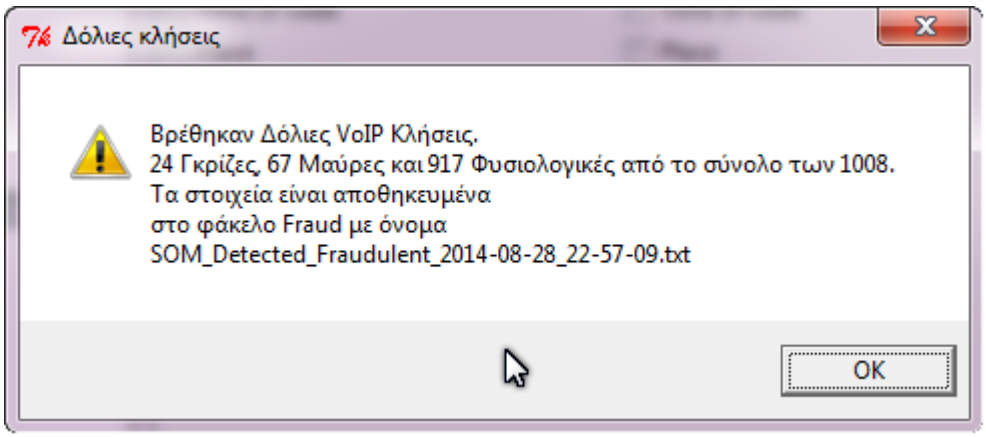
Εικόνα 5.26 Οπτική απεικόνιση των δεδομένων που θα επεξεργαστεί η εφαρμογή. 80% δόλιες VoIP κλήσεις.



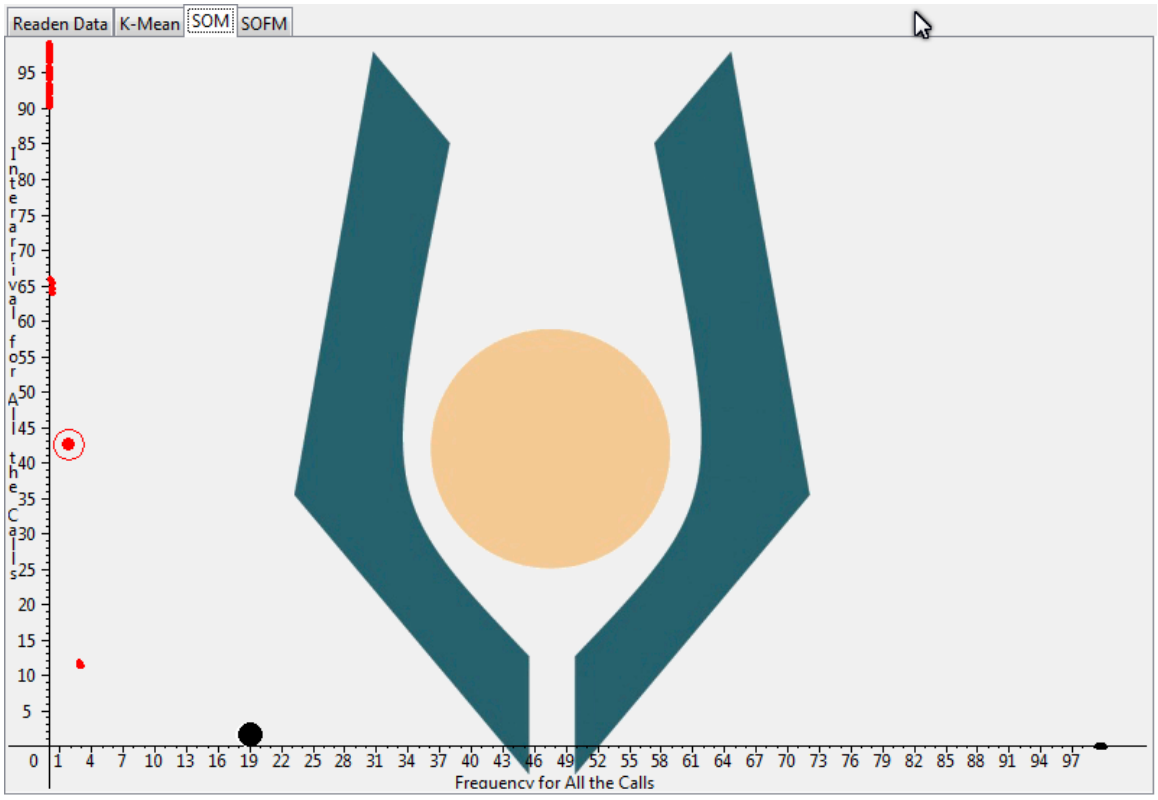
Εικόνα 5.27 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.



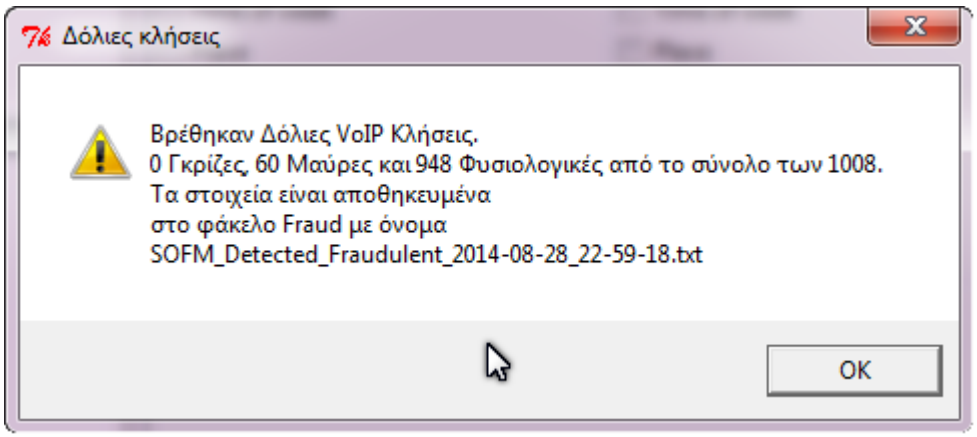
Εικόνα 5.28 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο K-Mean. 80% δόλιες κλήσεις



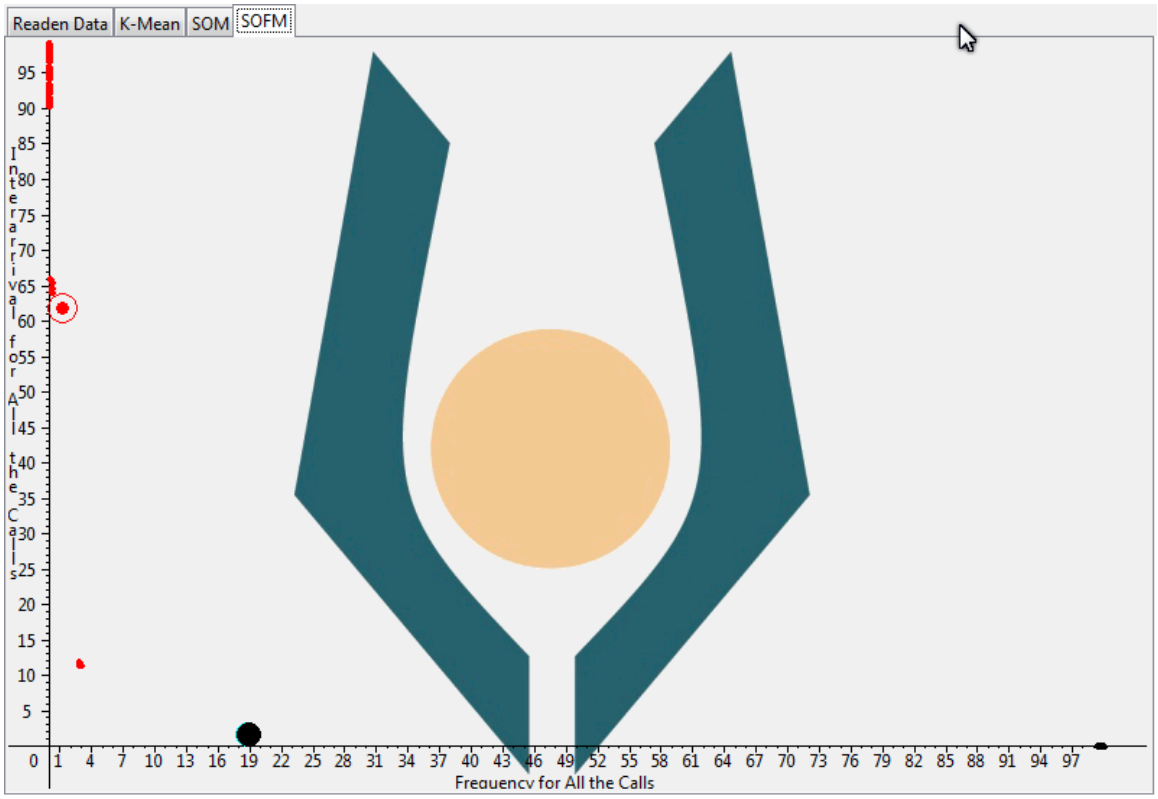
Εικόνα 5.29 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.



Εικόνα 5.30 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOM. 80% δόλιες κλήσεις

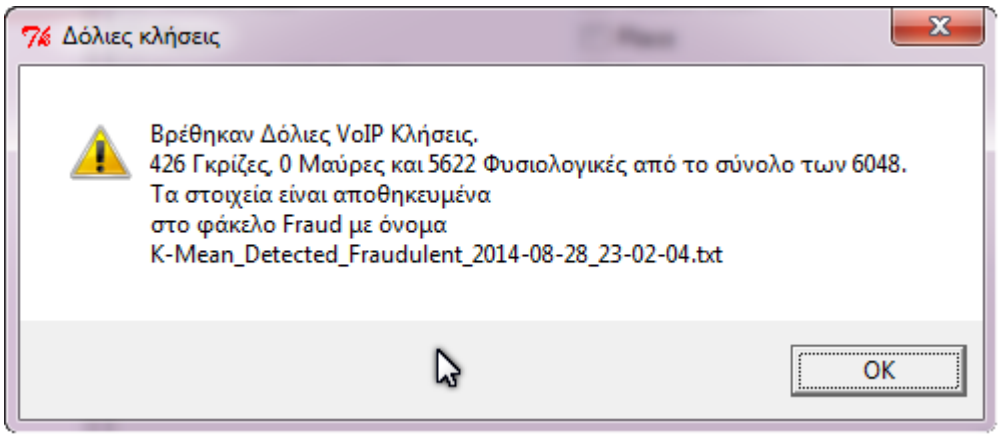


Εικόνα 5.31 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

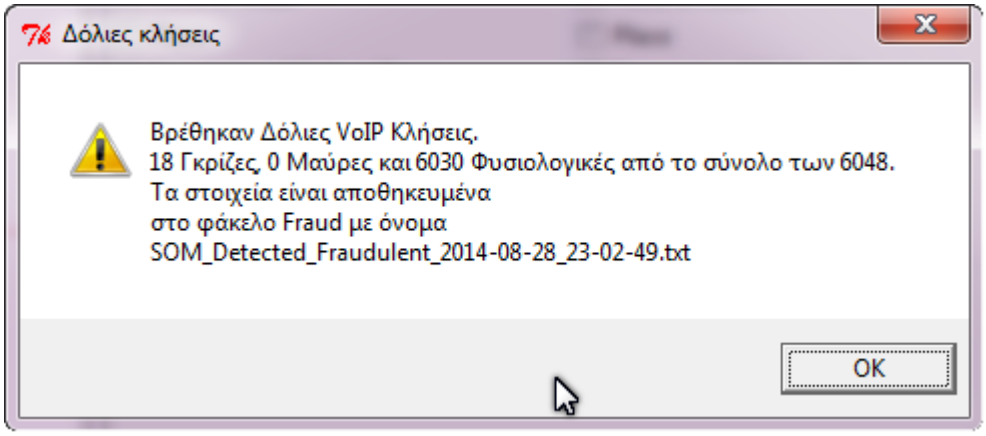


Εικόνα 5.32 Οπτική απεικόνιση της συσταδοποίησης των δεδομένων με τον Αλγόριθμο SOFM. 80% δόλιες κλήσεις

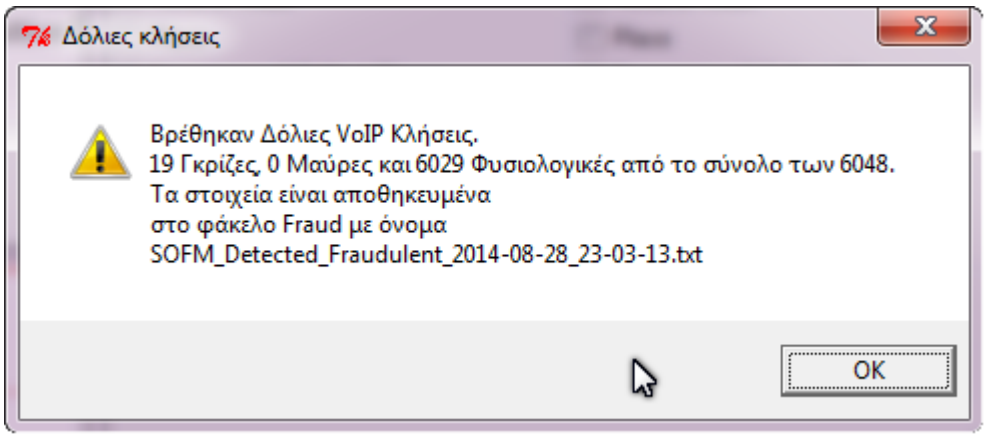
Όλα τα παραπάνω σενάρια επαναλαμβάνονται χρησιμοποιώντας τέσσερις συστάδες αντί για οκτώ. Η εφαρμογή για κάθε σενάριο που προαναφέραμε θα εμφανίσει στο χρήστη τα ακόλουθα μηνύματα. Η οπτική απεικόνιση των δεδομένων που θα χειριστεί η εφαρμογή δεν επαναλαμβάνεται παρακάτω.



Εικόνα 5.33 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

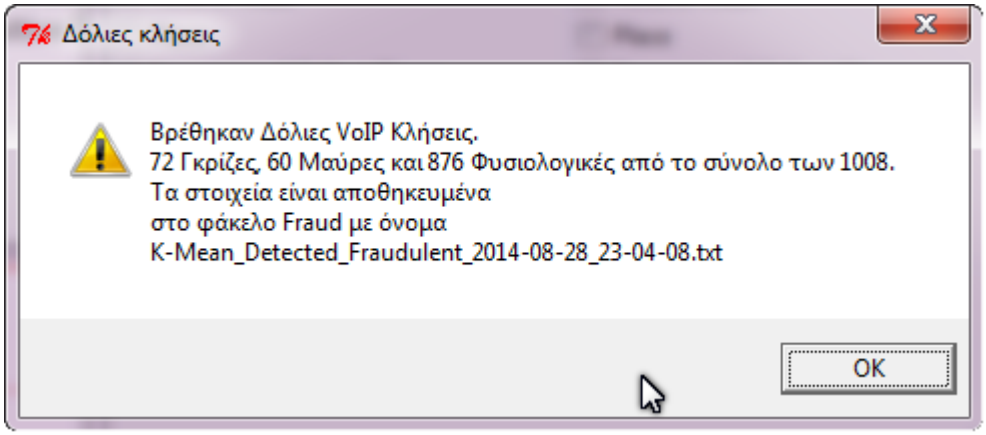


Εικόνα 5.34 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

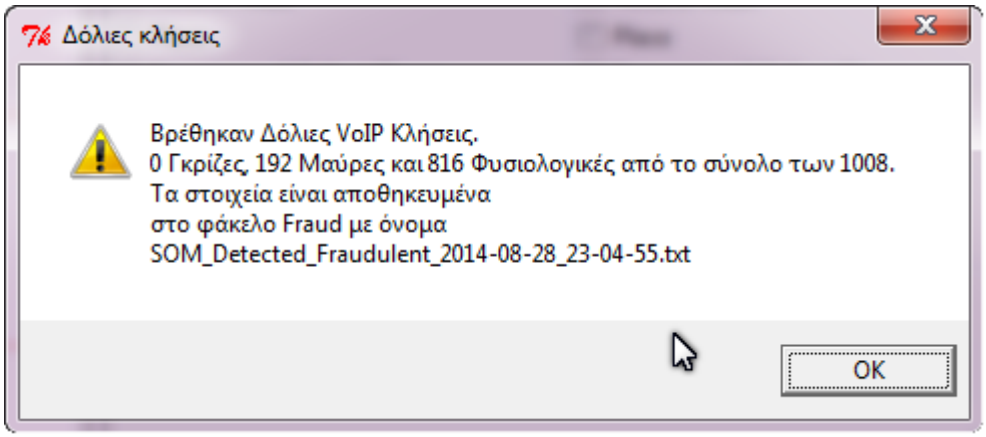


Εικόνα 5.35 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

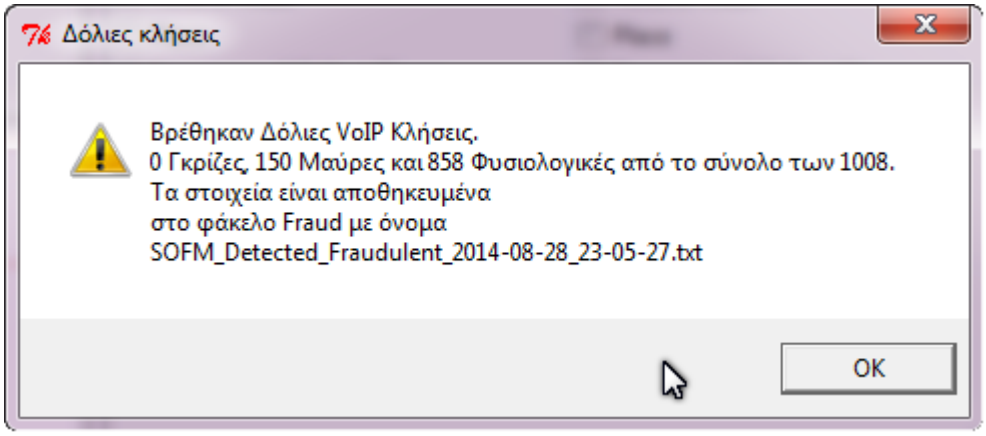
Σενάριο με τέσσερις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 30%



