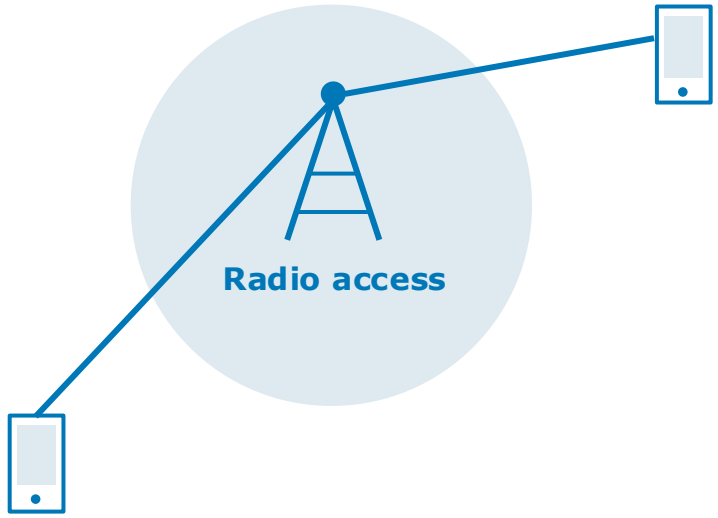
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

# Μ301: Παλιές και νέες προκλήσεις στα δίκτυα κινητών επικοινωνιών

## Διάλεξη #4: Προκλήσεις διαχείρισης πόρων και ετερογενούς κίνησης

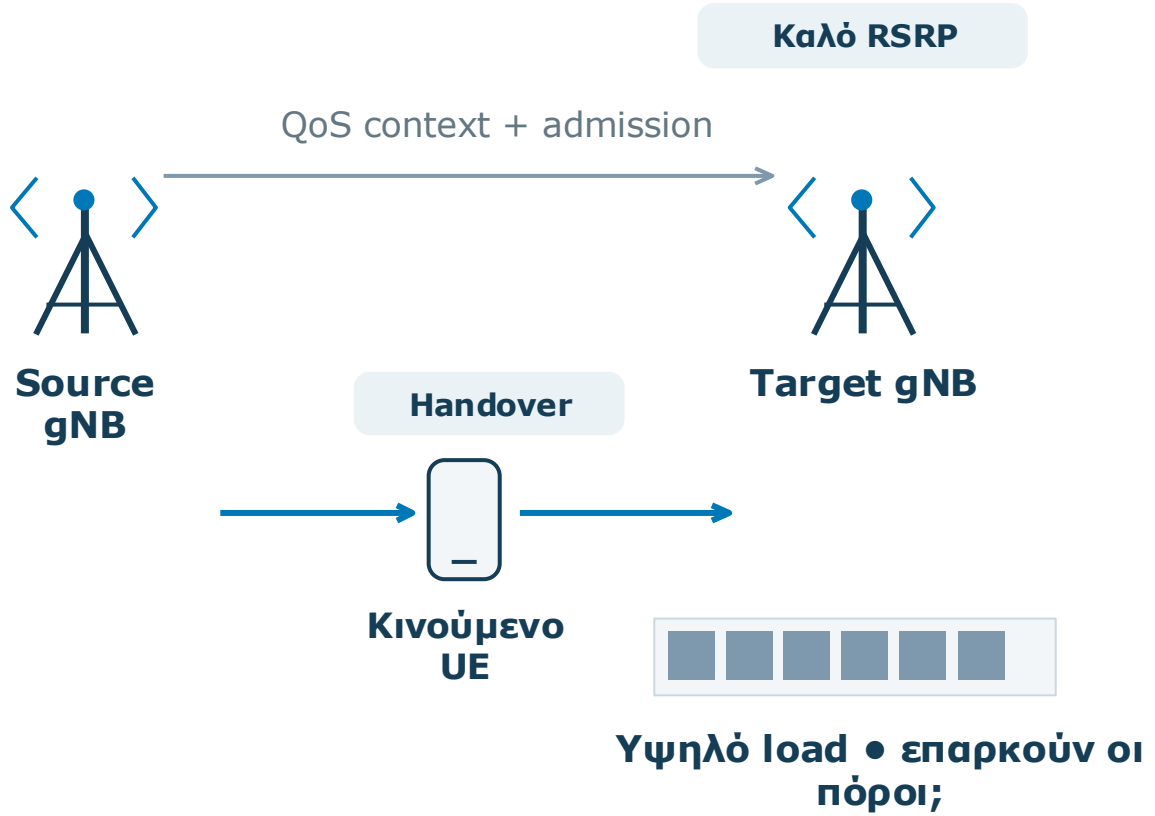



Σωκράτης Μπαρμπουνάκης
Τμήμα Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ
Χειμερινό εξάμηνο 2026-2027

# Τι κρατάμε από την κινητικότητα

Το κατάλληλο target χρειάζεται και διαθέσιμους πόρους

- Το καλό RSRP δεν εξασφαλίζει διαθέσιμη χωρητικότητα: το target cell μπορεί να χρησιμοποιεί ήδη μεγάλο μέρος των radio πόρων του για άλλα UEs.
- Κατά την προετοιμασία του handover, το target gNB εκτελεί admission control, εξετάζοντας αν μπορεί να υποστηρίξει τις απαιτήσεις QoS του UE με τους διαθέσιμους πόρους.
- Μετά το handover, ο scheduler του target gNB κατανέμει πόρους στα QoS flows, λαμβάνοντας υπόψη τις απαιτήσεις τους και την υπόλοιπη κίνηση του cell.



