

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



Πρόγραμμα Μεταπτυχιακών Σπουδών

Τμήματος Διδακτικής της Τεχνολογίας και Ψηφιακών Συστημάτων

Διπλωματική εργασία

Multimodal υπηρεσίες για την ubiquitous
παροχή εφαρμογών πολυμέσων

Σωτηρόπουλος Γιάννης

Επιβλέπων καθηγητής: Δεμέστιχας Παναγιώτης

Σεπτέμβριος 2006

Περίληψη

Η διπλωματική εργασία πραγματεύεται το θέμα της ενοποίησης των διαφορετικών δικτύων κάτω από ένα ενιαίο δίκτυο το οποίο θα είναι προσβάσιμο από κάθε ηλεκτρονική συσκευή ανεξαρτήτως γεωγραφικής θέσης και χρόνου. Η δυνατότητα αυτή παρέχεται με την εξέλιξη των ασύρματων δικτύων και τη χρήση του πρωτοκόλλου IP ως συνδετικού κρίκου όλων των δικτύων (ενσύρματων και ασύρματων), δημιουργώντας ένα εικονικό δίκτυο που θα παρέχει υπηρεσίες στους χρήστες οποτεδήποτε και οπουδήποτε (ubiquitous services). Συνεπώς οδηγούμαστε σε έναν πλήρως δικτυωμένο κόσμο όπου η κάθε συσκευή θα μπορεί να επικοινωνήσει με οποιαδήποτε άλλη, ανεξαρτήτως των επιμέρους τεχνολογιών πρόσβασης (εφόσον βέβαια δεν τίθενται περιορισμοί ασφαλείας).

Σε αυτόν τον δικτυωμένο κόσμο οι υπηρεσίες πλέον δεν θα περιορίζονται μόνο στην μετάδοση φωνής και δεδομένων αλλά στην πραγματικού χρόνου μετάδοση ήχου και video. Τα πολυμεσικά δίκτυα θα προσφέρουν αυτές τις υπηρεσίες και σε συνδυασμό με την ενοποίηση των δικτύων η παροχή αυτών των υπηρεσιών θα γίνεται οποτεδήποτε και οπουδήποτε (multimedia ubiquitous services).

Επιπλέον η ενοποίηση των δικτύων απαιτεί να υπάρξουν και αλλαγές στον τρόπο με τον οποίο οι χρήστες αλληλεπιδρούν με τις συσκευές για την χρήση των υπηρεσιών. Πιο συγκεκριμένα οι χρήστες θα πρέπει να μπορούν να αλληλεπιδρούν με την συσκευή με τον τρόπο που τους βολεύει καλύτερα και αναλόγως το είδος της υπηρεσίας που χρησιμοποιούν. Η επίτευξη αυτών των στόχων πραγματοποιείται με την χρήση πολύτροπων

εφαρμογών (multimodal applications) που επιτρέπουν την χρήση πολλών διαφορετικών συσκευών και τρόπων αλληλεπίδρασης για την κατανάλωση των υπηρεσιών.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Πίνακας περιεχομένων

1	ΕΙΣΑΓΩΓΗ	7
1.1	ΕΙΣΑΓΩΓΗ.....	7
1.2	ΔΟΜΗ ΤΟΥ ΣΥΓΓΡΑΜΜΑΤΟΣ.....	10
2	ΣΥΓΚΛΙΣΗ ΔΙΚΤΥΩΝ (NETWORKS CONVERGENCE)	12
2.1	ΕΙΣΑΓΩΓΗ.....	12
2.2	COMPOSITE RADIO ENVIRONMENT (CRE).....	13
2.3	ΕΦΑΡΜΟΓΗ IP ΠΡΩΤΟΚΟΛΛΟΥ ΣΤΑ ΑΣΥΡΜΑΤΑ ΔΙΚΤΥΑ (WIRELESS IP).....	15
2.4	MOBILE INTERNET.....	19
2.4.1	Ορισμός.....	19
2.4.2	Απαιτήσεις για την ανάπτυξη του Mobile Internet.....	24
3	ΔΙΚΤΥΑ ΠΟΛΥΜΕΣΩΝ	27
3.1	ΕΙΣΑΓΩΓΗ.....	27
3.2	ΠΟΛΥΜΕΣΙΚΑ ΔΕΔΟΜΕΝΑ.....	28
3.3	ΔΙΚΤΥΑ ΠΟΛΥΜΕΣΩΝ.....	30
3.4	ΠΑΡΟΧΗ ΠΟΛΥΜΕΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	31
3.5	ΤΕΧΝΙΚΗ STREAMING.....	36
3.5.1	TCP/UDP Streaming.....	42
3.5.1.1	TCP Streaming.....	42
3.5.1.2	UDP Streaming.....	44
3.5.2	RTSP Streaming.....	45
3.5.3	RTP Streaming.....	47
3.6	MOBILE STREAMING.....	48
3.7	ΠΡΟΚΛΗΣΕΙΣ ΣΤΟΝ ΣΧΕΔΙΑΣΜΟ ΔΙΚΤΥΩΝ ΠΟΛΥΜΕΣΩΝ.....	51
3.7.1	Προβλήματα.....	52
3.7.1.1	Συνεχή ροή ήχου και video.....	52
3.7.1.2	Μεταβαλλόμενοι παράμετροι.....	53
3.7.2	Λύσεις.....	55
3.7.2.1	Capacity.....	57

3.7.2.2 Time	58
3.7.2.3 Space	59
3.7.2.4 Quality	61
3.7.2.5 Complexity	62
4 MULTIMODAL ΕΦΑΡΜΟΓΕΣ.....	63
4.1 ΕΙΣΑΓΩΓΗ.....	63
4.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΟΛΥΤΡΟΠΩΝ ΕΦΑΡΜΟΓΩΝ	66
4.3 ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ MULTIMODAL ΕΦΑΡΜΟΓΩΝ	68
4.3.1 Γενικές αρχιτεκτονικές	68
4.4 ΤΕΧΝΟΛΟΓΙΕΣ ΦΩΝΗΣ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΠΟΛΥΤΡΟΠΩΝ ΕΦΑΡΜΟΓΩΝ.....	69
4.4.1 Γενικά.....	69
4.4.2 Αναγνώριση ομιλίας	70
4.4.3 Σύνθεση ομιλίας	72
4.4.4 Συστήματα διαλόγων ομιλίας.....	74
4.5 ΠΡΟΚΛΗΣΕΙΣ ΣΤΗΝ ΑΝΑΠΤΥΞΗ ΠΟΛΥΤΡΟΠΩΝ ΕΦΑΡΜΟΓΩΝ	75
4.5.1 Σχεδιασμός πολύτροπων εφαρμογών.....	75
4.5.2 Υλοποίηση πολυτροπων εφαρμογών.....	75
4.5.3 Αβεβαιότητα στις επιλογές του χρήστη.....	76
4.6 ΛΟΓΙΣΜΙΚΟ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΑΣΥΡΜΑΤΩΝ ΠΟΛΥΤΡΟΠΩΝ ΕΦΑΡΜΟΓΩΝ ΦΩΝΗΣ	77
4.6.1 VoiceXML	77
4.6.2 XHTML + Voice (X+V)	78
5 ΑΠΟΤΕΛΕΣΜΑΤΑ.....	81
5.1 ΕΙΣΑΓΩΓΗ.....	81
5.2 ΠΟΛΥΤΡΟΠΕΣ ΕΦΑΡΜΟΓΕΣ ΓΙΑ ΤΗΝ ΠΑΡΟΧΗ UBIQUITOUS ΥΠΗΡΕΣΙΩΝ.....	82
5.3 ΑΝΑΠΤΥΞΗ ΠΟΛΥΤΡΟΠΩΝ ΕΦΑΡΜΟΓΩΝ ΜΕ ΣΥΝΔΥΑΣΜΟ ΤΕΧΝΟΛΟΓΙΩΝ ΦΩΝΗΣ ΚΑΙ ΓΡΑΦΙΚΩΝ ΔΙΕΠΑΦΩΝ.....	84
5.3.1 Γενικά.....	84
5.3.2 Επιλογή ή μη χρήσης τεχνολογιών φωνής.....	84
5.3.3 Σχεδιαστικές επιλογές.....	85
5.3.3.1 Αποφυγή της μετατροπής γραφικών διεπαφών σε φωνητικές.....	86

5.3.3.2	Παράμετροι για την επιτυχή ανάπτυξη πολύτροπων εφαρμογών	88
5.3.3.2.1	Παροδικότητα του λόγου (Speech transience).....	88
5.3.3.2.2	Απόδοση αναγνώρισης φωνής	88
5.3.3.2.3	Ευελξία έναντι ακρίβειας κατά την αναγνώριση φωνής	89
5.3.3.2.4	Απόκριση και καθυστέρηση (Feedback and Latency)	89
5.3.3.2.5	Διάλογοι (prompts)	90
5.3.3.2.6	Χειρισμός λαθών	90
5.4	ΑΝΑΠΤΥΞΗ ΠΟΛΥΤΡΟΠΗΣ ΕΦΑΡΜΟΓΗΣ	90
5.4.1	Γενικά	90
5.4.2	Περιγραφή εφαρμογής και λειτουργίες	91
5.4.3	Αρχιτεκτονική εφαρμογής και τεχνολογίες	92
5.4.4	Εννοιολογικό μοντέλο	93
5.4.5	Παραδείγματα χρήσης	94
6	ΣΥΜΠΕΡΑΣΜΑΤΑ	98
7	ΒΙΒΛΙΟΓΡΑΦΙΑ	100

1

Εισαγωγή

1.1 Εισαγωγή

Η ραγδαία ανάπτυξη των ασύρματων δικτύων και των δικτύων κινητής τηλεφωνίας και η υψηλή αποδοχή από τους τελικούς χρήστες [1, 2] των υπηρεσιών που προσφέρουν, καθιστούν τα ασύρματα δίκτυα και τα δίκτυα κινητής επικοινωνίας στις δυναμικότερα αναπτυσσόμενες τεχνολογίες στον χώρο της πληροφορικής και των τηλεπικοινωνιών.

Ομοίως οι ποικίλες υπηρεσίες που παρέχει το Διαδίκτυο (πλοήγηση στον παγκόσμιο ιστό, αποστολή και λήψη ηλεκτρονικού ταχυδρομείου, video και audio streaming κτλ) έχουν συνεισφέρει σημαντικά στην ραγδαία εξάπλωση του [3-5].

Άμεσο επακόλουθο της ανάπτυξης αυτών των δυο τεχνολογιών είναι η ανάγκη ενοποίησης τους με στόχο την παροχή υπηρεσιών του Διαδικτύου σε ασύρματες και κινητές συσκευές. Για παράδειγμα οι χρήστες των κινητών και των ασύρματων συσκευών επιθυμούν να έχουν πρόσβαση στο ηλεκτρονικό τους ταχυδρομείο, να surfάρουν στο Διαδίκτυο καθώς

κινούνται , να κατεβάζουν μουσική και video κτλ. Οι υπηρεσίες αυτές είναι πλέον αναγκαίες για τους τελικούς χρήστες εφόσον τους παρέχουν την δυνατότητα να εκτελέσουν τις εργασίες που επιθυμούν όχι μόνο από τον χώρο εργασίας τους ή το σπίτι αλλά οποτεδήποτε και οπουδήποτε επιθυμούν (ubiquitous services) [6].

Η ανάγκη αυτή οδηγεί στο ότι πολλές διαφορετικές τεχνολογίες διαφορετικών δικτύων (ενσύρματα, ασύρματα και δίκτυα κινητής τηλεφωνίας) κατευθύνονται προς την ενοποίηση, και τίθεται ως στόχος η ύπαρξη ενός ενιαίου δικτύου το οποίο θα προσφέρει διαφανώς (ανεξαρτήτως του δικτύου πρόσβασης) υπηρεσίες στους τελικούς χρηστές (4G Networks). Για να οδηγηθούμε σε αυτό τον στόχο πλέον μπορεί να ειπωθεί πως οι διαφορετικές τεχνολογίες δικτύων αναπτύσσονται παράλληλα και όχι ανεξάρτητα η μία από την άλλη [7]. Τα πλεονεκτήματα που θα προκύψουν από αυτή την σύγκλιση είναι:

- Οι τελικοί χρήστες θα έχουν δυνατότητα να αποκτήσουν πρόσβαση στις δικτυακές υπηρεσίες οποτεδήποτε και οπουδήποτε (ubiquitous service) εφόσον πλέον οι υπηρεσίες είναι ανεξάρτητες από τις επιμέρους τηλεπικοινωνιακές υποδομές. Για παράδειγμα ένας χρήστης θα μπορεί να διαβάζει το ηλεκτρονικό του ταχυδρομείο είτε μέσω ενός σταθερού τερματικού (πχ προσωπικός υπολογιστής) είτε μέσω ενός ασύρματου τερματικού (πχ PDA) είτε να ακούει τα μηνύματα του μέσω του κινητού του τηλεφώνου. Επίσης ο χρήστης θα έχει την δυνατότητα διαφορετικών τρόπων αλληλεπίδρασης με τη δικτυακή υπηρεσία (Πολύτροπη αλληλεπίδραση - Multimodal interaction [8-10] μέσω του ίδιου τερματικού. Για παράδειγμα θα

μπορεί να εκτελέσει μια δοσοληψία με μια τράπεζα είτε μέσω του γραφικού περιβάλλοντος του PDA είτε με την χρήση φωνητικών εντολών από το κινητό του τηλέφωνο. Ακόμα περισσότερο δυο οι περισσότεροι διαφορετικοί τρόποι αλληλεπίδρασης θα μπορούν να χρησιμοποιηθούν παράλληλα και συγχρονισμένα για την εκτέλεση της επιθυμητής εργασίας.

- Η πρόσβαση στο Internet θα γίνεται με την μέγιστη δυνατή ταχύτητα αναλόγως της γεωγραφικής θέσης και της ταχύτητας κίνησης. Για παράδειγμα σε ένα σημείο που υπάρχουν δυο διαφορετικά δίκτυα πρόσβασης (πχ 802.11b/11Mbps [11] και UMTS/384Kbps [12]) θα πρέπει αυτόματα η συσκευή να επιλέγει το δίκτυο πρόσβασης που προσφέρει μεγαλύτερους ρυθμούς μετάδοσης δεδομένων. Φυσικά αυτό απαιτεί από την τερματική συσκευή να μπορεί να έχει δυνατότητα πρόσβασης σε πολλά διαφορετικά δίκτυα.
- Καθώς η ασύρματη δικτύωση επιτρέπει την εκτέλεση των εργασιών εν κίνηση θα πρέπει να παρέχεται ομαλή μετάβαση από το ένα δίκτυο πρόσβασης στο άλλο χωρίς να διακόπτεται η εργασία που εκτελείτε. Επιπλέον το κάθε δίκτυο θα πρέπει να γνωρίζει κάθε στιγμή τη θέση του τερματικού έτσι ώστε να δρομολογεί τις εισερχόμενες κλήσεις.
- Δυνατότητα μετάδοσης ήχου και video σε πραγματικό χρόνο μέσα από κατάλληλα δίκτυα πολυμέσων (Multimedia networks).
- Οι δικτυακές υπηρεσίες θα είναι χαλαρά συνδεδεμένες με τα επιμέρους στοιχεία του δικτύου, προσφέροντας την δυνατότητα ανάπτυξης και αξιολόγησης των νέων υπηρεσιών σε μικρό χρονικό διάστημα και με μικρότερο κόστος. Περαιτέρω, οι υπηρεσίες που δεν

έχουν τα αναμενόμενα οφέλη μπορούν εύκολα να προσαρμόζονται στις νέες απαιτήσεις.

Ο στόχος της σύγκλισης των δικτύων δεν είναι εύκολα υλοποιήσιμος και ήδη υπάρχει πλήθος ερευνητικών δημοσιεύσεων που προσπαθούν να λύσουν τα ήδη υπάρχοντα προβλήματα και να προτείνουν νέες λύσεις και κατευθύνσεις. Σκοπός του παρόντος συγγράμματος είναι να συμβάλει προς αυτή την κατεύθυνση κυρίως στα ζητήματα της ενοποίησης των δικτύων, στην παροχή πολυμεσικών δεδομένων μέσω κατάλληλων δικτύων πολυμέσων και στην ανάπτυξη πολύτροπων εφαρμογών (Multimodal applications) για την παροχή υπηρεσιών οπουδήποτε και οποτεδήποτε (ubiquitous service).

1.2 Δομή του συγγράμματος

Η πτυχιακή εργασία αποτελείται από 7 ενότητες :

Στην ενότητα 1 παρουσιάζεται το γενικό πλαίσιο της διπλωματικής και καθορίζονται τα κύρια θέματα τα οποία θα αναλυθούν διεξοδικά στις επόμενες ενότητες.

Στην ενότητα 2 παρουσιάζεται το θέμα της σύγκλισης των δικτύων για την δημιουργία ενός ενιαίου δικτύου το οποίο θα προσφέρει οπουδήποτε και οποτεδήποτε διαδικτυακές υπηρεσίες στον χρήστη.

Στην ενότητα 3 γίνεται αναφορά στα δίκτυα πολυμέσων και στις τεχνικές παροχής πολυμεσικών δεδομένων στους χρήστες. Επιπλέον αναγνωρίζονται τα προβλήματα στην μετάδοση πολυμεσικών δεδομένων

και αναπτύσσεται ένα πλαίσιο διαστάσεων το οποίο μπορεί να εξαλείψει αυτά τα προβλήματα.

Στην ενότητα 4 παρουσιάζονται οι πολύτροπες εφαρμογές. Κύριος βάρος δίνεται στην ανάπτυξη πολύτροπων εφαρμογών που βασίζονται σε τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας. Επιπλέον αναλύονται οι δυσκολίες στην ανάπτυξη αυτής της μορφής των εφαρμογών.

Στην ενότητα 5 παρουσιάζονται τα αποτελέσματα της ερευνητικής δουλειάς που ενοποιούν τις τεχνολογίες και τις προσεγγίσεις που αναλύθηκαν στις προηγούμενες ενότητες με στόχο την παροχή πραγματικά ubiquitous υπηρεσιών. Επιπλέον παρέχεται ένα πλαίσιο για την ανάπτυξη πολύτροπων εφαρμογών.

Στην ενότητα 6 παρουσιάζονται συμπεράσματα και επεκτάσεις των θεμάτων που αναπτύχθηκαν στις προηγούμενες ενότητες.

Στην ενότητα 7 παρέχονται οι βιβλιογραφικές αναφορές που χρησιμοποιήθηκαν κατά την συγγραφή της διπλωματικής.

2

Σύγκλιση δικτύων (Networks convergence)

2.1 Εισαγωγή

Ο όρος της σύγκλισης των δικτύων περιλαμβάνει τόσο τα δίκτυα τηλεπικοινωνιών όσο και το Διαδίκτυο. Στόχος είναι η δημιουργία μιας ασύρματης υποδομής η οποία θα δρα ως ο ενδιάμεσος για την παροχή υπηρεσιών στους τελικούς χρήστες ανεξαρτήτως της τεχνολογίας των δικτύων πρόσβασης. Αποτέλεσμα αυτού είναι οι χρήστες να έχουν συνεχή πρόσβαση σε διαδικτυακές υπηρεσίες ανεξαρτήτως χρόνου και φυσικής τοποθεσίας (ubiquitous service)[6].

Βασικό κλειδί για αυτή την ενοποίηση είναι η αποδοχή του πρωτοκόλλου IP [9, 10] το οποίο θα χρησιμοποιηθεί ως το μέσο για την παροχή των προσφερόμενων υπηρεσιών.

Στόχος της παρόντος κεφαλαίου είναι να αναλύσει της σύγκλιση των δικτύων για την δημιουργία ενός ενιαίου δικτύου, τις τεχνολογίες, τις προκλήσεις στην ανάπτυξη και τα οφέλη από την σύγκλιση των δικτύων.

2.2 Composite Radio Environment (CRE)

Η έννοια του Composite Radio Environment (CRE) προσδιορίζει πως διαφορετικά δίκτυα μπορούν να συνεργαστούν σε μια ασύρματη υποδομή έτσι ώστε οι τηλεπικοινωνιακοί πάροχοι να μπορούν να προσφέρουν στους τελικούς χρήστες υπηρεσίες με υψηλές ταχύτητες και υψηλή ποιότητα υπηρεσίας (QoS).

Βασικό στοιχείο για να μπορούν τα τερματικά να λειτουργούν σε διαφορετικές τεχνολογίες δικτύων είναι η έννοια του reconfiguration [13]. Η έννοια του reconfiguration επιτρέπει στα τερματικά και στα στοιχεία του δικτύου να επιλέγουν και να προσαρμόζονται δυναμικά στην πιο κατάλληλη τεχνολογία πρόσβασης για να χειριστούν με τον βέλτιστο τρόπο τις απαιτήσεις τις εκάστοτε υπηρεσίας αναλόγως της γεωγραφικής θέσης και της χρονικής στιγμής που προσφέρεται η υπηρεσία.

Σήμερα υπάρχουν πολυάριθμες τεχνολογίες πρόσβασης (Radio Access Technology) που χρησιμοποιούνται στις ασύρματες τηλεπικοινωνίες. Οι τεχνολογίες αυτές μπορούν να κατηγοριοποιηθούν σε τέσσερις κατηγορίες:

- **Δίκτυα κινητής τηλεφωνίας:** Συμπεριλαμβάνουν δίκτυα 2^{ης} γενιάς όπως το Global System for Mobile Communications (GSM) [14], δίκτυα 2.5^{ης} γενιάς όπως το Enhanced Data Rates for Global Evolution (EDGE) και το General Packet Radio Service (GPRS) [15]. Δίκτυα 3^{ης} γενιάς όπως το Universal Mobile Telecommunications Service (UMTS) [12] και το CDMA-200 [12] και δίκτυα 4^{ης} γενιάς τα οποία θα προσφέρουν μεταφορά δεδομένων μέχρι και 100Mbit/s

- **Ευρυζωνικά ασύρματα δίκτυα και ασύρματα δίκτυα:**
Συμπεριλαμβάνουν δίκτυα Broadband Radio Access Networks (BRAN) [16] όπως το HiperLan2 και ασύρματα τοπικά δίκτυα Wireless Local Area Networks (WLAN) [11]
- **Δορυφορικά δίκτυα και δίκτυα εκπομπής video** όπως Digital Video Broadcasting (DVB) [17]
- **Ad-hoc δίκτυα και δίκτυα αισθητήρων**

Αν και τα δίκτυα 4^{ης} γενιάς είναι ανοιχτά σε νέες τεχνολογίες πρόσβασης, τα υπάρχοντα δίκτυα παραμένουν ελκυστικά από άποψη αποδοτικότητας παροχής υπηρεσιών. Ο κύριος στόχος είναι τα διαφορετικά δίκτυα πρόσβασης (GPRS, UMTS BRAN/WLAN, DVB κτλ) να αποτελούν συστατικά μιας ενιαίας ασύρματης υποδομής. Σε κάθε περίπτωση οι τηλεπικοινωνιακοί πάροχοι θα μπορούν να κατέχουν πολλά διαφορετικά συστατικά του ενιαίου δικτύου και θα μπορούν να συνεργαστούν αρμονικά με άλλους τηλεπικοινωνιακούς παρόχους. Επίσης θα μπορούν να βασίζονται σε εναλλακτικά δίκτυα πρόσβασης για να επιτύχουν τις επιθυμητές ταχύτητες και επιθυμητά επίπεδα ποιότητας υπηρεσίας.

Τα τερματικά των χρηστών θα μπορούν να επιλέγουν αυτόματα το κατάλληλο ασύρματο δίκτυο βάση των απαιτήσεων των υπηρεσιών και της απόδοσης του δικτύου. Συνεπώς τα διαφορετικά δίκτυα πρόσβασης χρησιμοποιούνται πλέον ως συμπληρωματικά το ένα προς το άλλο για να επιτύχουν το καλύτερο δυνατό αποτέλεσμα τόσο βάσης απόδοσης όσο και κόστους. Για παράδειγμα στην περίπτωση που είναι διαθέσιμα δυο δίκτυα που προσφέρουν διαφορετικές ταχύτητες μεταφοράς δεδομένων το τερματικό του χρήστη μπορεί να επιλέξει την φθηνότερη υπηρεσία με την

μεγαλύτερη ταχύτητα ανάλογα με την συνδρομή του χρήστη στον τηλεπικοινωνιακό πάροχο. Η δυνατότητα αυτή βασίζεται στην έννοια του reconfiguration το οποίο επιτρέπει στα στοιχεία του δικτύου να επιλεγούν με διαφάνεια και ασφάλεια το κατάλληλο δίκτυο πρόσβασης που ικανοποιεί τις απαιτήσεις της υπηρεσίας την συγκεκριμένη χρονική στιγμή. Επιπλέον οι τεχνολογίες πρόσβασης που επιλέγονται δυναμικά από το τερματικό δεν περιορίζονται μόνο σε αυτές που είναι από πριν εγκατεστημένες στο τερματικό, αλλά υπάρχει δυνατότητα να κατεβεί και να εγκατασταθεί αυτόματα στο τερματικό του χρήστη κατάλληλο λογισμικό για την χρήση νέων τεχνολογιών πρόσβασης [18].

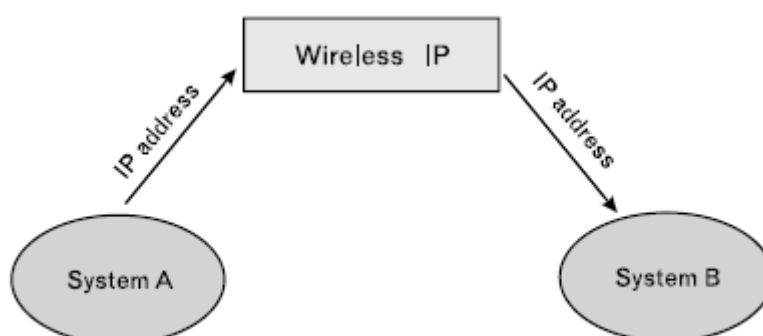
Η ενοποίηση των επιμέρους συστατικών του CRE επιτυγχάνεται μέσω μιας διαχειριστικής πλατφόρμας (Management System For CRE) η οποία υπάρχει ως συστατικό σε κάθε επιμέρους δίκτυο και διαχειρίζεται σε καθένα από αυτά τις διαφορετικές τεχνολογίες πρόσβασης. Το δίκτυο κορμού που θα ενοποιεί όλα αυτά τα δίκτυα του CRE αποτελεί ένα IP δίκτυο που μαζί με κατάλληλα πρωτοκόλλα (πχ Mobile IP [19]) επιτρέπει την διατήρηση της συνδεσιμότητας ανεξαρτήτως των επιμέρους τεχνολογιών πρόσβασης.

2.3 Εφαρμογή IP πρωτοκόλλου στα ασύρματα δίκτυα (Wireless IP)

Καθώς η τηλεπικοινωνιακή κίνηση που προκύπτει από υπηρεσίες δεδομένων ξεπερνά πλέον την κίνηση των υπηρεσιών φωνής οι τηλεπικοινωνιακοί πάροχοι συνειδητοποιούν ότι η παροχή όλων των υπηρεσιών (φωνή και δεδομένα) μέσα από δίκτυα μεταγωγής πακέτων είναι οικονομικότερη αντί της χρήσης δικτύων μεταγωγής κυκλώματος. Η εξέλιξη αυτή άρχισε να υλοποιείται στα δίκτυα 2.5^{ns} και συνεχίζεται στα

δίκτυα 3^{ης} γενιάς μέσω της χρήσης του πρωτοκόλλου IP. Επιπρόσθετα το πρωτόκολλο IP αποτελεί το κύριο πρωτόκολλο του Διαδικτύου και συνεπώς έχει δοκιμαστεί επιτυχώς για την αποτελεσματικότητά του εφόσον καταφέρνει να εξυπηρετήσει εκατομμύρια χρήστες. Άμεσο επακόλουθο είναι και η μεταφορά του στα ασύρματα δίκτυα ως Wireless IP.

Το Wireless IP συνδυάζει τις δυνατότητες διαδίκτυωσης του επιπέδου δικτύου με την δυνατότητα που παρέχουν οι ασύρματες συσκευές για παροχή υπηρεσιών οπουδήποτε και οποτεδήποτε επιτρέποντας την παροχή υπηρεσιών ανεξαρτήτως της τεχνολογίας του δικτύου πρόσβασης όπως φαίνεται παρακάτω:



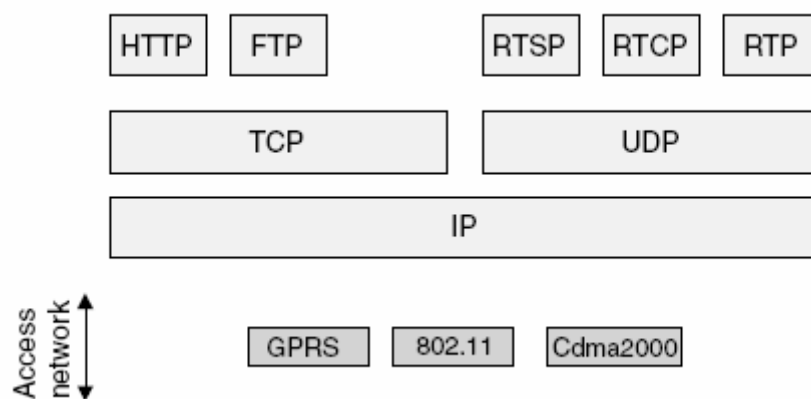
Σχήμα 2.1 Wireless IP

Συνεπώς συνδυάζοντας τα πλεονεκτήματα που προσφέρει το πρωτόκολλο IP και οι τεχνολογίες ασύρματων δικτύων δίνει την δυνατότητα για ανάπτυξη πολυμεσικών υπηρεσιών χαμηλού κόστους όπως για παράδειγμα υπηρεσιών VoIP [20].

Ο στόχος για παροχή από άκρο σε άκρο συνδεσιμότητας ώθησε την αποδοχή του πρωτοκόλλου IP ως ένα ενδιάμεσο επίπεδο το οποίο θα μπορεί να υποστηρίξει μια πληθώρα διαφορετικών τεχνολογιών πρόσβασης

του επιπέδου ζεύξης δεδομένων. Γενικότερα ο στόχος είναι ότι πρωτόκολλα βασισμένα στο IP χρησιμοποιούνται για όλες τις λειτουργίες, όπως η μεταφορά δεδομένων, η σηματοδότηση σε επίπεδο εφαρμογής, η ασφάλεια και η ποιότητα υπηρεσίας (QoS). Με την εφαρμογή αυτής της προσέγγισης οι χρήστες θα μπορούν να εκτελούν πολυμεσικές εφαρμογές στα τερματικά τους, οι οποίες θα υποστηρίζουν πολλαπλές τεχνολογίες πρόσβασης (WLAN, WPAN, UMTS, CDMA2000 κλπ). Αυτή η υποστήριξη του πρωτοκόλλου IP από άκρο σε άκρο (all-IP) οδηγεί στην εξέλιξη των δικτύων 3^{ης} γενιάς στα δίκτυα 4^{ης} γενιάς.