Εικόνα 5.36 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

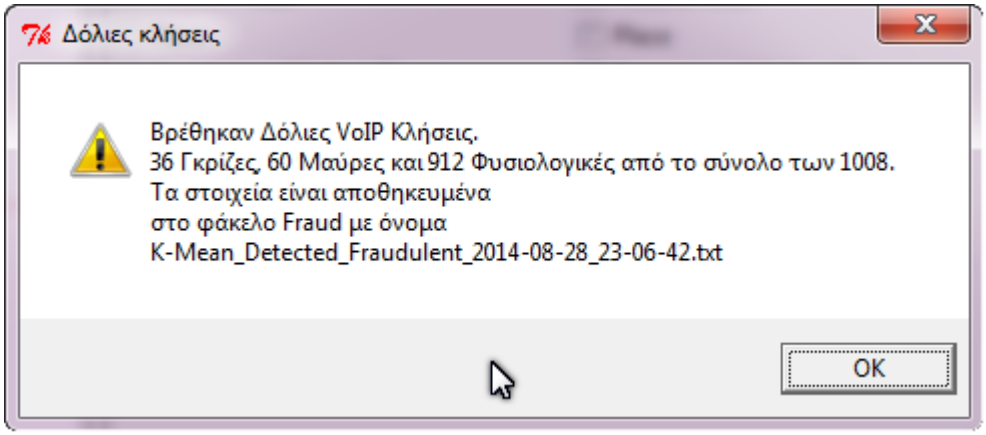


Εικόνα 5.37 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

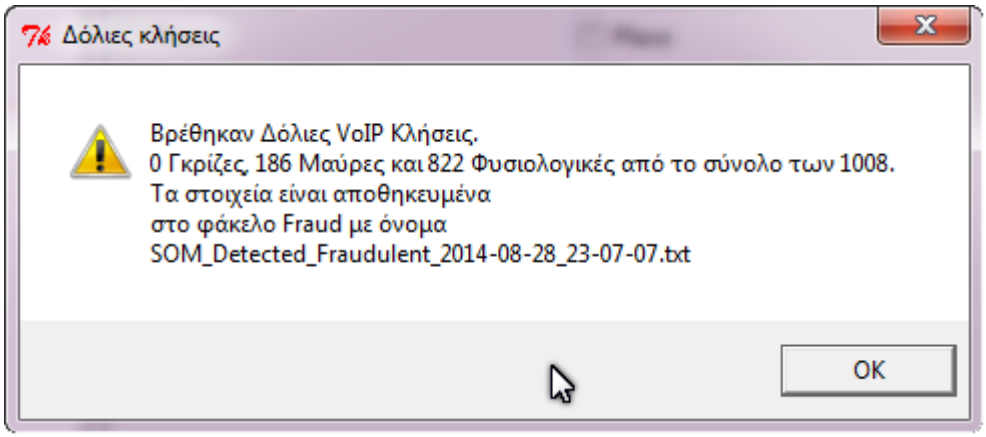


Εικόνα 5.38 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

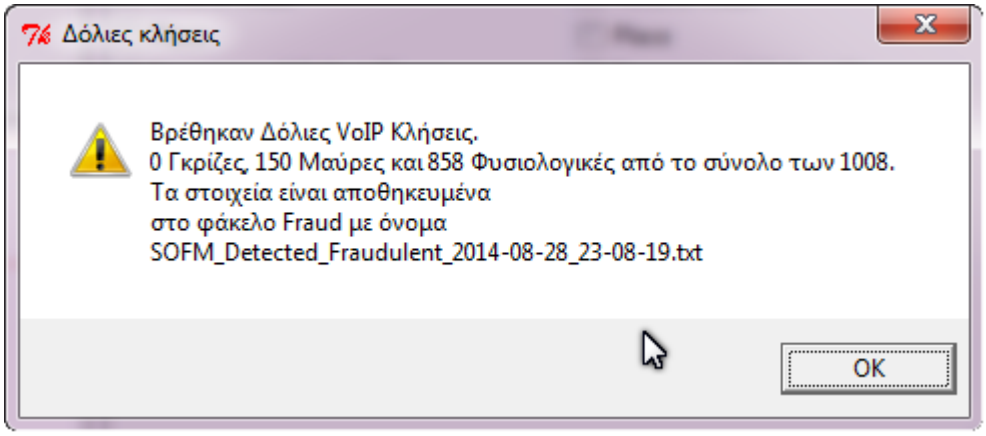
Σενάριο με τέσσερις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 60%



Εικόνα 5.39 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

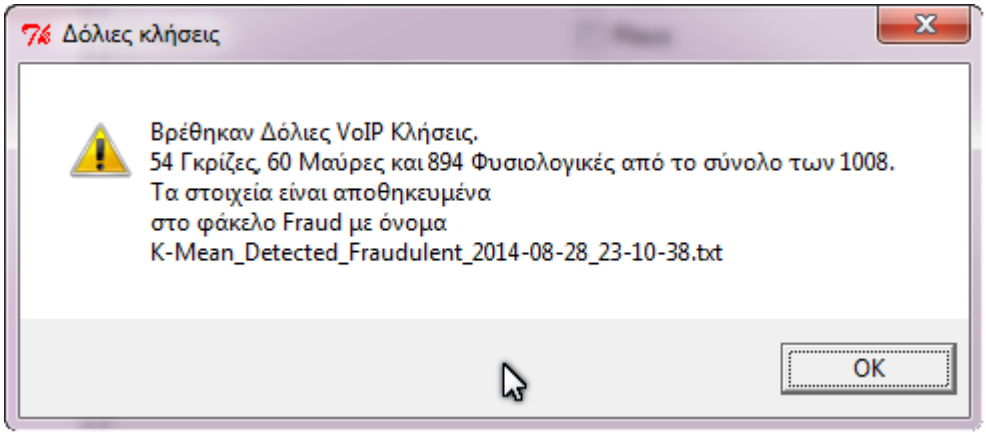


Εικόνα 5.40 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

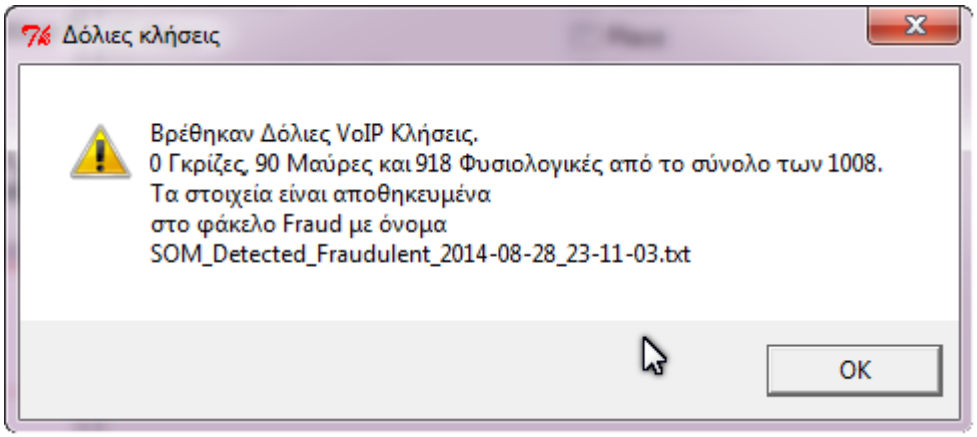


Εικόνα 5.41 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

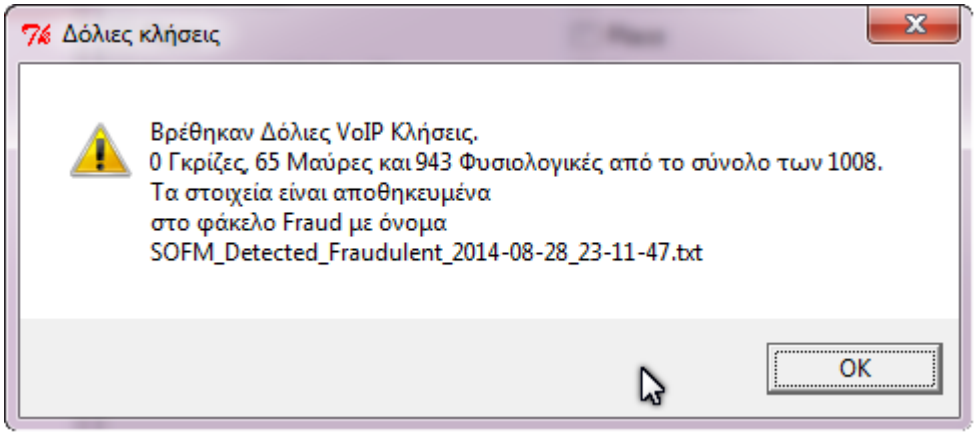
Σενάριο με τέσσερις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 80%



Εικόνα 5.42 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

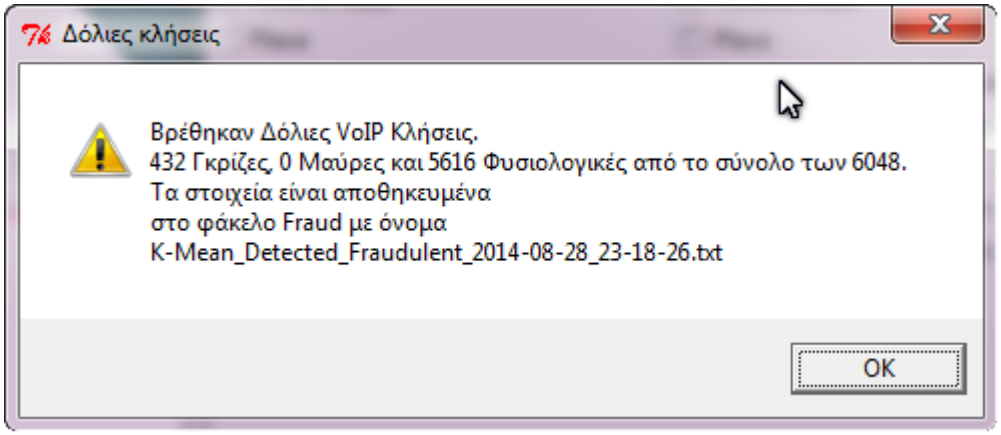


Εικόνα 5.43 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

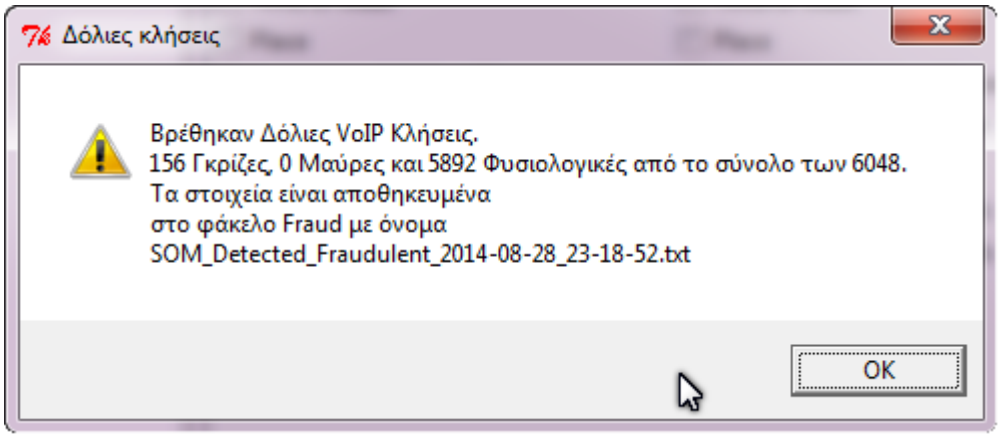


Εικόνα 5.44 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

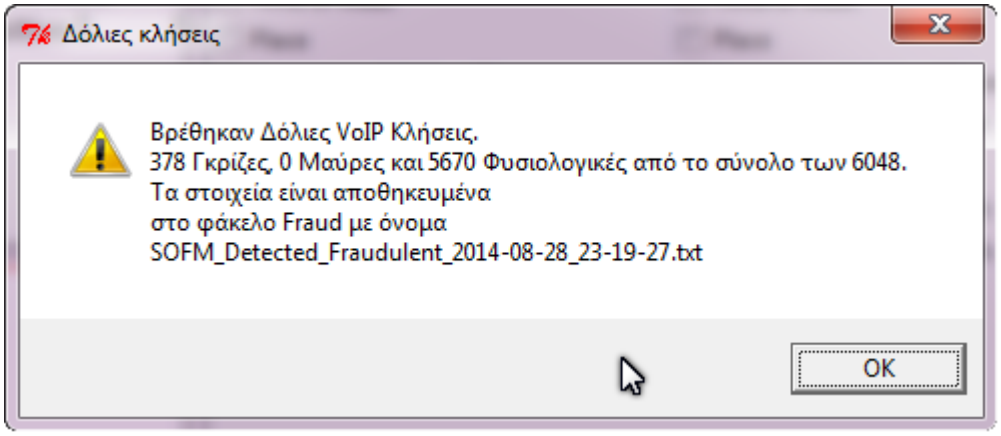
Σενάριο με τρεις συστάδες και καθόλου δόλιες VoIP κλήσεις