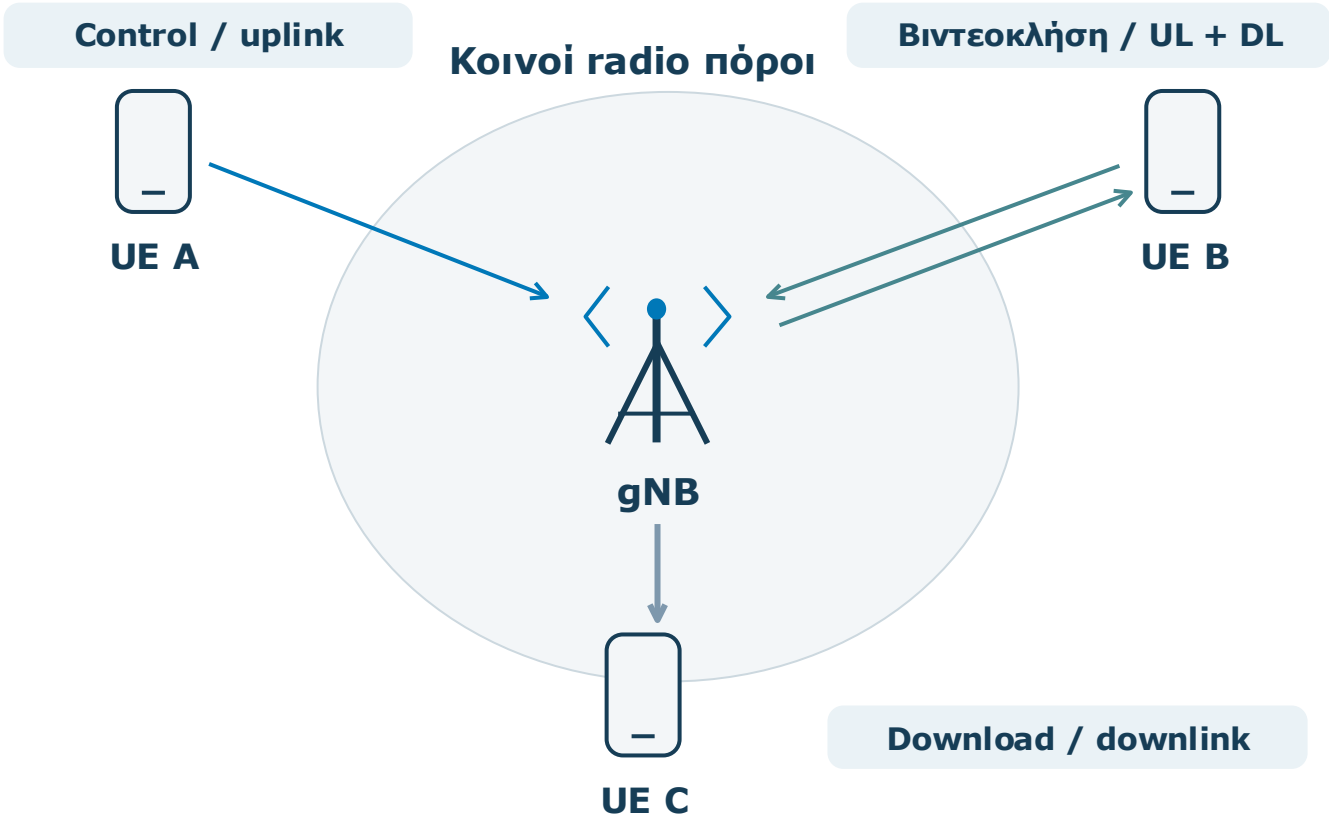

**Σήμερα εξετάζουμε πώς μοιράζεται η χωρητικότητα μέσα στο cell.**



TS 38.300 · 3GPP TS 23.501

# Διαφορετική κίνηση στο ίδιο cell

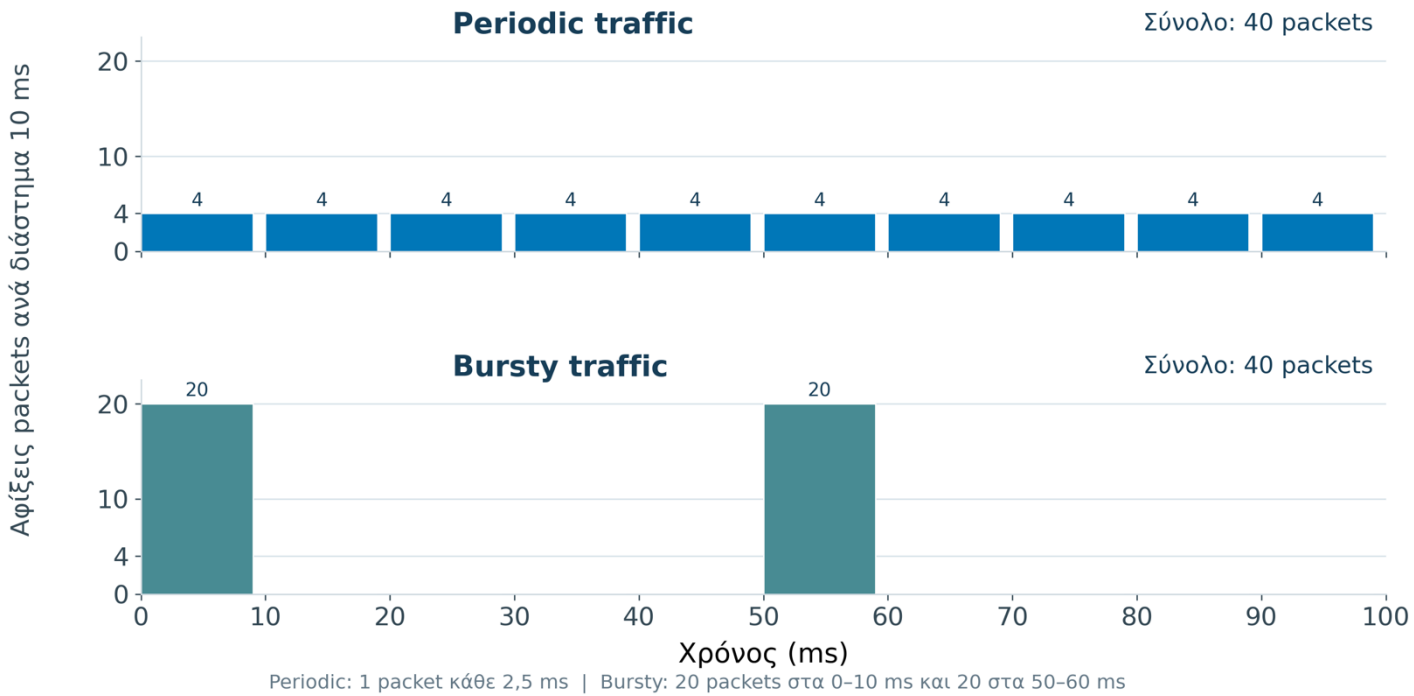
- Μια βιντεοκλήση χρειάζεται έγκαιρη και σχετικά σταθερή παράδοση πακέτων, ενώ ένα download μπορεί συνήθως να ανεχθεί μεγαλύτερη καθυστέρηση.
- Ένα packet με εντολή ελέγχου μπορεί να είναι μικρό, αλλά χρονικά κρίσιμο: π.χ. η εντολή ακινητοποίησης ενός ρομπότ καθώς πλησιάζει ένα εμπόδιο πρέπει να φτάσει έγκαιρα.
- Οι διαφορετικές υπηρεσίες μοιράζονται τους ίδιους radio πόρους και ανταγωνίζονται για χρόνο, συχνότητα, ισχύ και χωρικούς πόρους.