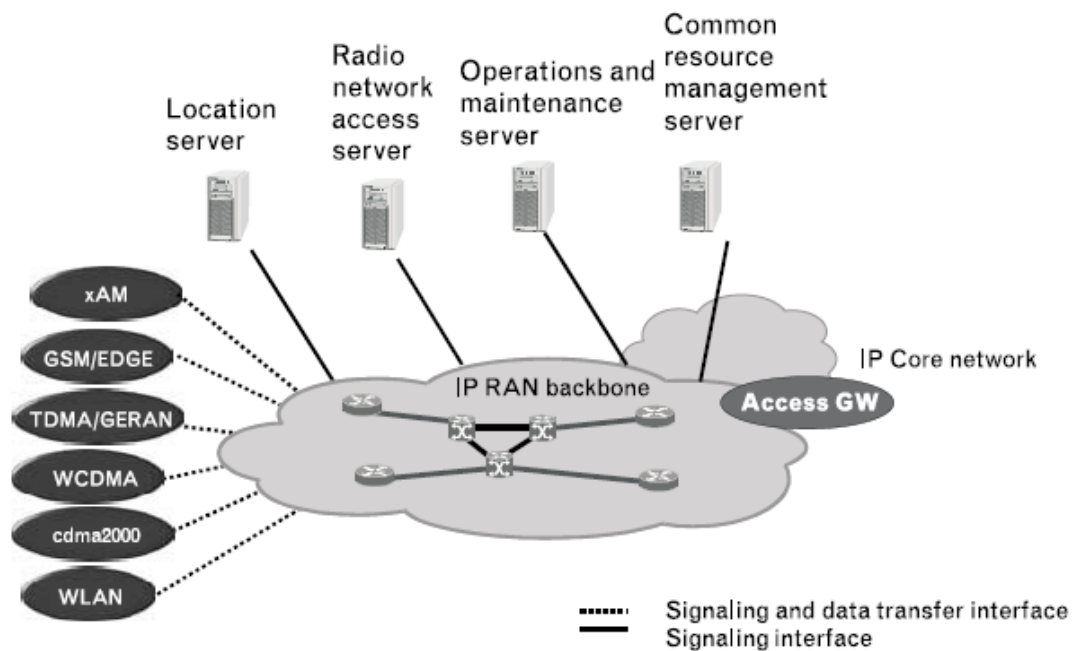
Στην all-IP αρχιτεκτονική όλες οι υπηρεσίες προσφέρονται από επίπεδα πάνω από το IP όπως αυτά προσδιορίζονται από την εκάστοτε στοίβα πρωτοκόλλων. Η προσέγγιση αυτή παρουσιάζεται παρακάτω:



Σχήμα 2.2 Στοίβα πρωτοκόλλων για all-IP δίκτυα

Επιπλέον ενώ τα συστήματα 3^{ης} γενιάς υποστηρίζουν μόνο κινητικότητα όσο αφορά το τερματικό του χρήστη (terminal mobility) τα δίκτυα 4^{ης} γενιάς επιτρέπουν κινητικότητα τόσο ως προς το τερματικό όσο και κινητικότητα μεταξύ πολλών διαφορετικών δικτύων πρόσβασης (πχ WLAN,

ad hoc networks, δίκτυα κινητών επικοινωνιών κλπ). Η αρχιτεκτονική ενός all-IP δικτύου παρουσιάζεται παρακάτω:



Σχήμα 2.3 All-IP αρχιτεκτονική

Παρατηρούμε πως η πρόσβαση πραγματοποιείται με τη χρήση διαφορετικών ασύρματων τεχνολογιών (πχ GSM,WLAN) ενώ η μεταγωγή των πακέτων εκτελείται από το δίκτυο κορμού ασύρματης πρόσβασης με την χρήση του πρωτοκόλλου IP (IP Remote Access Network Backbone). Επιπλέον οι διάφορες λειτουργίες και οι υπηρεσίες του δικτύου παρέχονται από ειδικούς εξυπηρετητές (πχ Location server, Resource management server). Τέλος η all-IP αρχιτεκτονική υπονοεί στο ότι οι χρήστες μετακινούνται συνεχώς και ότι το δίκτυο μπορεί να υποστηρίξει όλους τους τύπους κίνησης ανεξαρτήτως απαιτήσεων ποιότητας υπηρεσίας (QoS).

2.4 Mobile Internet

2.4.1 Ορισμός

Γενικά όρος Mobile Internet χαρακτηρίζει την μεταφορά των εφαρμογών και των υπηρεσιών του Διαδικτύου στο χώρο των ασύρματων επικοινωνιών [21]. Όμως η μεταφορά αυτή συνοδεύεται από τα παρακάτω προβλήματα:

- Το πρόβλημα της επιλογής της καλύτερης ασύρματης τεχνολογίας για την παροχή υπηρεσιών Διαδικτύου
- Αν επιλεγμένη ασύρματη τεχνολογία είναι κατάλληλη για υπηρεσίες με διαφορετικές απαιτήσεις, όπως για παράδειγμα η υπηρεσία ηλεκτρονικού ταχυδρομείου και η υπηρεσία μετάδοσης video (video broadcasting)
- Πως οι διαθέσιμες υπηρεσίες παρέχονται οπουδήποτε και οποτεδήποτε (ubiquitous access) στον τελικό χρήστη ακόμα και όταν το τερματικό μεταβαίνει σε μια περιοχή η οποία δεν παρέχει την συγκεκριμένη υπηρεσία. Στην περίπτωση που χρησιμοποιούνται παράλληλα πολλές διαφορετικές ασύρματες τεχνολογίες πρόσβασης όπως δίκτυα κινητών επικοινωνιών και τοπικά ασύρματα δίκτυα για την επίλυση του παραπάνω προβλήματος, τίθεται το πρόβλημα πως όλες αυτές οι επιμέρους τεχνολογίες θα συνδυαστούν για την παροχή διαφανών υπηρεσιών προς τον χρήστη.
- Το πρόβλημα της βελτιστοποίησης των ασύρματων δικτύων με στόχο την υποστήριξη μεγάλου αριθμού χρηστών που επιθυμούν ασύρματες υπηρεσίες Διαδικτύου.

- Το πρόβλημα της ασφάλειας για υπηρεσίες που παρέχονται από διαφορετικούς τηλεπικοινωνιακούς παρόχους.

Τα προβλήματα αυτά είναι ένα μέρος των προβλημάτων τα οποία πρέπει να λυθούν για να γίνει πραγματικότητα το Mobile Internet. Το γεγονός ότι σήμερα μπορούμε να χρησιμοποιήσουμε τον φορητό υπολογιστή (Notebook) ή το PDA σε συνδυασμό με μία ασύρματη συσκευή (κινητό τηλέφωνο, ασύρματη κάρτα WLAN) για να συνδεόμαστε στο Διαδίκτυο, στο εταιρικό Δίκτυο και σε οποιοδήποτε άλλο βασισμένο στο IP πρωτόκολλο δίκτυο δεν σημαίνει πως το Mobile Internet είναι ήδη εφικτό. Στην πραγματικότητα το Mobile Internet είναι αρκετά πιο πολύπλοκο από ότι αυτές οι υπηρεσίες.

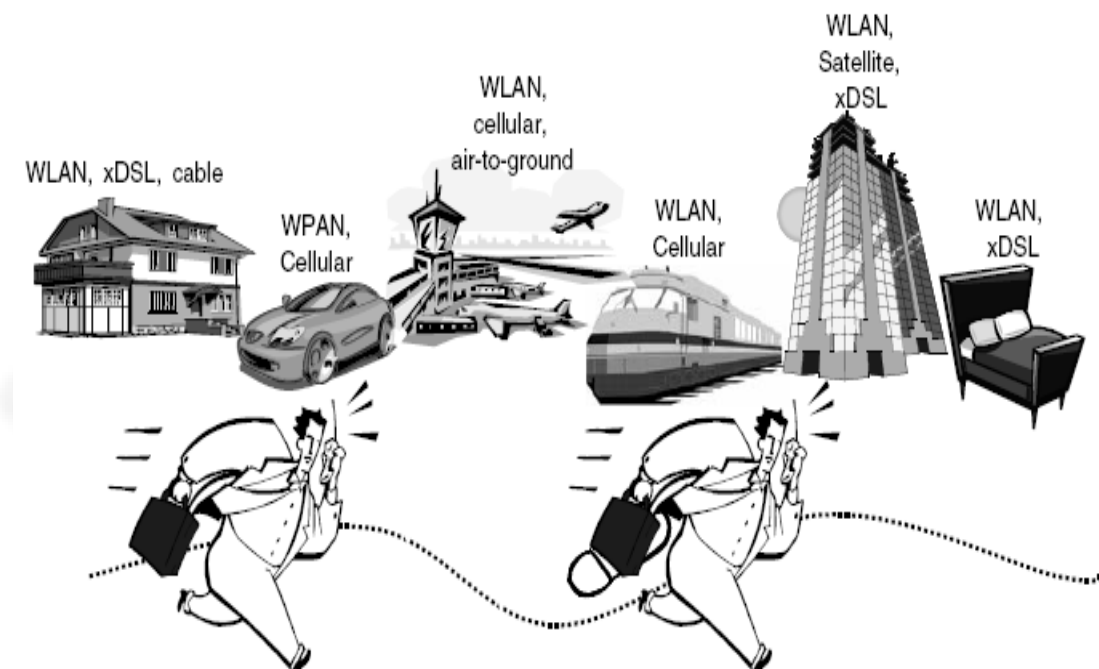
Πιο συγκεκριμένα το Διαδίκτυο είναι ένα δίκτυο εκατομμυρίων χρηστών που επικοινωνούν με μέσω καλά καθορισμένων πρωτοκόλλων. Συνεπώς και στο Mobile Internet είναι ανάγκη να υποστηρίζονται εκατομμύρια συνδεδεμένοι χρήστες. Οι λέξεις κλειδιά σε αυτή την απαίτηση είναι:

- Η υποστήριξη εκατομμυρίων χρηστών
- Οι χρήστες θα πρέπει να είναι συνεχώς συνδεδεμένοι (always connected)

Το ότι οι χρήστες είναι συνεχώς συνδεδεμένοι σημαίνει πως δεν θα πρέπει να εγκαθιδρύεται μια νέα σύνδεση πριν από κάθε δικτυακή λειτουργία ή μεταξύ διαφορετικών υπηρεσιών. Η υπόθεση είναι πως θα υπάρχει πάντα συνδεσιμότητα στις υπηρεσίες του Διαδικτύου και ότι το τερματικό του

χρήστη είναι πάντα διαθέσιμο στο δίκτυο μέσω μιας IP διεύθυνσης ανεξαρτήτως τοποθεσίας. Αυτό απαιτεί υψηλή διαχείριση κινητικότητας (extensive mobility management) και ομαλή μεταπομπή (handover) μεταξύ των διαφορετικών ασύρματων δικτύων. Η απαίτηση για πολλά διαφορετικά ασύρματα δίκτυα είναι αναγκαία επειδή κανένα ασύρματο δίκτυο από μόνο του δεν μπορεί να παρέχει παντού και πάντα υπηρεσίες (ubiquitous services) και συνεπώς κανένα ασύρματο δίκτυο από μόνο του δεν ικανοποιεί την απαίτηση για συνεχή συνδεσιμότητα.

Ένας τυπικός χρήστης που κάνει χρήση του Mobile Internet θα πρέπει να κινείται χωρίς πρόβλημα μεταξύ διαφορετικών ασύρματων δικτύων ή ακόμα μεταξύ ασύρματων και σταθερών δικτύων, τα οποία δεν χρησιμοποιούν την ίδια τεχνολογία πρόσβασης. Αυτό παρουσιάζεται παρακάτω:



Σχήμα 2.4 Μετάβαση μεταξύ διαφορετικών δικτύων με συνεχή συνδεσιμότητα

Αυτή η ομαλή μετάβαση υπονοεί πως η σύνδεση στο Διαδίκτυο μεταφέρεται από το ένα δίκτυο πρόσβασης στο άλλο χωρίς καμία ενέργεια από την πλευρά του χρήστη. Σαν αποτέλεσμα αυτό δημιουργεί ένα εικονικό ασύρματο δίκτυο στην πλευρά του χρήστη που παρέχει συνεχή συνδεσιμότητα με το Διαδίκτυο.

Η άλλη απαίτηση είναι η υποστήριξη εκατομμύριων χρηστών που συνδέονται ασύρματα στο Διαδίκτυο έτσι ώστε να υπάρχει συνέπεια με την τεχνολογία των σταθερών δικτύων τα οποία υποστηρίζουν εκατομμύρια χρήστες. Η απαίτηση αυτή δημιουργεί προβλήματα εύρους ζώνης και χωρητικότητας του δικτύου και δίνει την απάντηση γιατί χρειάζονται πολλαπλές ασύρματες τεχνολογίες.

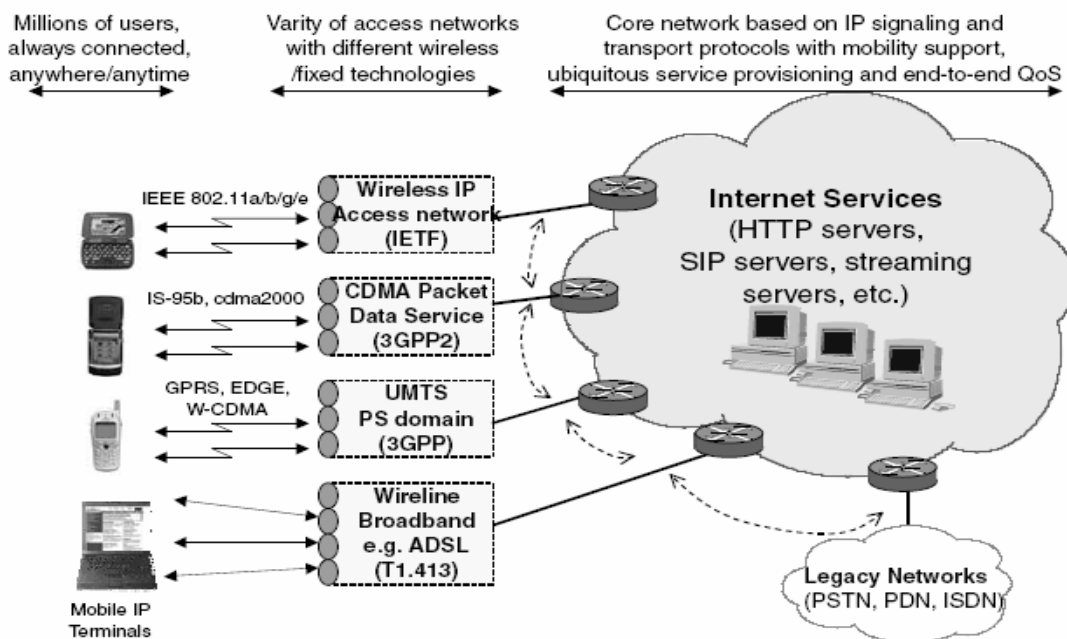
Τα προβλήματα της χωρητικότητας έχουν άμεσες συνέπειες σε πολλές σχεδιαστικές επιλογές. Για παράδειγμα η ασύρματη τεχνολογία που θα χρησιμοποιηθεί θα πρέπει να είναι φασματικά αποδοτική, θα πρέπει να υλοποιεί έναν αποτελεσματικό αλγόριθμο πρόσβασης που θα μπορεί να ικανοποιήσει ένα μεγάλο αριθμό χρηστών και θα μπορεί να αντιμετωπίζει καταστάσεις συμφόρησης.

Είναι σημαντικό να παρατηρήσουμε πως σήμερα οι περισσότερες ερευνητικές προσπάθειες και οι εργασίες προτυποποίησης στοχεύουν στην επίτευξη αυτών των στόχων. Στην πραγματικότητα τα δίκτυα τέταρτης γενιάς (4G) στοχεύουν στις απαιτήσεις του Mobile Internet με κυριότερες τις εξής:

- Αυξημένη χωρητικότητα
- Ποιότητα υπηρεσίας (QoS)

- Κινητικότητα μεταξύ διαφορετικών δικτύων (Mobility)
- Ασφάλεια
- Αύξηση της αποδοτικότητας του TCP/IP
- Ενοποίηση των διαφορετικών τεχνολογιών δικτύων σε ένα ενιαίο εικονικό δίκτυο

Παρόλο που οι ασύρματες επικοινωνίες είναι αυτές που έχουν άμεση σχέση με το Mobile Internet, είναι σημαντικό να αναφέρουμε πως και τα σταθερά δίκτυα και οι αντίστοιχες τεχνολογίες τους (xDSL, Cable modems) έχουν σημαντικό ρόλο στο Mobile Internet. Στην πραγματικότητα το όραμα του Mobile Internet συμπεριλαμβάνει και τις σταθερές και τις ασύρματες τεχνολογίες πρόσβασης (wireless-access and fixed-access technologies) καθώς και μεθόδους ενοποίησης τους σε ένα ενιαίο IP δίκτυο. Η αρχιτεκτονική αυτή παρουσιάζεται παρακάτω:



Σχήμα 2.5 Αρχιτεκτονική Mobile Internet

Ένα τυπικό σενάριο το οποίο δείχνει την σχέση μεταξύ σταθερών και ασύρματων δικτύων είναι όταν ο χρήστης κατεβάζει ένα αρχείο από το Διαδίκτυο χρησιμοποιώντας ADSL τεχνολογία πρόσβασης και όταν έχει ληφθεί το αρχείο κατά το μισό ο χρήστης παίρνει το φορητό υπολογιστή του και μετακινείται προς το γραφείο του με τη χρήση του αυτοκινήτου του. Σε ένα ιδανικό περιβάλλον Mobile Internet το κατέβασμα του αρχείου θα συνεχιστεί απρόσκοπτα καθώς ο χρήστης κινείται από το δίκτυο ADSL του σπιτιού σε ένα δίκτυο κινητής τηλεφωνίας καθώς αυτός κινείται προς το γραφείο του.

2.4.2 Απαιτήσεις για την ανάπτυξη του Mobile Internet

Για την επιτυχή ανάπτυξη του Mobile Internet χρειάζονται ένα πλήθος παραγόντων οι οποίοι παρουσιάζονται παρακάτω:

- Τα δίκτυα κινητών επικοινωνιών θα πρέπει να συμπεριλαμβάνουν ένα δίκτυο κορμού βασισμένο στο IP πρωτόκολλο και πολλά ασύρματα δίκτυα πρόσβασης. Το κλειδί σε αυτή την προσέγγιση είναι πως η σηματοδότηση με το δίκτυο κορμού θα βασίζεται σε IP πρωτόκολλα και πως η αρχιτεκτονική θα είναι ανεξάρτητη από το δίκτυο πρόσβασης (UMTS, CDMA2000, WLAN κλπ). Συνεπώς οι ίδιες IP υπηρεσίες θα προσφέρονται διαφανώς σε κάθε συνδεδεμένο τερματικό. Ένα βασισμένο σε IP δίκτυο κορμού χρησιμοποιεί IP πρωτόκολλα για κάθε λειτουργία όπως για μεταφορά δεδομένων και δρομολόγηση. Η πρώτη εμπορική απόπειρα για την δημιουργία ενός τέτοιου δίκτυο είναι το IP Multimedia Core Network Subsystem (IMS) που προτυποποιήθηκε από το 3GG και 3GG2 [12]. Βέβαια ο

μακροπρόθεσμος στόχος είναι στην κατεύθυνση της ολοκληρωτικής υποστήριξης του IP πρωτοκόλλου από τα δίκτυα κινητών επικοινωνιών. Σε αυτή την προσέγγιση οι σταθμοί βάσης (base stations) σε ένα κυψελωτό δίκτυο κινητής επικοινωνίας θα είναι IP δρομολογητές και η διαχείριση των συνόδων των χρηστών θα βασίζεται σε IP πρωτοκόλλα

- Υποστήριξη πολυμεσικών εφαρμογών οι οποίες θα βασίζονται σε σηματοδοσία επιπέδου εφαρμογής όπως για παράδειγμα μέσω της χρήσης του Session Initiation Protocol (SIP)[22] και του πρωτοκόλλου Hypertext Transfer Protocol (HTTP)[23] πρωτοκόλλων. Τα πρωτόκολλα αυτά υποστηρίζονται από το IP Multimedia Core Network Subsystem (IMS)[6]
- Ύπαρξη ποιότητας υπηρεσίας από άκρο σε άκρο (end to end QoS) για την υποστήριξη απαιτητικών πολυμεσικών εφαρμογών.
- Τα Mobile Internet τερματικά θα πρέπει να υποστηρίζουν πολλαπλές τεχνολογίες πρόσβασης κατά μήκος του τηλεπικοινωνιακού φάσματος μέσω τεχνικών reconfiguration. Επίσης θα πρέπει να υποστηρίζεται η ομαλή μετάβαση από την μια τεχνολογία πρόσβασης στην άλλη.
- Οι απαιτήσεις για ασφάλεια είναι ένα κρίσιμο θέμα στην ανάπτυξη του Mobile Internet. Το πρόβλημα της ασφάλειας σχετίζεται τόσο με τις επιμέρους τεχνολογίες ασύρματης πρόσβασης (πχ WLAN Security) όσο και κατά την διάρκεια των μεταπομπών (handoffs) από το ένα δίκτυο πρόσβασης στο άλλο.

- Ανάγκη ύπαρξης πρωτοκόλλων πιστοποίησης και εξακρίβωσης στοιχείων για τον προσδιορισμό και της χρέωση κάθε χρήστη (Authentication, Authorization, and Accounting (AAA) protocols).
- Υποστήριξη WPAN δικτύων (IEEE 802.15 [24], Bluetooth [25])
- Μεταβίβαση στο πρωτόκολλο IPv6 [9] εφόσον εκατομμύρια χρηστών θα επιθυμούν συνεχή συνδεσιμότητα στο δίκτυο.

3

Δίκτυα πολυμέσων

3.1 Εισαγωγή

Η βελτίωση των ασύρματων δικτύων όσο αναφορά τις υποστηριζόμενες ταχύτητες, την υπολογιστική ισχύ και την αυξανόμενη χωρητικότητα των τερματικών (ιδιαίτερα στα κινητά τηλέφωνα και τα PDA), επιτρέπει την μεταφορά πολυμεσικών δεδομένων, όπως υψηλής ποιότητα ήχο και video, αντί για την μέχρι πρότινος μεταφορά κυρίως κειμένου και εικόνων.

Η υποστήριξη πολυμεσικών υπηρεσιών θέτει νέες προκλήσεις έναντι των κλασικών υπηρεσιών δεδομένων όπως η πλοήγηση στο Διαδίκτυο, η αποστολή και λήψη email κλπ. Πιο συγκεκριμένα οι πολυμεσικές εφαρμογές απαιτούν για την αποστολή και την λήψη περιεχομένου μεγάλο αποθηκευτικό χώρο και εύρος ζώνης. Επιπλέον η ενοποίηση των δικτύων και η ανάπτυξη του Mobile Internet θα οδηγήσει σε έναν τεράστιο αριθμό χρηστών που θα απαιτούν ταυτόχρονα να έχουν στο κινητό, τον υπολογιστή τους και σε κάθε άλλη διαθέσιμη συσκευή που προσφέρει συνδεσιμότητα στο Mobile Internet πολυμεσικό περιεχόμενο υψηλής

ποιότητας και χωρίς καθυστερήσεις [26, 27]. Συνεπώς γίνεται αντιληπτό πως η ανάπτυξη επεκτάσιμων, αξιόπιστων και χαμηλού κόστους πολυμεσικών δικτύων είναι ιδιαίτερα δύσκολη.

Στα παρακάτω ενότητες θα αναλυθούν οι έννοιες των πολυμεσικών δεδομένων και τα σημεία κλειδιά που επιτρέπουν την παροχή πολυμεσικού περιεχομένου (multimedia content delivery) σε ασύρματες συσκευές υπό το πλαίσιο της ενοποίησης των δικτύων και την ανάπτυξη του Mobile Internet.

3.2 Πολυμεσικά δεδομένα

Τα πολυμεσικά δεδομένα αποτελούν δεδομένα του ίδιου ή διαφορετικού τύπου τα οποία παρουσιάζονται μαζί στον τελικό χρήστη. Οι κυριότεροι τύποι πολυμεσικών δεδομένων είναι το κείμενο, τα γραφικά, οι εικόνες, ο ήχος και το video. Μπορούμε να κατατάξουμε αυτούς τους τύπους σε δυο κατηγορίες:

- **Discrete media:** Αναφέρονται σε πολυμεσικά δεδομένα τα οποία δεν έχουν χρονικούς περιορισμούς όσο αναφορά την παρουσίαση τους. Για παράδειγμα στην περίπτωση της ανάκτησης μιας εικόνας από έναν εξυπηρετητή ιστού (Web server) η παρουσίαση της στον φυλλομετρητή του χρήστη (Web browser) μπορεί να διαρκέσει από λίγα δευτερόλεπτα έως περισσότερα αναλόγως το διαθέσιμο εύρος ζώνης του δικτύου και το μέγεθος της εικόνας. Όπως γίνεται κατανοητό ο Web browser ανακτά την εικόνα σε μεταβαλλόμενο κάθε φορά χρόνο και την παρουσιάζει στον χρήστη. Σε αυτή την περίπτωση στόχος είναι να μειωθεί ο χρόνος ανάκτησης της εικόνας

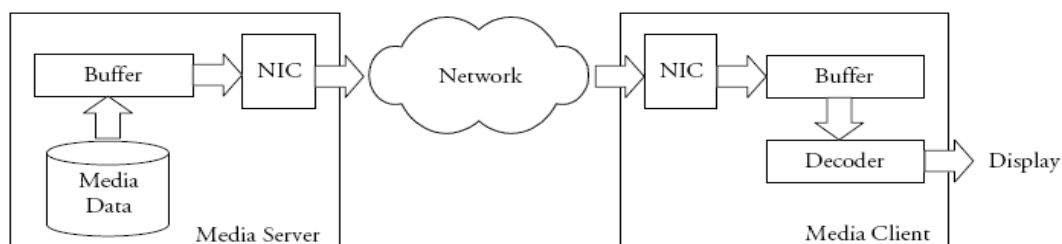
στο ελάχιστο αλλά εφόσον η εικόνα ανακτηθεί και αποκωδικοποιηθεί σωστά και μετέπειτα παρουσιαστεί στο χρήστη η διαδικασία θεωρείται επιτυχής. Με άλλα λόγια δεν υπάρχει κανένας περιορισμός που απαιτεί τα δεδομένα να ανακτηθούν και να παρουσιαστούν εντός ενός χρονικού διαστήματος. Για το λόγο αυτό η δικτυακή κίνηση που προκύπτει από την μετάδοση τέτοιου τύπου πολυμεσικών δεδομένων ονομάζεται ελαστική (elastic traffic)

- **Continuous media:** Αυτά τα πολυμεσικά δεδομένα αντιθέτως από τα Discrete media έχουν σαφείς χρονικούς περιορισμούς όσο αναφορά την παρουσίαση τους. Χαρακτηριστικό παράδειγμα τέτοιων δεδομένων είναι ο ήχος και το video. Για παράδειγμα, τα δεδομένα video κωδικοποιούνται υπό την μορφή πλαισίων (frames) έτσι ώστε να παρουσιαστούν με μια συγκεκριμένη συχνότητα, όπως 25 πλαίσια ανά δευτερόλεπτο (Frame Per Second –FPS-) για την κωδικοποίηση PAL και 29.9 FPS για την κωδικοποίηση NTSC. Για αυτό το λόγο για να μπορεί μια ταινία video να παρουσιαστεί σωστά θα πρέπει εκτός από το να ληφθεί σωστά να αποκωδικοποιηθεί και να παρουσιαστεί με την συχνότητα που έχει κωδικοποιηθεί. Η απαίτηση αυτή θέτει χρονικούς περιορισμούς στην λήψη των δεδομένων. Στην περίπτωση που οι χρονικοί περιορισμοί παραβιαστούν, η ποιότητα του video μειώνεται, κάτι που γίνεται φανερό από αργές κινήσεις και κολλήματα της ταινίας. Συνεπώς στην περίπτωση αυτή έχουμε μη ελαστική δικτυακή κίνηση (inelastic traffic) εφόσον θα πρέπει να τηρηθούν οι χρονικοί περιορισμοί.

Επομένως η πρόκληση στην παροχή πολυμεσικών δεδομένων γενικότερα και πιο συγκεκριμένα στην παροχή Continuous media είναι να διασφαλιστούν οι χρονικοί περιορισμοί. Αντιθέτως με τα Discrete media η ακεραιότητα των δεδομένων είναι λιγότερο σημαντική εφόσον ακόμα και το χάσιμο ορισμένων πολυμεσικών δεδομένων επιτρέπει στον χρήστη να συνεχίσει να παρακολουθεί επιτυχώς τον ήχο ή το video. Επιπλέον τις περισσότερες φορές οι εφαρμογές πολυμέσων συνδυάζουν πολλαπλές ροές πολυμεσικών δεδομένων οι οποίες εμπεριέχουν διαφορετικά πολυμεσικά δεδομένα τα οποία θα πρέπει να συγχρονιστούν για την επιτυχή παρουσίαση τους στον χρήστη. Σε ένα τέτοιο σύστημα πολυμέσων εκτός από ότι είναι αναγκαίο να φροντίσουμε για την ικανοποίηση των χρονικών περιορισμών μόνο για τη μία ροή δεδομένων θα πρέπει να υπάρξει και μέριμνα και για τον συγχρονισμό μεταξύ των επιμέρους ροών.

3.3 Δίκτυα πολυμέσων

Ένα δίκτυο πολυμέσων έχει στόχο την παροχή πολυμεσικών δεδομένων όπως ήχος και video μέσα από ένα τηλεπικοινωνιακό δίκτυο στους τελικούς χρήστες. Στην παρακάτω εικόνα παρουσιάζεται μια υψηλού επιπέδου αρχιτεκτονική για ένα δίκτυο πολυμέσων:



Media Server:

- Media Data
- Storage subsystem
- Buffers
- Network
- Processor

Media Client:

- Media Data
- Media Codec
- Buffers
- Network
- Processor

Σχήμα 3.1 Client – server αρχιτεκτονική συστήματος πολυμέσων

Η αρχιτεκτονική βασίζεται στο μοντέλο client – server για παροχή πολυμεσικών δεδομένων από άκρο σε άκρο. Στην πλευρά του εξυπηρετητή υπάρχουν αποθηκευμένα πολυμεσικά δεδομένα σε κατάλληλη μορφή. Ο εξυπηρετητής πολυμέσων δέχεται τις αιτήσεις του πελάτη και με την χρήση κατάλληλων πρωτοκόλλων αποστέλλει τα πολυμεσικά δεδομένα στον πελάτη. Με τη σειρά του ο πελάτης αποθηκεύει τα δεδομένα στην μνήμη ή τον σκληρό δίσκο, τα αποκωδικοποιεί και τα παρουσιάζει στον χρήστη.