Εικόνα 5.45 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

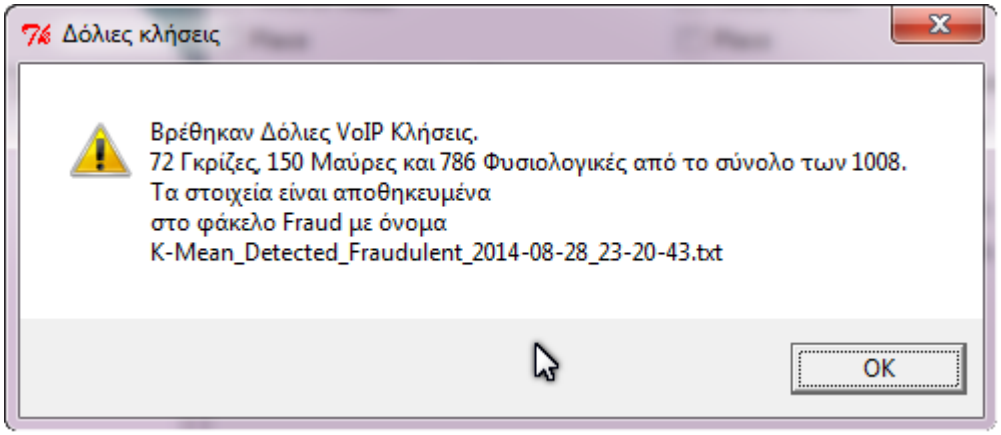


Εικόνα 5.46 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

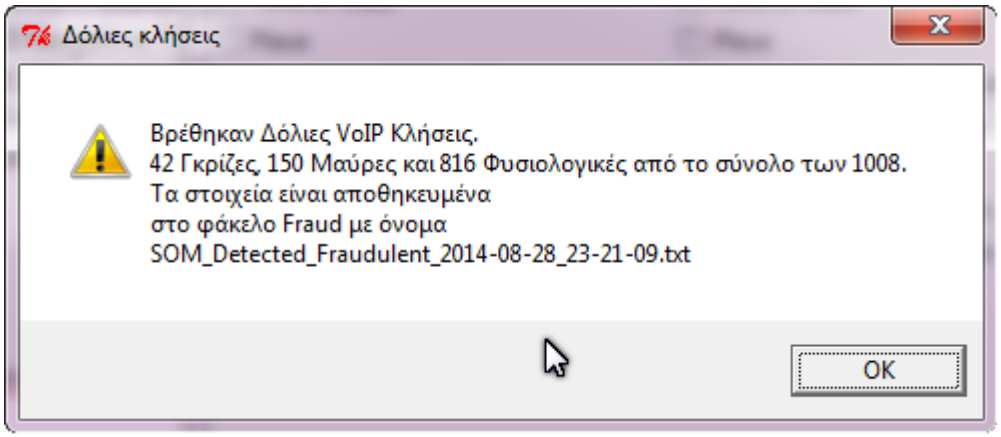


Εικόνα 5.47 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

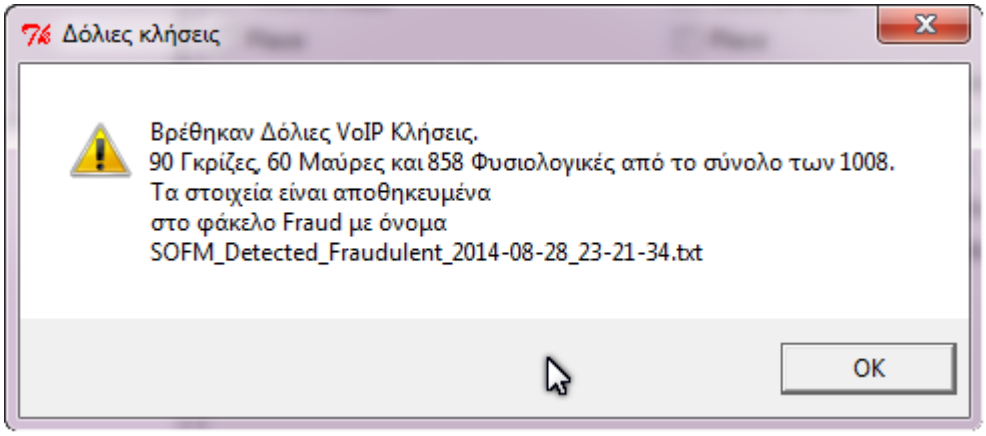
Σενάριο με τρεις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 30%



Εικόνα 5.48 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

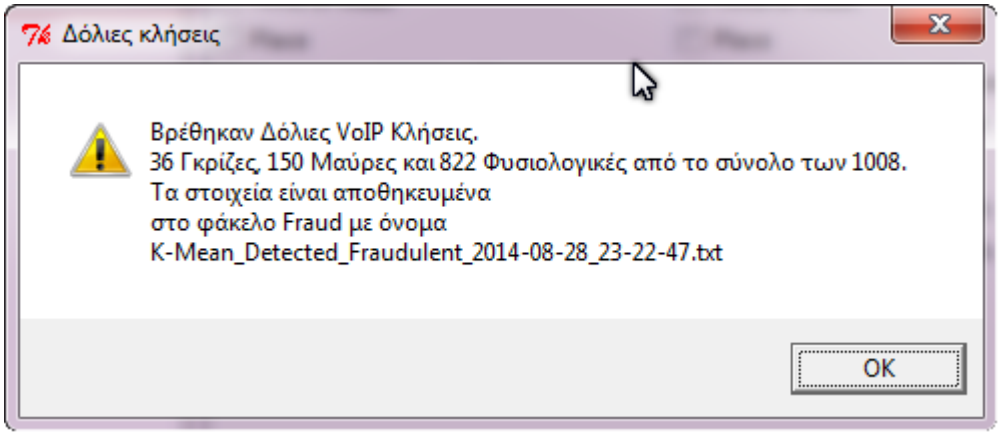


Εικόνα 5.49 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

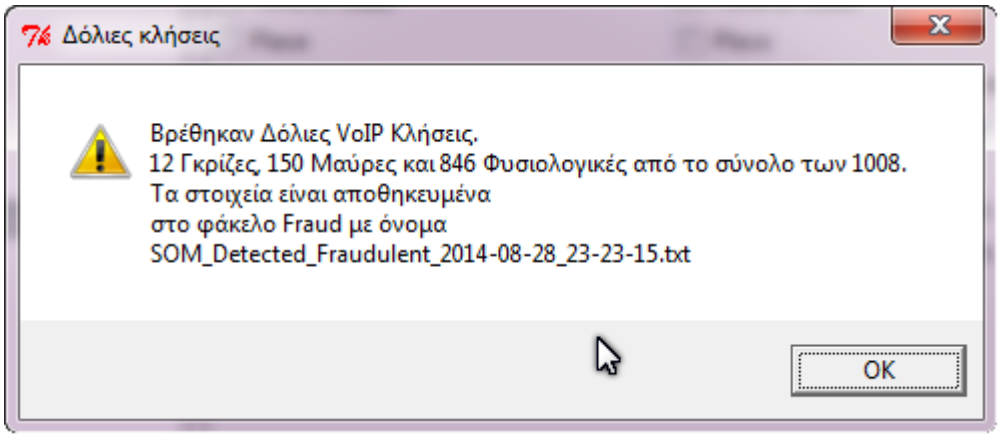


Εικόνα 5.50 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

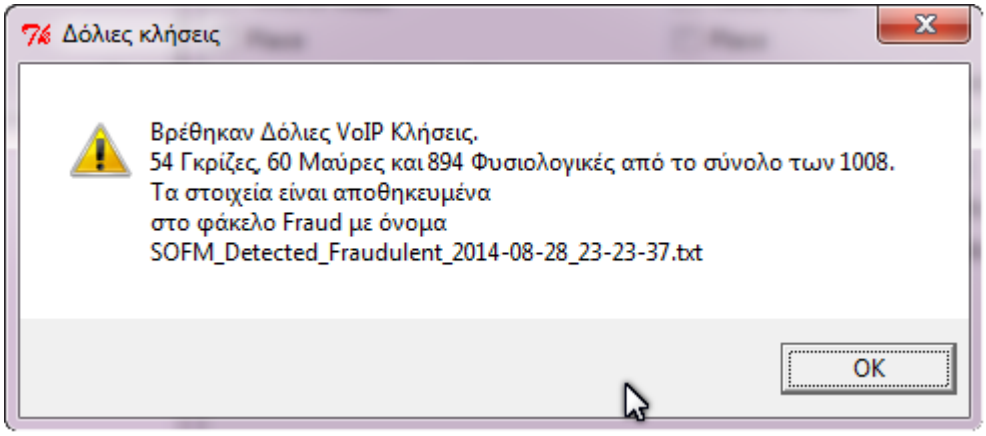
Σενάριο με τρεις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 60%



Εικόνα 5.51 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.

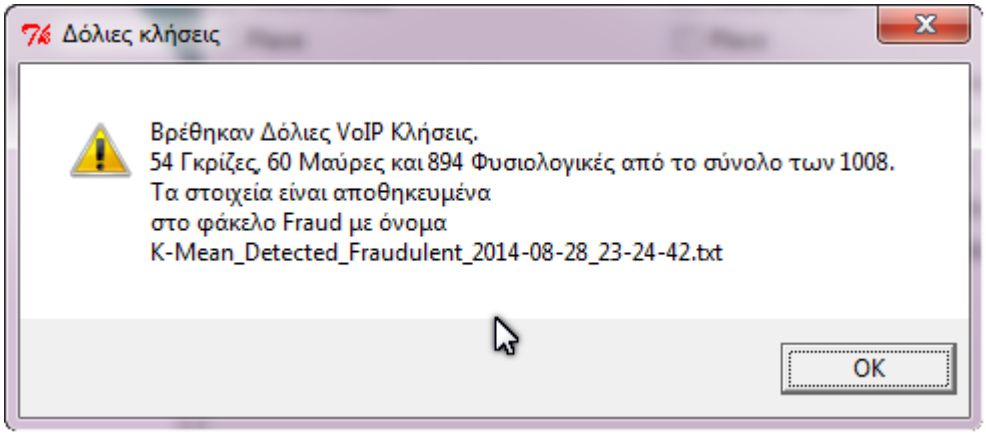


Εικόνα 5.52 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.

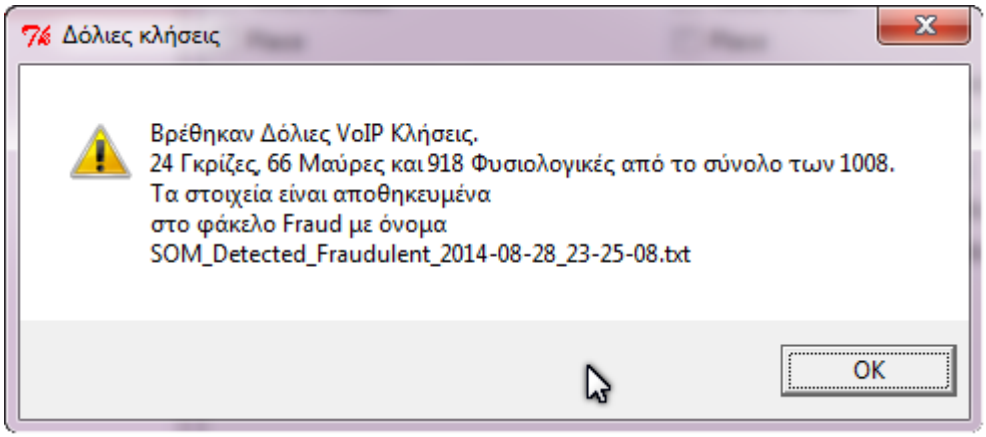


Εικόνα 5.53 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

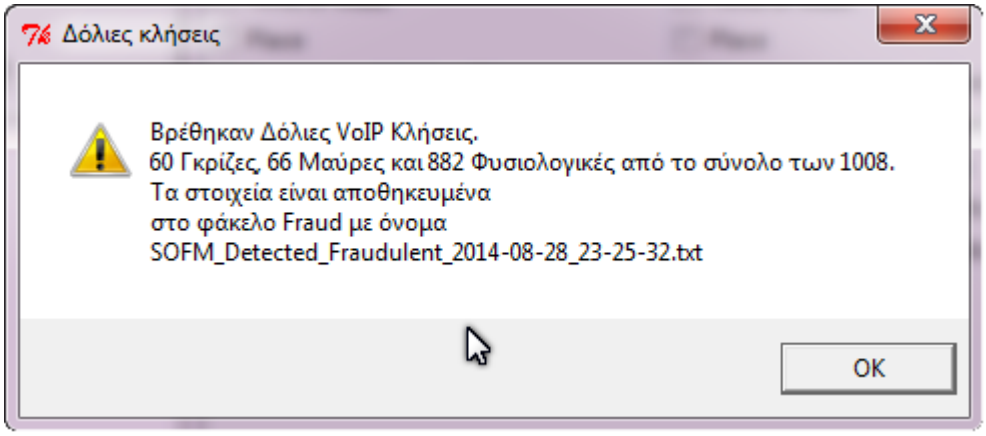
Σενάριο με τρεις συστάδες και ποσοστό δόλιων VoIP κλήσεων στο 80%



Εικόνα 5.54 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο K-Mean.



Εικόνα 5.55 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOM.



Εικόνα 5.56 Ενημερωτικό μήνυμα σχετικά με την επιτυχία αναγνώρισης δόλιων κλήσεων με τον Αλγόριθμο SOFM.

5.2 Αποτελέσματα

Στο υπό-τμήμα αυτό θα παρουσιαστούν τα αποτελέσματα του εκπαιδευμένου μηχανισμού ανίχνευσης δόλιων κλήσεων. Ακολουθούν οι πίνακες με τα αποτελέσματα από τη χρήση των τριών σεναρίων που αναφέρθηκαν παραπάνω χρησιμοποιώντας τρεις διαφορετικούς αριθμούς συστάδων (οκτώ συστάδες, τέσσερις συστάδες και τρεις συστάδες).

	Συστάδες 8, Φυσιολογικές Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	432	0	0	5616	0	6048
SOM	0	161	0	0	5887	0	6048
SOFM	0	21	0	0	6027	0	6048

Πίνακας 5.1. Αποτελέσματα με βαθμό Συσταδοποίησης 8. Δεδομένα εισόδου: Φυσιολογικές VoIP κλήσεις

	Συστάδες 8, 30% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	72	150	0	750	36	1008
SOM	0	42	150	0	780	36	1008
SOFM	0	0	150	0	822	36	1008

Πίνακας 5.2. Αποτελέσματα με βαθμό Συσταδοποίησης 8. Δεδομένα εισόδου: 30% Δόλιες VoIP κλήσεις

	Συστάδες 8, 60% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	36	150	0	786	36	1008
SOM	0	12	150	1	809	36	1008
SOFM	0	0	150	0	822	36	1008

Πίνακας 5.3. Αποτελέσματα με βαθμό Συσταδοποίησης 8. Δεδομένα εισόδου: 60% Δόλιες VoIP κλήσεις

	Συστάδες 8, 80% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	54	60	0	870	24	1008
SOM	0	24	60	7	893	24	1008
SOFM	0	0	60	0	924	24	1008

Πίνακας 5.4. Αποτελέσματα με βαθμό Συσταδοποίησης 8. Δεδομένα εισόδου: 80% Δόλιες VoIP κλήσεις

	Συστάδες 4, Φυσιολογικές Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	426	0	0	5622	0	6048
SOM	0	18	0	0	6030	0	6048
SOFM	0	19	0	0	6029	0	6048

Πίνακας 5.5. Αποτελέσματα με βαθμό Συσταδοποίησης 4. Δεδομένα εισόδου: Φυσιολογικές VoIP κλήσεις

	Συστάδες 4, 30% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	72	60	0	750	126	1008
SOM	0	0	186	6	816	0	1008
SOFM	0	0	150	0	822	36	1008

Πίνακας 5.6. Αποτελέσματα με βαθμό Συσταδοποίησης 4. Δεδομένα εισόδου: 30% Δόλιες VoIP κλήσεις

	Συστάδες 4, 60% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	36	60	0	786	126	1008
SOM	0	0	186	0	822	0	1008
SOFM	0	0	150	0	822	36	1008

Πίνακας 5.7. Αποτελέσματα με βαθμό Συσταδοποίησης 4. Δεδομένα εισόδου: 60% Δόλιες VoIP κλήσεις

	Συστάδες 4, 80% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	54	60	0	870	24	1008
SOM	0	0	84	6	918	0	1008
SOFM	0	0	60	5	919	24	1008

Πίνακας 5.8. Αποτελέσματα με βαθμό Συσταδοποίησης 4. Δεδομένα εισόδου: 80% Δόλιες VoIP κλήσεις

	Συστάδες 3, Φυσιολογικές Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	432	0	0	5616	0	6048
SOM	0	156	0	0	5892	0	6048
SOFM	0	378	0	0	5670	0	6048

Πίνακας 5.9. Αποτελέσματα με βαθμό Συσταδοποίησης 3. Δεδομένα εισόδου: Φυσιολογικές VoIP κλήσεις

	Συστάδες 3, 30% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	72	150	0	750	36	1008
SOM	0	42	150	0	780	36	1008
SOFM	0	90	60	0	732	126	1008

Πίνακας 5.10. Αποτελέσματα με βαθμό Συσταδοποίησης 3. Δεδομένα εισόδου: 30% Δόλιες VoIP κλήσεις

	Συστάδες 3, 60% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	36	150	0	786	36	1008
SOM	0	12	150	0	810	36	1008
SOFM	0	54	60	0	768	126	1008

Πίνακας 5.11. Αποτελέσματα με βαθμό Συσταδοποίησης 3. Δεδομένα εισόδου: 60% Δόλιες VoIP κλήσεις

	Συστάδες 3, 80% Δόλιες Κλήσεις						
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης		
	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σωστή Επιλογή	Λάθος Επιλογή	Σύνολο Εγγραφών
K-Mean	0	54	60	0	870	24	1008
SOM	0	24	60	6	894	24	1008
SOFM	0	60	60	6	858	24	1008

Πίνακας 5.12. Αποτελέσματα με βαθμό Συσταδοποίησης 3. Δεδομένα εισόδου: 80% Δόλιες VoIP κλήσεις

Από τους παραπάνω Πίνακες εξάγονται οι μέσοι όροι με τα ποσοστά επιτυχίας και αποτυχίας που είχε ο κάθε Αλγόριθμος. Οι επόμενοι τρεις Πίνακες θα δείξουν τα αποτελέσματα για κάθε διαφορετικό βαθμό συσταδοποίησης που χρησιμοποιήθηκε.