3.4 Παροχή πολυμεσικών δεδομένων

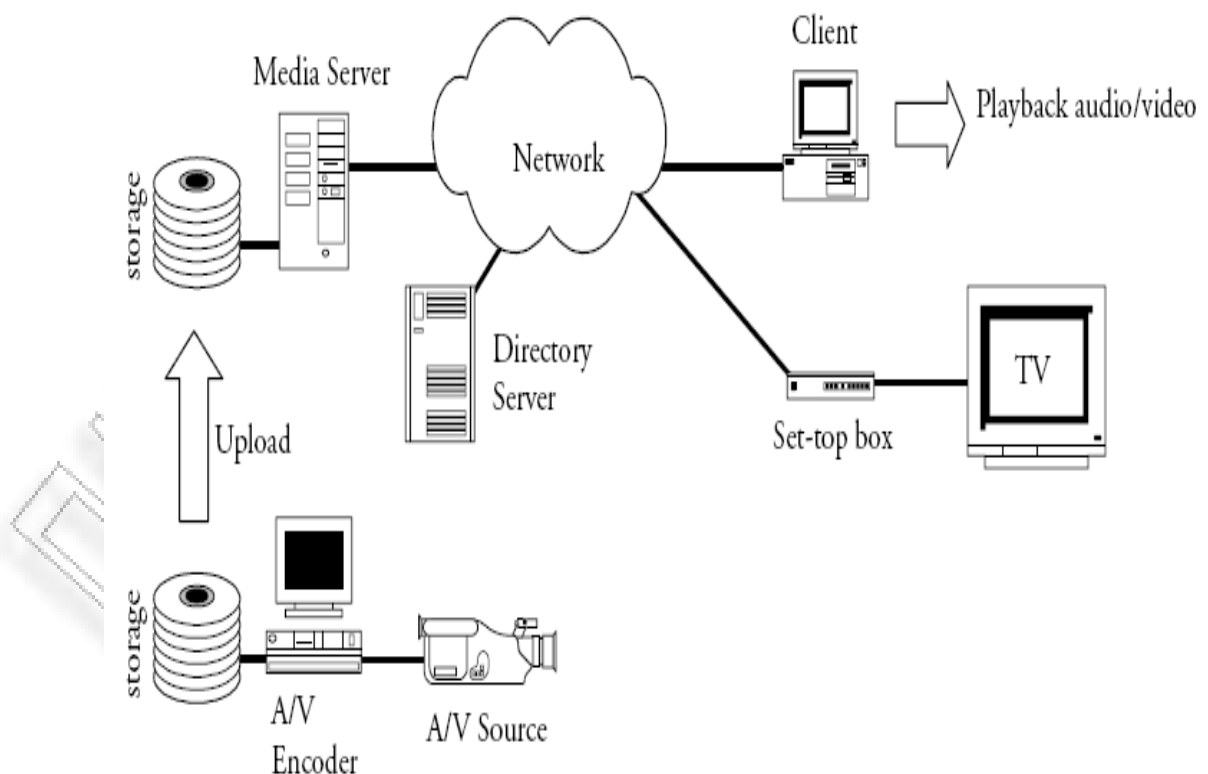
Η μετάδοση πολυμεσικών δεδομένων που ανήκουν στην κατηγορία Discrete media δεν παρουσιάζει κάποιο ιδιαίτερο ενδιαφέρον εφόσον ισχύει η συνθήκη του απαιτούμενου εύρους ζώνης για τα πολυμεσικά δεδομένα που μεταδίδονται. Συνεπώς παρακάτω θα αναλυθεί η παροχή πολυμεσικών δεδομένων που ανήκουν στην κατηγορία Continuous Media δηλαδή η παροχή ήχου και video. Μπορούμε να διακρίνουμε τρεις κατηγορίες:

- **Παροχή αποθηκευμένου ήχου και video (Stored Audio και Video):** Σε αυτή την κατηγορία, οι πελάτες ζητούν την παροχή ήχου και video από εξυπηρετητές που έχουν ήδη προ-αποθηκεύσει τον ήχο και το video σε κατάλληλη μορφή. Για παράδειγμα ο εξυπηρετητής πολυμέσων μπορεί να αποθηκεύει αρχεία ήχου που εμπεριέχουν μουσική, ηχητικές συνεντεύξεις κλπ. Ομοίως ο εξυπηρετητής πολυμέσων μπορεί να αποθηκεύει σε αρχεία video ταινίες, διαλέξεις από συνέδρια, music video clip κλπ. Οι εφαρμογές αυτής της κατηγορίας έχουν τις παρακάτω ιδιαιτερότητες:

- Η προ-αποθήκευση του ήχου και του video επιτρέπει στον χρήστη να διακόψει (pause), να επιστρέψει στην αρχή (rewind) και να προχωρήσει (forward) το σημείο αναπαραγωγής του ήχου ή του video. Ο χρόνος από την στιγμή που ο χρήστης ενεργοποιήσει μία τέτοια λειτουργία έως την στιγμή που αυτή η λειτουργία θα περατωθεί πρέπει να είναι από ένα έως δέκα δευτερόλεπτα για να υπάρχει αποδεκτή αλληλεπιδραστικότητα.
- Υπάρχει δυνατότητα μετάδοσης των δεδομένων με την τεχνολογία streaming όπου η εφαρμογή του πελάτη ξεκινάει την παρουσίαση του video ή το παίξιμο του ήχου λίγα δευτερόλεπτα μετά την αρχή της λήψης του αρχείου. Αυτό σημαίνει πως η εφαρμογή του πελάτη ξεκινάει την παρουσίαση του video ή το παίξιμο του ήχου μέχρι το σημείο που έχει ληφθεί στιγμή ενώ παράλληλα συνεχίζει να κατεβάζει

τα μετέπειτα δεδομένα του ήχου ή του video. Με αυτήν την τεχνική δεν χρειάζεται να κατέβει όλο το αρχείο που εμπεριέχει τον ήχο ή το video με αποτέλεσμα να αποφεύγεται αυτή η καθυστέρηση και η αναπαραγωγή του ήχου ή το video να ξεκινά άμεσα.

Ένα παράδειγμα που ανήκει στην παραπάνω κατηγορία είναι η παροχή ταινιών κατά απαίτηση (Video On Demand). Στην περίπτωση αυτή ο χρήστης μπορεί να επιλέξει να δει κάποια ταινία από μια συλλογή ταινιών που προσφέρονται από ένα εξυπηρετητή ταινιών. Μια γενική αρχιτεκτονική παρουσιάζεται παρακάτω:



Σχήμα 3.2 Video on demand

Οι εφαρμογές αυτές είναι πολύ πιο ανθεκτικές όσο αφορά τον χρόνο εκκίνησης του παιχνιδιού του ήχου και του video από ότι οι αλληλεπιδραστικές εφαρμογές που θα εξεταστούν παρακάτω.

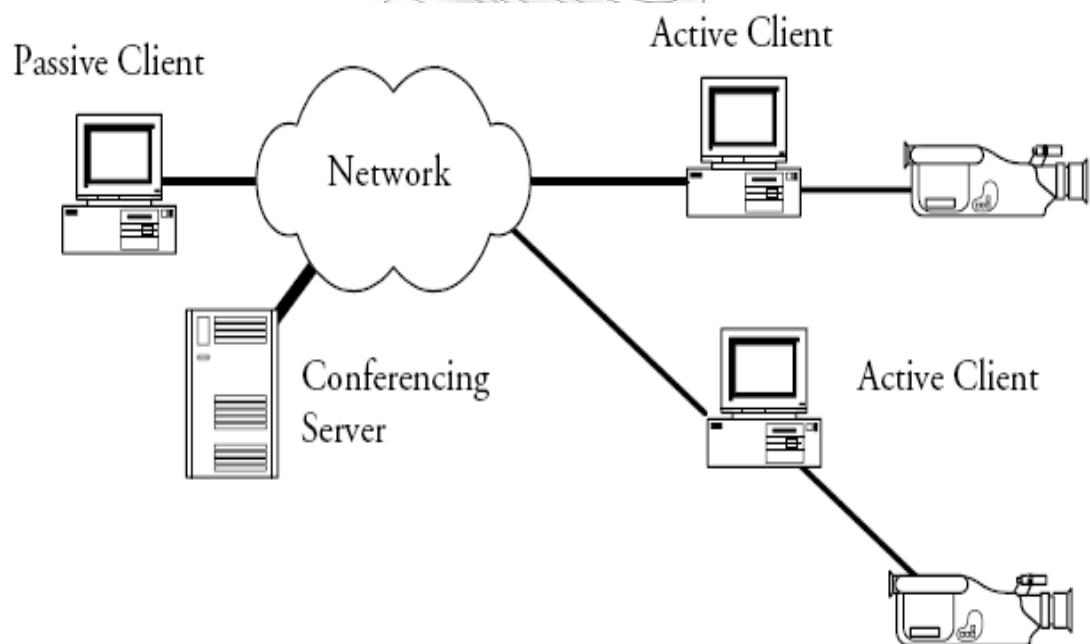
▪ **Παροχή πολυμεσικών δεδομένων σε πραγματικό χρόνο (Live**

Audio and Video): Στην κατηγορία αυτή ανήκουν εφαρμογές οι οποίες είναι παρόμοιες με την παραδοσιακή εκπομπή των προγραμμάτων της τηλεόρασης και του ραδιόφωνου με την κύρια διαφορά ότι η μετάδοση πραγματοποιείται μέσω IP δικτύων όπως το Διαδίκτυο και το Mobile Internet. Οι εφαρμογές αυτές επιτρέπουν στον χρήστη να λαμβάνει ζωντανά προγράμματα της τηλεόρασης και του ραδιόφωνου από όλα τα μέρη του κόσμου μέσω της τεχνικής streaming.

Σε αυτή την περίπτωση εφόσον ο ήχος και το video δεν είναι προ-αποθηκευμένα ο χρήστης δεν μπορεί να εκτελέσει την λειτουργία forward αλλά μπορεί κάνοντας χρήση τοπικής αποθήκευσης των λαμβανομένων δεδομένων να εκτελέσει τις λειτουργίες pause και rewind.

Ακόμα επειδή τις περισσότερες φορές οι ζωντανές τηλεοπτικές και ραδιοφωνικές μεταδόσεις έχουν πολλούς χρήστες οι οποίοι λαμβάνουν το ίδιο περιεχόμενο, η μετάδοση του περιεχομένου μπορεί να χρησιμοποιήσει τεχνικές IP Multicasting. Μια εναλλακτική προσέγγιση είναι η μετάδοση να γίνεται με πολλές ανεξάρτητες συνόδους (Unicast streams)

- **Αλληλεπιδραστικές εφαρμογές πολυμέσων πραγματικού χρόνου (Real time interactive Audio and Video):** Οι εφαρμογές αυτής της κατηγορίας επιτρέπουν στους χρήστες να χρησιμοποιούν ήχο και video για να αλληλεπιδρούν μεταξύ τους σε πραγματικό χρόνο. Στις αλληλεπιδραστικές εφαρμογές μεταξύ πολλών ένας χρήστης μπορεί να μιλήσει ή να κινηθεί οποιαδήποτε στιγμή. Χαρακτηριστικές εφαρμογές αυτής της κατηγορίας είναι το Internet Phone, το Video Conferencing και οι ανερχόμενες τεχνολογίες Voice Over IP (VoIP) [20]. Μια γενική αρχιτεκτονική για την διεξαγωγή Video Conferencing είναι:



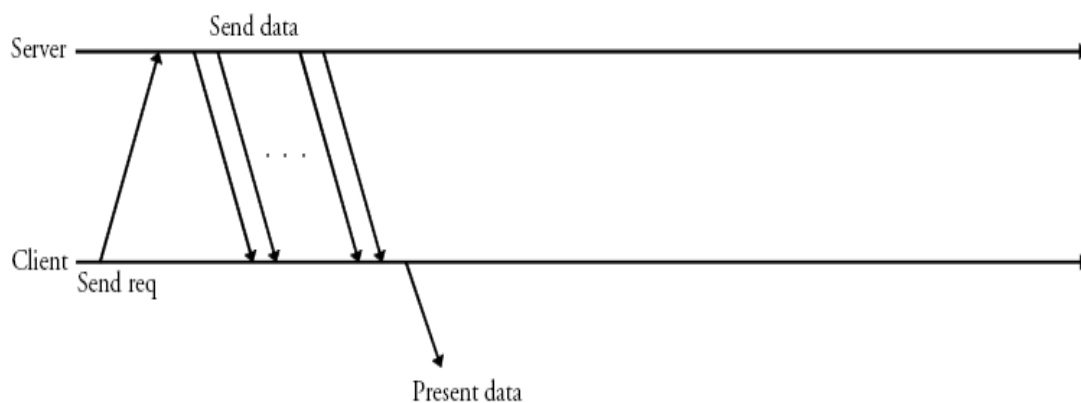
Σχήμα 3.3 Video conferencing

Για να είναι αποδεκτή η ποιότητα μιας τέτοιας εφαρμογής η καθυστέρηση από την στιγμή που ο χρήστης μιλά ή κινείται μέχρι αυτή η ομιλία ή η κίνηση να παρουσιαστεί στους υπόλοιπους

χρήστες θα πρέπει να είναι μικρότερη από μερικές εκατοντάδες ms. Για παράδειγμα στην περίπτωση του Internet Phone, καθυστερήσεις στην φωνή (από την στιγμή που ο χρήστης μιλά μέχρι την στιγμή που θα ακουστεί η φωνή του στο άλλο χρήστη) μικρότερες των 150ms δεν θα γίνουν αντιληπτές από τον χρήστη, καθυστέρηση μεταξύ 150ms και 400ms είναι αποδεκτές και τέλος καθυστερήσεις που ξεπερνούν τα 400ms δεν επιτρέπουν την συνομιλία μεταξύ των χρηστών[28].

3.5 Τεχνική Streaming

Το κατέβασμα των δεδομένων (Download data) αποτελεί την πιο γνωστή μέθοδος μεταφοράς δεδομένων από έναν εξυπηρετητή σε έναν πελάτη. Το μοντέλο αυτό παρουσιάζεται παρακάτω:

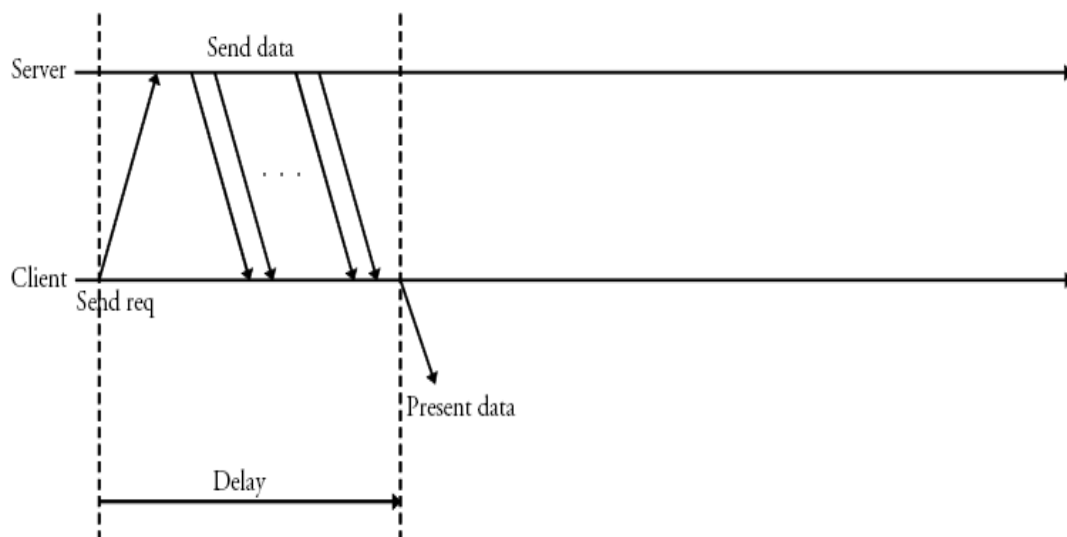


Σχήμα 3.4 Αλληλεπίδραση πελάτη-εξυπηρετητή για το κατέβασμα δεδομένων

Στην περίπτωση αυτή ο πελάτης πρώτα στέλνει ένα αίτημα στον εξυπηρετητή για την ανάκτηση δεδομένων και ο εξυπηρετητής με τη σειρά του στέλνει τα δεδομένα στον πελάτη μέσω δικτύου χρησιμοποιώντας

κλασσικά πρωτόκολλα επιπέδου εφαρμογής και μεταφοράς (πχ FTP/TCP, HTTP/TCP κλπ).

Το βασικό μειονέκτημα σε αυτή την προσέγγιση είναι πως τα δεδομένα θα πρέπει να κατεβούν στο σύνολο τους και να αποθηκευτούν είτε στην μνήμη είτε στον σκληρό δίσκο του πελάτη και μετά να παρουσιαστούν στον πελάτη. Σε αυτή την περίπτωση ο υπολογισμός της καθυστέρησης, αγνοώντας τους επιμέρους υπολογιστικούς χρόνους (processing time) μπορεί να υπολογιστεί από το μέγεθος του μεταδιδόμενου αρχείου και από τον ρυθμό με τον οποίο μεταδίδεται στο δίκτυο. Αυτό φαίνεται παρακάτω:

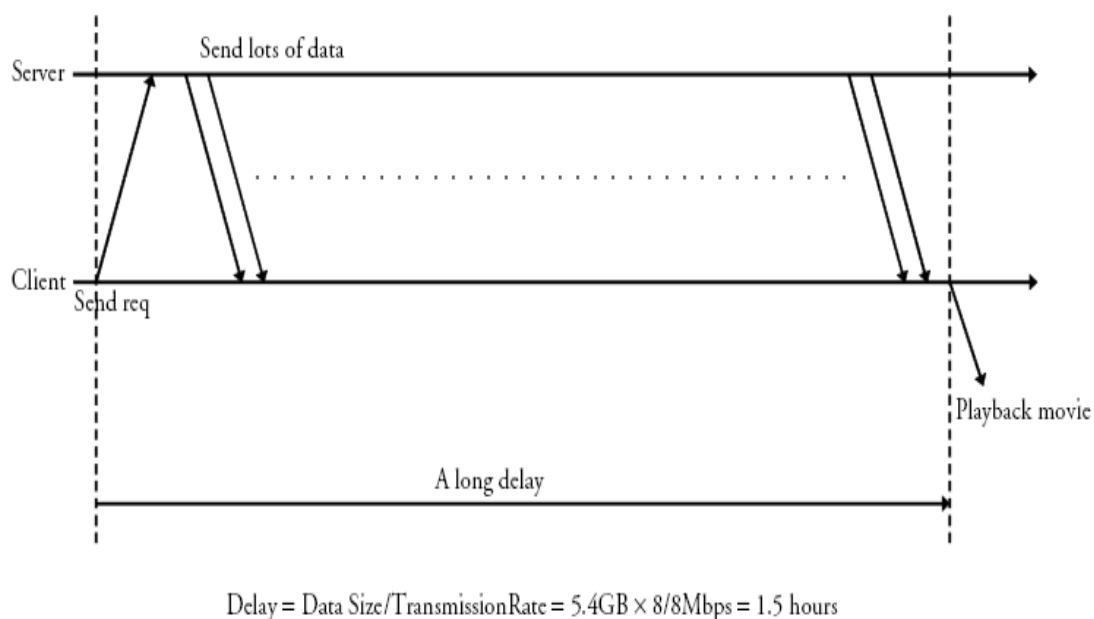


$$\text{Delay} = \text{Data Size} / \text{Transmission Rate}$$

Σχήμα 3.5 Υπολογισμός καθυστέρησης

Στην περίπτωση όπου τα δεδομένα έχουν μικρό μέγεθος (κείμενο, συμπιεσμένες εικόνες), όπως οι κλασσικές ιστοσελίδες του Παγκοσμίου Ιστού, η καθυστέρηση θα είναι μικρή. Στην περίπτωση όμως του τα δεδομένα είναι video και ήχος τα αρχεία ήχου και video είναι πολύ μεγάλα

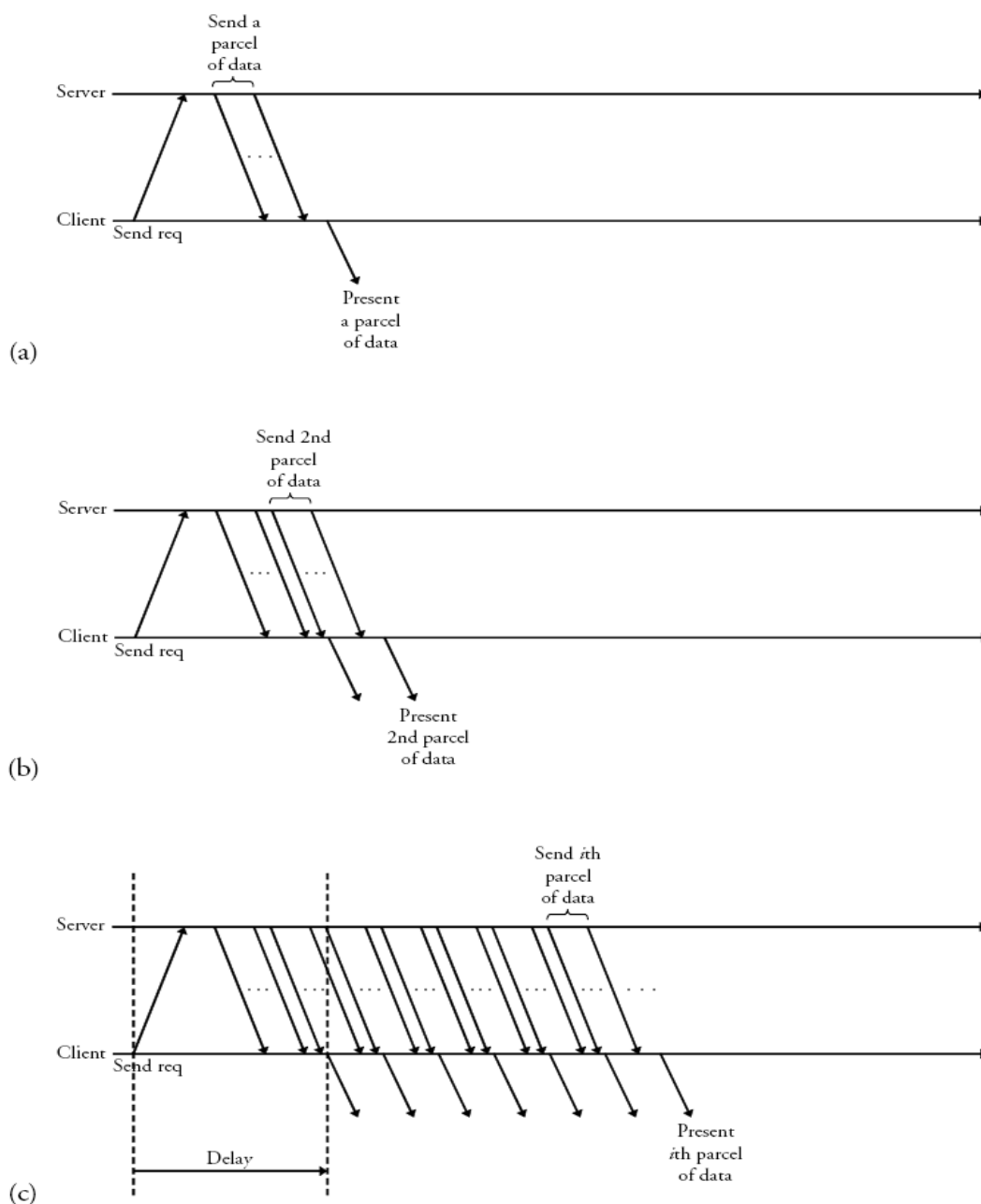
με αποτέλεσμα η καθυστέρηση να είναι υπερβολικά μεγάλη. Για παράδειγμα μια τυπική ταινία video δυο ωρών κωδικοποιημένη κατά το πρότυπο MPEG2 έχει μέγεθος περίπου 5.4GB. Στην περίπτωση που προσπαθήσουμε να μεταδώσουμε την ταινία αυτή ακόμα και μέσω ευρυζωνικού δικτύου με ταχύτητες της τάξης των 8Mbps η αποστολή της ταινίας απαιτεί 1.5 ώρες όπως παρουσιάζεται παρακάτω:



Σχήμα 3.6 Παράδειγμα υπολογισμού καθυστέρησης

Συνεπώς γίνεται κατανοητό πως η προσέγγιση αυτή δεν είναι ιδανική για μετάδοση video και ήχου εφόσον ο χρόνος που θα πρέπει να αναμένει ο πελάτης πριν την αναπαραγωγή της ταινίας είναι υπερβολικά μεγάλος. Το πρόβλημα αυτό μπορεί να λυθεί αν ληφθεί υπόψη πως ο ήχος και το video μπορούν να αποκωδικοποιηθούν και να παρουσιαστούν στον χρήστη χωρίς να έχει κατεβεί το σύνολο του αρχείου. Για παράδειγμα μια ταινία video αποτελείται από πλαίσια (frames), τα οποία μπορούν να αναπαραχθούν

εφόσον έχουν ληφθεί τα δεδομένα κάθε πλαισίου και όχι το σύνολο των πλαισίων που αποτελούν ολόκληρη την ταινία. Το πλεονέκτημα αυτό επιτρέπει την εξέλιξη του κλασσικού μοντέλου κατεβάσματος ενός ολόκληρου αρχείου στο streaming μοντέλο όπου τα δεδομένα αναπαράγονται καθώς η λήψη των δεδομένων είναι σε εξέλιξη. Η προσέγγιση αυτή παρουσιάζεται παρακάτω:



Σχήμα 3.7 Τεχνική streaming

Πιο συγκεκριμένα ο πελάτης μετά την αποστολή του αιτήματος για την λήψη των δεδομένων ήχου ή video περιμένει για την λήψη του πρώτου τμήματος δεδομένων. Εφόσον το λάβει αρχίζει να το αναπαραγάγει στην συσκευή του χρήστη καθώς παράλληλα λαμβάνει το δεύτερο τμήμα δεδομένων. Η διαδικασία αυτή συνεχίζεται μέχρι να ολοκληρωθεί η αποστολή των δεδομένων. Επομένως ο χρόνος εκκίνησης αναπαραγωγής μειώνεται στο ελάχιστο.

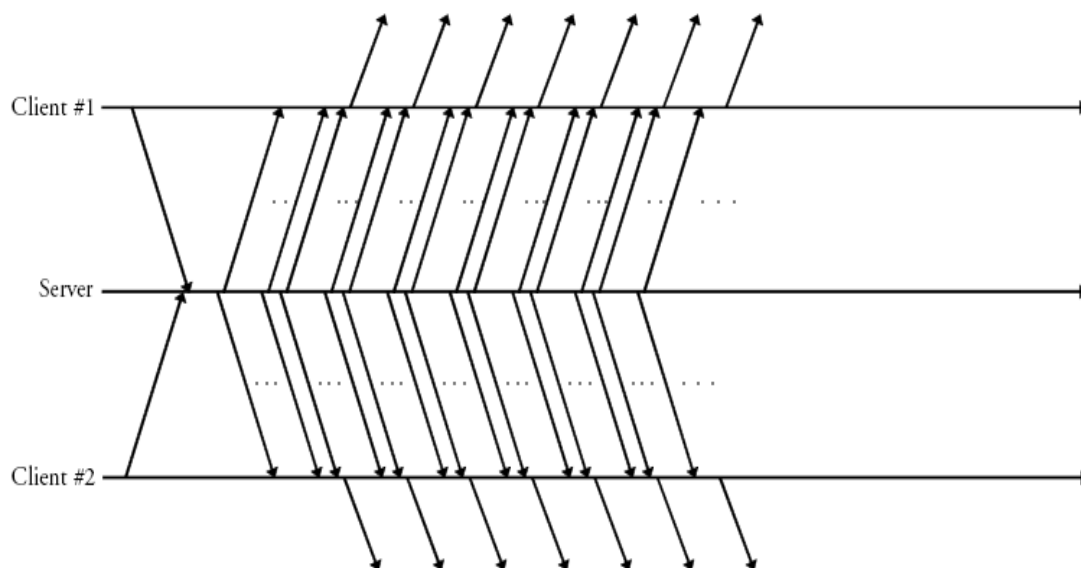
Για να γίνει εφικτή η μετάδοση των δεδομένων μέσω streaming είναι αναγκαίες δυο προϋποθέσεις:

- Το αντικείμενο που μεταδίδεται θα πρέπει να μπορεί να διασπαστεί σε επιμέρους τμήματα τα οποία από μόνα τους ή κάνοντας χρήση των προηγούμενων τμημάτων που έχουν ληφθεί και των επόμενων τμημάτων που θα ληφθούν να μπορούν να αποκωδικοποιηθούν και να παρουσιαστούν στον χρήστη. Στην περίπτωση του ήχου και του video η απαίτηση αυτή ικανοποιείται.
- Για να διασφαλιστούν ο χρονικοί περιορισμοί της αναπαραγωγής του ήχου και του video θα πρέπει να διασφαλιστεί ότι το κάθε τμήμα μπορεί να παραδοθεί στον δέκτη πριν από την χρονική στιγμή που θα γίνει η αναπαραγωγή του.

Επιπρόσθετα οι εξυπηρετητές πολυμέσων θα πρέπει να μπορούν εξυπηρετούν ταυτόχρονα πολλούς πελάτες. Σε αυτή την περίπτωση αν ο εξυπηρετητής δεν εφαρμόζει τεχνικές streaming τότε μπορεί να εφαρμόσει δυο διαφορετικές τεχνικές που όμως κρίνονται ακατάλληλες:

- Η εξυπηρέτηση των πελατών να γίνεται σειριακά. Σε αυτή την περίπτωση όλοι οι πελάτες εκτός από τον πρώτο θα πρέπει να μπουν σε ουρά αναμονής, κάτι το οποίο αυξάνει ακόμα περισσότερο την καθυστέρηση έναρξης αναπαραγωγής του ήχου ή του video.
- Η εξυπηρέτηση των πελατών να γίνεται παράλληλα. Σε αυτή την περίπτωση το εύρος ζώνης του δικτύου μειώνετε αναλόγως του πλήθους των χρηστών με αποτέλεσμα των κατέβασμα των δεδομένων να γίνεται με πιο αργό ρυθμό.

Αντιθέτως τεχνολογία streaming δεν πάσχει από αυτά τα προβλήματα εφόσον η αναπαραγωγή του ή ήχου ή του video γίνεται άμεσα μόλις ληφθεί το πρώτο τμήμα δεδομένων όπως φαίνεται παρακάτω:



Σχήμα 3.8 Streaming σε πολλούς χρήστες

3.5.1 TCP/UDP Streaming

Τα βασικά πρωτόκολλα του Διαδικτύου στο επίπεδο μεταφοράς είναι το TCP [29] και το UDP [30]. Παρακάτω θα εξεταστεί η δυνατότητα της χρήσης της τεχνολογίας streaming με αυτά τα πρωτόκολλα.

3.5.1.1 TCP Streaming

Το TCP είναι ένα πρωτόκολλο με σύνδεση το οποίο παρέχει διόρθωση σφαλμάτων, έλεγχο ροής και έλεγχο συμφόρησης [31, 32]. Το αν το TCP είναι κατάλληλο για streaming εξαρτάται από το εύρος ζώνης, τα χαρακτηριστικά του δικτύου και την επιθυμητή ποιότητα υπηρεσίας. Αν το εύρος ζώνης του δικτύου είναι πολύ μεγάλο σε σχέση με τις απαιτήσεις της εφαρμογής streaming, τότε το TCP θα δουλέψει ικανοποιητικά τις περισσότερες φορές. Για παράδειγμα στην περίπτωση που ο δέκτης ζητήσει δεδομένα ήχου ή video από έναν εξυπηρετητή του παγκοσμίου ιστού (Web server), ο δέκτης εγκαθιδρύει μια σύνδεση TCP με τον εξυπηρετητή ιστού και με τη χρήση του πρωτοκόλλου HTTP [23] αποστέλλει ένα HTTP GET Request στον εξυπηρετητή. Εν συνέχεια ο εξυπηρετητής ανακτά το αντικείμενο που του ζητήθηκε (αρχείο ήχου ή video) και το αποστέλλει ως πολλαπλά HTTP μηνύματα (HTTP Replies) μέσω της ίδιας σύνδεσης TCP. Έπειτα ο Web browser μετά την λήψη μέρους του αρχείου, το αποκωδικοποιεί και το παρουσιάζει στον χρήστη.

Σε αυτή την περίπτωση ο εξυπηρετητής ιστού στην πραγματικότητα δεν χρησιμοποιεί τεχνολογία streaming αλλά αποστέλλει πακέτα TCP όσο πιο γρήγορα μπορεί. Εν συνέχεια ο δέκτης έχοντας αρχίσει να λαμβάνει δεδομένα μπορεί να αρχίζει την αναπαραγωγή του ήχου ή του video χωρίς να περιμένει να κατέβει το σύνολο των δεδομένων. Για όσο η μετάδοση

των δεδομένων μπορεί να ξεπερνά τον ρυθμό με τον οποίο έχουν κωδικοποιηθεί τα δεδομένα (πχ για το video το FPS), η αναπαραγωγή του ήχου ή του video θα είναι ομαλή και χωρίς προβλήματα.

Το streaming με αυτή την τεχνική έχει τα εξής πλεονεκτήματα:

- Ο πάροχος των υπηρεσιών πολυμέσων δεν πρέπει να αγοράσει εξειδικευμένους εξυπηρετητές πολυμέσων εφόσον ένας εξυπηρετητής ιστού είναι αρκετός.
- Η ανάπτυξη εφαρμογών πολυμέσων είναι απλή εφόσον χρησιμοποιούνται τεχνολογίες παγκοσμίου ιστού (Web)
- Η ευρεία υποστήριξη του πρωτοκόλλου HTTP από ασύρματες συσκευές μέσω του πρωτοκόλλου WAP [33] επιτρέπει την μεταφορά της τεχνολογίας και στα ασύρματα δίκτυα.

Όμως η τεχνική αυτή εμφανίζει και αρκετά μειονεκτήματα με κυριότερο τη χαμηλή απόδοση. Στο επίπεδο εφαρμογής ο εξυπηρετητής Παγκοσμίου Ιστού δεν έχει σχεδιαστεί με στόχο την παροχή χρονικά ευαίσθητων δεδομένων όπως τα πολυμεσικά δεδομένα με αποτέλεσμα να μην μπορεί να διατηρήσει μια ομαλή ροή δεδομένων σε περιπτώσεις υψηλού φόρτου. Στο επίπεδο μεταφοράς το TCP παρέχει μηχανισμούς που σε καμία περίπτωση δεν διευκολύνουν την μετάδοση χρονικά ευαίσθητων δεδομένων. Για παράδειγμα ο αλγόριθμος ελέγχου συμφόρησης αναγκάζει την ταχύτητα μετάδοσης να αυξάνεται σιγά σιγά μετά την εγκατάσταση της σύνδεσης (slow-start algorithm [31, 32]), ανεξαρτήτως των απαιτήσεων εύρους ζώνης της υπηρεσίας. Ακόμα ο έλεγχος λαθών που παρέχει το TCP

επιβάλλει την σωστή και σε σειρά μετάδοση των δεδομένων βάση αριθμών ακολουθίας (sequence numbers). Αυτό σημαίνει πως στην περίπτωση που ένα TCP πακέτο χαθεί, ο αποστολέας θα πρέπει να το ξαναστείλει μέχρι να λάβει απόκριση (ACK) από τον δέκτη ότι το έλαβε. Όμως σε μια εφαρμογή πολυμεσικών δεδομένων η λειτουργία αυτή δεν είναι επιθυμητή εφόσον τα τμήματα του ήχου ή του video θα πρέπει να αναπαραχθούν σε προκαθορισμένα χρονικά διαστήματα. Επομένως στην περίπτωση που ένα τμήμα αναμεταδοθεί και φτάσει στον δέκτη μετά το χρονικό διάστημα που έπρεπε, το τμήμα αυτό δεν χρησιμοποιείται στον δέκτη ενώ παράλληλα έχουμε καταναλώσει άσκοπα εύρος ζώνης για την μετάδοσή του. Ένα ακόμα πρόβλημα είναι πως το TCP για κάθε πακέτο που χάνεται θεωρεί πως αυτό οφείλεται στην συμφόρηση του δικτύου. Για να μετριάσει την συμφόρηση μειώνει τον ρυθμό μετάδοσης δεδομένων (reduce congestion window size [31, 32]) με αποτέλεσμα να αυξάνει ακόμα περισσότερο την καθυστέρηση μετάδοσης.

3.5.1.2 UDP Streaming

Το UDP είναι ένα πρωτόκολλο χωρίς σύνδεση το οποίο δεν πάσχει από τα προβλήματα του TCP, καθώς πρόκειται για ένα απλό πρωτόκολλο μεταφοράς δεδομένων που δεν υποστηρίζει έλεγχο ροής, έλεγχο αποτυχίας μετάδοσης και έλεγχο συμφόρησης. Επομένως το πρωτόκολλο UDP δεν εισάγει πρόσθετη καθυστέρηση κατά την μετάδοση επιτρέποντας την ομαλή μετάδοση χρονικά ευαίσθητων δεδομένων.

Ωστόσο κατά την μετάδοση χρονικά ευαίσθητων δεδομένων είναι επιθυμητό να γίνεται έλεγχος ροής και χειρισμός των πακέτων που χάνονται. Το σημαντικό σε αυτές τις λειτουργίες είναι να λαμβάνονται

υπόψη οι χρονικοί περιορισμοί και οι απαιτήσεις σε εύρος ζώνης των δεδομένων που μεταδίδονται. Για τον σκοπό αυτό έχουν αναπτυχθεί ειδικά πρωτόκολλα πάνω από το επίπεδο του UDP. Τα πρωτόκολλα αυτά αναλαμβάνουν να εκτελέσουν λειτουργίες που απαιτούνται για την μετάδοση πολυμεσικών δεδομένων ενώ χρησιμοποιούν το UDP για την μετάδοση των δεδομένων. Τα πρωτόκολλα αυτά αναλύονται στις επόμενες ενότητες.

3.5.2 RTSP Streaming

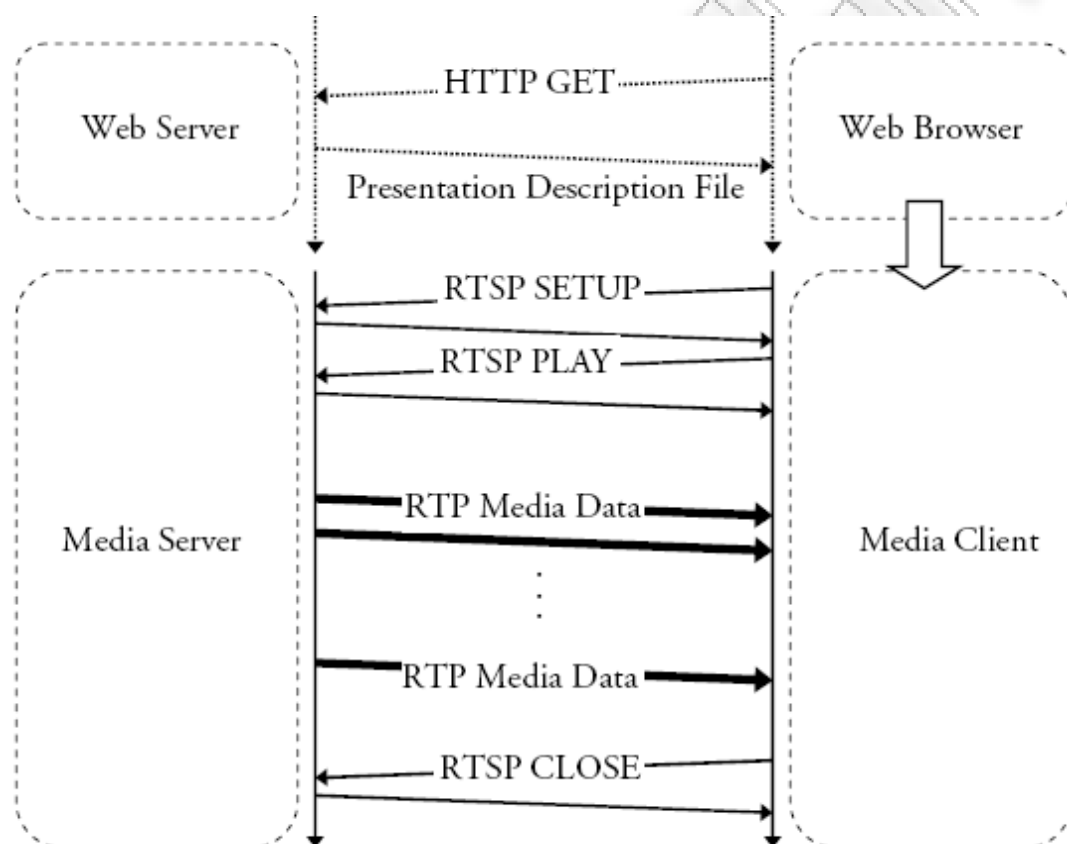
Το RTSP (Real-Time Streaming Protocol) [34] είναι ένα πρωτόκολλο επιπέδου εφαρμογής που επιτρέπει στον δέκτη να ελέγχει την μετάδοση πολυμεσικών δεδομένων. Δηλαδή παρέχει λειτουργίες διακοπής/εκκίνησης (pause/resume) , καθορισμό του σημείου από που θα ξεκινήσει η αναπαραγωγή (playback repositioning), μετάβαση σε επόμενο τμήμα του ήχου ή του video (forward) και επανεκκίνηση της αναπαραγωγής (rewind). Το RTSP πρωτόκολλο παρέχει μόνο αυτές τις λειτουργίες και δεν προκαθορίζει το πρωτόκολλο που θα χρησιμοποιηθεί για την πραγματική μετάδοση των δεδομένων.

Για την πραγματική μετάδοση των δεδομένων μπορεί να χρησιμοποιηθεί το TCP, το UDP είτε οποιοδήποτε άλλο πρωτόκολλο όπως το RTP που περιγράφεται στην επόμενη ενότητα. Η μετάδοση των μηνυμάτων ελέγχου του RTSP χρησιμοποιεί είτε το TCP είτε το UDP και γίνεται σε διαφορετικό κανάλι από το κανάλι μεταφοράς των πολυμεσικών δεδομένων.

Η σύνταξη του RTSP ομοιάζει με την σύνταξη του HTTP αλλά η λειτουργία του διαφέρει σημαντικά. Το RTSP είναι ένα πρωτόκολλο το οποίο απαιτεί από τον εξυπηρετητή να κρατά πληροφορίες για την κατάσταση των

συνδέσεων RTSP που έχει ανοίξει (stateful protocol) με τον πελάτη. Για παράδειγμα ο εξυπηρετητής κρατά πληροφορίες για τον αν ο πελάτης είναι στην φάση αρχικοποίησης, αν έχει αρχίσει την αναπαραγωγή ή εάν έχει διακόψει την εφαρμογή.

Ένα τυπικό σενάριο χρήσης του RTSP παρουσιάζεται και αναλύεται παρακάτω:



Σχήμα 3.9 RTSP Streaming

Αρχικά ο πελάτης ανακτά ένα ειδικό αρχείο το Presentation Description File (PDF) μέσω του πρωτοκόλλου HTTP. Το PDF περιέχει αναφορές σε ένα ή περισσότερα αρχεία ήχου ή video. Επιπλέον περιέχει πληροφορίες για το κάθε αρχείο ήχου και video, όπως για παράδειγμα τον αλγόριθμο

συμπίεσης για να μπορεί η εφαρμογή αναπαραγωγής να χρησιμοποιήσει τον κατάλληλο αλγόριθμο αποσυμπίεσης. Εν συνεχεία ο πελάτης αποστέλλει ένα μήνυμα RTSP SETUP για να αρχικοποιήσει την μετάδοση των πολυμεσικών δεδομένων και ο εξυπηρετητής του απαντά με ένα μήνυμα RTSP OK. Μετά ο πελάτης αποστέλλει ένα μήνυμα RTSP PLAY για να ξεκινήσει η πραγματική μετάδοση των δεδομένων. Ο εξυπηρετητής του απαντά με ένα μήνυμα RTSP OK και αρχίζει την μετάδοση των δεδομένων με την χρήση του πρωτοκόλλου RTP. Κατά την διάρκεια της μετάδοσης ο πελάτης μπορεί να αποστείλει μηνύματα ελέγχου όπως το RTSP PAUSE, το RTSP RECORD κλπ για να εκτελέσει τις αντίστοιχες λειτουργίες. Τέλος ο πελάτης στέλνει το μήνυμα ελέγχου RTSP CLOSE για να τερματίσει την σύνδεση.

3.5.3 RTP Streaming

Το RTP πρωτόκολλο (Real-time Transport Protocol) [35] παρέχει την δυνατότητα μεταφοράς δεδομένων πραγματικού χρόνου (ήχος και video). Το RTP χρησιμοποιεί το πρωτόκολλο UDP για την μεταφορά δεδομένων, το οποίο του παρέχει την δυνατότητα να ανοίγει πολλές διαφορετικές συνδέσεις από τον ίδιο δέκτη (με τη χρήση διαφορετικών UDP θυρών). Επίσης το UDP παρέχει δυνατότητα μεταφοράς δεδομένων τόσο σε συνδέσεις unicast όσο και multicast. Παράλληλα με το RTP έχει οριστεί και το πρωτόκολλο RTCP (RTP Control Protocol) [35] το οποίο παρέχει λειτουργίες ελέγχου όπως συγχρονισμός, αναφορά στατιστικών κλπ.

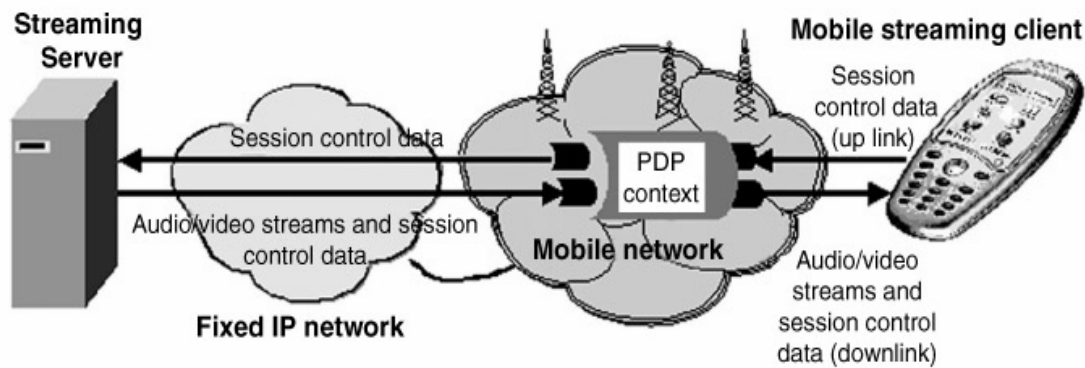
Θα πρέπει να τονιστεί πως το RTP πρωτόκολλο δεν παρέχει έλεγχο της ποιότητας υπηρεσίας και έλεγχο της κατανομής του απαιτούμενου εύρους δικτύου που απαιτείται για την εκάστοτε υπηρεσία. Το RTP προδιαγράφει το

γενικό πλαίσιο, όπως την δομή των πεδίων επικεφαλίδας των πακέτων και τον τύπο των αναφορών απόδοσης της μετάδοσης που παρέχονται μέσω του πρωτοκόλλου RTCP (πχ αριθμός χαμένων πακέτων κλπ). Αυτό επιτρέπει στον κάθε δικτυακό πάροχο να υλοποιήσει τις λειτουργίες ποιότητας υπηρεσίας που ταιριάζουν καλύτερα στην υπηρεσία και το δίκτυο του. Επίσης το RTP [35] δεν προσδιορίζει πως τα πολυμεσικά δεδομένα αποθηκεύονται μέσα στο πακέτο RTP. Το πρόβλημα αυτό λύνεται με άλλες προτυποποιήσεις που αναφέρονται στο [36].

Το RTP παρέχει την δυνατότητα στον αποστολέα των πολυμεσικών δεδομένων να αλλάξει δυναμικά την κωδικοποίηση που χρησιμοποιεί κατά το μέσο μιας RTP συνόδου. Η δυνατότητα αυτή είναι ιδιαίτερα σημαντική εφόσον επιτρέπει την συμπίεση των πολυμεσικών δεδομένων σε δεδομένα χαμηλότερης ροής σε περίπτωση συμφόρησης του δικτύου. Επιπλέον με την χρήση αριθμών ακολουθίας (sequence numbers) σε κάθε πακέτο ο δέκτης είναι δυνατό να συγχρονίζει τα RTP πακέτα ακόμα και αν αυτά έρχονται εκτός σειράς. Τέλος στην περίπτωση που μια ροή πολυμέσων εμπεριέχει πολλές επιμέρους ροές πολυμέσων (για παράδειγμα μια ταινία συμπιεσμένη βάση του προτύπου MPEG εμπεριέχει και video και ήχο), οι επιμέρους ροές πολυμέσων πρέπει να σταλούν σε διαφορετικές RTP συνόδους και μέσω κατάλληλων RTCP μηνυμάτων να συγχρονιστούν.

3.6 Mobile streaming

Η παροχή πολυμεσικών δεδομένων με την εισαγωγή των all-IP δικτύων δεν διαφέρει σε τίποτα από την παροχή πολυμεσικών δεδομένων στο Διαδίκτυο. Παρακάτω παρουσιάζεται και αναλύεται η αρχιτεκτονική για την παροχή πολυμεσικών δεδομένων σε ασύρματες συσκευές:



Σχήμα 3.10 Mobile streaming system

Τα συστατικά που συνθέτουν το Mobile streaming system είναι:

- **Streaming server:** ο οποίος μπορεί να είναι εγκατεστημένος εντός του IP δικτύου του παρόχου ασύρματων υπηρεσιών είτε εκτός του δικτύου του παρόχου (πχ στο Διαδίκτυο). Η τοποθεσία του streaming server είναι ιδιαίτερα σημαντική για τα επίπεδα της ποιότητας υπηρεσίας. Για παράδειγμα αν ο streaming server βρίσκεται στο Διαδίκτυο ο πάροχος ασύρματων υπηρεσιών δεν μπορεί να ρυθμίσει το επίπεδο της ποιότητας υπηρεσίας της σύνδεσης μεταξύ του streaming server και του ασύρματου δικτύου βάση των απαιτήσεων των χρηστών.
- **Ασύρματο δίκτυο (Mobile network):** το οποίο μεταφέρει την κίνηση μεταξύ του streaming server και του τερματικού του χρήστη. Η λογική σύνδεση μεταξύ του streaming server και του τερματικού του χρήστη ονομάζεται PDP Context (Packet Data Protocol) το οποίο περιέχει πληροφορίες για την σύνδεση επιτρέποντας την χρήση διάφορων φυσικών καναλιών (δίκτυα πρόσβασης) για την μεταφορά δεδομένων στο uplink και downlink. Αυτή η υψηλού επιπέδου προσέγγιση είναι δυνατή λόγω της ενοποίησης των δικτύων και την

αποδοχή του πρωτοκόλλου IP όπως περιγράφηκε στο Κεφάλαιο 2 εφόσον πλέον οι υπηρεσιών παροχής πολυμεσικών δεδομένων είναι ανεξάρτητες των τεχνολογιών πρόσβασης.

- **Τερματικό χρήστη (Mobile streaming client):** το οποίο διατηρεί την ασύρματη ζεύξη με το ασύρματο δίκτυο κάνοντας χρήση του των πληροφοριών του PDP Context για την μεταφορά δεδομένων. Τα δεδομένα που λαμβάνονται στο downlink είναι πολυμεσικά δεδομένα καθώς και πληροφορίες για έλεγχο και συγχρονισμό των πολυμεσικών δεδομένων ενώ τα δεδομένα που αποστέλλονται κατά το uplink περιέχουν δεδομένα ελέγχου και αναφορές για την ποιότητα υπηρεσίας. Ο streaming server λαμβάνοντας υπόψη τις αναφορές ποιότητας υπηρεσίας μπορεί να ενεργήσει κατάλληλα για να βελτιώσει την παρεχόμενη ποιότητα υπηρεσίας.

Βέβαια παρά τις ομοιότητες της παροχής πολυμεσικών δεδομένων μεταξύ της παροχής πολυμέσων στο Διαδίκτυο και στα ασύρματα δίκτυα, τα τελευταία εισάγουν και κάποιες νέες παραμέτρους οι οποίες θα πρέπει να ληφθούν υπόψη:

- **Ποιότητα ραδιοζεύξης (Radio link quality):** Στα ασύρματα δίκτυα κάτω από συγκεκριμένες καταστάσεις μπορεί να υπάρξουν υψηλοί ρυθμοί λαθών (Bit Error Rate –BER-) όπως για παράδειγμα κατά την διαδικασία των μεταπομπών ή στην περίπτωση που το τερματικό του χρήστη λαμβάνει ασθενές σήμα (πχ όταν το τερματικό βρίσκεται μέσα σε μια γέφυρα ή πίσω από κτίριο). Οι παράγοντες

αυτοί μπορούν να οδηγήσουν στην αλλοίωση ή το χάσιμο των πακέτων επιβαρύνοντας την μετάδοση των πολυμεσικών δεδομένων.

- **Κινητικότητα(Mobility):** Τα ασύρματα δίκτυα παρέχουν στον χρήστη την δυνατότητα να μπορεί αλλάζει γεωγραφική θέση χωρίς να διακόπτεται η υπηρεσία η οποία εκτελεί. Η δυνατότητα αυτή απαιτεί να λαμβάνει χώρα η διαδικασία της μεταπομπής είτε εντός του ίδιου δικτύου (πχ από κυψέλη σε κυψέλη) είτε από ένα δίκτυο σε ένα άλλο (πχ από GSM σε WLAN). Όμως η διαδικασία της μεταπομπής μπορεί να προκαλέσει διακοπή της υπηρεσία για μικρό χρονικό διάστημα οδηγώντας σε καθυστέρηση ή χάσιμο πακέτων από το τερματικό του χρήστη. Επιπλέον όταν για παράδειγμα το τερματικό του χρήστη μετακινείται σε μια νέα κυψέλη, η κυψέλη αυτή μπορεί να μην μπορεί να παρέχει το ίδιο εύρος ζώνης που παρείχε η προηγούμενη.

Οι αντιμετώπιση αυτών των προβλημάτων λύνεται με την χρήση των πρωτοκόλλων streaming που περιγράφηκαν παραπάνω και τη διαχείριση διαφόρων παραμέτρων (διαστάσεων) που εξετάζονται στην παρακάτω ενότητα.

3.7 Προκλήσεις στον σχεδιασμό δικτύων πολυμέσων

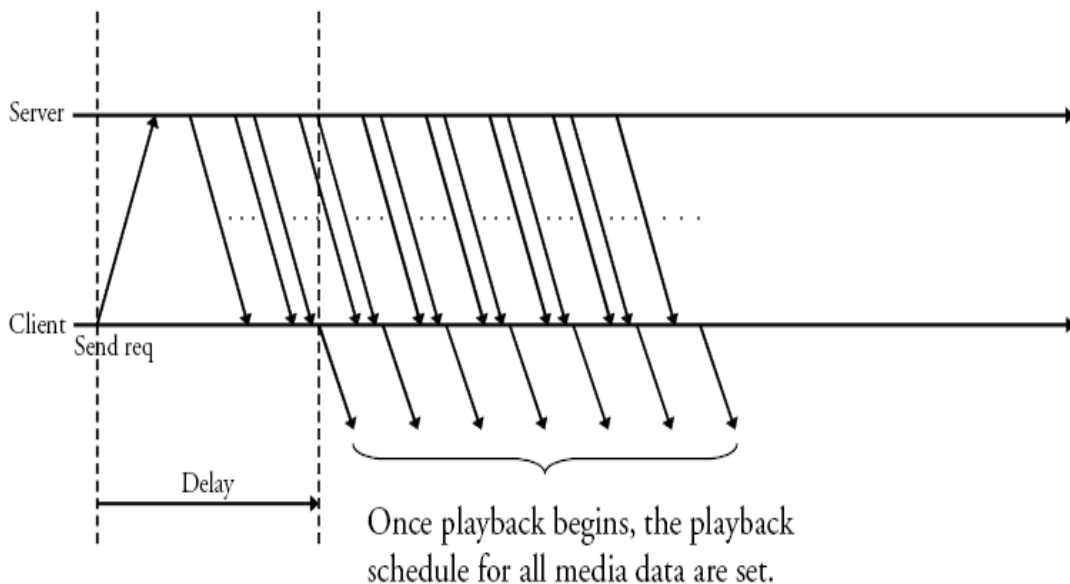
Ο σχεδιασμός ενός δικτύου πολυμέσων αντιμετωπίζει αρκετές δυσκολίες. Στόχος της παρούσας ενότητας είναι να παρουσιάσει αυτά τα προβλήματα και να δώσει λύσεις.

3.7.1 Προβλήματα

Τα κυριότερα προβλήματα παρουσιάζονται παρακάτω:

3.7.1.1 Συνεχή ροή ήχου και video

Η τεχνική streaming απαιτεί από τα τμήματα (πχ video frames) του μεταδιδόμενου ήχου ή video να φτάνουν στην εφαρμογή πελάτη σε καθορισμένους χρόνους έτσι ώστε να υπάρχει συνεχή ροή κατά την αναπαραγωγή. Κρίσιμος παράγοντας σε αυτή την λειτουργία είναι ο χρόνος που θα αρχίσει η αναπαραγωγή μετά την εκκίνηση της διαδικασίας streaming. Όπως παρουσιάζεται παρακάτω μόλις ξεκινήσει η αναπαραγωγή, ο χρόνος αναπαραγωγής των υπολοίπων τμημάτων είναι προκαθορισμένος (σε αυτή την περίπτωση αγνοούμε τις λειτουργίες αλληλεπίδρασης όπως διακοπή, επιστροφή στην αρχή κλπ)



Σχήμα 3.11 Προκαθορισμένοι χρόνοι αναπαραγωγής κατά την τεχνική streaming

Συνεπώς αν η εφαρμογή αναβάλει για κάποιο χρονικό διάστημα t την αναπαραγωγή των δεδομένων τότε και όλη η διαδικασία αναπαραγωγής θα καθυστερήσει κατά t επιτρέποντας να ληφθούν περισσότερα δεδομένα κατά την εκκίνηση και η αναπαραγωγή να είναι ομαλότερη στην συνέχεια.

3.7.1.2 Μεταβαλλόμενοι παράμετροι

Ένα δίκτυο πολυμέσων αποτελεί ένα πολύπλοκο σύστημα που εμπεριέχει πολλά επιμέρους στοιχεία. Μια γενική αρχιτεκτονική παρουσιάζεται στην εικόνα 3.1. Για ένα τυπικό σενάριο μετάδοσης video στους χρήστες ο εξυπηρετητής θα πρέπει να ανακτήσει τα δεδομένα από τον σκληρό δίσκο, να τα τοποθετήσει στην μνήμη και να τα αποστείλει μέσω του δικτύου με χρήση τεχνικής streaming. Από την πλευρά του ο πελάτης θα ανακτήσει τα δεδομένα μέσω της κατάλληλης δικτυακής διεπαφής (κάρτα δικτύου, ασύρματη κάρτα, κεραία κλπ) θα τα αποκωδικοποιήσει και θα τα παρουσιάσει στον χρήστη.