	Μέσος Όρος Ποσοστού Επιτυχίας/Αποτυχίας: Συστάδες 8					
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης	
	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας
K-Mean	0%	100%	100%	0%	97%	3%
SOM	0%	100%	97%	3%	97%	3%
SOFM	75%	25%	100%	0%	97%	3%

Πίνακας 5.13. Μέσος όρος επιτυχίας και αποτυχίας για κάθε Αλγόριθμο χρησιμοποιώντας οκτώ συστάδες.

	Μέσος Όρος Ποσοστού Επιτυχίας/Αποτυχίας: Συστάδες 4					
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης	
	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας
K-Mean	0%	100%	100%	0%	92%	8%
SOM	75%	25%	98%	2%	100%	0%
SOFM	75%	25%	98%	2%	97%	3%

Πίνακας 5.14. Μέσος όρος επιτυχίας και αποτυχίας για κάθε Αλγόριθμο χρησιμοποιώντας τέσσερις συστάδες.

	Μέσος Όρος Ποσοστού Επιτυχίας/Αποτυχίας: Συστάδες 3					
	Γκρίζα Ζώνη		Μαύρη Ζώνη		Εντός Ζώνης	
	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας	% Επιτυχίας	% Αποτυχίας
K-Mean	0%	100%	100%	0%	97%	3%
SOM	0%	100%	98%	2%	97%	3%
SOFM	0%	100%	98%	2%	92%	8%

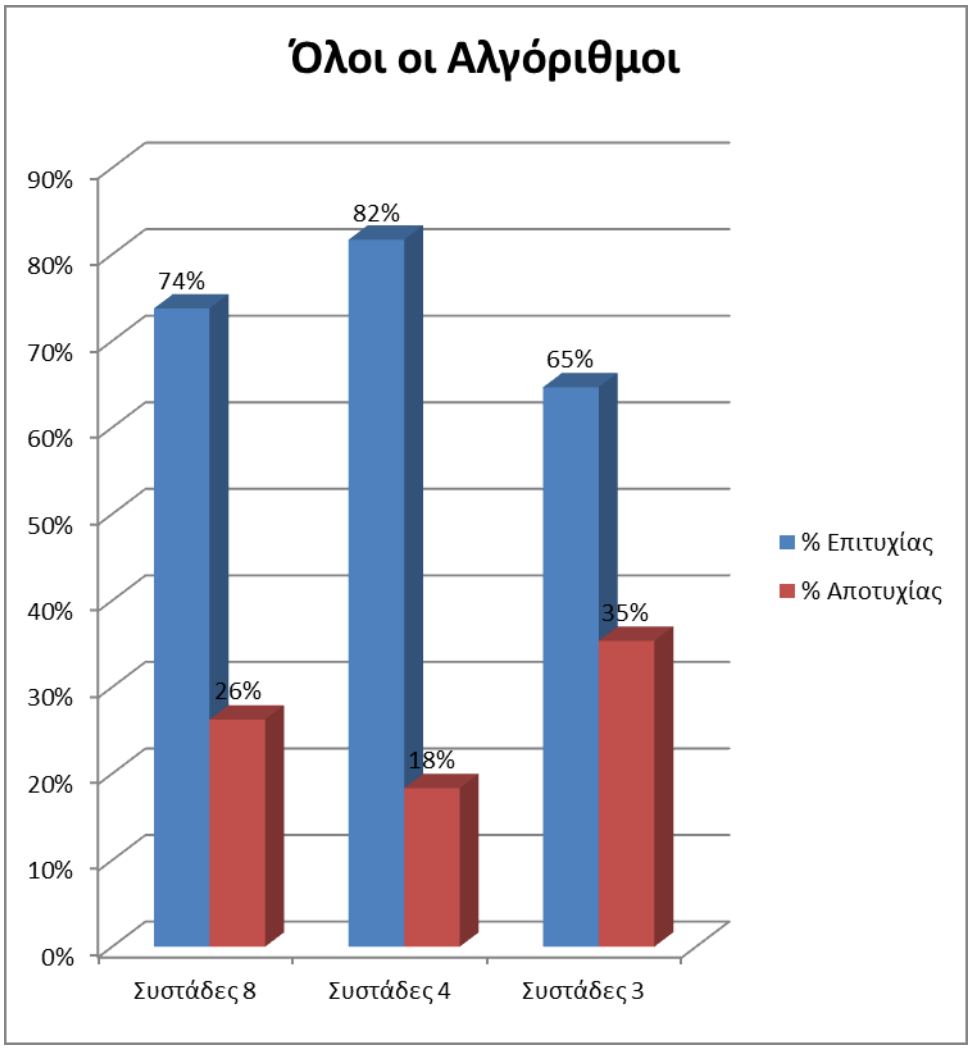
Πίνακας 5.15. Μέσος όρος επιτυχίας και αποτυχίας για κάθε Αλγόριθμο χρησιμοποιώντας τρεις συστάδες.

Από τους παραπάνω πίνακες μπορεί να γίνει αντιληπτό πως η χρήση τεσσάρων συστάδων δίνει καλύτερα αποτελέσματα έναντι των οκτώ και τριών. Στον παρακάτω Πίνακα και στην Εικόνα 5.57 φαίνεται η διαφορά τους.

Όλοι οι Αλγόριθμοι

	% Επιτυχίας	% Αποτυχίας
Συστάδες 8	74%	26%
Συστάδες 4	82%	18%
Συστάδες 3	65%	35%

Πίνακας 5.16. Μέσος όρος επιτυχίας και αποτυχίας για κάθε βαθμό συσταδοποίησης.



Εικόνα 5.57 Γραφική παράσταση ποσοστών επιτυχίας και αποτυχίας του βαθμού συσταδοποίησης στο σύνολο των Αλγορίθμων.

Ο Πίνακας 5.17, ο Πίνακας 5.18 και ο Πίνακας 5.19 παρουσιάζουν τα ποσοστά επιτυχίας και αποτυχίας που είχε κάθε Αλγόριθμος στον αντίστοιχο βαθμό συσταδοποίησης.

Συστάδες 8

	% Επιτυχίας	% Αποτυχίας
K-Mean	66%	34%
SOM	65%	35%
SOFM	91%	9%

Πίνακας 5.17. Μέσος όρος επιτυχίας και αποτυχίας κάθε Αλγορίθμου με οκτώ συστάδες.

Συστάδες 4

	% Επιτυχίας	% Αποτυχίας
K-Mean	64%	36%
SOM	91%	9%
SOFM	90%	10%

Πίνακας 5.18. Μέσος όρος επιτυχίας και αποτυχίας κάθε Αλγορίθμου με οκτώ συστάδες.

Συστάδες 3

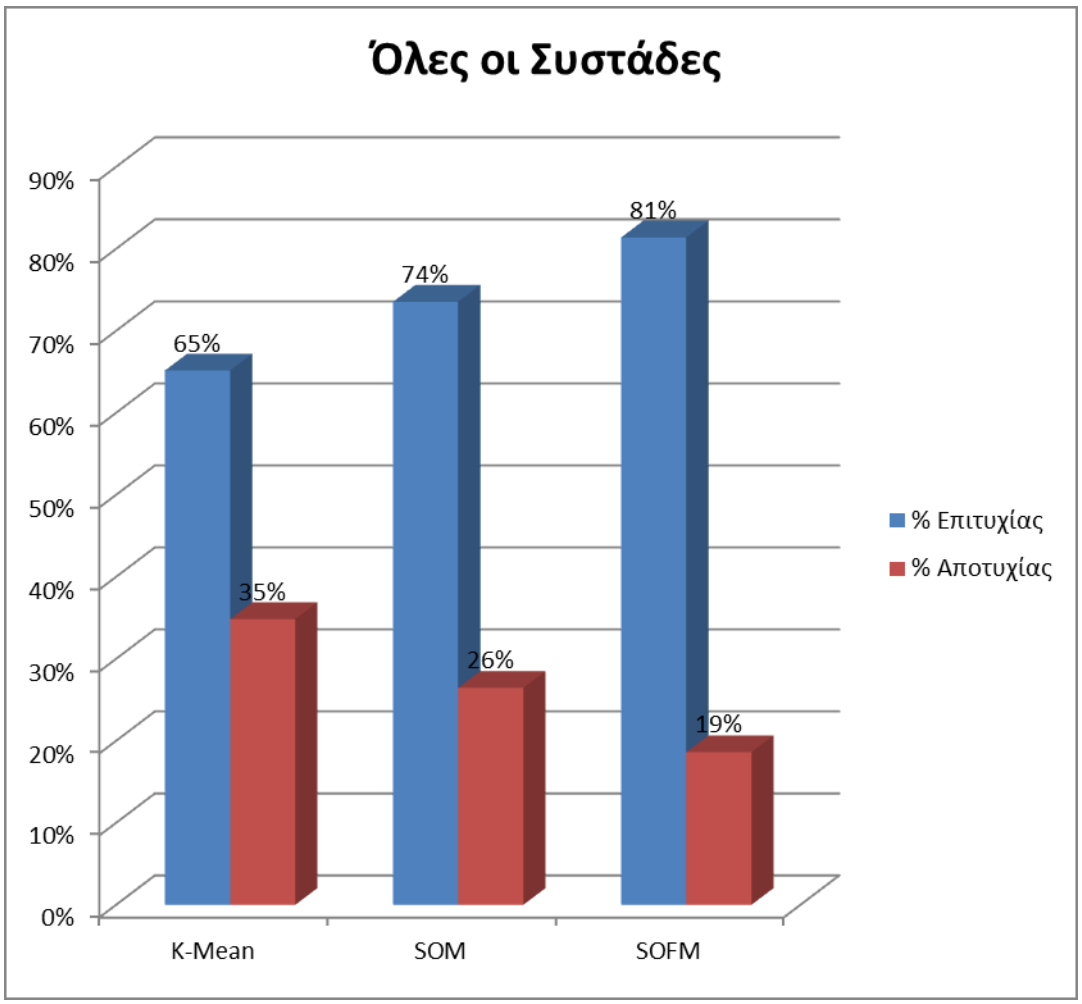
	% Επιτυχίας	% Αποτυχίας
K-Mean	66%	34%
SOM	65%	35%
SOFM	63%	37%

Πίνακας 5.19. Μέσος όρος επιτυχίας και αποτυχίας κάθε Αλγορίθμου με οκτώ συστάδες.

Από τους Αλγόριθμους K-Mean, SOM και SOFM το καλύτερο ποσοστό επιτυχίας και αποτυχίας συγκεντρώνει ο SOFM, όπως πολύ καλύτερα φαίνεται στον επόμενο πίνακα και εικόνα με τη γραφική τους παράσταση.

Όλες οι Συστάδες

	% Επιτυχίας	% Αποτυχίας
K-Mean	65%	35%
SOM	74%	26%
SOFM	81%	19%



Εικόνα 5.58 Γραφική παράσταση ποσοστών επιτυχίας και αποτυχίας των αλγορίθμων στο σύνολο των σεναρίων.

Από το σύνολο των αποτελεσμάτων που είδαμε παραπάνω μπορούμε να θεωρήσουμε πως η χρήση του Αλγόριθμου SOFM με τέσσερις συστάδες μπορεί να δώσει πολύ καλά αποτελέσματα στην ανάγκη της ανίχνευσης δόλιων κλήσεων. Απομονώνοντας το σενάριο που δίνεται στον Πίνακας 5.7 θα παρατηρήσουμε ότι ο Αλγόριθμος SOM πέτυχε το 100% της απόλυτης επιτυχίας. Αυτό το ποσοστό όμως αποτελεί ένα μεμονωμένο συμβάν που δεν επαναλαμβάνεται στο σύνολο των σεναρίων που χρησιμοποιήσαμε στην παρούσα εργασία.

Η διαφορά των ποσοστών επιτυχίας και αποτυχίας των Αλγορίθμων SOM και SOFM είναι μικρή, έναντι του K-Mean, και θα μπορούσε να ανατραπεί χρησιμοποιώντας άλλα σενάρια κι άλλα προφίλ σε μία άλλη εργασία. Στο επόμενο και τελευταίο κεφάλαιο δίνεται μία σύνοψη της εργασίας αυτής και κάποιες μελλοντικές σκέψεις για την εξέλιξή της.

Κεφάλαιο 6
Επίλογος

Ο άνθρωπος ποτέ δε σταμάτησε την αναζήτηση για κάτι καλύτερο από αυτό που έχει. Ο χώρος των επιστημών συνεχώς προβληματίζεται και δίνει νέες λύσεις ή νέες απαντήσεις στα ζητήματα που θέτει. Σε μας δίνεται η δυνατότητα να απολαμβάνουμε αυτές τις νέες λύσεις. Παλαιότερα οι άνθρωποι επικοινωνούσαν με νοήματα, στη συνέχεια με τη φωνή τους και σιγά σιγά επέκτειναν τον τρόπο της επικοινωνίας ως προς την απόσταση, με τις φωτιές, το φως του ήλιου ή και τον ήχο για άμεση επικοινωνία. Το μεγάλο άλμα γίνεται με την εφεύρεση του τηλεφώνου. Η χρήση του όμως από εμάς τους πολίτες έχει ένα κόστος. Κάποιοι συμπολίτες μας μη μπορώντας να ανταπεξέλθουν στο κόστος αυτό ή από άλλους προσωπικούς τους λόγους, οδηγούνται στην παράνομη χρήση αυτής της υπηρεσίας. Συνεχώς βρίσκουν νέους τρόπους και διεισδύουν στα τηλεφωνικά συστήματα.

Με την παρούσα εργασία προσπαθήσαμε να προχωρήσουμε ένα βήμα στην επισήμανση της παράνομης ή δόλιας χρήσης των VoIP συστημάτων. Επιτακτική ανάγκη στην εργασία αυτή ήταν η χρήση CDRs. Η έλλειψή τους οδήγησε στη δημιουργία μίας νέας εφαρμογής που θα προσομοιώνει κλήσεις ενός παρόχου με Πανελλαδική εμβέλεια. Η κύρια εφαρμογή εισάγει τα δεδομένα που παράγει η γεννήτρια τυχαίων κλήσεων κι εξάγει πληροφορίες που τις ανακοινώνει στο χρήστη της.

Για τη δημιουργία των εφαρμογών χρησιμοποιήθηκε μία νέα σχετικά γλώσσα (από το 2000 κι έπειτα εξελίσσεται με γοργό ρυθμό). Η περιγραφική γλώσσα Python, η οποία είναι δυνατό να στηρίξει τις εφαρμογές της σε πολλά λειτουργικά συστήματα.

Η εφαρμογή «Ανίχνευση Δόλιων Κλήσεων» ενσωματώνει τρεις Αλγορίθμους που χρησιμοποιούνται για τη συσταδοποίηση. Οι Αλγόριθμοι αυτοί είναι οι K-Mean, Self-Organizing Maps και Self-Organizing Feature Maps. Η εργασία αυτή κάνει μία σύγκριση μεταξύ τους στην αποδοτικότητα που έχουν να ανιχνεύουν δόλιες κλήσεις. Βελτίωση της εφαρμογής μπορεί να αποτελέσει η προσθήκη νέων δυνατοτήτων στις ρυθμίσεις της χρήσης των αλγορίθμων, όπως είναι για παράδειγμα ένα πλαίσιο που θα ορίζει το πλήθος των συστάδων που θα χρησιμοποιούνται στην ανίχνευση δόλιων κλήσεων.

Σε μελλοντική εργασία θα μπορούσαν επίσης να συμπεριληφθούν κι άλλοι αλγόριθμοι, όπως ο Ιεραρχικός ή ο Fuzzy c-means.

Επίσης, θέτοντας νέα σενάρια σχετικά με την αλλαγή του προφίλ των κλήσεων σε κάποιους ακριβούς προορισμούς, ή ακόμη εισάγοντας τα προφίλ των χρηστών – πελατών των υπηρεσιών VoIP, θα μπορούσαμε να δούμε την αντίδραση του μηχανισμού ανίχνευσης δόλιων κλήσεων.

Βιβλιογραφία

[01] S. Hofbauer, J. Beckers, G. Quirchmayr. «A Lightweight Privacy Preserving Approach for Analyzing Communication Records to Prevent VoIP Attacks Using Toll Fraud as an Example» Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, vol., no., pp.992,997, 25-27 June 2012.