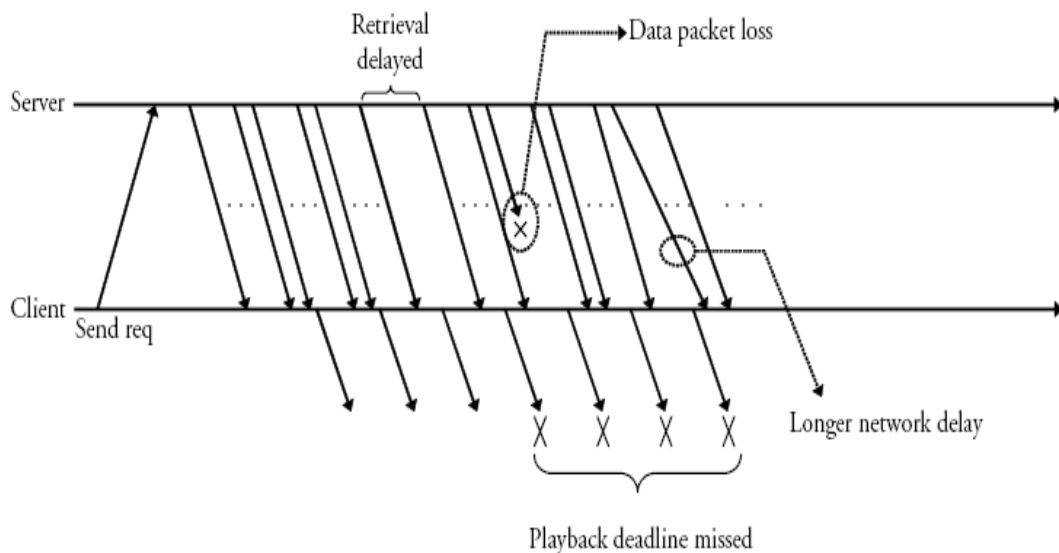
Όλες αυτές οι λειτουργίες έχουν μεταβαλλόμενη χρονική διάρκεια περάτωσης το οποίο καθιστά την μετάδοση των τμημάτων του ήχου ή του video δύσκολη. Πιο συγκεκριμένα οι μεταβαλλόμενες χρονικές καθυστερήσεις μπορεί να προκύψουν από:

- Εφόσον τα δεδομένα έχουν συμπιεστεί ο χρόνος που απαιτείται από την εφαρμογή πελάτη να αποκωδικοποιήσει τα δεδομένα δεν είναι προκαθορισμένος.
- Εφόσον τα δεδομένα ανακτώνται από συσκευές αποθήκευσης (πχ σκληροί δίσκοι) οι χρόνοι ανάκτησης των δεδομένων μπορούν να διαφέρουν αναλόγως τα χαρακτηριστικά της συσκευής, τον βαθμό

χρήσης της αποθηκευτικής συσκευής από άλλες εφαρμογές και τον τρόπο αποθήκευσης των δεδομένων.

- Η καθυστέρηση της μετάδοσης των δεδομένων στο δίκτυο είναι μεταβλητή και αυξομειώνεται αναλόγως το διαθέσιμο εύρος ζώνης και την συμφόρηση του δικτύου. Για δίκτυα που υποστηρίζουν υπηρεσίες ποιότητας υπηρεσίας (QoS) η καθυστέρηση μπορεί να προσδιοριστεί από πριν αλλά σε δημόσια δίκτυα όπως το Διαδίκτυο η καθυστέρηση είναι μεταβαλλόμενη και σε μεγάλο βαθμό απροσδιόριστη.

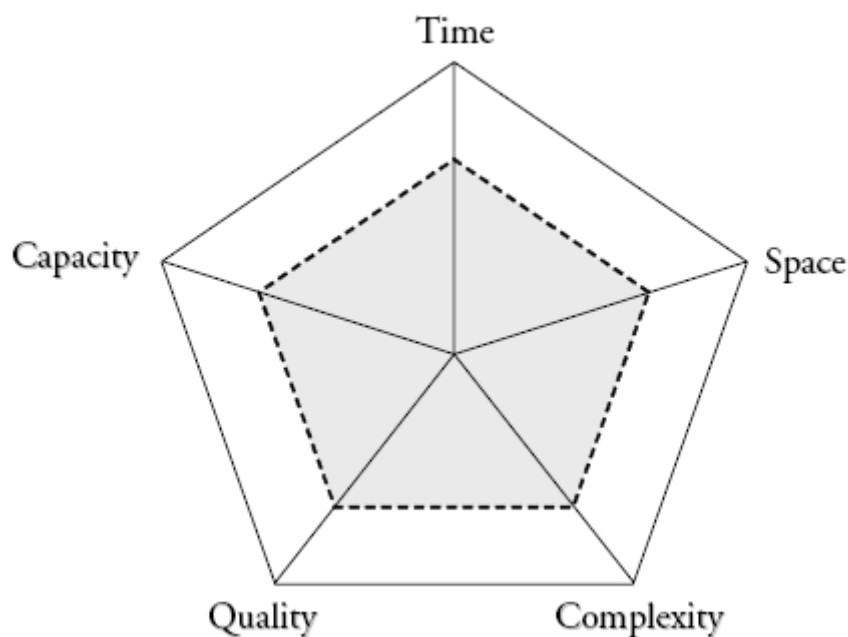
Αποτέλεσμα αυτών των χρονικών διακυμάνσεων είναι να μην επιτρέπεται η ομαλή αναπαραγωγή εφόσον τμήματα του ήχου ή του video μπορούν να φτάσουν εκτός του χρονικού πλαισίου το οποίο έχει τεθεί ως περιορισμός για επιτυχή αναπαραγωγή όπως φαίνεται παρακάτω:



Σχήμα 3.12 Χρονικές διακυμάνσεις στην άφιξη των τμημάτων και αποτυχία ομαλής αναπαραγωγής

3.7.2 Λύσεις

Στην προσπάθεια να επιλυθούν τα παραπάνω προβλήματα συχνά υπάρχουν αλληλοσυγκρουόμενοι παράμετροι οι οποίοι αναλόγως των απαιτήσεων πρέπει να σταθμιστούν με διαφορετικό βαθμό προτεραιότητας. Οι παράμετροι αυτοί μπορούν να χωριστούν σε πέντε διαστάσεις όπως φαίνεται και αναλύεται παρακάτω:



Σχήμα 3.13 Διαστάσεις αλληλοσυγκρουόμενων παραμέτρων

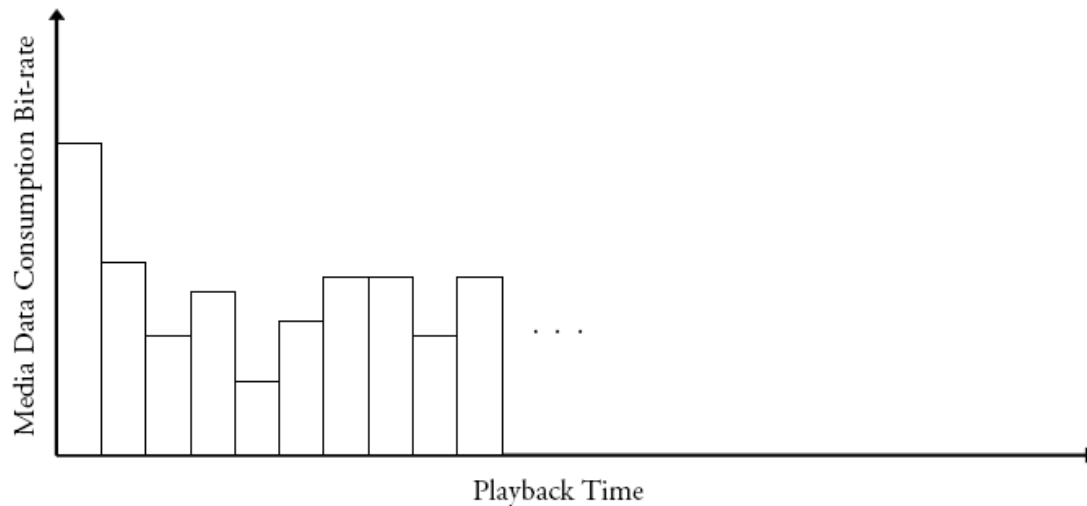
Οι διαστάσεις είναι:

- **Capacity:** Αναφέρεται στον βαθμό χρησιμοποίησης του διαθέσιμου εύρους ζώνης του δικτύου (network utilization) και τον βαθμό διαμεταγωγής δεδομένων εισόδου/εξόδου του συστήματος πολυμέσων (I/O throughput)

- **Time:** Αναφέρεται στον χρόνο εκκίνησης της αναπαραγωγής του ήχου ή του video από την εφαρμογή του πελάτη και τον χρόνο απόκρισης του εξυπηρετητή πολυμέσων στα αιτήματα του πελάτη.
- **Space:** Περιλαμβάνει τα χαρακτηριστικά των συσκευών αποθήκευσης
- **Quality:** Αναφέρεται στην ποιότητα της παρεχόμενης υπηρεσίας και την ποιότητα των πολυμεσικών δεδομένων που προσφέρονται στον τελικό χρήστη
- **Complexity:** Περιλαμβάνει ζητήματα υπολογιστικής πολυπλοκότητας και τον βαθμό δυσκολίας της ανάπτυξης του δικτύου πολυμέσων

Ο κύριος στόχος μας είναι να βρεθεί η εξάρτηση των παραπάνω παραμέτρων με την μετάδοση μέσω δικτύου ήχου ή video από τον εξυπηρετητή πολυμέσων στον πελάτη. Για να βρεθεί αυτή εξάρτηση θα τεθούν διάφορες παραδοχές. Αρχικά θεωρούμε ότι η μετάδοση των πολυμεσικών δεδομένων γίνεται με την τεχνική streaming και ότι τα πολυμεσικά δεδομένα μπορούν να διαχωριστούν σε σταθερού χρόνου τμήματα των T δευτερολέπτων με αποτέλεσμα η διάρκεια αναπαραγωγής του κάθε τμήματος από την εφαρμογή του χρήστη να είναι η ίδια. Για παράδειγμα μια ταινία που πρόκειται να μεταδοθεί μπορεί να αποτελείται από πλαίσια (frames) διάρκειας 30 δευτερολέπτων και συνεπώς το κάθε τμήμα θα πρέπει να φτάνει στον πελάτη σε 30 δευτερόλεπτα ή λιγότερο για να έχουμε μια ομαλή αναπαραγωγή. Αντιθέτως το μέγεθος του κάθε τμήματος δεν είναι το ίδιο. Επιπλέον ορίζουμε ως r_i τον ρυθμό αποστολής

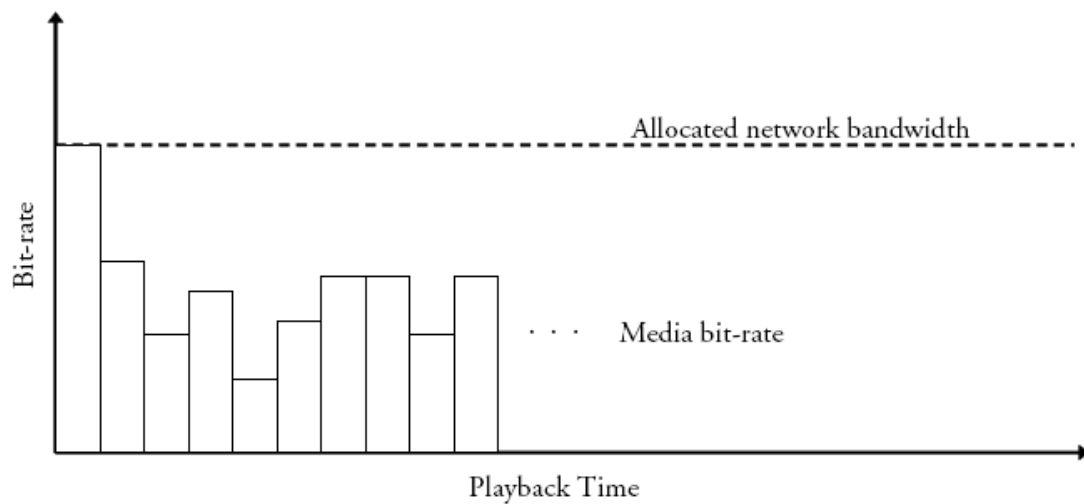
(Bit-rate) με τον οποίο το κάθε τμήμα i αποστέλλεται στον πελάτη. Συνεπώς έχουμε ένα σύστημα μεταβλητού μεγέθους τμημάτων αλλά σταθερού χρόνου αναπαραγωγής όπως παρουσιάζεται παρακάτω:



Σχήμα 3.14 Ροή πολυμέσων με μεταβλητού μεγέθους τμήματα

3.7.2.1 Capacity

Θα θεωρήσουμε για χάρη απλότητας ότι η ροή πολυμεσικών δεδομένων είναι η μόνη κίνηση που υπάρχει στο δίκτυο, το οποίο έχει πεπερασμένο εύρος ζώνης. Η πρώτη φανερή λύση για την απρόσκοπτη μετάδοση των πολυμεσικών δεδομένων είναι ότι εφόσον γνωρίζουμε από πριν τον ρυθμό που απαιτείται για την ομαλή αναπαραγωγή κάθε τμήματος, είναι να κατανείμουμε τόσο εύρος ζώνης στην σύνδεση που μπορεί να υποστηρίξει τον μεγαλύτερο ρυθμό αναπαραγωγής όλων των τμημάτων. Δηλαδή το εύρος ζώνης θα ισούται με $C = \max \{r_i \mid \forall i\}$ όπως φαίνεται παρακάτω:

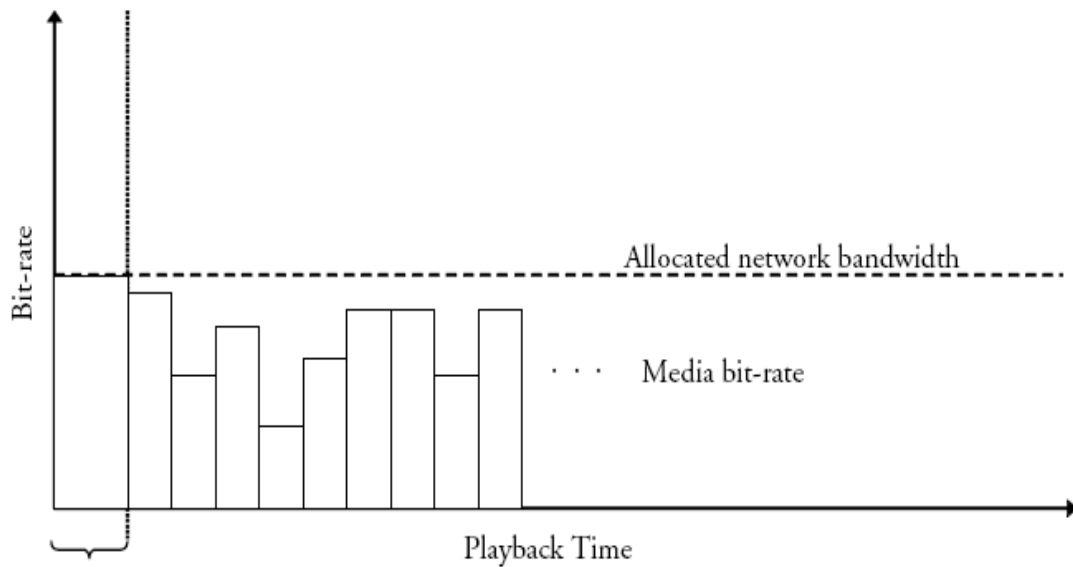


Σχήμα 3.15 Κατανομή εύρους ζώνης βάση του υψηλότερου ρυθμού αναπαραγωγής

Το βασικό μειονέκτημα αυτής της προσέγγισης είναι το εύρος ζώνης του δικτύου θα μένει αχρησιμοποίητο για όλα τα τμήματα του ήχου ή του video εκτός του τμήματος που έχει το μέγιστο ρυθμό.

3.7.2.2 Time

Η αδυναμία της προηγούμενης μεθόδου μπορεί να λυθεί κάνοντας χρήση της διάστασης του χρόνου. Παρατηρώντας στο σχήμα 3.16 ότι το αρχικό τμήμα είναι αυτό που έχει τον υψηλότερο ρυθμό μπορούμε να το μεταδώσουμε με μικρότερο ρυθμό. Αυτό υπονοεί πως το πρώτο τμήμα θα απαιτήσει περισσότερο από T δευτερόλεπτα για να φτάσει στον δέκτη και επομένως η αναπαραγωγή θα πρέπει να καθυστερήσει. Η προσέγγιση αυτή παρουσιάζεται παρακάτω:

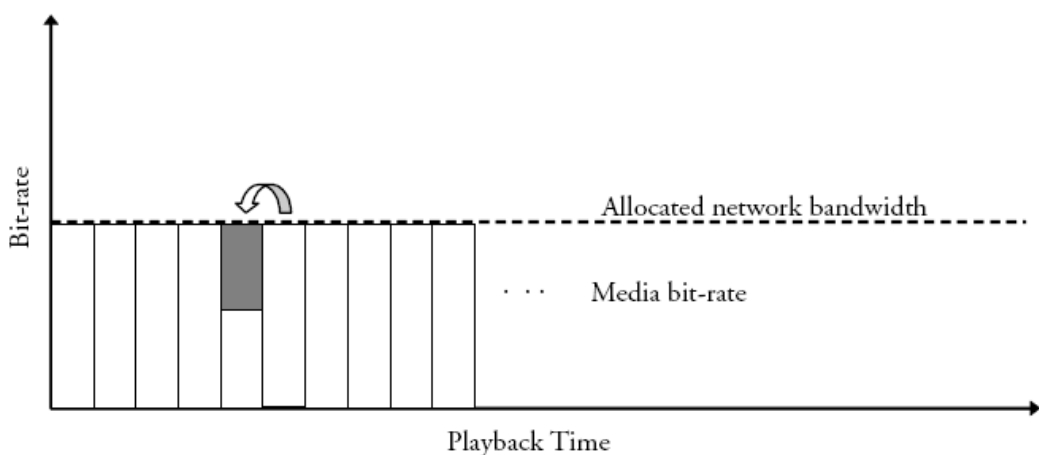
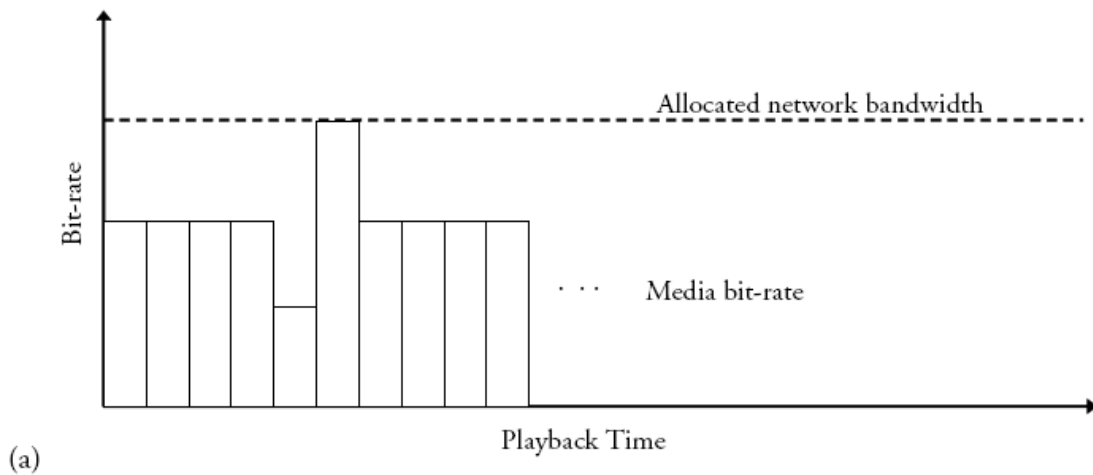


Extra startup delay

Σχήμα 3.16 Καθυστέρηση έναρξης της αναπαραγωγής με μείωση του διαθέσιμου εύρους ζώνης

3.7.2.3 Space

Και στις δυο παραπάνω προσεγγίσεις παρατηρούμε πως μεγάλο μέρος του δικτύου παραμένει ανεκμετάλλευτο. Όμως μπορούμε να εκμεταλλευτούμε το μη χρησιμοποιούμενο εύρος δικτύου επιτρέποντας στον εξυπηρετητή πολυμέσων να αποστείλει τμήματα τα οποία δεν έχουν ακόμα σειρά αναπαραγωγής όπως φαίνεται παρακάτω:



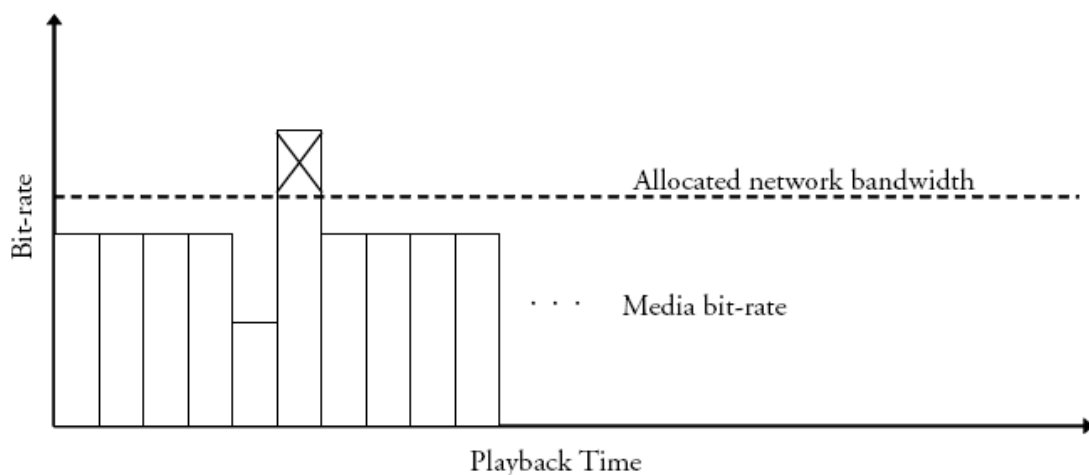
Media data transmitted ahead of playback time will need to be buffered at the client.

Σχήμα 3.17 Αποστολή τμημάτων εκτός σειράς για εξοικονόμηση εύρους ζώνης

Σε αυτή την περίπτωση εφόσον κατά την διάρκεια μετάδοσης του 5^ο τμήματος υπάρχει διαθέσιμο εύρος ζώνης ο εξυπηρετητής πολυμέσων μπορεί να αποστείλει ένα τμήμα του επόμενου τμήματος χωρίς αυτό να επηρεάσει τον χρόνο μετάδοσης του 5^ο τμήματος (χρόνος T). Βέβαια στην πράξη η προσέγγιση αυτή ισούται με το να στείλουμε αμέσως μετά το 5^ο τμήμα το 6^ο. Αυτό ισχύει εφόσον ο χρόνος που απαιτείται για την αποστολή του 5^ο τμήματος είναι μικρότερος του T .

3.7.2.4 Quality

Σε όλες τις παραπάνω προσεγγίσεις υπήρξε η απαίτηση ότι τα πολυμεσικά δεδομένα φτάνουν στον δέκτη στο σύνολο τους. Αν εξαλείψουμε αυτή την απαίτηση επιτρέποντας την αφαίρεση δεδομένων από τα τμήματα τότε επιτυγχάνουμε την μείωση του απαιτούμενου εύρους ζώνης θυσιάζοντας την παρεχόμενη ποιότητα όπως παρουσιάζεται παρακάτω:



Σχήμα 3.18 Μη αποστολή μέρους των πολυμεσικών δεδομένων

Όμως με ημιτελή δεδομένα το λογισμικό αναπαραγωγής του δέκτη δεν θα μπορεί να αποκωδικοποιήσει και να αναπαράγει τον ήχο ή το video. Όμως με κατάλληλους αλγόριθμους κωδικοποίησης το πρόβλημα αυτό εξαλείφεται. Ένας εναλλακτικός τρόπος η δυναμική μετατροπή των τμημάτων σε τμήματα μικρότερου ρυθμού αναπαραγωγής πριν αυτά μεταδοθούν στον δέκτη. Για παράδειγμα στην εικόνα 3.18 εφόσον γνωρίζουμε πως το 6^ο τμήμα απαιτεί ρυθμό μετάδοσης μεγαλύτερο από αυτόν που παρέχεται από το δίκτυο μπορούμε να το κωδικοποιήσουμε με μικρότερο ρυθμό μετάδοσης.

3.7.2.5 Complexity

Τις περισσότερες φορές τα πολυμεσικά δεδομένα που μεταδίδονται είναι συμπιεσμένα. Η επιλογή του αλγόριθμου συμπίεσης είναι κρίσιμης σημασίας. Για παράδειγμα οι πλέον γνωστοί αλγόριθμοι συμπίεσης Video είναι το MPEG2, το MPEG4 και το H.264. Οι αλγόριθμοι αυτοί όσο αφορά τον βαθμό της συμπίεσης που παρέχουν έναντι της ποιότητας video ταξινομούνται από το μικρότερο προς το μεγαλύτερο με τη σειρά που αναφέρθηκαν παραπάνω. Όμως η κωδικοποίηση και η αποκωδικοποίηση με βάση το πρότυπο H.264 απαιτεί υψηλότερη υπολογιστική ισχύ από ότι οι άλλοι δυο αλγόριθμοι. Συνεπώς επιλέγοντας αυτόν τον αλγόριθμο συμπίεσης θα μειωθεί μεν το απαιτούμενο εύρος δικτύου αλλά θα αυξηθεί η πολυπλοκότητα στην συμπίεση και την αποσυμπίεση. Το πρόβλημα αυτό γίνεται περισσότερο φανερό στις αλληλεπιδραστικές εφαρμογές πραγματικού χρόνου εφόσον η συμπίεση και η αποσυμπίεση θα πρέπει να γίνεται σε πραγματικό χρόνο.

4

Multimodal εφαρμογές

4.1 Εισαγωγή

Οι πολύτροπες εφαρμογές (multimodal applications) αναπτύχθηκαν για να εκμεταλλευτούν την φυσική ικανότητα των ανθρώπων να μπορούν να αντλούν πληροφορίες και να ανταποκρίνονται σε αυτές μέσω πολλών διαφορετικών αισθήσεων (όραση, ακοή, αφή, γεύση και οσμή). Πιο συγκεκριμένα οι πολύτροπες εφαρμογές επιτρέπουν την χρήση πολλαπλών φυσικών καναλιών επικοινωνίας (modalities), όπως η φωνή (πχ φωνητικές εντολές), η αφή (πχ πλοήγηση μέσω οθόνων αφής), για να διευκολύνουν την αλληλεπίδραση ανθρώπου μηχανής (Human Computer Interaction) [37].

Τα φυσικά κανάλια επικοινωνίας (modalities) αναφέρονται στην φυσική δυνατότητα του ανθρώπου να αντιλαμβάνεται το περιβάλλον και μπορούν να διαχωριστούν στα παρακάτω:

- Οπτικό (Visual)
- Ακουστικό (Auditive)
- Αισθητό - Απτό (Tactile)

Οι πολύτροπες εφαρμογές επιτρέπουν αλληλεπιδράσεις που συνδυάζουν περισσότερες της μιας αίσθησης είτε στο κανάλι εισόδου (από τον άνθρωπο στην μηχανή) είτε στο κανάλι εξόδου (από την μηχανή στον άνθρωπο) και την χρήση συσκευών επικοινωνίας (πχ μικρόφωνο, πληκτρολόγιο, ποντίκι κλπ) και από τις δύο μεριές. Επίσης οι διαφορετικοί τρόποι αλληλεπίδρασης δεν παρέχονται ανεξάρτητα ο ένας από τον άλλον αλλά παρέχονται ταυτόχρονα και συγχρονισμένα.

Η χρήση πολύτροπων διεπαφών δίνει πολλά πλεονεκτήματα που παρουσιάζονται παρακάτω:

- Υψηλή ταχύτητα και αυξανόμενη αποτελεσματικότητα στην αλληλεπίδραση χρήστη-συσσκευής. Για παράδειγμα η παράλληλη αλληλεπίδραση, όπως η δυνατότητα χρήσης και του πληκτρολογίου και της φωνής για την εκτέλεση εντολών από την συσκευή, επιτρέπει στους χρήστες να έχουν γρηγορότερη πρόσβαση και απόκριση στις πληροφορίες που επιστρέφονται από την συσκευή.
- Παροχή ποικιλίας επιλογών στον χρήστη για το πώς θα αλληλεπιδράσουν με την συσκευή. Για παράδειγμα στην περίπτωση που ο χρήστης οδηγεί μπορεί να χρησιμοποιήσει φωνητικές εντολές για την εκτέλεση των ενεργειών που θέλει να πραγματοποιήσει ενώ αντίθετα αν ο χρήστης βρίσκεται σε ένα θορυβώδες περιβάλλον που

κάνει αναποτελεσματική την χρήση φωνητικών εντολών μπορεί να χρησιμοποιήσει την γραφική διεπαφή. Η εναλλαγή αυτή μεταξύ των πολλαπλών διεπαφών επιτρέπει την αποτελεσματικότερη αλληλεπίδραση μεταξύ χρήστη-συσσκευής με τα λιγότερα δυνατά λάθη. Εν αντιθέσει οι εφαρμογές που προσφέρουν μόνο έναν τρόπο διεπαφής δεν επιτρέπουν την καθολική χρήση τους σε όλα τα περιβάλλοντα. Για παράδειγμα στην περίπτωση που ο χρήστης οδηγεί και παρέχεται μόνο γραφική διεπαφή στο κινητό τηλέφωνο, δεν θα ήταν δυνατό (ή μάλλον θα ήταν επικίνδυνο) να εκτελέσει τις ενέργειες που επιθυμεί.

Αυτά τα πλεονεκτήματα των πολύτροπων εφαρμογών τις καθιστούν ιδιαίτερα δημοφιλής στον χώρο των ασύρματων συσκευών και στην επίτευξη της παροχής υπηρεσιών παντού και πάντα (ubiquitous service) [38]. Αυτό ισχύει εφόσον καθώς οι ασύρματες συσκευές μικραίνουν σε μέγεθος η αλληλεπίδραση των χρηστών με αυτές καθίσταται συνεχώς δυσκολότερη με την χρήση των κλασσικών διεπαφών όπως το πληκτρολόγιο. Η λύση σε αυτό το πρόβλημα μπορεί να επιτευχθεί με την χρήση πολύτροπων διεπαφών.

Κύριος στόχος της παρούσας ενότητας είναι να εστιάσει στην ανάπτυξη ασύρματων πολύτροπων εφαρμογών με έμφαση κυρίως στις τεχνολογίες φωνής (αναγνώριση φωνής και σύνθεση φωνής) σε συνδυασμό με την ανάπτυξη και τον συγχρονισμό τους με κλασσικές γραφικές διεπαφές (GUI).

Βέβαια θα πρέπει να αναφερθεί πως πέρα από τις τεχνολογίες φωνής και τις γραφικές διεπαφές (GUI) υπάρχουν αρκετές ανερχόμενες τεχνολογίες

που προσφέρουν πολύτροπη αλληλεπίδραση που προέρχονται κυρίως από τον τομέα της τεχνητής όρασης και της τεχνητής νοημοσύνης. Για παράδειγμα υπάρχουν συστήματα τα οποία επιτρέπουν την αναγνώριση ανθρώπινων εκφράσεων και χειρονομιών (recognition of facial expressions and gestures) [39], ανάλυση κίνησης των χειλιών (lip movement analysis) και ανίχνευση της κίνησης του ματιού (gaze tracking) [40]. Οι παραπάνω τεχνολογίες δεν θα αναλυθούν περαιτέρω αλλά απλώς θα χρησιμοποιηθούν ως παραδείγματα όπου κρίνεται απαραίτητο για να δείξουν τις δυνατότητες που προσφέρουν οι πολύτροπες εφαρμογές.

4.2 Χαρακτηριστικά πολύτροπων εφαρμογών

Οι πολύτροπες εφαρμογές δεν αποτελούνται απλώς από την τοποθέτηση πολλαπλών διακριτών τρόπων αλληλεπίδρασης στην διεπαφή του χρήστη αλλά βασικός στόχος τους είναι να ενοποιήσουν αυτούς τους διαφορετικούς τρόπους αλληλεπίδρασης προς όφελος του τελικού χρήστη. Ο συνδυασμός των διαφορετικών τρόπων αλληλεπίδρασης μπορεί να λάβει πολλές μορφές και να ικανοποιήσει πολλούς ρόλους. Τα κυριότερα χαρακτηριστικά των πολύτροπων εφαρμογών που λαμβάνουν χώρα κατά την ενοποίηση των διαφορετικών τρόπων αλληλεπίδρασης είναι:

- **Πλεονασμός (Redundancy):** Πρόκειται για τη χρήση δυο ή περισσότερων τρόπων αλληλεπίδρασης για την έξοδο ή την είσοδο της ίδιας πληροφορίας. Για παράδειγμα υπάρχει δυνατότητα να απεικονίζονται οι πληροφορίες μέσω μιας γραφικής διεπαφής (GUI) και να ακούγονται ταυτόχρονα από τα ηχεία της συσκευής του χρήστη (έξοδος πληροφορίας). Αντίστοιχα ο χρήστης μπορεί να

αλληλεπιδρά με τη χρήση του ποντικιού πατώντας κατάλληλα κουμπιά στην γραφική διεπαφή ή ταυτόχρονα να εκτελεί τις ίδιες λειτουργίες μέσω φωνητικών εντολών (είσοδος πληροφορίας). Η χρήση «εφεδρικών» τρόπων αλληλεπίδρασης χρησιμοποιείται για να κάνει αποδοτικότερη και με λιγότερα λάθη την αλληλεπίδραση [41].

- **Συμπληρωματικότητα (Complementarity):** Δυσή ή περισσότεροι τρόποι αλληλεπίδρασης χαρακτηρίζονται ως συμπληρωματικοί όταν ο καθένας από αυτούς οδηγεί στην έξοδο ή την είσοδο ενός μέρους της πληροφορίας αλλά το σύνολο τους οδηγεί στην είσοδο ή την έξοδο του συνόλου της πληροφορίας. Η συμπληρωματικότητα χαρακτηρίζεται από υψηλή ευελιξία εφόσον επιτρέπει στον χρήστη να επιλέξει το μέσο επικοινωνίας το οποίο είναι καλύτερο για τον κάθε τύπο πληροφορίας.
- **Αποσαφηνισμός (disambiguation):** Συνήθως η χρήση ενός μόνο τρόπου αλληλεπίδρασης μπορεί να οδηγήσει σε αμφιβολία για το τι πρέπει να παραχθεί ως είσοδος ή ως έξοδος. Αντιθέτως αν δυο ή περισσότεροι τρόποι αλληλεπίδρασης συνδυαστούν μπορούν να οδηγήσουν σε μια σαφή απόφαση για το τι πρέπει να παραχθεί ως είσοδος ή ως έξοδος. Στην περίπτωση αυτή σταθμίζονται όλες οι υπό αμφιβολία είσοδοι οι έξοδοι για να οδηγηθεί στο σύστημα στην καλύτερη δυνατή λύση.
- **Υποστήριξη (Support):** Περιγράφει την δυνατότητα ενός δευτερεύοντος τρόπου αλληλεπίδρασης να υποστηρίξει έναν άλλο πρωτεύοντα τρόπο αλληλεπίδρασης για να κάνει ευκολότερη και αποδοτικότερη την αλληλεπίδραση ανθρώπου μηχανής. Για παράδειγμα συχνά η αναγνώριση φωνής μπορεί να συνοδεύεται από

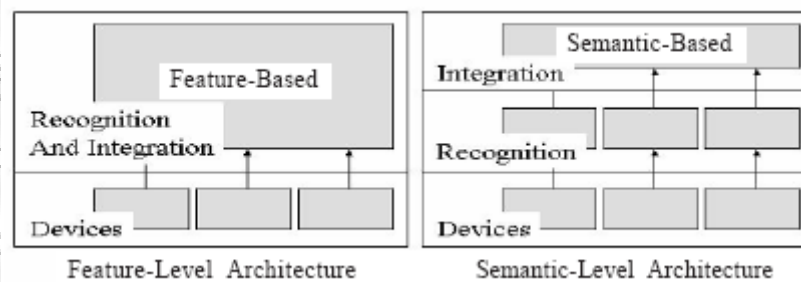
αναγνώριση χειρονομιών (recognition of gestures) για να κάνει ομαλότερη την αλληλεπίδραση ανθρώπου μηχανής.

- **Διαμόρφωση (Modulation):** Διαμόρφωση συμβαίνει όταν το αποτέλεσμα ενός τρόπου αλληλεπίδρασης μπορεί να μεταβάλλει το αποτέλεσμα ενός άλλου τρόπου αλληλεπίδρασης. Για παράδειγμα στην περίπτωση μιας πολύτροπης εφαρμογής αποστολής ηλεκτρονικού ταχυδρομείου το κείμενο που θα σταλθεί μπορεί να έχει διαφορετική μορφή γραμματοσειράς (έντονα γράμματα, κανονικά γράμματα) αναλόγως του τόνου της φωνής του αποστολέα (έντονη ομιλία, κανονική ομιλία).

4.3 Αρχιτεκτονικές λογισμικού για την ανάπτυξη multimodal εφαρμογών

4.3.1 Γενικές αρχιτεκτονικές

Για να επιτευχθεί επιτυχώς ο συνδυασμός διαφορετικών τρόπων αλληλεπίδρασης στην διεπαφή του χρήστη έχουν προταθεί δυο βασικές αρχιτεκτονικές λογισμικού που παρουσιάζονται και αναλύονται παρακάτω:



Σχήμα 4.1 Αρχιτεκτονικές πολύτροπων εφαρμογών

- **Feature level architecture:** Η ενοποίηση των διαφορετικών τρόπων αλληλεπίδρασης πραγματοποιείται κατά την φάση της

επεξεργασίας των πληροφοριών εισόδου ή εξόδου των επιμέρους τρόπων αλληλεπίδρασης. Οι αρχιτεκτονικές αυτές θεωρούνται κατάλληλες όταν υπάρχει συγχρονισμός και ισχυρή σχέση μεταξύ των επιμέρους τρόπων αλληλεπίδρασης, όπως για παράδειγμα μεταξύ της αναγνώριση φωνής και της ανάλυσης της κίνησης των χειριών [41].

- **Semantic level architecture:** Οπού οι πληροφορίες εισόδου ή εξόδου του κάθε τρόπου αλληλεπίδρασης επεξεργάζονται ανεξάρτητα και εν συνεχεία ενοποιούνται. Οι αρχιτεκτονικές αυτές θεωρούνται κατάλληλες όταν παρατηρείται συμπληρωματικότητα ή ανάγκη αποσαφηνισμού (πχ αναγνώριση φωνής και χειρονομιών) [42].

4.4 Τεχνολογίες φωνής για την ανάπτυξη πολύτροπων εφαρμογών

4.4.1 Γενικά

Οι τεχνολογίες φωνής περιλαμβάνουν συστήματα αναγνώρισης ομιλίας και συστήματα σύνθεσης ομιλίας. Η χρήση τους όμως είναι ακόμα περιορισμένη λόγω τεχνικών περιορισμών, όπως οι μεγάλες απαιτήσεις χώρου για ηχητικές πληροφορίες και σε άλλα ακόμα προβλήματα της τεχνολογίας σύνθεσης ομιλίας και αναγνώρισης ομιλίας [43, 44].

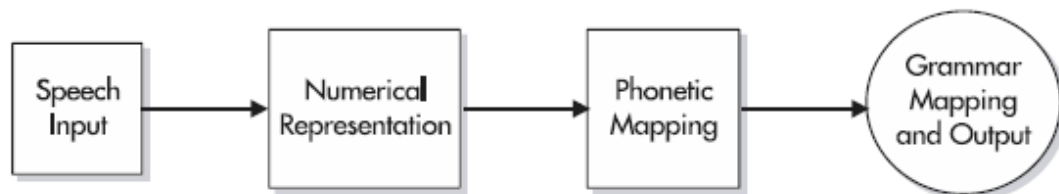
Σήμερα τα συστήματα αυτά έχουν εξελιχθεί από απλά συστήματα που μπορούσαν απλώς να αναγνωρίσουν απομονωμένες λέξεις με την χρήση ενός στατικού λεξικού λέξεων σε συστήματα που μπορούν να αναγνωρίσουν συνεχή λόγο ή επιτρέπουν τη διενέργεια φυσικών διαλόγων

για ένα προκαθορισμένο θέμα (πχ πληροφορίες για αφίξεις-αναχωρήσεις αεροπλάνων κλπ) με ένα αποδεκτό αριθμό λαθών . Βέβαια δεν παύουν να υπάρχουν ακόμα πλήθος προβλημάτων που βρίσκονται υπό έρευνα.

Συνεπώς οι τεχνολογίες φωνής αποτελούν ένα βασικό συστατικό των πολύτροπων εφαρμογών όσο αναφορά την χρήση του ακουστικού καναλιού κατά την αλληλεπίδραση ανθρώπου με τον υπολογιστή ή την ασύρματη συσκευή. Το ακουστικό κανάλι μπορεί να χρησιμοποιηθεί είτε σαν έξοδος του υπολογιστή ή της ασύρματης συσκευής (παραγωγή/απόκριση/ομιλία) είτε σαν είσοδος (αναγνώριση ομιλίας).

4.4.2 Αναγνώριση ομιλίας

Πρόκειται για την δυνατότητα να εισάγουμε πληροφορίες ή να δίδουμε εντολές σε υπολογιστές μέσω της φωνής. Τα βήματα της διαδικασίας αναγνώρισης ομιλίας παρουσιάζονται παρακάτω:



Σχήμα 4.2 Διαδικασία αναγνώρισης ομιλίας

Τα επιμέρους βήματα της διαδικασίας αναγνώρισης ομιλίας αναλύονται παρακάτω:

- **Speech input & Numerical Representation:** Αναφέρεται στην ψηφιοποίηση της ομιλίας του χρήστη.
- **Phonetic mapping:** Στο βήμα αυτό η ψηφιοποιημένη ομιλία συγκρίνεται με την προκαθορισμένη γραμματική της εφαρμογής και

αναλόγως του βαθμού ομοιότητας επιλέγεται η αντίστοιχη λέξη της γραμματικής.

Τα πλεονεκτήματα χρήσης φωνής στην αλληλεπίδραση ανθρώπου μηχανής είναι πολλά εφόσον η χρήση της ομιλίας είναι πιο εύκολη και φυσική και επίσης διευκολύνει την επικοινωνία ανθρώπων με ειδικές ανάγκες. Δυστυχώς η πρόοδος στην περιοχή αυτή δεν είναι τέτοια ώστε η ομιλία να αποτελεί τον συνήθη τρόπο επικοινωνίας ανθρώπου μηχανής. Αυτό οφείλεται στην δυσκολία του εγχειρήματος αυτόματης αναγνώρισης ομιλίας λόγω της πολυπλοκότητας της ανθρώπινης ομιλίας (μεταφορικός λόγος, μη δομημένος λόγος κλπ), λόγω της συχνής ύπαρξης περιβαλλοντικού θορύβου, την ιδιαιτερότητα στην χροιά της φωνής διαφορετικών ομιλητών κλπ. Επιπλέον πολλοί ερευνητές πιστεύουν ότι οι χρήση της ομιλίας σε μεγάλη έκταση δεν θα γίνει ποτέ εφικτή ούτε και είναι επιθυμητή. Σύμφωνα με την άποψη αυτή οι εντολές φυσικής γλώσσας αν και είναι ο φυσικός τρόπος επικοινωνίας μεταξύ ανθρώπων δεν διευκολύνουν την επικοινωνία ανθρώπου μηχανής επειδή εισάγουν μεγάλο βαθμό αβεβαιότητας στον χρήστη [45].

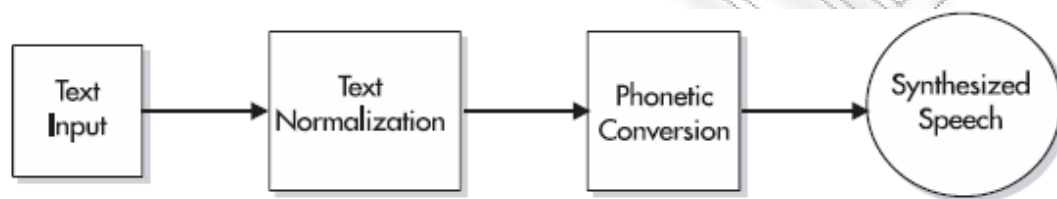
Όμως τα συστήματα αναγνώρισης ομιλίας μπορούν να χρησιμοποιηθούν επιτυχώς σε ειδικές εφαρμογές με περιορισμένα λεξιλόγια. Ανάλογα με την μορφή της ομιλίας τα συστήματα αυτά κατηγοριοποιούνται στα:

- Συστήματα αναγνώρισης διακριτών λέξεων
- Συστήματα αναγνώρισης συνεχούς λόγου

Επίσης τα συστήματα αναγνώρισης ομιλίας μπορούν να διακριθούν σε συστήματα που εξαρτώνται από τον ομιλητή, τα οποία εκπαιδεύονται να αναγνωρίζουν τις ιδιαιτερότητες της φωνής του συγκεκριμένου ομιλητή ή σε συστήματα ανεξάρτητα από τον ομιλητή.

4.4.3 Σύνθεση ομιλίας

Η σύνθεση ομιλίας παρουσιάζει παρόμοια προβλήματα με αυτά της αναγνώρισης ομιλίας που αναφέρθηκαν παραπάνω. Η διαδικασία της σύνθεσης ομιλίας ακολουθεί τα παρακάτω βήματα:



Σχήμα 4.3 Διαδικασία σύνθεσης ομιλίας

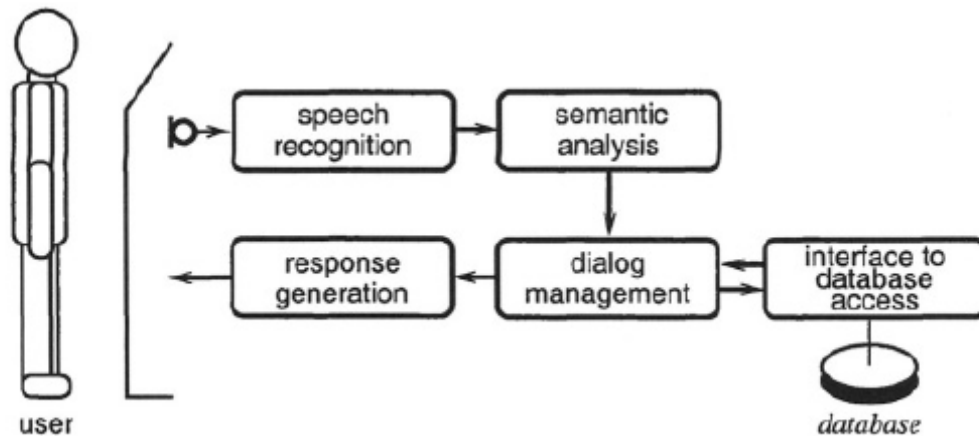
Τα επιμέρους βήματα της διαδικασίας σύνθεσης ομιλίας αναλύονται παρακάτω:

- **Text input & Text Normalization:** Η κύρια είσοδος είναι υπό μορφή κειμένου. Εν συνεχεία λαμβάνει χώρα η διαδικασία του Normalization όπου οι συντομεύσεις, τα αριθμητικά δεδομένα και τα σύμβολα μετατρέπονται σε μορφή κειμένου. Για παράδειγμα στην περίπτωση που το κείμενο εισόδου περιέχει το κείμενο «52%» μετατρέπεται στην πλήρη μορφή δηλαδή «52 τις εκατό»
- **Phonetic conversion:** όπου κάθε λέξη μετατρέπεται στο φωνητικό ισοδύναμο της. Υπάρχουν δυο μέθοδοι που χρησιμοποιούνται για αυτή τη λειτουργία:

- **Η διασύνδεση προ-ηχογραφημένων αποσπασμάτων ανθρώπινης ομιλίας** με τα αποσπάσματα αυτά να αφορούν είτε ολόκληρες προτάσεις, λέξεις οι στοιχειώδης συλλαβές. Σε αυτή την περίπτωση η σύνθεση της πρότασης γίνεται από τον υπολογιστή ή την ασύρματη συσκευή και παρατηρούνται προβλήματα όσο αφορά τον φυσικό τονισμό της ομιλίας. Αποτέλεσμα αυτού είναι η ομιλία να είναι μονότονη και χωρίς φυσικότητα. Όμως η μέθοδος αυτή είναι αποτελεσματική και χρησιμοποιείται σε περιπτώσεις που ο αριθμός των εκφερόμενων προτάσεων είναι περιορισμένος και καθορισμένος.
- **Η μέθοδος σύνθεσης ομιλίας από στοιχειώδη φωνήματα με βάση φωνητικούς κανόνες.** Τα φωνήματα είναι οι διακριτοί ήχοι που χρησιμοποιούνται σε μια γλώσσα και δεν αντιστοιχούν κατά ανάγκη στα γράμματα της γλώσσας. Για παράδειγμα το φώνημα *i* μπορεί να παραχθεί από διαφορετικούς συνδυασμούς γραμμάτων και διφθόγγων, ενώ το γράμμα *ψ* αντιστοιχεί σε δύο φωνήματα $\langle p,s \rangle$. Η ομιλία αυτή είναι ακόμα πιο συνθετική και πιο τεχνητή με το πλεονέκτημα όμως ότι δεν υπάρχει περιορισμός στον αριθμό λέξεων που μπορούν να παραχθούν. Και σε αυτή την περίπτωση ο τονισμός της ομιλίας είναι εξαιρετικά δύσκολος αφού προϋποθέτει κατανόηση της σημασιολογίας της πρότασης. Το ίδιο πρόβλημα υπάρχει και κατά τις παύσεις της ομιλίας (για παράδειγμα στην κατανόηση της μαθηματικής σχέσης $4X+5$ από τη $4(X+5)$)

4.4.4 Συστήματα διαλόγων ομιλίας

Τα συστήματα διαλόγων συνδυάζουν όλες τις τεχνολογίες φωνής για να δημιουργήσουν ένα ολοκληρωμένο σύστημα για την δημιουργία πολύτροπων εφαρμογών φωνής που βασίζονται σε φωνητικούς διαλόγους. Τα συστήματα διαλόγων περιέχουν λογισμικό για αναγνώριση ομιλίας, για σημασιολογική ανάλυση, για διαχείριση των διαλόγων, για αλληλεπίδραση με βάσεις δεδομένων και λογισμικό σύνθεσης φωνής όπως παρουσιάζεται παρακάτω:



Σχήμα 4.4 Σύστημα διαλόγων ομιλίας

Πιο συγκεκριμένα τα διακριτά συστατικά (πχ λέξεις ή γράμματα) που αποτελούν την ομιλία του χρήστη αναγνωρίζονται κατά την φάση της αναγνώρισης ομιλίας (speech recognition). Εν συνεχεία στην φάση της σημασιολογικής ανάλυσης γίνεται η ενοποίηση και κατανόηση των διακριτών συστατικών. Συνήθως η φάση της σημασιολογικής ανάλυσης είναι αυξητική καθώς ο χρήστης και ο υπολογιστής αλληλεπιδρούν μέσα από τους φωνητικούς διαλόγους. Στην περίπτωση που οι πληροφορίες που έχουν ληφθεί μέχρι εκείνη την στιγμή δεν είναι επαρκής για να μπορέσει το σύστημα να εκτελέσει μια από τις λειτουργίες του, ο ελεγκτής διαλόγων

(dialog manager) αναλαμβάνει να ρωτήσει τον χρήστη για να αποσαφηνίσει την επόμενη ενέργεια. Τέλος οι πληροφορίες και οι αποκρίσεις του συστήματος παρέχονται φωνητικά ή οπτικά στον χρήστη (response generation)

4.5 Προκλήσεις στην ανάπτυξη πολύτροπων εφαρμογών

Η πρόσφατη ανάπτυξη των πολύτροπων εφαρμογών και η εξέλιξη στις τεχνολογίες αναγνώρισης (πχ αναγνώριση φωνής, αναγνώριση χειρονομιών κλπ) δημιούργησε πολλαπλές νέες προκλήσεις στον σχεδιασμό πολύτροπων εφαρμογών. Στις παρακάτω ενότητες παρουσιάζονται και αναλύονται αυτές οι προκλήσεις.

4.5.1 Σχεδιασμός πολύτροπων εφαρμογών

Ο σχεδιασμός συστημάτων που θα αξιοποιούν αυτές τις νέες δυνατότητες αλληλεπίδρασης είναι ιδιαίτερα δύσκολος. Αυτό οφείλεται στην έλλειψη γνώσης για το πώς οι διαφορετικοί τρόποι αλληλεπίδρασης μπορούν να ενοποιηθούν με τον καλύτερο δυνατό τρόπο σε μια διεπαφή με αποτέλεσμα να δημιουργούνται διεπαφές μη φιλικές προς τον χρήστη. Επίσης το πρόβλημα γίνεται ακόμα χειρότερο εφόσον οι περισσότερες μελέτες για την κατανόηση της ενοποίησης των διαφορετικών μορφών αλληλεπίδρασης προέρχονται από τον ερευνητικό χώρο της ψυχολογίας παρέχοντας μόνο ποιοτικά στοιχεία [37].

4.5.2 Υλοποίηση πολυτροπων εφαρμογών

Η ανάπτυξη πολύτροπων εφαρμογών είναι ιδιαίτερα δύσκολη και εμφανίζει πολλές διαφορές σε σχέση με τις παραδοσιακές γραφικές διεπαφές[46]. Τα κυριότερα προβλήματα παρουσιάζονται παρακάτω:

- Ανάγκη επεξεργασίας πληροφοριών που προέρχονται από πολλά διαφορετικά κανάλια αλληλεπίδρασης
- Συγχρονισμός των διαφορετικών καναλιών αλληλεπίδρασης τα οποία λειτουργούν παράλληλα (modalities fusion)
- Διαχωρισμός και παρουσίαση των πληροφοριών εξόδου για κάθε κανάλι αλληλεπίδρασης (modality fission)
- Χειρισμός των λαθών του χρήστη

4.5.3 Αβεβαιότητα στις επιλογές του χρήστη

Η φυσική αλληλεπίδραση, όπως η χρήση της φωνής ή των χειρονομιών βασίζονται σε τεχνολογίες που είναι επιρρεπής σε λάθη. Για παράδειγμα τα συστήματα αναγνώρισης φωνής εξαρτώνται άμεσα από το μέγεθος του λεξικού που περιλαμβάνει τις λέξεις ή φράσεις που θα αναγνωρίζονται, από την ποιότητα του φωνητικού σήματος καθώς και από παραμέτρους της φωνής (τόνος, χροιά κλπ)[47]. Επιπρόσθετα ο διαχωρισμός του σήματος από τον θόρυβο παραμένει μια μεγάλη πρόκληση στα συστήματα αναγνώρισης φωνής τα οποία είναι υπερβολικά επιρρεπείς στον θόρυβο από το περιβάλλον και από την ύπαρξη περισσότερων του ενός χρηστών. Τέλος η γλώσσα που υποστηρίζουν τα περισσότερα συστήματα αναγνώρισης φωνής είναι η αγγλική με αποτέλεσμα να τίθενται προβλήματα μεταφερσιμότητας σε πληθυσμούς που ομιλούν διαφορετική γλώσσα.

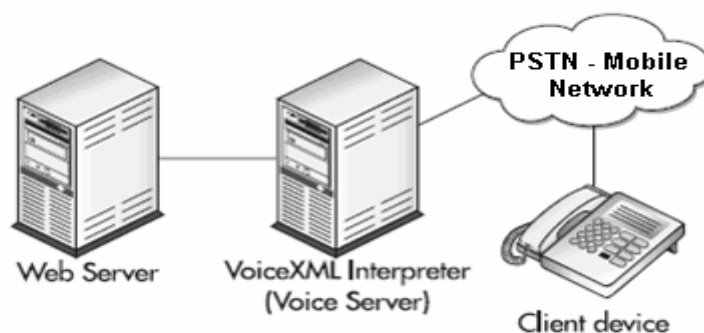
4.6 Λογισμικό για την ανάπτυξη ασύρματων πολύτροπων εφαρμογών φωνής

4.6.1 VoiceXML

Η VoiceXML [48] αποτελεί μια γλωσσά βασισμένη στην XML η οποία παρέχει την δυνατότητα ανάπτυξης αλληλεπιδραστικών εφαρμογών φωνής. Πιο συγκεκριμένα επιτρέπει την πρόσβαση σε πληροφορίες μέσω της χρήσης του τηλεφώνου ή άλλης συσκευής που επιτρέπει φωνητικές κλήσεις. Για παράδειγμα η VoiceXML επιτρέπει την δημιουργία εφαρμογών όπως:

- Φωνητική κλήση (voice activated dialing)
- Φωνητικό email (voice email)
- Φωνητικά portals πληροφοριών (πχ φωνητική υπηρεσία πρόγνωσης καιρού, ενημέρωσης κλπ)
- Φωνητικό CRM εξυπηρέτησης πελατών

Η αρχιτεκτονική των εφαρμογών VoiceXML ομοιάζει με αυτή των εφαρμογών διαδικτύου (Web applications) όπως παρουσιάζεται παρακάτω:



Σχήμα 4.5 Αρχιτεκτονική εφαρμογών VoiceXML

Στις εφαρμογές VoiceXML τα έγγραφα VoiceXML αποθηκεύονται σε έναν κλασσικό Web Server. Επιπρόσθετα στην αρχιτεκτονική υπάρχει ο Voice server ο οποίος αναλαμβάνει να ερμηνεύσει όλες τις φωνητικές εντολές που λαμβάνει από τον χρήστη και να παρέχει σε αυτόν φωνητικές αποκρίσεις. Παρακάτω απεικονίζεται VoiceXML κώδικας :

```
<?xml version = "1.0"?>
<vxml version = "2.0">
  <form id = "datecheck">
    <field name = "datetype" type = "date">
      <prompt>
        Please enter your birth date.
      </prompt>
      <filled>
        <prompt>
          Your birth date is
          <value expr = "datetype" />
        </prompt>
      </filled>
    </field>
  </form>
</vxml>
```

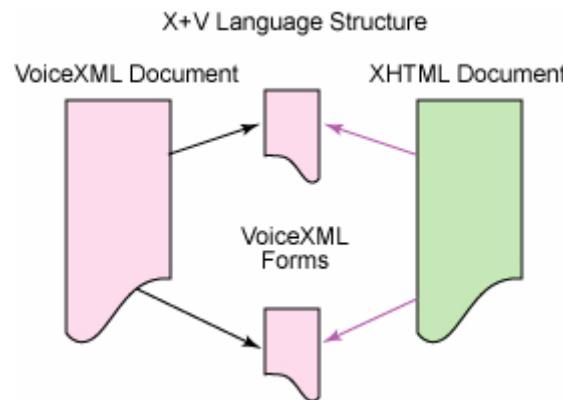
Σχήμα 4.6 Κώδικας VoiceXML.

4.6.2 XHTML + Voice (X+V)

Η XHTML+Voice (X+V) [49] αποτελεί μια γλώσσα σήμανσης που επιτρέπει την ανάπτυξη πολύτροπων εφαρμογών στο Διαδίκτυο. Η X+V επιτρέπει την ανάπτυξη ιστοσελίδων που παρέχουν την δυνατότητα αλληλεπίδρασης είτε μέσω φωνητικών εντολών είτε μέσω γραφικής διεπαφής. Η X+V επιτυγχάνει την υποστήριξη φωνητικής αλληλεπίδρασης ενσωματώνοντας στην γλώσσα σήμανσης XHTML που χρησιμοποιείται για την ανάπτυξη ιστοσελίδων την γλώσσα VoiceXML. Πιο συγκεκριμένα η V+X επιτρέπει να προστεθεί φωνητική υποστήριξη σε κάθε στοιχείο της XHTML. Για παράδειγμα η συμπλήρωση ενός πεδίο κειμένου μιας ιστοσελίδας μπορεί να γίνεται είτε μέσω της γραφικής διεπαφής είτε μέσω της φωνής. Επεκτείνοντας αυτή την λειτουργία σε όλα τα συστατικά της XHTML ο

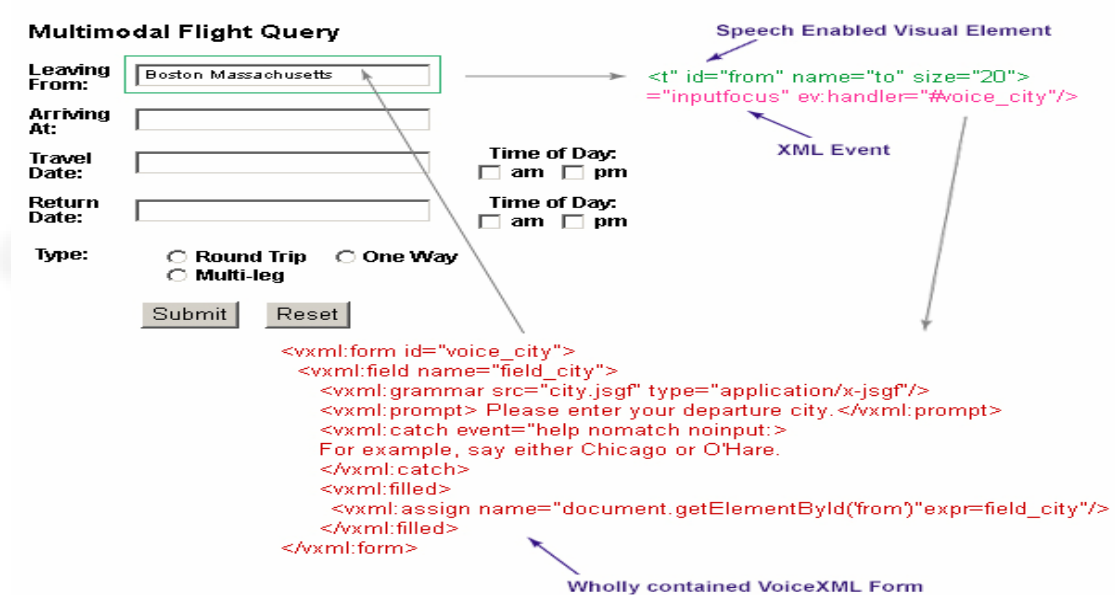
δικτυακός τόπος μπορεί να ελέγχεται πλήρως με την χρήση φωνητικών εντολών δημιουργώντας μια πλήρης πολύτροπη εφαρμογή.