[02] J. Rosenberg, H. Schulzrinne, G. Camarillo, et al. «RFC 3261-SIP: Session Initiation Protocol». 2002, [Online]. Ανακτημένο στις 11-05-14 στο διαδικτυακό τόπο http://www.ietf.org/rfc/rfc3261.txt.

[03] Lu Zheng, Peng Taoxin. «The VoIP intrusion detection through a LVQ-based neural network». International Conference for Internet Technology and Secured Transactions, CFP0981I-CDR, 1-6, 9-12 Nov. 2009.

[04] D. Hoffstadt, E. Rathgeb, M. Liebig, R. Meister, Y. Rebahi, T.Q. Thanh. «A comprehensive framework for detecting and preventing VoIP fraud and misuse » Computing, Networking and Communications (ICNC), 2014 International Conference on, vol., no., pp.807, 813, 3-6 Feb. 2014.

[05] AD. Keromytis. «A Comprehensive Survey of Voice over IP Security Research» Communications Surveys & Tutorials, IEEE, vol.14, no.2, pp.514, 537, Second Quarter 2012.

[06] B. Goode. «Voice Over Internet Protocol (VOIP) ». Proc. IEEE 90, 1495-1517, Sept. 2002.

[07] H. Schulzrinne. Received from http://www.cs.columbia.edu/~hgs/internet/.

[08] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson. «RTP: A transport protocol for realtime applications». IETF RFC 1889, 1996. Received from ftp://ftp.ietf.org/rfc/rfc1889.txt.

[09] J.F. Kurose, K.W. Ross. «Computer Networking: a Top-down Approach Featuring the Internet Sixth Edition». Ανακτημένο στις 11 Μαΐου 2014 από το δικτυακό τόπο http://www.pdfiles.com/pdf/files/English/Networking/Computer_Networking_A_Top-Down_Approach.pdf.

[10] ITU-T Recommendation H.323. «Packet-based multimedia communications systems». International Telecommunication Union, 1997.

[11] M. Arango, A. Dugan, I. Elliott, C. Huitema, S. Pickett. «Media gateway control protocol (MGCP) Version 1.0». IETF RFC 2705, 1999.

[12] N. Greene, M. Ramalho, B. Rosen. «Media gateway control protocol architecture and requirements». IETF RFC 2805, 2000.

[13] F. Cuervo, N. Greene, A. Rayhan, C. Huitema, B. Rosen, J. Segers. «Megaco Protocol Version 1.0». IETF RFC 3015, 2000.

[14] J. Mitola III. «Cognitive Radio An Integrated Agent Architecture for Software Defined Radio». Dissertation, Royal Institute of Technology (KTH), May 2000.

[15] S. Haykin. «Cognitive radio: brain-empowered wireless communications». IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp. 201–202, Feb. 2005.

[16] ITU-T. «Recommendation H.225.0–Call signalling protocols and media stream packetization for packet-based multimedia communication systems». December 2009.

[17] ITU-T. «Recommendation H.245–Control protocol for multimedia communication». May 2011.

[18] ITU-T: Recommendation H.323–Packet-based multimedia communications systems, December 2009.

[19] CISCO. « Understanding MGCP Interactions with Cisco CallManager». Ανακτημένο στις 14 Μαρτίου 2014 από το δικτυακό τόπο http://www.cisco.com/c/en/us/support/docs/voice/media-gateway-control-protocol-mgcp/44130-understanding-mgcp.html.

[20] ITU-T, Recommendation H.248.1–Gateway control protocol: Version 3, March 2013.

[21] E. Belmekki, B. Raouyane, A. Belmekki, M. Bellafkih. «Secure SIP signalling service in IMS network». Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on, vol., no., pp.1,7, 7-8 May 2014.

[22] Choi Seung-Han, Min Sun-Ho, Seo Chang-Ho. «The multipurpose IP phone architecture for home service». Advanced Communication Technology (ICACT), 2012 14th International Conference on, vol., no., pp.637,641, 19-22 Feb. 2012.

[23] P. Thermos, A. Takanen. «Securing VoIP Networks: Threats, Vulnerabilities, and Countermeasures». Pearson Education, Inc., 2008. Received from http://www.safaribooksonline.com/library/view/securing-voip-networks/9780321437341/ch03.html#ch03fn02.

[24] Μ. Γαβριηλίδου, Π. Λαμπροπούλου, Κ. Αγγελάκος. «Ερμηνευτικό Λεξικό Νέας Ελληνικής. Α', Β', Γ' Γυμνασίου». Έκδοση Α, Οργανισμός Εκδόσεως Διδακτικών Βιβλίων, Αθήνα, 2012.

[25] M. Gruber, C. Schanes, F. Fankhauser, T. Grechenig. «Voice calls for free: How the black market establishes free phone calls — Trapped and uncovered by a VoIP Honeynet». Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on , vol., no., pp.205,212, 10-12 July 2013

[26] S. Rao Vallabhaneni. «Wiley CIA Exam Review 2013, Internal Audit Knowledge Elements». John Wiley & Sons, Inc., Canada, 2013, Ανακτήθηκε στις Απριλίου 2014 από τον δικτυακό τόπο http://books.google.gr/books?id=vSQiyrlbDIMC&printsec=frontcover

[27] Communications Fraud Control Association. «2013 Global Fraud Loss Survey». Ανακτήθηκε στις 27 Απριλίου 2014 από τον δικτυακό τόπο http://www.cfca.org/pdf/survey/CFCA2013GlobalFraudLossSurvey-pressrelease.pdf.

[28] Google. «What is SIPVicious tool suite?». Ανακτήθηκε στις 27 Απριλίου 2014 από τον δικτυακό τόπο https://code.google.com/p/sipvicious/

[29] Z. Avdagic, A. Midzic. «The effects of combined application of SOM, ANFIS and Subtractive Clustering in detecting intrusions in computer networks». Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on , vol., no., pp.1435,1440, 26-30 May 2014.

[30] U. Bashir, M. Chachoo. «Intrusion detection and prevention system: Challenges & opportunities». Computing for Sustainable Global Development (INDIACom), 2014 International Conference on , vol., no., pp.806,809, 5-7 March 2014.

[31] United States Attorney's Office. «Two Fraudulent Telephone Service Wholesalers Sentenced to Prison for $4.4 Million VoIP Fraud Scheme». Federal Bureau of Investigation (FBI), 15Μαΐου 2012. Ανακτήθηκε στις 29 Μαρτίου 2014 από τον δικτυακό τόπο http://www.fbi.gov/newark/press-releases/2012/two-fraudulent-telephone-service-wholesalers-sentenced-to-prison-for-4.4-million-voip-fraud-scheme.

[32] R. Wilonsky. « Three ringleaders of local “cybercrime conspiracy” given long prison sentences, forced to pay back many millions». The Dallas Morning News Inc., 25 Μαΐου 2012. Ανακτήθηκε στις 29 Μαρτίου 2014 από τον δικτυακό τόπο http://crimeblog.dallasnews.com/2012/05/three-ringleaders-of-local-cybercrime-conspiracy-given-long-prison-sentences-forced-to-pay-back-many-millions.html/

[33] K. Yufeng, L. Chang-Tien, S. Sirwongwattana, H. Yo-Ping. «Survey of fraud detection techniques». Networking, Sensing and Control, 2004 IEEE International Conference on , vol.2, no., pp.749,754 Vol.2, 2004.

[34] M.A. Bihina Bella, J.H.P. Eloff, M.S. Olivier. «A fraud management system architecture for next-generation networks». Forensic Science International, Volume 185, Issues 1–3, Pages 51-58, ISSN 0379-0738, 10 March 2009,

[35] U. Murad, G. Pinkas. «Unsupervised Profiling for Identifying Superimposed Fraud». Principles of Data Mining and Knowledge Discovery, Springer Berlin / Heidelberg, Pages 251-261, 08 June, 2004.

[36] D.E. Denning. «An Intrusion-Detection Model». Software Engineering, IEEE Transactions on , vol.SE-13, no.2, pp.222,232, Feb. 1987

[37] P.-N. Tan, M. Steinbach, V. Kumar, «Introduction to Data Mining», Addison Wesley, 2006

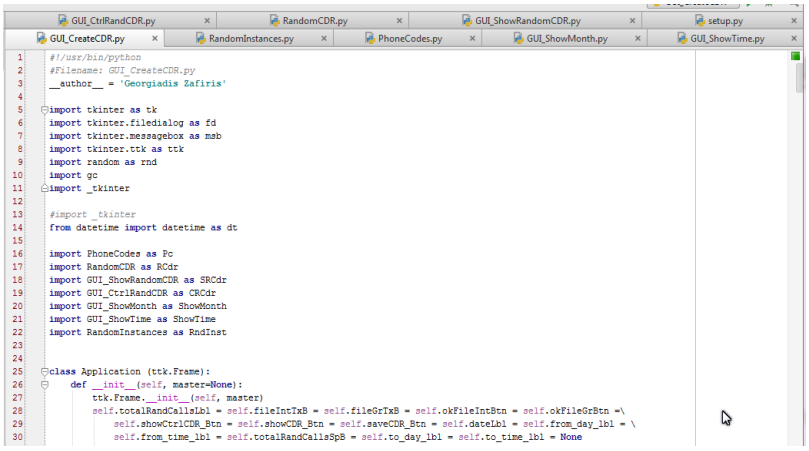
[38] ITU-T: Recommendation E.164 Dialling Procedures, Geneva, 2011,

[39] CISCO, «Cisco TelePresence Conferencing Call Detail Records – File Format Reference Guide», D14663.13, June , 2014. Ανακτήθηκε στις 01 Ιουλίου 2014 από τον δικτυακό τόπο http://www.cisco.com/c/dam/en/us/td/docs/telepresence/infrastructure/mcu/admin_guide/cisco_telepresence_infrastructure_cdr_reference_guide.pdf

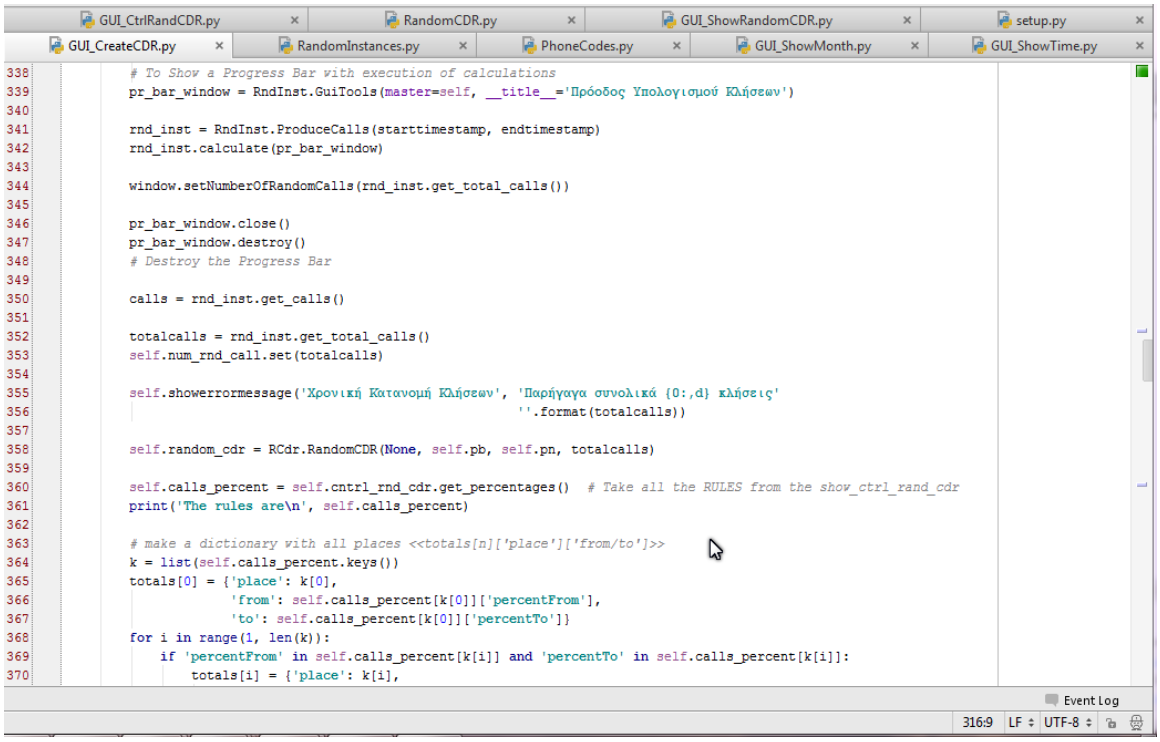
Παράρτημα Α
Τμήματα Κώδικα Εφαρμογής

Στο Παράρτημα Α θα δοθούν επιμέρους τμήματα του κώδικα της περιγραφικής γλώσσας Python στην έκδοση 3.3 με τα οποία υλοποιήθηκαν οι εφαρμογές «Παραγωγή Αρχείου Τυχαίων Κλήσεων» και «Ανίχνευση Δόλιων Κλήσεων»

Α.1 GUI (Graphical User Interface) της Εφαρμογής «Παραγωγή Αρχείου Τυχαίων Κλήσεων»

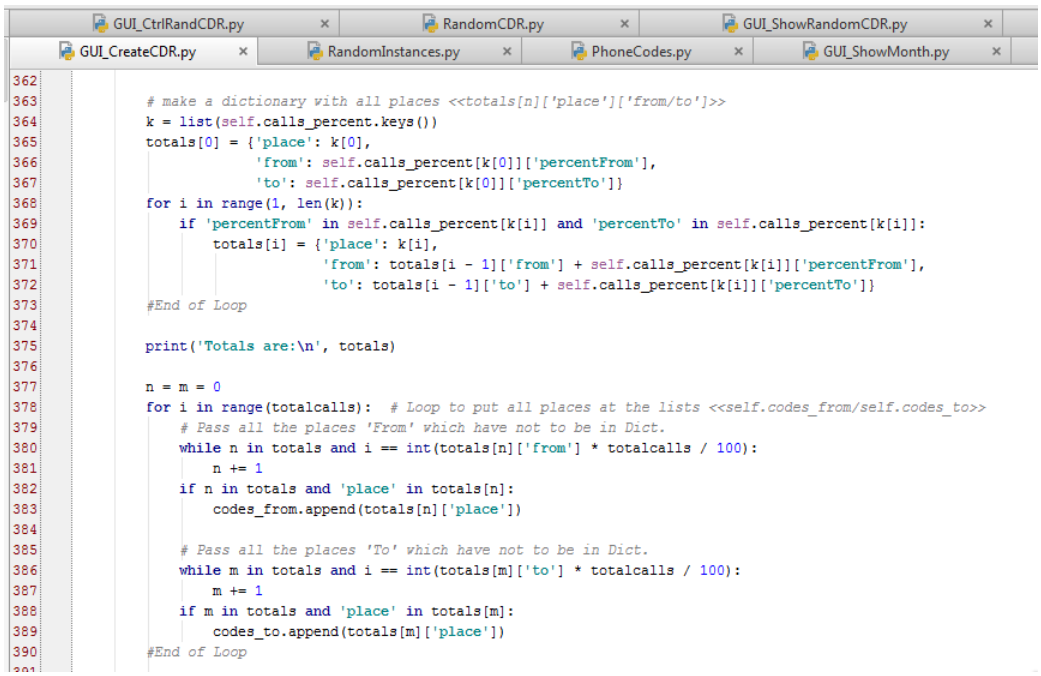


Εικόνα Α.1 Σύνδεση της εφαρμογής με εξωτερικές βιβλιοθήκες για τη δημιουργία GUI.