Η αρχιτεκτονική της X+V που απεικονίζεται παρακάτω παρουσιάζει τον συνδυασμό των γραφικών συστατικών με τα φωνητικά συστατικά:



Σχήμα 4.7 Η γλώσσα σήμανσης X+V

Η ένωση των γραφικών και φωνητικών συστατικών πραγματοποιείται μέσω κατάλληλων γεγονότων (events) [50] τα οποία προσδιορίζουν ποιο φωνητικό κομμάτι θα εκτελεστεί για κάθε γραφικό συστατικό. Η συσχέτιση αυτή παρουσιάζεται παρακάτω:



Σχήμα 4.8 Συστατικά της X+V (XHTML, VoiceXML, Events)

Η εκτέλεση της X+V γίνεται με την χρήση καταλλήλων voice browser μεταξύ των οποίων ο Opera [51] και ο Netfront [52] ο οποίος υποστηρίζετε από PDA και ασύρματες συσκευές.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

5

Αποτελέσματα

5.1 Εισαγωγή

Η σύγκλιση των δικτύων για παροχή πολυμεσικών και μη υπηρεσιών αποτελεί τον στόχο των δικτύων 4^{ης} γενιάς. Όμως αν και φαινομενικά η ενοποίηση των δικτύων μπορεί να προσφέρει υπηρεσίες οπουδήποτε και οποτεδήποτε (ubiquitous services), οι υπηρεσίες αυτές δεν είναι πραγματικά ubiquitous. Για να επιτευχθεί η προσέγγιση αυτή προτείνεται η χρήση πολύτροπων εφαρμογών που προέρχονται από τον χώρο της αλληλεπίδρασης ανθρώπου μηχανής στις συσκευές ασύρματης πρόσβασης.

Το συμπέρασμα είναι πως ο συνδυασμός των τεχνολογιών ασύρματων δικτύων και οι πολύτροπες εφαρμογές μπορούν να οδηγήσουν στην πραγματική παροχή ubiquitous υπηρεσιών. Επιπλέον παρακάτω αναλύονται προβλήματα και λύσεις στην ανάπτυξη πολύτροπων εφαρμογών φωνής στις ασύρματες συσκευές εφόσον η ανάπτυξη τέτοιων εφαρμογών διαφέρει σημαντικά από την ανάπτυξη κλασσικών γραφικών εφαρμογών.

5.2 Πολύτροπες εφαρμογές για την παροχή ubiquitous υπηρεσιών

Η ενοποίηση των δικτύων προσφέρει κατά ένα μέρος ubiquitous υπηρεσίες.

Στην περίπτωση της ανεξαρτησίας δικτύου πρόσβασης οι συσκευές και τα δίκτυα κάνουν χρήση του πρωτοκόλλου IP, το οποίο αποτελεί τον συνδετικό κρίκο όλων των επιμέρους συστατικών για την δημιουργία ενός ενιαίου δικτύου.

Στην περίπτωση της ανεξαρτησίας συσκευής η εκάστοτε συσκευή που παρέχει συνδεσιμότητα στο δίκτυο θα πρέπει να είναι ικανή να παρέχει το σύνολο των υπηρεσιών του δικτύου ανεξαρτήτως γεωγραφικής θέσης και χρόνου. Αυτό υπονοεί πως το δίκτυο θα πρέπει να επιτρέπει πολλαπλούς τρόπους αλληλεπίδρασης ικανούς να ανταπεξέλθουν σε κάθε κατάσταση. Για παράδειγμα στην περίπτωση που ο χρήστης οδηγεί και επιθυμεί να δει το ηλεκτρονικό του ταχυδρομείο θα μπορούσε να διακόψει την οδήγηση, να σταματήσει και να χρησιμοποιήσει την γραφική διεπαφή του κινητού για να ανακτήσει το ηλεκτρονικό του ταχυδρομείο. Όμως στην περίπτωση αυτή η υπηρεσία αυτή δεν είναι πραγματικά ubiquitous εφόσον αν και παρέχει ανεξαρτησία από το δίκτυο πρόσβασης (το κινητό θα μπορεί να επιλέξει το καταλληλότερο δίκτυο στην γεωγραφική θέση που βρίσκεται) δεν παρέχει την δυνατότητα στον χρήστη να εκτελέσει τη λειτουργία ανάκτησης του ηλεκτρονικού ταχυδρομείου ενώ κινείται. Η λύση σε αυτό το πρόβλημα είναι η υποστήριξη από το δίκτυο πολλαπλών τρόπων αλληλεπίδρασης με το τερματικό του χρήστη μέσω πολύτροπων εφαρμογών. Στο παράδειγμα της ανάκτησης του ηλεκτρονικού ταχυδρομείου η χρήση μιας πολύτροπης εφαρμογής που συνδυάζει και γραφικό περιβάλλον και παροχή της υπηρεσίας μέσω φωνής θα έλυσε το πρόβλημα.

Το πρόβλημα που περιγράφηκε παραπάνω δεν είναι μεμονωμένο εφόσον υπάρχει πλήθος παρόμοιων παραδειγμάτων που αποτρέπουν την παροχή ubiquitous υπηρεσιών στον τελικό χρήστη όπως:

- Στην περίπτωση ατόμων που η δουλειά τους απαιτεί να χρησιμοποιούν συνεχώς τα χέρια τους ενώ παράλληλα θα έπρεπε να εκτελούν και μια άλλη εργασία (πχ στην περίπτωση φωνητικών συστημάτων εκμάθησης συντήρησης και διόρθωσης μηχανημάτων όπου ο χρήστης πρέπει να χρησιμοποιεί τα χέρια του συνεχώς ενώ παράλληλα θα πρέπει να κατευθύνει την ροή των οδηγιών αναλόγως του προβλήματος [53])
- Στις περιπτώσεις ατόμων με ειδικές ανάγκες όπου απαιτείται η χρήση εναλλακτικών τρόπων αλληλεπίδρασης.
- Η δυσκολία χρήσης συσκευών από τον άνθρωπο που παρέχουν περιορισμένες δυνατότητες παρουσίασης των πληροφοριών (πχ μικρή οθόνη) και περιορισμένες δυνατότητες στον τρόπο εισαγωγής των πληροφοριών (πχ έλλειψη ποντικιού) καθιστούν την πολύτροπη αλληλεπίδραση απαραίτητη.

Η γενικότητα του προβλήματος δείχνει πως το πρόβλημα είναι υπαρκτό και χρήζει άμεσης λύσης. Συνεπώς για να γίνει εφικτή η παροχή πραγματικών ubiquitous υπηρεσιών στα ασύρματα δίκτυα δεν θα πρέπει να λαμβάνεται υπόψη μόνο η κατεύθυνση από το δίκτυο προς το τερματικό του χρήστη αλλά και η κατεύθυνση από το τερματικό του χρήστη προς το δίκτυο. Για την πραγματική παροχή ubiquitous υπηρεσιών απαιτείται η χρήση

πολύτροπων τρόπων αλληλεπίδρασης. Μόνο ο συνδυασμός αυτών των δυο συστατικών (ασύρματα δίκτυα, πολύτροπες εφαρμογές) μπορεί να οδηγήσει σε πραγματικά ubiquitous εφαρμογές.

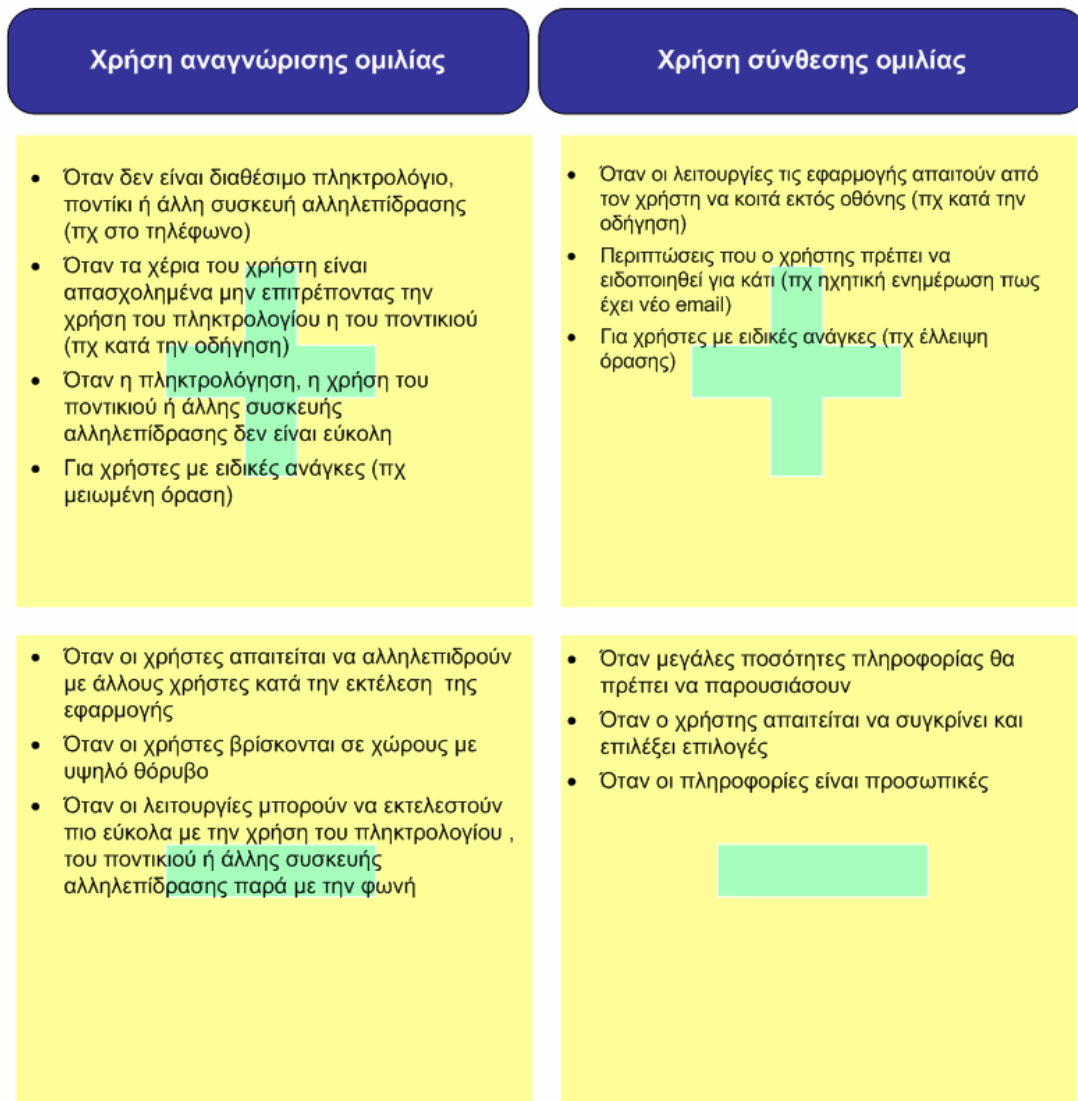
5.3 Ανάπτυξη πολυτρόπων εφαρμογών με συνδυασμό τεχνολογιών φωνής και γραφικών διεπαφών

5.3.1 Γενικά

Η εισαγωγή φωνητικών διαλόγων σε μια γραφική εφαρμογή αποτελεί τον πλέον γνωστό τύπο πολύτροπων εφαρμογών. Επιπλέον η φωνητική αλληλεπίδραση θεωρείται ως η πλέον κατάλληλη για τα δίκτυα κινητών επικοινωνιών εφόσον η μετάδοση φωνής αποτελεί την κύρια λειτουργία αυτών των δικτύων. Όμως η ανάπτυξη τέτοιων εφαρμογών είναι ιδιαίτερα δύσκολη εφόσον δεν υπάρχουν καλά καθορισμένοι κανόνες για τον σχεδιασμό και την ανάπτυξη τέτοιων εφαρμογών.

5.3.2 Επιλογή ή μη χρήσης τεχνολογιών φωνής

Βασικός παράγοντας για την επιτυχή ανάπτυξη μιας πολύτροπης εφαρμογής είναι να αποφασισθεί αν θα χρησιμοποιηθούν η όχι τεχνολογίες φωνής. Οι τεχνολογίες φωνής δεν θα πρέπει να θεωρούνται πανάκεια εφόσον η χωρίς σκέψη χρήση τους μπορεί να οδηγήσει σε μη εύχρηστες εφαρμογές. Παρακάτω παρουσιάζονται γενικοί κανόνες για το πότε είναι κατάλληλη η χρήση τεχνολογιών αναγνώρισης και σύνθεσης φωνής:



Σχήμα 5.1 Παράγοντες για την επιλογή χρήσης ή μη τεχνολογιών φωνής

5.3.3 Σχεδιαστικές επιλογές

Εφόσον αποφασιστεί πως η χρήση τεχνολογιών ομιλίας είναι κατάλληλες για το υπό ανάπτυξη σύστημα θα πρέπει να αποφασιστεί ο τρόπος με τον οποίο οι τεχνολογίες ομιλίας θα εισαχθούν επιτυχώς στην πολύτροπη εφαρμογή. Παρακάτω παρουσιάζονται κανόνες που θα πρέπει να ακολουθηθούν για την ανάπτυξη επιτυχών πολύτροπων εφαρμογών

5.3.3.1 Αποφυγή της μετατροπής γραφικών διεπαφών σε φωνητικές

Ο σχεδιασμός μιας πολύτροπης εφαρμογής φωνής θα πρέπει να λάβει υπόψη του πολλές παραμέτρους και σε καμία περίπτωση δεν θα πρέπει να θεωρηθεί πως αποτελεί απλώς μια μετατροπή της γραφικής διεπαφής. Αυτό ισχύει εφόσον οι γραφικές εφαρμογές δεν μπορούν να αναπαραστήσουν πάντα με ακρίβεια το λεξιλόγιο ή τις βασικές έννοιες που χρησιμοποιούν οι άνθρωποι όταν ομιλούν για την συγκεκριμένη εφαρμογή. Για παράδειγμα σε μια γραφική εφαρμογή δημιουργίας χρονοδιαγράμματος (schedule application) το ημερολόγιο αναπαριστά τις ημερομηνίες σε ημέρες, μήνες και έτη. Αντιθέτως όταν ένας άνθρωπος καθορίζει το πρόγραμμα του με τη χρήση ομιλίας χρησιμοποιεί έννοιες όπως «την επόμενη Τρίτη», «μεθαύριο» κλπ. Εξαιτίας αυτού η εφαρμογή που κάνει χρήση τεχνολογιών ομιλίας θα πρέπει να συμπεριλάβει στην γραμματική της αυτές τις εκφράσεις για να γίνει προσιτή και φιλική στον χρήστη. Αν όμως ο σχεδιαστής βασιστεί στην γραφική διεπαφή αυτοί οι παράμετροι δεν θα ληφθούν υπόψη με αποτέλεσμα η ευχρηστία της φωνητικής εφαρμογής να μειωθεί.

Η οργάνωση της πληροφορίας είναι ακόμα ένα ζήτημα που θα πρέπει να εξεταστεί. Αυτό γιατί απεικονίσεις πληροφοριών που είναι καλές σε γραφικές διεπαφές, στις φωνητικές διεπαφές μπορεί να αποτυγχάνουν τελείως. Για παράδειγμα μια τυπική θυρίδα ηλεκτρονικού ταχυδρομείου έχει την παρακάτω μορφή:

<i>Sender</i>	<i>Subject</i>	<i>Date & Time</i>	<i>Size</i>
Arlene Rexford	Learn about Java	Mon Oct 28 11:23	2K
Shari Jackson	Re: Boston rumors	Fri Jul 18 09:32	3K
Hilary Binda	Change of address	Wed Jul 16 12:59	1K
Arlene Rexford	Class Openings	Tue Jul 21 12:35	8K
George Fitz	Re: Boston rumors	Tue Jul 21 12:46	1K

Σχήμα 5.2 Θυρίδα ηλεκτρονικού ταχυδρομείου

Η προσπάθεια να διαβαστεί το σύνολο της πληροφορίας όπως ακριβώς παρουσιάζεται και να αναπαραχθεί ηχητικά δεν είναι η καλύτερη δυνατή λύση επειδή καταρχήν είναι χρονοβόρο και δεύτερον ο χρήστης δεν μπορεί να συγκρατήσει το σύνολο της πληροφορίας που θα ακούσει. Η λύση στην παραπάνω εφαρμογή είναι να αναπαράγεται ηχητικά μόνο το όνομα του αποστολέα και το θέμα του ηλεκτρονικού ταχυδρομείου και κατόπιν απαίτησης του χρήστη να αναπαράγονται ηχητικά τα πεδία της ημερομηνίας και του μεγέθους. Μια ακόμα καλύτερη λύση είναι το μηνύματα να κατηγοριοποιούνται ανά αποστολέα και η ηχητική απόκριση να έχει την μορφή «Έχετε δυο μηνύματα από τον Σωτηρόπουλο Γιάννη». Με αυτό τον τρόπο περιορίζουμε το πλήθος των αποκρίσεων διευκολύνοντας τον χρήστη να κάνει την επιλογή του.

Επόμενως γίνεται σαφές πως η δημιουργία μιας εφαρμογής που κάνει χρήση τεχνολογιών ομιλίας θα πρέπει να έχει ως βασικό άξονα την φυσική ομιλία και όχι την γραφική διεπαφή της εφαρμογής.

5.3.3.2 Παράμετροι για την επιτυχή ανάπτυξη πολύτροπων εφαρμογών

Υπάρχουν αρκετοί παράμετροι οι οποίοι θα πρέπει να ληφθούν υπόψη κατά την ανάπτυξη πολύτροπων εφαρμογών. Οι παράμετροι αυτοί εξετάζονται παρακάτω:

5.3.3.2.1 Παροδικότητα του λόγου (Speech transience)

Ο λόγος είναι παροδικός εφόσον όταν λεχθεί ή ακουστεί δεν συγκρατείται σε κάποιο αποθηκευτικό χώρο. Αντιθέτως οι γραφικές διεπαφές παραμένουν στην οθόνη του υπολογιστή ή της ασύρματης συσκευής αναμένοντας από τον χρήστη να εκτελέσει κάποια λειτουργία.

Έτσι οι χρήστες μπορούν να συγκρατήσουν μόνο ένα μικρό μέρος των πληροφοριών που είπαν ή ακούσαν. Βάση αυτού θα πρέπει να αποφεύγεται η είσοδος ή η έξοδος μεγάλου όγκου πληροφορίας εφόσον ο χρήστης είναι αδύνατο να τις συγκρατήσει. Όμως η ιδιότητα αυτή έχει και πλεονεκτήματα. Εφόσον ο άνθρωπος μπορεί να ακούει και να βλέπει ταυτόχρονα, η ομιλία μπορεί επιτυχώς να χρησιμοποιηθεί για να τραβήξει την προσοχή του χρήστη για ένα σημαντικό γεγονός (πχ λήψη νέου email)

5.3.3.2.2 Απόδοση αναγνώρισης φωνής

Τα συστήματα αναγνώρισης φωνής δεν είναι τέλεια και συχνά κάνουν λάθη κατά την αναγνώριση. Αυτό οδηγεί συχνά τους χρήστες να σχηματίζουν ένα λάθος μοντέλο για τον τρόπο με τον οποίο λειτουργεί το σύστημα. Για παράδειγμα αν ο χρήστης πει "Read the next message" και το σύστημα αναγνώρισης φωνής αναγνωρίσει την έκφραση "Repeat the message", η εφαρμογή θα επαναλάβει το μήνυμα οδηγώντας τον χρήστη να πιστέψει πως η επιλογή "Read the next message" δεν είναι σωστή. Για να λυθεί το

πρόβλημα αυτό θα πρέπει να παρέχεται μέσα από την γραφική διεπαφή το σύνολο των επιτρεπτών φωνητικών εντολών ή/και να μπορεί ο χρήστης να τις ακούσει.

5.3.3.2.3 Ευελιξία έναντι ακρίβειας κατά την αναγνώριση φωνής

Ένα καλό πολύτροπο σύστημα επιτρέπει στους χρήστες να εκτελούν τις ίδιες λειτουργίες μέσα από διαφορετικές φωνητικές εντολές. Οι φωνητικές εντολές θα πρέπει να περιλαμβάνουν τόσο φυσικές εκφράσεις όσο και ποικιλία προκαθορισμένων λέξεων. Η προσέγγιση αυτή θα πρέπει να τηρηθεί εφόσον οι χρήστες έχουν διαφορετική αντίληψη για το πώς θα πρέπει να δουλεύει το σύστημα (πχ άλλοι χρήστες ικανοποιούνται με το να πουν μόνο μια λέξη κλειδί ενώ άλλοι επιθυμούν να πουν μια ολόκληρη πρόταση)

5.3.3.2.4 Απόκριση και καθυστέρηση (Feedback and Latency)

Στους φωνητικούς διαλόγους ο χρόνος απόκρισης του συστήματος αναγνώρισης ή σύνθεσης φωνής είναι κρίσιμος. Για παράδειγμα στην περίπτωση που οι χρήστες διατυπώσουν μια φωνητική εντολή και δεν λάβουν άμεσα απόκριση από το σύστημα θεωρούν ότι το σύστημα δεν τους άκουσε και προσπαθούν να ξανα-υποβάλουν την φωνητική εντολή. Αυτό οδηγεί είτε στο να χάσουν την απόκριση του συστήματος (εφόσον εκείνη την στιγμή θα μιλάνε) είτε να προκληθεί λάθος στην αναγνώριση. Για το λόγο αυτό οι χρήστες θα πρέπει να λαμβάνουν συνεχώς απόκριση από το σύστημα για να μπορεί να αντιληφθεί που βρίσκεται και τι θα γίνει μετά.

Επιπλέον ο χρήστης θα πρέπει να επιβεβαιώνει σημαντικές ενέργειες. Για παράδειγμα ασφαλές και σωστό θα ήταν ο χρήστης να επιβεβαιώσει την

κατάθεση ενός ποσού αλλά θα ήταν υπερβολικό να επιβεβαιώσει την επιστροφή στο αρχικό μενού των λειτουργιών.

5.3.3.2.5 Διάλογοι (prompts)

Οι φωνητικοί διάλογοι στα πολύτροπα συστήματα μπορούν να συνδυαστούν με την γραφική διεπαφή. Στην προσέγγιση αυτή οι χρήστες μπορούν να πουν το κείμενο που βλέπουν στην γραφική διεπαφή όπως για παράδειγμα τα στοιχεία μιας λίστας.

5.3.3.2.6 Χειρισμός λαθών

Ο χειρισμός των λαθών είναι άμεσος στις πολύτροπες εφαρμογές εφόσον αν συμβεί κάποιο λάθος κατά την αναγνώριση φωνής δεν θα παρουσιαστούν αποτελέσματα την περιοχή κειμένου που θα εμφανίζονταν τα αποτελέσματα της αναγνώρισης.

5.4 Ανάπτυξη πολύτροπης εφαρμογής

5.4.1 Γενικά

Στόχος της εφαρμογής είναι να εφαρμόσει τα αποτελέσματα που αναλύθηκαν παραπάνω για την ανάπτυξη μιας πολύτροπης εφαρμογής που συνδυάζει γραφικό περιβάλλον με τεχνολογίες ομιλίας για την παροχή ubiquitous πολυμεσικών ή μη δεδομένων στον υπολογιστή και σε ασύρματες συσκευές

5.4.2 Περιγραφή εφαρμογής και λειτουργίες

Η εφαρμογή παρέχει την δυνατότητα στους χρήστες να αναζητήσουν και να επιλέξουν προγράμματα της τηλεόρασης που επιθυμούν να δουν. Πιο συγκεκριμένα η εφαρμογή παρέχει τις παρακάτω λειτουργίες:

- Αναζήτηση προγραμμάτων της τηλεόρασης ανά κανάλι και ημερομηνία και ώρα προβολής (Browse by channel and time)
- Προβολή των προγραμμάτων που μεταδίδονται αυτή την στιγμή (Show now playing programmes)
- Προβολή των προγραμμάτων ανά είδος και κατηγορία (Browse by genres). Βάση αυτού τα προγράμματα διαχωρίζονται σε είδη και το κάθε είδος σε επιμέρους κατηγορίες. Για παράδειγμα αν διαχωρίσουμε τα προγράμματα βάση του περιεχομένου τους (είδος) τότε οι αντίστοιχες κατηγορίες μπορεί να είναι ψυχαγωγία, αθλητικά κλπ
- Αναζήτηση βάση λέξεων κλειδιών (Search by keyword), οι οποίες περιγράφουν το κάθε πρόγραμμα.
- Διαχείριση των ταινιών που έχουν επιλεγεί (Recording manager). Ο χρήστης με αυτή την επιλογή μπορεί να αναπαράγει τις ταινίες που έχει επιλέξει αν είναι διαθέσιμες ή να τις διαγράψει

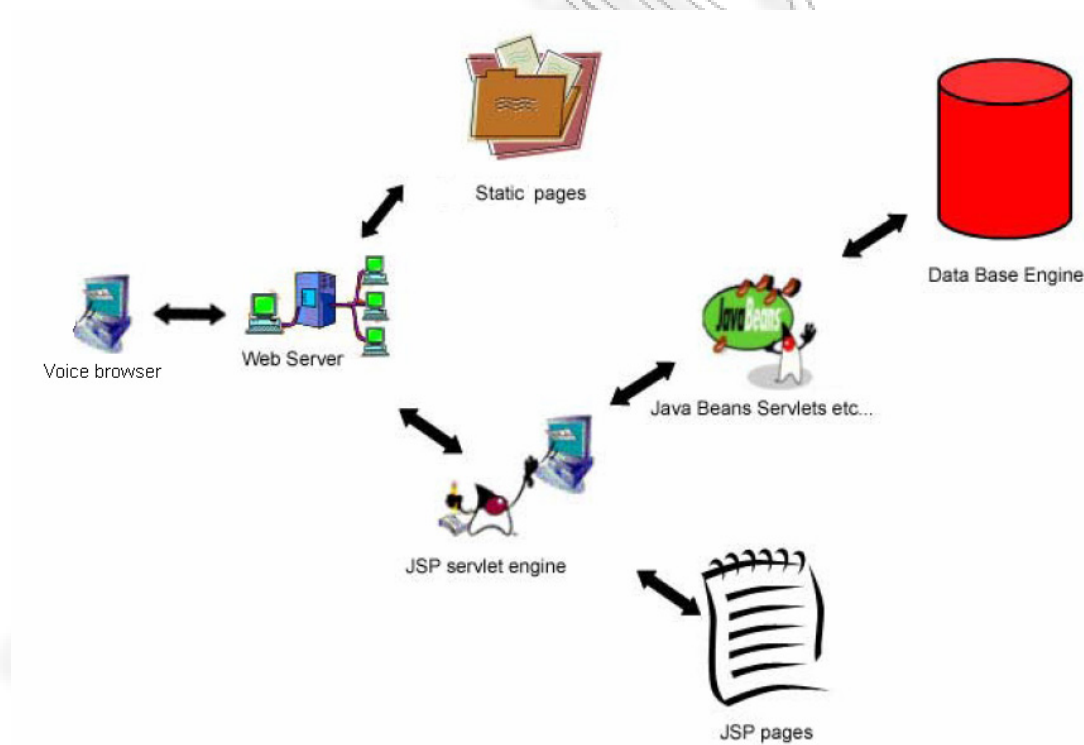
Γενικότερα η διαδικασία που ακολουθεί ο χρήστης είναι να αναζητήσει τις ταινίες που τον ενδιαφέρουν (με την χρήση των τεσσάρων πρώτων επιλογών που περιγράφηκαν παραπάνω), να δει πληροφορίες για τις ταινίες που τον ενδιαφέρουν και αν επιθυμεί να τις δει να τις επιλέξει. Εν

συνεχία με την χρήση του Recording Manager μπορεί είτε να αναπαράγει αυτές τις ταινίες είτε να τις διαγράψει.

Όλες οι παραπάνω λειτουργίες μπορούν να εκτελεστούν είτε μέσω του γραφικού περιβάλλοντος είτε μέσω φωνητικών εντολών. Επιπλέον ο χρήστης μπορεί να δει το σύνολο των διαθέσιμων φωνητικών εντολών στην οθόνη ή να τις ακούσει κάνοντας χρήση της φωνητικής εντολής help.

5.4.3 Αρχιτεκτονική εφαρμογής και τεχνολογίες

Η εφαρμογή αναπτύχθηκε με την γλώσσα σήμανσης X+V [54] η οποία περιγράφηκε στην ενότητα 4.6.2. Η αρχιτεκτονική παρουσιάζεται παρακάτω:



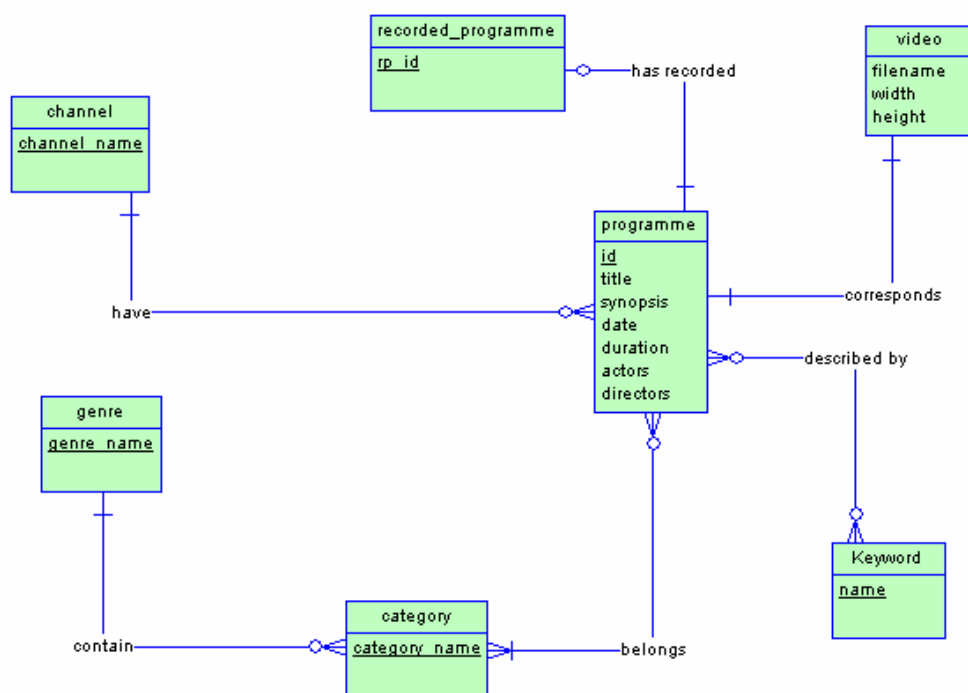
Σχήμα 5.3 Αρχιτεκτονική εφαρμογής

Η δημιουργία των δυναμικών ιστοσελίδων (που εμπεριέχουν τμήματα του κώδικα VoiceXML για την δημιουργία της φωνητικής αλληλεπίδρασης και

τμήματα κώδικα XHTML για την γραφική αναπαράσταση της εφαρμογής) έγινε με την χρήση της γλώσσας ανάπτυξης δυναμικών ιστοσελίδων JSP. Ο voice browser που χρησιμοποιήθηκε είναι ο Opera ενώ για την μετάδοση του video με τεχνολογίες streaming χρησιμοποιήθηκε ο Flash player ο οποίος υποστηρίζεται και από ασύρματες συσκευές [55].