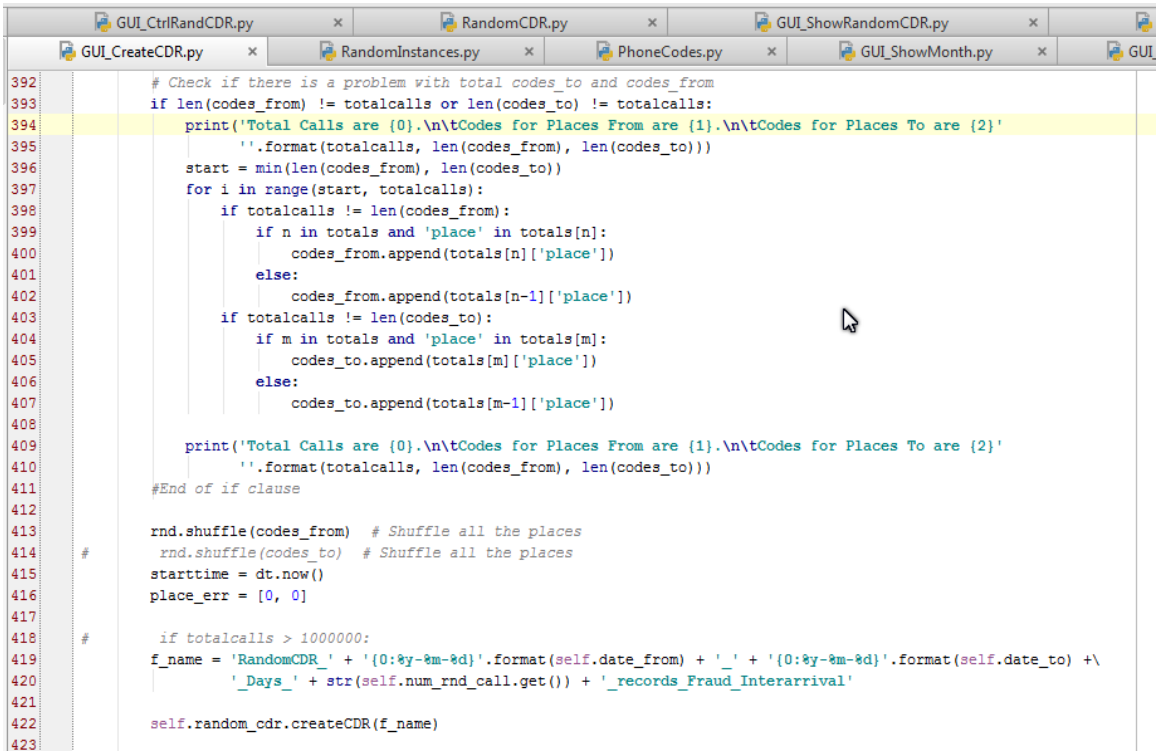


Εικόνα Α.2 Κλήση της διαδικασίας «calculate» Γραμμή 342, για τη δημιουργία κλήσεων σε τυχαίους χρόνους.

Στην παραπάνω εικόνα καλείται η διαδικασία «calculate» και στη συνέχεια στη γραμμή 360 και 361 καλούνται κι εμφανίζονται στη γραμμή εντολών τα ποσοστά των κλήσεων, προς και από περιοχές της Ελλάδος και του Εξωτερικού, όπως τελικά τα έχει διαμορφώσει ο χρήστης της εφαρμογής

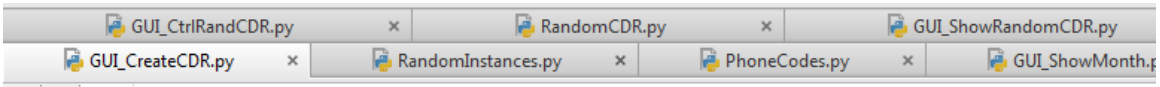


Εικόνα Α.3 Αντιστοίχηση των τηλεφωνικών κλήσεων στα ποσοστά των περιοχών που όρισε ο χρήστης




```
392         # Check if there is a problem with total codes_to and codes_from
393         if len(codes_from) != totalcalls or len(codes_to) != totalcalls:
394             print('Total Calls are {0}.\n\tCodes for Places From are {1}.\n\tCodes for Places To are {2}'
395                   ''.format(totalcalls, len(codes_from), len(codes_to)))
396             start = min(len(codes_from), len(codes_to))
397             for i in range(start, totalcalls):
398                 if totalcalls != len(codes_from):
399                     if n in totals and 'place' in totals[n]:
400                         codes_from.append(totals[n]['place'])
401                     else:
402                         codes_from.append(totals[n-1]['place'])
403                 if totalcalls != len(codes_to):
404                     if m in totals and 'place' in totals[m]:
405                         codes_to.append(totals[m]['place'])
406                     else:
407                         codes_to.append(totals[m-1]['place'])
408 
409             print('Total Calls are {0}.\n\tCodes for Places From are {1}.\n\tCodes for Places To are {2}'
410                   ''.format(totalcalls, len(codes_from), len(codes_to)))
411         #End of if clause
412 
413         rnd.shuffle(codes_from)  # Shuffle all the places
414 #        rnd.shuffle(codes_to)  # Shuffle all the places
415         starttime = dt.now()
416         place_err = [0, 0]
417 
418 #        if totalcalls > 1000000:
419         f_name = 'RandomCDR_' + '{0:%y-%m-%d}'.format(self.date_from) + '_' + '{0:%y-%m-%d}'.format(self.date_to) +\
420                  '_Days_' + str(self.num_rnd_call.get()) + '_records_Fraud_Interarrival'
421 
422         self.random_cdr.createCDR(f_name)
423 
```

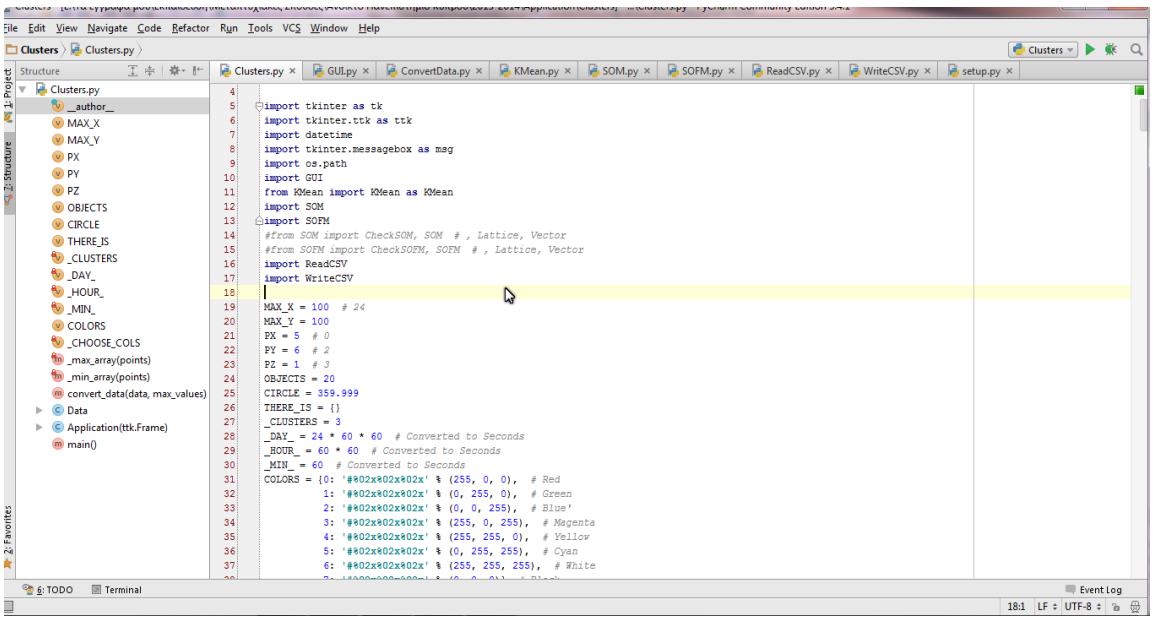
Εικόνα Α.4 Έλεγχος για την ομαλή κατανομή των κλήσεων, τυχαία ανακατανομή των καλούντων τηλεφωνικών κλήσεων και ορισμός του ονόματος με το οποίο θα αποθηκευθούν τα CDRs.




```
422         self.random_cdr.createCDR(f_name)
423 
424         rows_to_write = {}
425         for i in range(totalcalls):
426 #            aa = codes_from
427 #            bb = codes_to
428 #            cc = i
429 #            dd = totalcalls
430 #            ee = calls[i]
431 #            ff = window
432 #            gg = starttime
433 #            hh = place_err
434             place_err = self.produce_random_call(codes_from, codes_to, i, totalcalls, calls[i],
435                                                  window, starttime, place_err)
436             rows_to_write[i % 100] = calls[i]['Call']
437             calls.pop(i)
438             if i % 100 == 99:  # Used 99 to write the first 100 records starting from 0 to 99
439                 self.random_cdr.appendCDR(f_name, rows_to_write)
440                 rows_to_write = {}
441 
442         self.random_cdr.appendCDR(f_name, rows_to_write)  # Write the last data
443         #End of Loop
444 
445         self.showerrormessage('Αποθήκευση Κλήσεων', 'Συνολικά έχουν αποθηκευθεί επιτυχώς '
446                                                     '{0:,d} τηλεφωνικές κλήσεις.'.format(totalcalls))
```

Εικόνα Α.5 Βρόγχος επανάληψης για την προσθήκη τηλεφωνικών αριθμών, Γραμμή 434 και 435, και αποθήκευσή τους σε αρχείο, γραμμή 439

A.2 GUI (Graphical User Interface) της Εφαρμογής «Ανίχνευση Δόλιων Κλήσεων»



Εικόνα A.6 Σύνδεση της εφαρμογής με εξωτερικές βιβλιοθήκες για τη δημιουργία GUI.

```
1213
1214
1215    def main():
1216        title = 'Ανίχνευση Δόλιων Κλήσεων'
1217        root = tk.Tk()
1218
1219        app = Application(root, title)
1220
1221        app.mainloop()
1222    # End of main
1223
1224    if __name__ == '__main__':
1225        main()
1226    # End of if
```

Εικόνα A.7 Δημιουργία αντικειμένου που βλέπει στην κύρια εφαρμογή. Γραμμή 1,219.

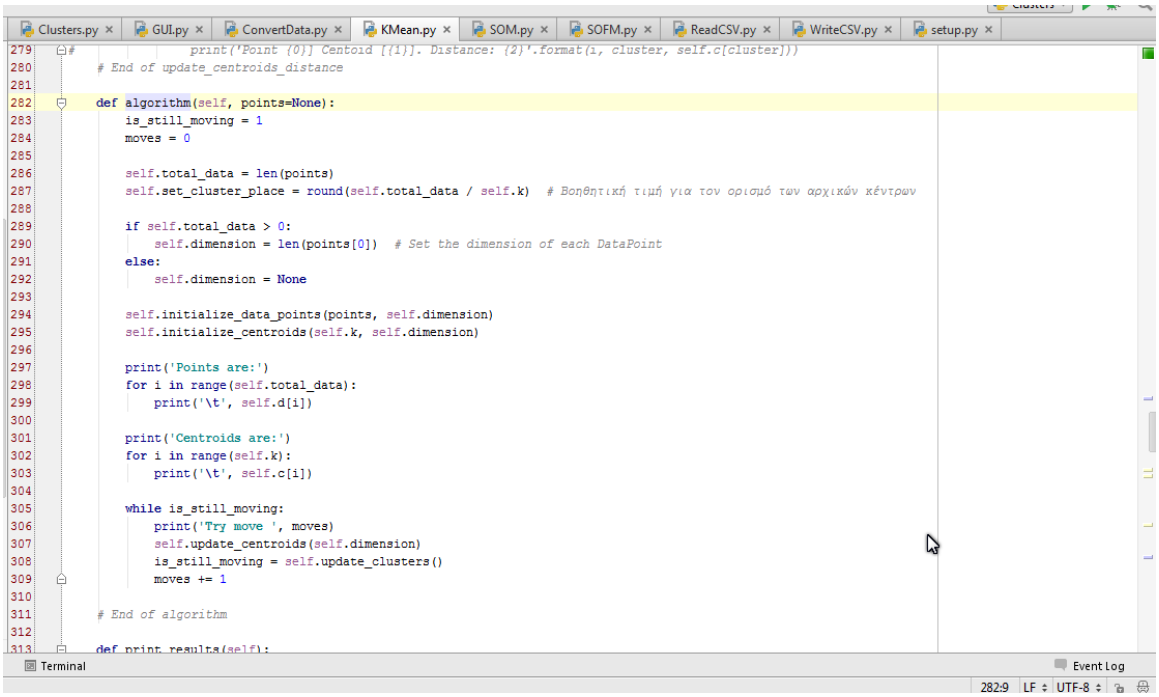
```
class Application(ttk.Frame):
    DIMENSION_X = 800
    DIMENSION_Y = 600
    _WIDTH = 800
    _HEIGHT = 550

    def __init__(self, parent, title):
        ttk.Frame.__init__(self, parent)

        self.img = tk.PhotoImage(file='.'+'\\Images\\university_40.gif')  # reference PhotoImage in local variable
        self.img_thief = tk.PhotoImage(file='.'+'\\Images\\stop_thief_small.gif')
        self.img_detect = tk.PhotoImage(file='.'+'\\Images\\finding_fraud.gif')
        self.img_normal = tk.PhotoImage(file='.'+'\\Images\\university_alpha.gif')
#        scale_w = new_width/old_width
#        scale_h = new_height/old_height
#        photoImg.zoom(scale_w, scale_h)

        s = ttk.Style()
        s.configure('Image.TLabel', image=self.img)
        s.configure('Image_Thief.TLabel', image=self.img_thief)
        s.configure('Image.TButton', image=self.img)
        s.configure('Image.TNotebook', image=self.img)
        s.configure('Image.TPanedWindow', image=self.img)

        root_size = tuple([self.DIMENSION_X, self.DIMENSION_Y])
        x = round(parent.winfo_screenwidth()/2 - root_size[0]/2)
        y = round(parent.winfo_screenheight()/2 - root_size[1]/2)
        parent.geometry('{0}x{1}+{2}+{3}'.format(str(root_size[0]), str(root_size[1]), str(x), str(y)))

        parent.update()

        self.parent = parent
```

Εικόνα Α.8 Δομητής της κύριας εφαρμογής. Ορίζονται τα χαρακτηριστικά του GUI.

```
    def create_widgets(self):
        #Hold place at the Frame for the objects
        for i in range(OBJECTS):
            self.top.rowconfigure(i, weight=1)
            self.top.columnconfigure(i, weight=1)

            self.rowconfigure(i, weight=1)
            self.columnconfigure(i, weight=1)
        #End of the Loop

        top_frame = ttk.PanedWindow(self)  # , width=1024, height=600)

        top_frame.rowconfigure(0, weight=1)
        top_frame.columnconfigure(0, weight=1)

        self.tab_frame = ttk.Notebook(top_frame, style='Image.TNotebook')

        self.tab_frame.add(self.pane_data(), text='Χρήση Δεδομένων')
        self.tab_frame.add(self.pane_train(), text='Εκπαίδευση')
        self.tab_frame.add(self.pane_show_results(), text='Εμφάνιση Αποτελεσμάτων')

        return_btn = ttk.Button(top_frame, text="Εξοδος", command=self.exit_application)

        self.tab_frame.grid(row=0, column=0, sticky=tk.S+tk.N+tk.W+tk.E)
        return_btn.grid(row=1, column=0, sticky=tk.N+tk.S)  # columnspan=3,
        top_frame.grid(row=0, column=0, sticky=tk.N+tk.S+tk.W+tk.E)

        top_frame.update()
#        self.update()
    # End of create_widgets
```

Εικόνα Α.9 Διαδικασία (Procedure) που δημιουργεί τα αντικείμενα του GUI με κλήσεις σε άλλες διαδικασίες εντός της κύριας κλάσης (Class Application).

A.2 Αλγόριθμος Ανίχνευσης Κλήσεων Fraud



Εικόνα Α.10 Διαδικασία «algorithm» του αρχείου K-Mean.py που εκτελεί τα βήματα εκπαίδευσης σύμφωνα με τον Αλγόριθμο K-Mean.

Στον Αλγόριθμο K-Mean χρησιμοποιείται η ευκλείδεια απόσταση για την ομαδοποίηση των δεδομένων σε συστάδες. Τμήμα του κώδικα δίνεται στην επόμενη εικόνα.

```
187     @staticmethod
188     def distance(p1=DEFAULT_POINT_2D, p2=DEFAULT_POINT_2D):
189         if len(p1) == len(p2):
190             total = 0
191             for i in range(len(p1)):
192 #                 print(p1, p2)
193                 total += pow(p1[i]-p2[i], 2)
194             d = math.sqrt(total)
195         else:
196             d = -1
197
198         return float(d)
199
```

Εικόνα Α.11 Διαδικασία για την εύρεση της απόστασης δύο σημείων n διαστάσεων με την ευκλείδεια μέθοδο.

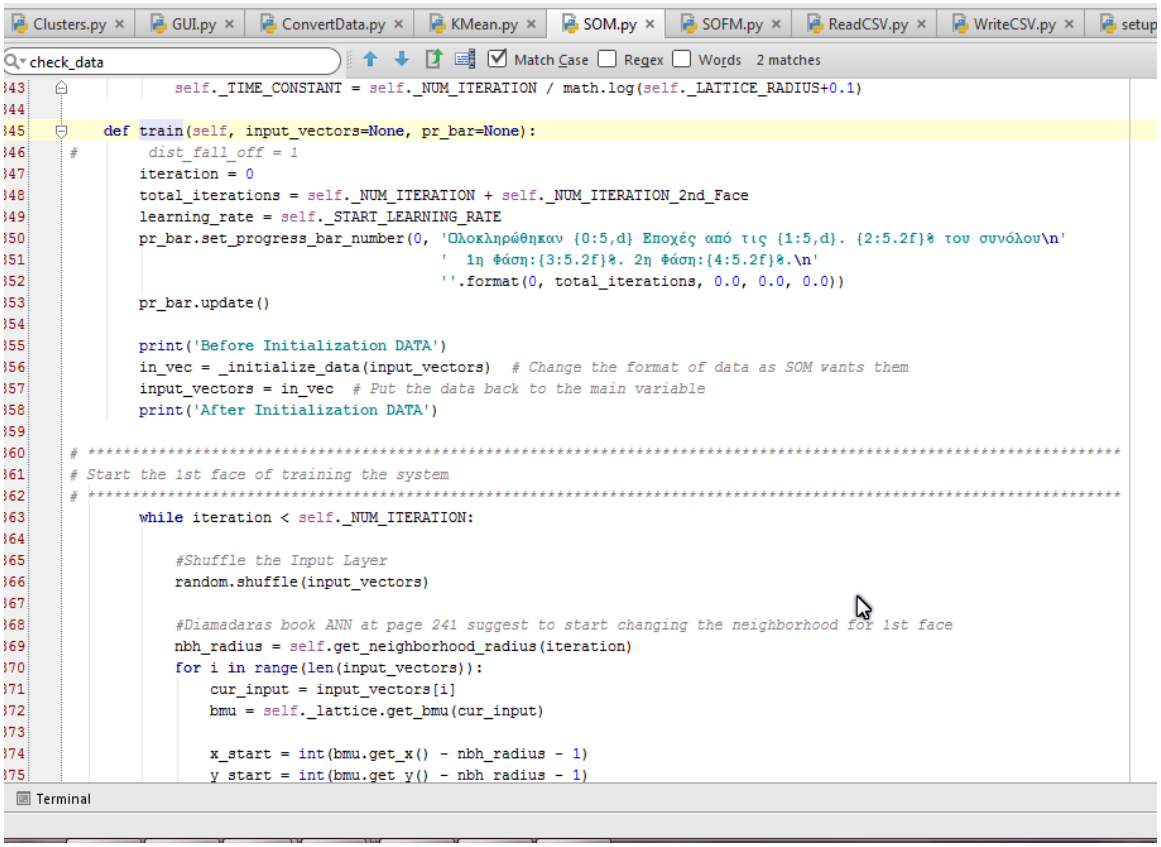
```
244     def update_clusters(self):
245         is_still_moving = False
246 
247         for i in range(self.total_data):  # Ορίζει την ομάδα στην οποία θα ανήκει το κάθε point
248             # Ορίζει ως αρχική μικρότερη απόσταση του κάθε point την 1η απόσταση από το 1ο κέντρο
249             temp_distance = self.distance(self.d[i].get_point(), self.c[0].get_centroid().get_point())
250             best_min = temp_distance
251             #best_min = sys.float_info.max
252             current_cluster = 0
253 
254             for j in range(1, self.k):  # Ψάχνει να βρει τη μικρότερη απόσταση από τα υπόλοιπα κέντρα
255                 temp_distance = self.distance(self.d[i].get_point(), self.c[j].get_centroid().get_point())
256                 # Εάν τη βρει ορίζει ως νέα ομάδα του point την ομάδα στην οποία βρίσκεται το κέντρο
257                 if best_min > temp_distance:
258                     best_min = temp_distance
259                     current_cluster = j
260             # Ελέγχει εάν το point ανήκει σε μία ομάδα ή έχει αλλάξει ομάδα
261             if (self.d[i].get_cluster() == -1) or (self.d[i].get_cluster() != current_cluster):
262                 self.d[i].set_cluster(current_cluster)  # Ορίζει τη νέα ομάδα στην οποία θα ανήκει το point
263 #                print('Point [{0}]. {1}'.format(i, self.d[i]))
264                 is_still_moving = True  # Ενημερώνει το σύστημα πως έγινε μία αλλαγή
265 
266         return is_still_moving
267     # End of update_clusters
268 
```

Εικόνα Α.12 Διαδικασία ελέγχου για εναλλαγή των εισόδων στις συστάδες.

Για κάθε προσπάθεια στην οποία δεν έχουν σταθεροποιηθεί τα δεδομένα εισόδου ο Αλγόριθμος K-Mean ενημερώνει τα κέντρα των συστάδων σύμφωνα με τον κώδικα που δίνεται παρακάτω.

```
225     def update_centroids(self, dimension=None):
226         total_point = []
227         for i in range(dimension):
228             total_point.append(0)
229         total_in_cluster = 0
230 
231         for i in range(self.k):  # Ενημερώνει όλες τις ομάδες (clusters)
232             for j in range(self.total_data):  # Ελέγχει όλα τα points
233                 if self.d[j].get_cluster() == i:  # εάν ανήκουν στην τρέχουσα ομάδα
234                     for k in range(dimension):
235                         total_point[k] += self.d[j].get_point_index(k)  # Αθροίζει όλα τα σημεία των points
236                     total_in_cluster += 1  # δίνει το σύνολο των points εντός της ομάδας
237             if total_in_cluster > 0:
238                 new_data = []
239                 for j in range(dimension):
240                     new_data.append(total_point[j] / total_in_cluster)  # Ορίζε το σημείο της νέας θέσης του κέντρου
241                 self.c[i].set_centroid(DataPoint(dimension, new_data))  # Ορίζε τη νεα θέση του κέντρου της ομάδας
242     # End of update_centroids
243 
```

Εικόνα Α.13 Διαδικασία ενημέρωσης των κέντρων των συστάδων του K-Mean Αλγόριθμου.



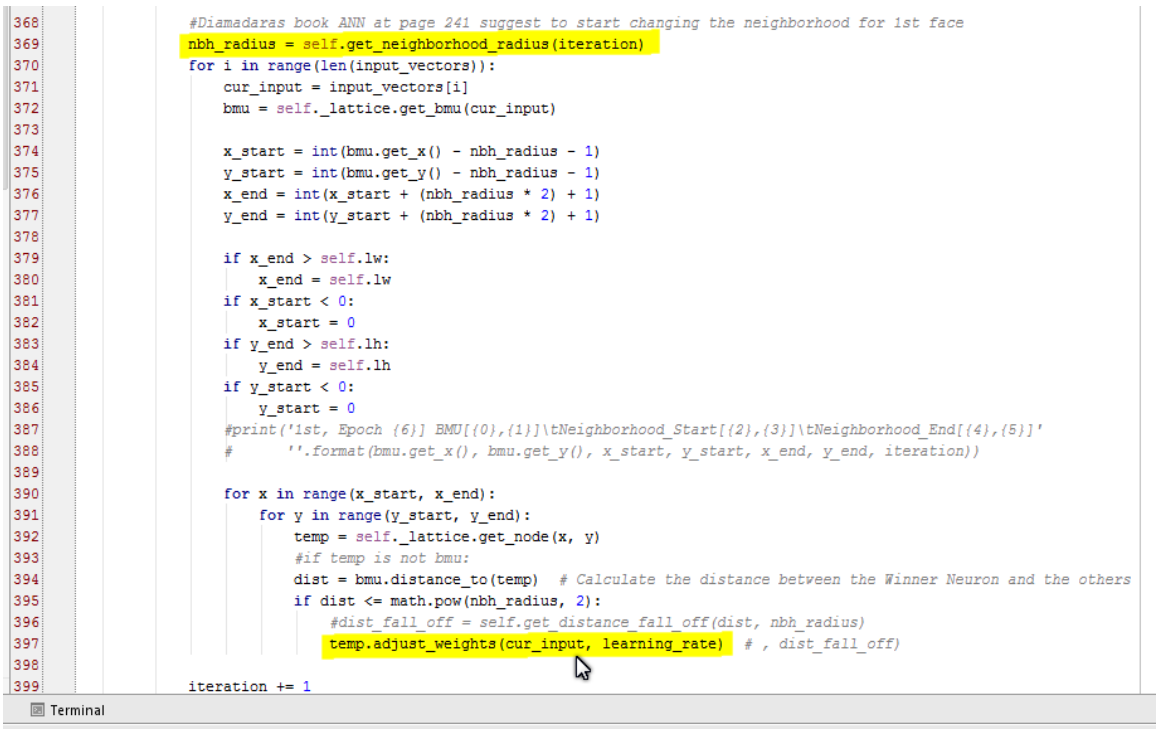
Εικόνα Α.14 Τμήμα του κώδικα εκπαίδευσης του Αλγόριθμου SOM

Σύμφωνα με το βιβλίο Τεχνητά νευρωνικά δίκτυα του κου Διαμαντάρα Κωνσταντίνου, καθηγητή του Τ.Ε.Ι. Θεσσαλονίκης, ο Αλγόριθμος SOM οφείλει να έχει δύο φάσεις εκπαίδευσης. Στην πρώτη φάση το σύστημα εκπαιδεύεται χωρίς να αλλάζει ο ρυθμός εκπαίδευσης αλλά να αλλάζει η περιοχή της γειτονιάς που θα εκπαιδεύεται. Στη δεύτερη φάση προτείνεται από το βιβλίο η γειτονιά να μη συμμετέχει στη φάση της εκπαίδευσης αλλά να αλλάζει ο ρυθμός της εκπαίδευσης με φθίνουσα σειρά.

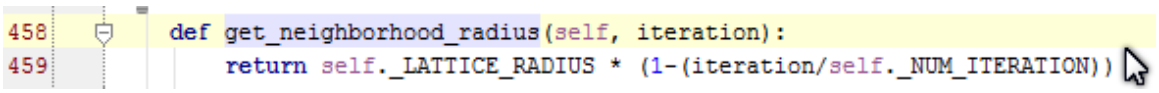
Για την πρώτη φάση το βιβλίο αναφέρει το πλήθος των εποχών να μην είναι παραπάνω από 200, σε αντίθεση με τη δεύτερη φάση όπου προτείνεται το σύνολο των εποχών να είναι το γινόμενο του συνόλου των Νευρώνων του εξωτερικού επιπέδου με το 500.

Όπως και στον αλγόριθμο K-Mean έτσι κι εδώ για την απόσταση μεταξύ των διανυσμάτων της εισόδου και των Νευρώνων, αλλά και των Νευρώνων μεταξύ τους, χρησιμοποιείται η ευκλείδεια απόσταση.

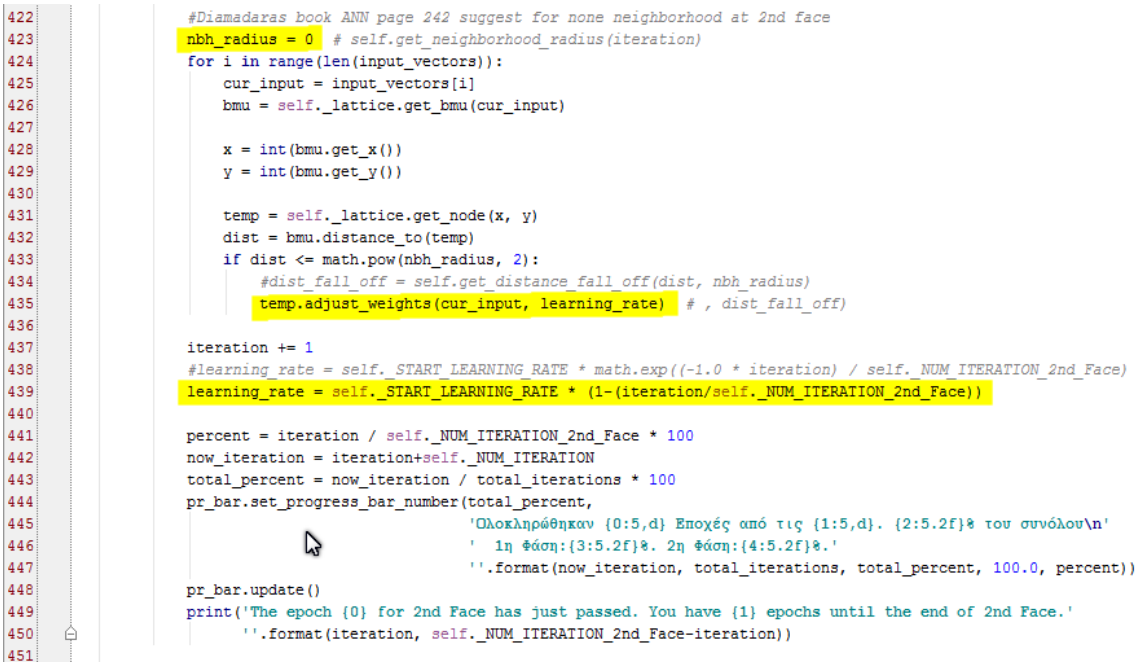
Η διαφορά του Αλγόριθμου SOM από τον SOFM έγκειται στον τρόπο υπολογισμού του ρυθμού εκπαίδευσης και της εκπαίδευσης των βαρών κάθε Νευρώνα. Οι διαφορές τους δίνονται στις παρακάτω εικόνες με τα αντίστοιχα τμήματα του κώδικα.



Εικόνα Α.15 Τμήμα του κώδικα εκπαίδευσης του Αλγόριθμου SOM



Εικόνα Α.16 Τμήμα του κώδικα υπολογισμού της ακτίνας της γειτονιάς του Αλγόριθμου SOM



Εικόνα Α.17 Τμήμα του κώδικα υπολογισμού του ρυθμού εκπαίδευσης του Αλγόριθμου SOM

```
214     def adjust_weights(self, vector_in, learning_rate, distance_fall_off=None):
215         """A procedure for update all the weights of this Node
216
217         :param vector_in: The Vector at the Input Layer
218         :param learning_rate: The Learning Rate number for this epoch
219     #        :param distance_fall_off:The rate for this epoch
220         """
221         for i in range(len(self._weights)):
222             wt = self._weights[i]
223             vw = vector_in[i]
224             if distance_fall_off is not None:
225                 wt += distance_fall_off * learning_rate * (vw - wt)
226             else:
227                 wt += learning_rate * (vw - wt)
228             self._weights.__setitem__(i, wt)
229
```

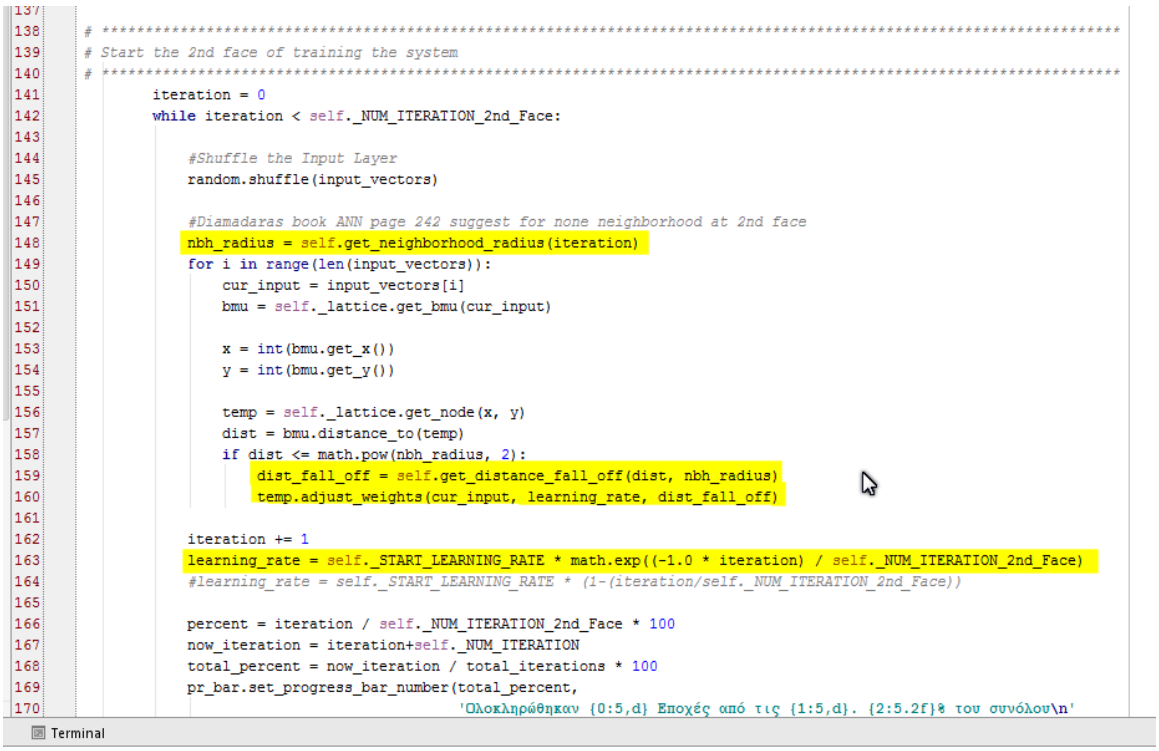
Εικόνα Α.18 Τμήμα του κώδικα ανανέωσης των βαρών κάθε νευρώνα του Αλγόριθμου SOM

```
93          #Diamadaras book ANN at page 241 suggest to start changing the neighborhood for 1st face
94          nbh_radius = self.get_neighborhood_radius(iteration)
95          for i in range(len(input_vectors)):
96              cur_input = input_vectors[i]
97              bmu = self._lattice.get_bmu(cur_input)
98
99              x_start = int(bmu.get_x() - nbh_radius - 1)
100             y_start = int(bmu.get_y() - nbh_radius - 1)
101             x_end = int(x_start + (nbh_radius * 2) + 1)
102             y_end = int(y_start + (nbh_radius * 2) + 1)
103
104             if x_end > self.lw:
105                 x_end = self.lw
106             if x_start < 0:
107                 x_start = 0
108             if y_end > self.lh:
109                 y_end = self.lh
110             if y_start < 0:
111                 y_start = 0
112             #print('1st, Epoch {6}] BMU[{0},{1}]\tNeighborhood_Start[{2},{3}]\tNeighborhood_End[{4},{5}]'
113             #      ''.format(bmu.get_x(), bmu.get_y(), x_start, y_start, x_end, y_end, iteration))
114
115             for x in range(x_start, x_end):
116                 for y in range(y_start, y_end):
117                     temp = self._lattice.get_node(x, y)
118                     #if temp is not bmu:
119                     dist = bmu.distance_to(temp)  # Calculate the distance between the Winner Neuron and the others
120                     if dist <= math.pow(nbh_radius, 2):
121                         dist_fall_off = self.get_distance_fall_off(dist, nbh_radius)
122                         temp.adjust_weights(cur_input, learning_rate, dist_fall_off)
123
124         iteration += 1
```