5.4.4 Εννοιολογικό μοντέλο

Για την αποθήκευση των δεδομένων χρησιμοποιήθηκε το σύστημα διαχείρισης βάσεων δεδομένων MySQL. Παρακάτω παρουσιάζεται το εννοιολογικό μοντέλο της βάσης δεδομένων της εφαρμογής:



Σχήμα 5.4 Εννοιολογικό μοντέλο

Η σημασία των οντοτήτων και των πεδίων είναι προφανής και δεν θα εξεταστούν περαιτέρω. Παρακάτω αναλύονται οι σχέσεις μεταξύ των οντοτήτων:

channel have programme: Ένα κανάλι παρέχει μηδέν ή περισσότερα προγράμματα ενώ το κάθε πρόγραμμα ανήκει αποκλειστικά σε ένα κανάλι. Η δεύτερη απαίτηση δεν σημαίνει πως ένα κανάλι δεν μπορεί να προσφέρει την ίδια ταινία αλλά ότι η ταινία αυτή θα πρέπει να καταχωρηθεί ξανά στην οντότητα programme.

programme has recorded recorded_programme: κάθε πρόγραμμα ενός καναλιού μπορεί να καταχωρείται ως εγγεγραμμένο καμία ή μια φορά.

programme corresponds to video: κάθε πρόγραμμα ενός καναλιού αντιστοιχεί σε ένα αρχείο video και κάθε αρχείο video αντιστοιχεί σε ένα πρόγραμμα.

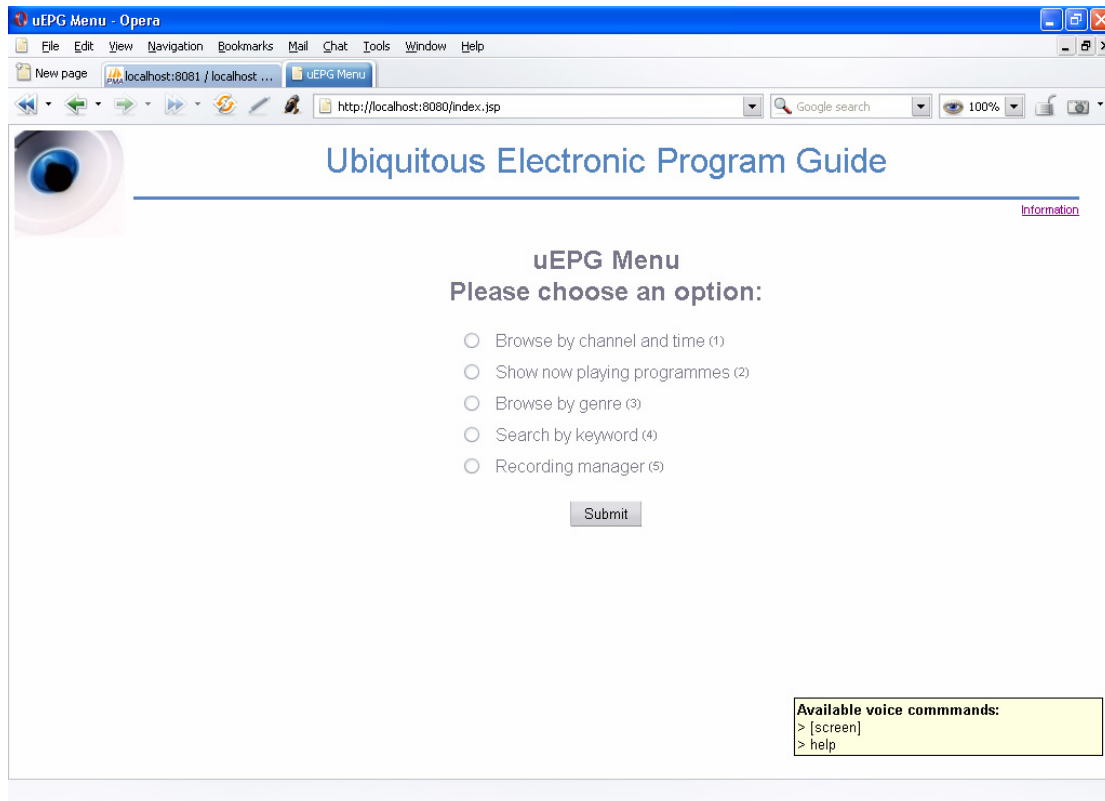
programme belongs to category: Ένα πρόγραμμα ανήκει σε μία ή περισσότερες κατηγορίες ενώ μια κατηγορία μπορεί να μην περιλαμβάνει κανένα πρόγραμμα ή να περιλαμβάνει πολλά προγράμματα.

genre contain category. Κάθε είδος προγράμματος περιλαμβάνει καμία ή περισσότερες κατηγορίες ενώ κάθε κατηγορία ανήκει ακριβώς σε ένα είδος προγράμματος.

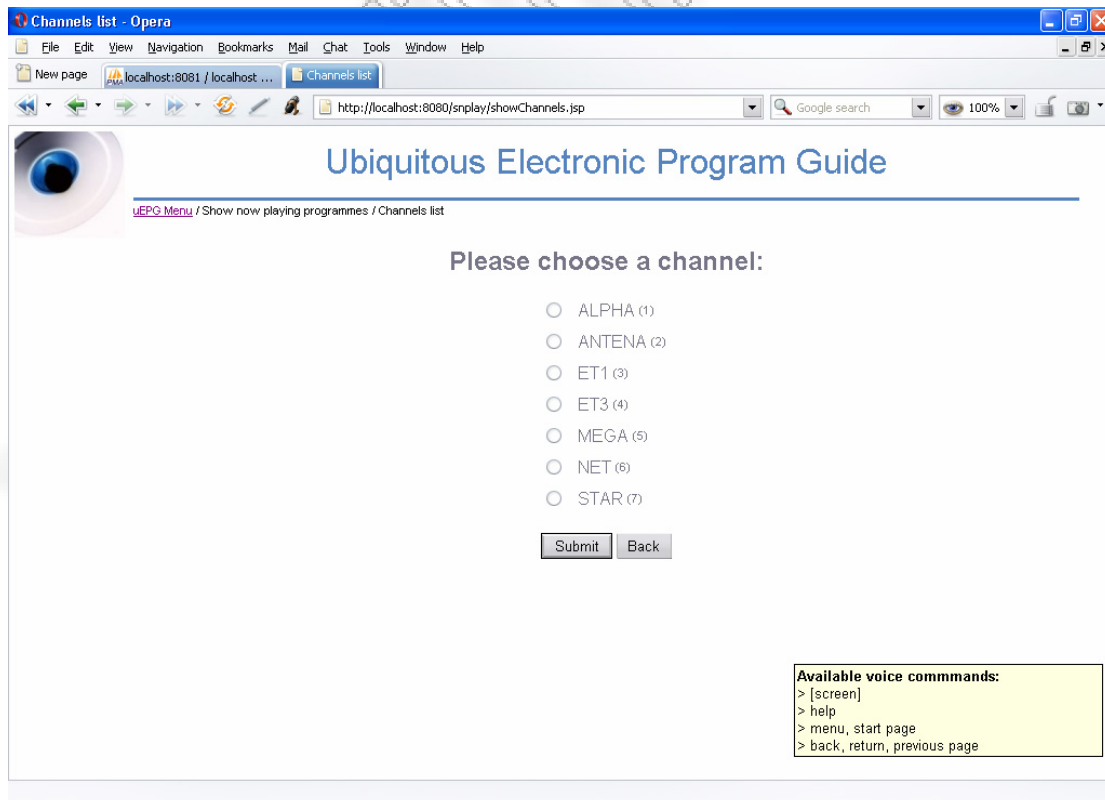
programme described by keyword: Κάθε πρόγραμμα μπορεί να περιγράφεται από καμία ή περισσότερες λέξεις κλειδιά και κάθε λέξη κλειδί μπορεί να περιγράφει πολλά προγράμματα

5.4.5 Παραδείγματα χρήσης

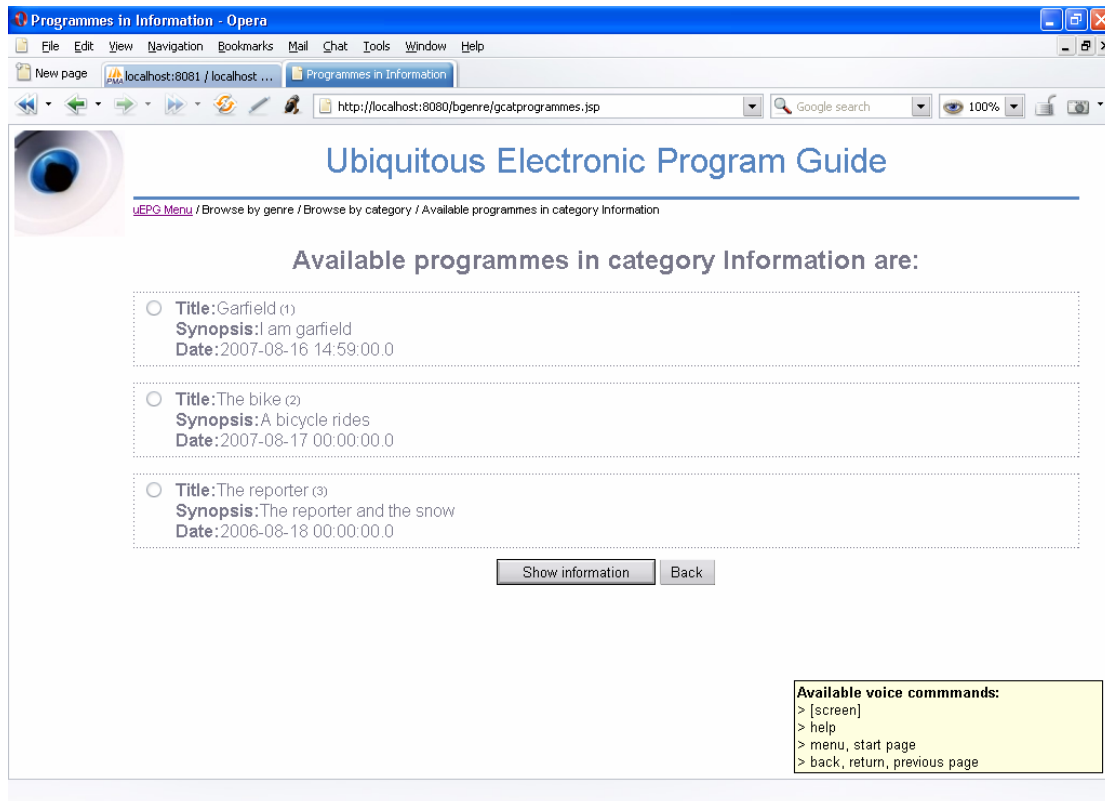
Παρακάτω θα δοθούν μερικές εικόνες από την εφαρμογή:



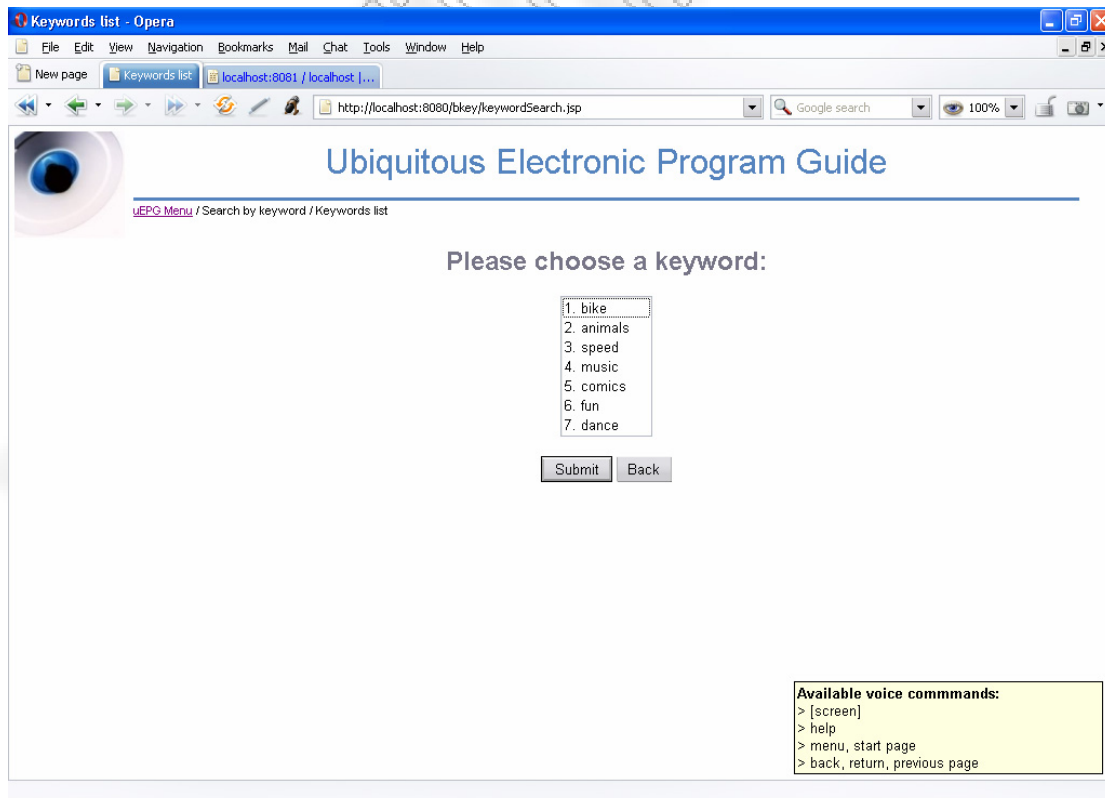
Σχήμα 5.5 Αρχική σελίδα εφαρμογής



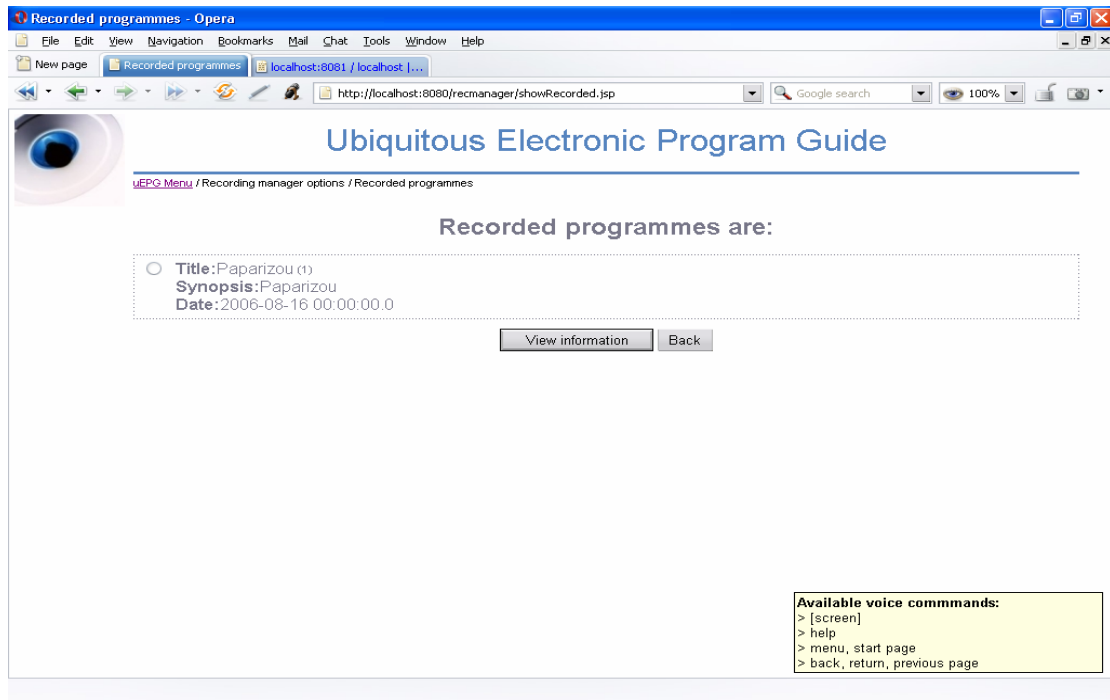
Σχήμα 5.6 Επιλογή καναλιού



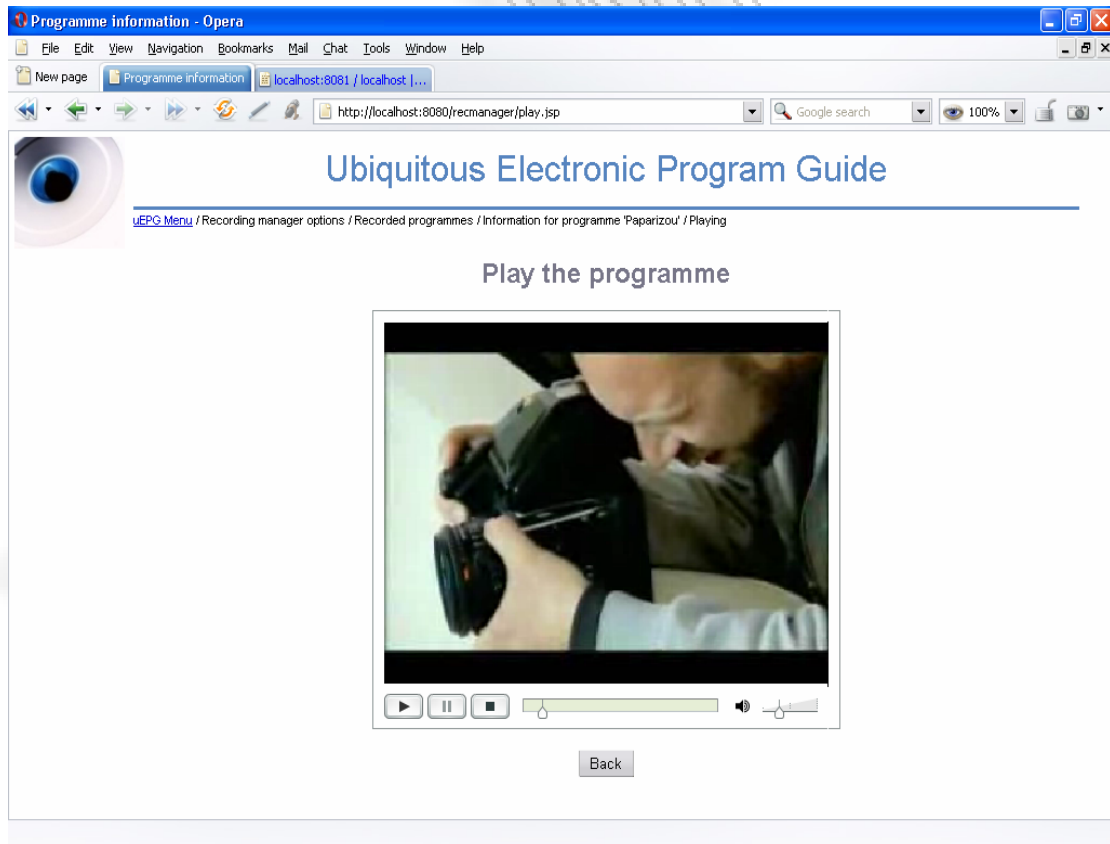
Σχήμα 5.7 Παρουσίαση προγραμμάτων στη κατηγορία Information



Σχήμα 5.8 Επιλογή προγραμμάτων βάση λέξεων κλειδιών



Σχήμα 5.9 Διαθέσιμα προγράμματα προς αναπαραγωγή



Σχήμα 5.10 Αναπαραγωγή προγράμματος

6

Συμπεράσματα

Η ενοποίηση των δικτύων για την παροχή Διαδικτυακών υπηρεσιών σε συνδυασμό με την ανάπτυξη πολυτρόπων εφαρμογών θα αλλάξει ριζικά το τοπίο στις σύγχρονες τηλεπικοινωνίες. Η δυνατότητα παροχής στους χρήστες υπηρεσιών οπουδήποτε και οποτεδήποτε αυτοί επιθυμούν παρέχει ανεξάντλητες ευκαιρίες δημιουργίας νέων υπηρεσιών. Πλέον τομείς όπως το εμπόριο, η ψυχαγωγία, η παροχή υπηρεσιών υγείας θα βασίζονται στα ασύρματα δίκτυα για την αποτελεσματική και γρήγορη παροχή υπηρεσιών στους χρήστες.

Οι πολύτροπες διεπαφές θα αποτελέσουν μια νέα κλάση ευφυών συστημάτων πολυμέσων οι οποίες θα επεκτείνουν τις δυνατότητες της αλληλεπίδρασης ανθρώπου μηχανής προσφέροντας πλεονεκτήματα όπως φιλικότητα και φυσικότητα κατά την αλληλεπίδραση. Εν τω μεταξύ καθώς οι τεχνολογίες αναγνώρισης όπως η αναγνώριση φωνής και η τεχνητή όραση συνεχίζουν να βελτιώνονται η πολύτροπες διεπαφές θα επεκτείνονται και τελικά θα αντικαταστήσουν πλήρως τις παραδοσιακούς τρόπους αλληλεπίδρασης (πχ πληκτρολόγιο, ποντίκι κλπ).

Για να είναι επιτυχής αυτή η μετάβαση σημαντική έρευνα θα πρέπει να γίνει τόσο στις τεχνολογίες ενοποίησης των δικτύων όσο και στην συμπεριφορά του χρήστη κατά την αλληλεπίδραση με πολύτροπες εφαρμογές με στόχο να παραχθούν κανόνες που θα βοηθούν τους σχεδιαστές στην ανάπτυξη καλών και ανθεκτικών πολύτροπων εφαρμογών. Η εξέλιξη προς σε αυτή την κατεύθυνση έχει ακόμα αρκετό δρόμο να διανύσει και ήδη ο στόχος ανάπτυξης ενός ενιαίου δικτύου για την παροχή ubiquitous πολυμεσικών και μη υπηρεσιών μέσω πολύτροπων εφαρμογών έχει ήδη τεθεί από τους ερευνητές ως στόχος για την επόμενη δεκαετία.

7

Βιβλιογραφία

1. *mobileYouth.org Wireless Statistics*,
http://www.mobileyouth.org/stats_search.php
2. *World mobile statistics*,
http://www.w2forum.com/i/World_Mobile_Statistics_Free
3. *Internet usage statistics - The Big Picture*,
<http://www.internetworldstats.com/stats.htm>
4. *Internet Usage in Europe*,
<http://www.internetworldstats.com/stats4.htm>
5. *Internet Usage in the European Union*,
<http://www.internetworldstats.com/stats9.htm>
6. Perry M, O'Hara K, Sellen A, Brown B, and Harper R, "*Dealing with mobility: Under standing access anytime, anywhere*", ACM Transactions on Computer-Human Interaction, 2001. 8(4): p. 323-347
7. Margherita Pagani, "*3G Wireless Market Attractiveness: Dynamic Challenges for Competitive Advantages*", 2003
8. *Multimodal. Why IBM? - Leadership in multimodal*,
<http://www.ibm.com/pvc/multimodal>
9. S. Deering and R. Hinden, "*Internet Protocol, Version 6 (IPv6) Specification (RFC 2460)*", December 1998

10. Jon Postel, *IP/IPv4: Internet Protocol Overview (RFC 791)*.
September 1981
11. U. Varshney, "The status and future of 802.11-based WLANs", *IEEE Comput.*, 2003. 36(6): p. 102-105
12. *3rd Generation Partnership Project (3GPP)*, www.3gpp.org
13. Jon Inouye, Jim Binkley, and Jonathan Walpole, "Dynamic Network Reconfiguration Support for Mobile Computers", *Mobile Computing and Networking*, 1997
14. M. Mouly and M.-B. Pautet, "The GSM System for Mobile Communications", 1992
15. R. Kalden, I. Meirick, and M. Meyer, "Wireless internet access based on GPRS", *IEEE Pers. Commun*, 2000. 7(2): p. 8-18
16. J. Khun-Jush, P. Schramm, G. Malmgren, and J. Torsner, "HiperLAN2: broadband wireless communications at 5 GHz", *IEEE Commun. Mag.*, 2002. 40(6): p. 130-137
17. *Digital Video Broadcasting (DVB)*, www.dvb.org
18. J Mitola and G Maguire Jr, "Cognitive radio: making software radios more personal", *IEEE Pers. Commun*, 1999. 6(4): p. 13-18
19. C. Perkins, "IP Mobility Support for IPv4 (RFC3220)", January 2002
20. B. Goode, "Voice over Internet Protocol (VoIP)", *Proceedings of the IEEE*, September 2002. 90(9): p. 1495-1517
21. Abigail J. Sellen and Rachel Murphy, "The Future of the Mobile Internet: Lessons from Looking at Web Use", *Hewlett-Packard Research Labs*, 2002
22. RFC 3261, *SIP: Session Initiation Protocol*, 3GPP TS 22.228 v5.6.0 and v5.9.0. 2002
23. R. Fielding, J. Gettys, J. Mogul, H. Nielsen, and T. Berners-Lee, "Hypertext Transfer Protocol – HTTP/1.1, RFC 2068", January 1997
24. *IEEE 802.15 Working Group for WPAN*, <http://www.ieee802.org/15/>
25. *Bluetooth.com, The Official Bluetooth Wireless Info Site*,
<http://www.bluetooth.com/bluetooth/>
26. Tatsuya Suda, "Providing ubiquitous multimedia services in future network environments", 1999-2000 for MICRO Project 99-113
27. Sethuraman Panchanathan, "Ubiquitous Multimedia: Bridging the Digital Divide", *Phillips Research*

28. D.D. Vleeschauer, J. Janssen, and G.H. Petit, *"Delay and Distortion Bounds for Packetized Voice Calls of Traditional"*, PSTN Quality, Proceedings of the 1st IP-Telephony Workshop (IPTel 2000), 12–13 April 2000
29. J. Postel, *"Transmission Control Protocol, RFC 793"*, September 1981.
30. J. Postel, *"User Datagram Protocol, RFC 768"*, August 1980
31. M. Allan, V. Paxson, and W. Stevens, *"TCP Congestion Control, RFC 2581"*, April 1999
32. V. Jacobson, *"Congestion Avoidance and Control"*, Computer Communication Review, Aug. 1988 18(4): p. 314-329
33. WAP FORUM, <http://www.wapforum.org/>
34. H. Schulzrinne, A. Rao, and R. Lanphier, *"Real Time Streaming Protocol (RTSP), RFC 2326"*, April 1998
35. H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *"RTP: A Transport Protocol for Real-Time Applications, RFC 3550"*, July 2003
36. H. Schulzrinne and S. Casner, *"RTP Profile for Audio and Video Conferences with Minimal Control, RFC 3551"*, July 2003
37. Marie-Luce Bourguet, *"An overview of multimodal interaction techniques and applications"*, in *Encyclopedia of Human Computer Interaction*, C. Ghaoui, 2006, Idea group
38. Georg Schneider, Safia Djennane, Thai-Lai Pham, and Stuart Goose, *"Multimodal Multi Device UIs for Ubiquitous Access to Multimedia Gateways"*
39. Lisetti C, Nasoz F, LeRouge C, Ozyer O, and Alvarez K, *"Developing Multimodal Intelligent Affective Interfaces for Tele-Home Health Care"*, International Journal of Human-Computer Studies, 2002
40. Qiang Ji and Zhiwei Zhu, *"Eye and Gaze Tracking for Interactive Graphic Display"*, Machine Vision and Applications, 2002
41. Duchnowski P, Meier U, and Waibel A, *"See me, hear me: Integrating automatic speech recognition and lipreading"*, Proceeding of the International Conference on Spoken Language Processing, 1994 Yokohama, Japan
42. Nigay L and Coutaz J, *"A generic platform for addressing the multimodal challenge"*, Proceedings of the Conference on Human Factors in Computing Systems, 1995

43. Patterson E., K Gurbuz, S Tufekci Z, and Gowdy J. N, "*Noise-Based Audio-Visual Fusion for Robust Speech Recognition*", *Proceedings of the International Conference on Auditory-Visual Speech Processing*, Scheelsminde, Denmark Scheelsminde, Denmark
44. DaviesDavid Richard Llewellyn, "*Representing time in automated speech recognition*", in *CSLAB, Research School of Information Sciences and Engineering ANU*. 2003
45. Shneiderman B, "*Designing the user interface, 3rd edition*", 1998
46. Bourguet M. L, "*Software design and development of multimodal interaction*", *Proceedings of the IFIP 18th World Computer Congress Topical Days*, 2004
47. Halverson C, Horn D, Karat C, and Karat J, "*The beauty of errors: Patterns of error correction in desktop speech systems*", *Proceedings of the Seventh IFIP TC13 International Conference on Human-Computer Interaction*, 1999 Edinburgh, Scotland
48. VoiceXML Forum, <http://www.voicexml.org/>
49. XHTML+Voice Profile 1.1, <http://www.ibm.com/software/pervasive/multimodal/x+v/11/spec.htm>
50. Steven Pemberton. *XML Events for HTML Authors*, <http://www.w3.org/TR/xml-events/>
51. *Opera Multimodal Browser*, <http://www.opera.com/products/devices/multimodal/>
52. *NetFront*, <http://www.access.co.jp/english/products/nf.html>
53. Luca Tummolini, Andrea Lorenzon, Giancarlo Bo, and Roberto Vaccaro, "*iTutor: a Wireless and Multimodal Support to Industrial Maintenance Activities*", in *Giunti Interactive Labs S.r.l.*
54. IBM, "*XHTML+Voice Programmer's Guide Version 1.0*"
55. *Adobe - Mobile & Devices*, <http://www.adobe.com/mobile/>