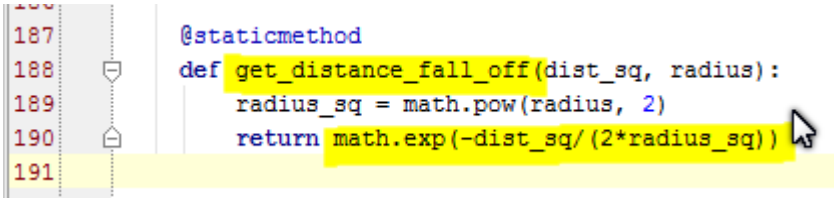
Εικόνα Α.19 Τμήμα του κώδικα εκπαίδευσης του Αλγόριθμου SOFM

```
183     def get_neighborhood_radius(self, iteration):
184         #return self._LATTICE_RADIUS * (1-(iteration/self._NUM_ITERATION))
185         return self._LATTICE_RADIUS * math.exp(-iteration/self._TIME_CONSTANT)
186
```

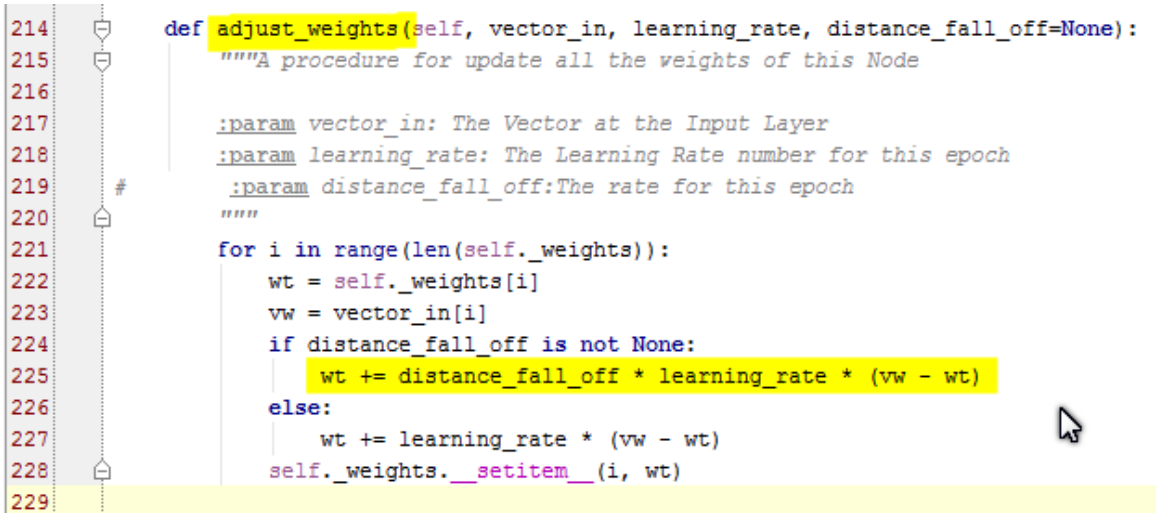
Εικόνα Α.20 Τμήμα του κώδικα υπολογισμού της ακτίνας της γειτονιάς του Αλγόριθμου SOFM



Εικόνα Α.21 Τμήμα του κώδικα υπολογισμού του ρυθμού εκπαίδευσης του Αλγόριθμου SOFM



Εικόνα Α.22 Τμήμα του κώδικα υπολογισμού του ρυθμού ενημέρωσης των βαρών κάθε Νευρώνα του Αλγόριθμου SOFM



Εικόνα Α.23 Τμήμα του κώδικα ανανέωσης των βαρών κάθε νευρώνα του Αλγόριθμου SOFM

Παράρτημα Β

Πρότυπο Ε.164 και δεδομένα που διακινούνται στις εφαρμογές

Στο Παράρτημα Β παρουσιάζονται: α) το πρότυπο Ε.164 που ορίζει τον τρόπο σχηματισμού τηλεφωνικών αριθμών για όλο τον κόσμο, β) οι κωδικοί όλων των περιοχών της Ελλάδος που οφείλουμε να πληκτρολογούμε ακόμη και για τοπικά τηλέφωνα, γ)

Β.1 Αρχείο με τους κανόνες που ορίζει το πρότυπο Ε.164

id	Zone	Country	CountryCode	InternationalPrefix	NationalPrefix	Code	MinNationalNumber	MaxNationalNumber	NationalNumber	UTC_DST*	Note
1	Asia	Afghanistan	93	0	0	0	9	9	9 digits	+4.30	
2	Europe_Russia	Albania	355	0	0	0	8	10	3 to 9 digits	+1/+2	
3	Rest	Algeria	213	0	0	0	8	9	8, 9 digits	1	
4	Rest	American Samoa	1	11	1	684	7	7	(684)+7 digits	-11	
5	Europe_Russia	Andorra	376	0	…	0	6	6	6, 8, 9 digits	+1/+2	
6	Europe_Russia	Andorra	376	0	…	0	8	9	6, 8, 9 digits	+1/+2	
7	Rest	Angola	244	0	0	0	9	9	9 digits	1	
8	Rest	Anguilla	1	11	1	264	7	7	(264)+7 digits	-4	
9	Rest	Antigua and	1	11	1	268	7	7	(268)+7 digits	-4	

		Barbuda									
10	Rest	Argentina	54	0	0	0	10	10	10 digits	-3	
11	Europe_Russia	Armenia	374	0	0	0	8	8	8 digits	+4/+5	
12	Rest	Aruba	297	0	…	0	7	7	7 digits	-4	
13	Australia	Australia	61	11	0	0	5	13	5 to 15 digits	+10/+11	12
15	Europe_Russia	Austria	43	0	0	0	4	13	4 to 13 digits	+1/+2	
16	Asia	Azerbaijan	994	0	0	0	8	9	8 to 9 digits	+4/+5	
17	Rest	Bahamas	1	11	1	242	7	7	(242)+7 digits	-5/-4	
18	Asia	Bahrain	973	0	…	0	8	8	8 digits	3	
19	Asia	Bangladesh	880	0	0	0	6	10	6 to 10 digits	6	
20	Rest	Barbados	1	11	1	246	7	7	(246)+7 digits	-4	
21	Europe_Russia	Belarus	375	810	8	0	9	10	9 to 10 digits	+2/+3	
22	Europe_Russia	Belgium	32	0	0	0	8	9	8 to 9 digits	+1/+2	2
23	Rest	Belize	501	0	…	0	7	7	7 digits	-6	
24	Rest	Benin	229	0	…	0	8	8	8 digits	1	
25	Rest	Bermuda	1	11	1	441	7	7	(441)+7 digits	-4/-3	
26	Asia	Bhutan	975	0	…	0	7	8	7 to 8 digits	6	
30	Rest	Botswana	267	0	…	0	7	8	7 to 8 digits	2	
31	Rest	Brazil	55	0	0	0	10	10	10 digits	-3/-2	18

Πίνακας Β.1 Εγγραφές με τους κανόνες που ορίζει το πρότυπο Ε.164

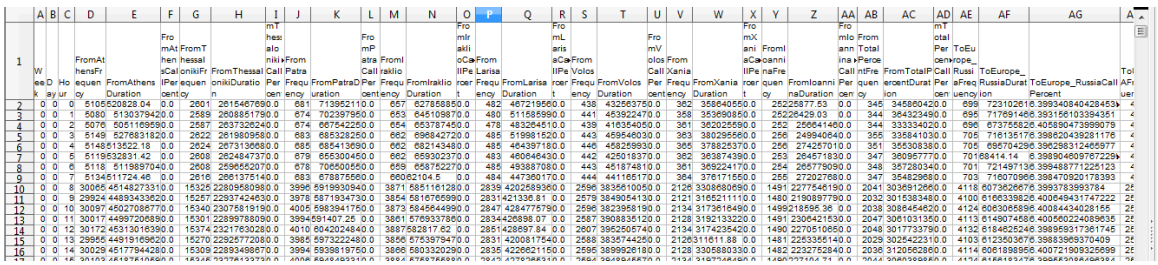
id	Νομός	Περιοχή	ΕΚΠ
1	Αιτωλοακαρνανία	ΜΕΣΟΛΟΓΓΙ	2631
2	Αιτωλοακαρνανία	ΑΙΤΩΛΙΚΟ	2632
3	Αιτωλοακαρνανία	ΝΑΥΠΑΚΤΟΣ	2634
4	Αιτωλοακαρνανία	ΜΑΤΑΡΑΓΚΑ	2635
5	Αιτωλοακαρνανία	ΑΓΡΙΝΙΟ	2641
6	Αιτωλοακαρνανία	ΑΜΦΙΛΟΧΙΑ	2642
7	Αιτωλοακαρνανία	ΒΟΝΙΤΣΑ	2643
8	Αιτωλοακαρνανία	ΘΕΡΜΟ	2644
9	Αιτωλοακαρνανία	ΦΥΤΕΙΕΣ	2646
10	Αιτωλοακαρνανία	Ν.ΧΑΛΚΙΟΠΟΥΛΟ	2647
11	Αργολίδος	ΑΡΓΟΣ	2751
12	Αργολίδος	ΝΑΥΠΛΙΟΝ	2752
13	Αργολίδος	ΛΥΓΟΥΡΙΟ	2753
14	Αργολίδος	ΚΡΑΝΙΔΙ	2754
15	Αρκαδίας	ΤΡΙΠΟΛΗ	271

Πίνακας Β.2 Εγγραφές από το αρχείο με τους τηλεφωνικούς κωδικούς αριθμούς όλων των περιοχών της Ελλάδος

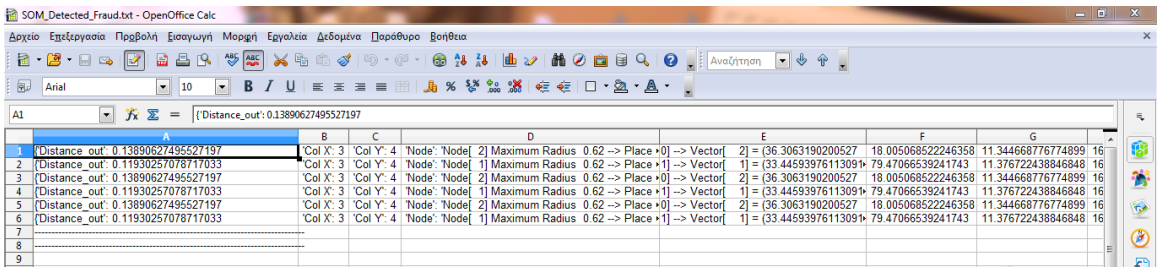
Β.2 Δεδομένα αποθηκευμένα σε αρχεία

StartTime	EndTime	Duration	CallerTelephone	CallerCountryCode	CallerCode	CallerMinNationalNumber	CallerMaxNationalNumber	CallerZone	CallerCountry	CallerCity	CallingTelephone	CallingCountryCode	CallingCode	CallingMinNationalNumber	CallingMaxNationalNumber	CallingZone	CallingCountry	CallingCity	Info
2014-02-01 10:00:02.279000	2014-02-01 10:02:54.223000	0:02:51.944000	302610765210	30	261	10	10	Europe	Greece	Αχαΐας	ΠΑΤΡΑ	302106070574	30	21	10	10	Europe	Greece	Αττικής	ΑΘΗΝΑ	FraudCall
2014-02-01 10:00:02.588000	2014-02-01 10:03:43.971000	0:03:41.383000	302315866510	30	231	10	10	Europe	Greece	Θεσσαλονίκης	ΘΕΣΣΑΛΟΝΙΚΗ	496537590	49	0	6	13	Europe_Russia	Germany			FraudCall
2014-02-01 10:00:02.902000	2014-02-01 10:01:44.541000	0:01:41.639000	302311624600	30	231	10	10	Europe	Greece	Θεσσαλονίκης	ΘΕΣΣΑΛΟΝΙΚΗ	302313832061	30	231	10	10	Europe	Greece	Θεσσαλονίκης	ΘΕΣΣΑΛΟΝΙΚΗ	FraudCall
2014-02-01 10:00:04.728000	2014-02-01 10:00:41.378000	0:00:36.650000	302310127948	30	231	10	10	Europe	Greece	Θεσσαλονίκης	ΘΕΣΣΑΛΟΝΙΚΗ	302310730766	30	231	10	10	Europe	Greece	Θεσσαλονίκης	ΘΕΣΣΑΛΟΝΙΚΗ	FraudCall

Πίνακας Β.3 Εγγραφές από το αρχείο με τις CDR εγγραφές



Πίνακας Β.4 Εγγραφές από το αρχείο με τα δεδομένα των τηλεφωνικών κλήσεων



Πίνακας B.5 Εγγραφές από το αρχείο με τις δόλιες κλήσεις.

B.3 Εξαγόμενα στοιχεία από αρχείο με CDRs ενός παρόχου VoIP υπηρεσιών

VoIP End Users
02:08:28

Call Inter-Arrival (Exponential)
Mean 6,08s
Variance 55,59s

Call Duration
Mean 114,27
Variance 36,904s
Exponential
Generalized Pareto (k; s) = (-0,39; 69,33)

Calls: 464161

ISP
Originating Distribution

Athens	46,7%
Thessaloniki	23,8%
Patras	6,2%
Irakleio	6%
Larissa	4,4%
Volos	4%
Chania	3,3%
Ioannina	2,3%
Other	3,3%

Destination ID Distribution
28% Local
22% Athens
20% Mobile
7% Thessaloniki
2% Patras
5% Other in Greece
16% International

- 40% European/Russia,
- 25%USA,
- 15% Asia,
- 5% Australia,
- 15%Rest)

Daily
08:00-16:00
Call Interarrival 56,48 ms (exponential)
Call Duration 149,92 sec (Weibull)
α = 1.5533 β = 166.15 γ = 17.213

Rest of the day:
Call Interarrival 328,48 ms (exponential)
Call Duration 100,83s (weibull)
α =0.83048 β = 125.17 γ = 18

Weekend
Call Interarrival: 2,59 sec (exponential)
Call Duration 104,12sec (weibull)
α =0.39613 β = 57.025 γ = 18

α=scale β=shape γ=location parameter

Call Interarrival: Είναι η χρονική διάρκεια που μεσολαβεί από μία κλήση που ξεκινά μέχρι την επόμενή της.

Call Duration: Είναι η χρονική διάρκεια μιας επιτυχημένης κλήσης από τη στιγμή που ξεκινά μέχρι να δωθεί το σήμα του τερματισμού της.