# Traffic profiles και χρονική μεταβλητότητα

Ο ίδιος μέσος ρυθμός μπορεί να κρύβει πολύ διαφορετικά bursts

- Το **packet size και το inter-arrival time** καθορίζουν τον **ρυθμό άφιξης δεδομένων στην ουρά** και, επομένως, τη ζήτηση για εξυπηρέτηση.
- Τα **bursts** προκαλούν παροδικές αιχμές κίνησης, οι οποίες μπορεί να **αυξήσουν** την ουρά και την **καθυστέρηση**, **ακόμη και όταν το μέσο load είναι χαμηλό.**
- Το **periodic traffic** κάνει τον χρόνο άφιξης των packets **πιο προβλέψιμο**, αλλά η ποιότητα του radio channel και ο χρόνος εξυπηρέτησής τους εξακολουθούν να μεταβάλλονται.



# Ενδεικτικά traffic profiles

Διαφορετικές εφαρμογές παράγουν κίνηση με διαφορετικό μέγεθος, ρυθμό και χρονική συμπεριφορά.

| Ενδεικτική υπηρεσία | Packet size | Inter-arrival time | Μέσος ρυθμός UL | Μέσος ρυθμός DL | Χαρακτηριστικό |
|---|---|---|---|---|---|
| Αισθητήρας IoT | 200 B | 10 s (ενδεικτικό, ποικίλει με την εφαρμογή) | 0,16 kbps | - | Μικρά, αραιά, περιοδικά μηνύματα |
| Έλεγχος ρομπότ | 100 B | 10 ms | 80 kbps | 80 kbps | Περιοδική ανταλλαγή με χρονικές απαιτήσεις |
| Φωνητική κλήση | 160 B | 20 ms | 64 kbps | 64 kbps | Συνεχής αμφίδρομη ροή |
| Βιντεοκλήση | ≈1.000 B | ≈4 ms | ≈2 Mbps | ≈2 Mbps | Μεταβλητή και αμφίδρομη κίνηση |
| Video streaming | ≈1.000 B | ≈1,6 ms | - | ≈5 Mbps | Κυρίως downlink, πιθανές αιχμές |
| File download | Μεταβλητό | Μεταβλητό | - | Μεταβλητός | Bursty κίνηση, αξιοποιεί τη διαθέσιμη χωρητικότητα |

# Απαιτήσεις QoS και εμπειρία εφαρμογής



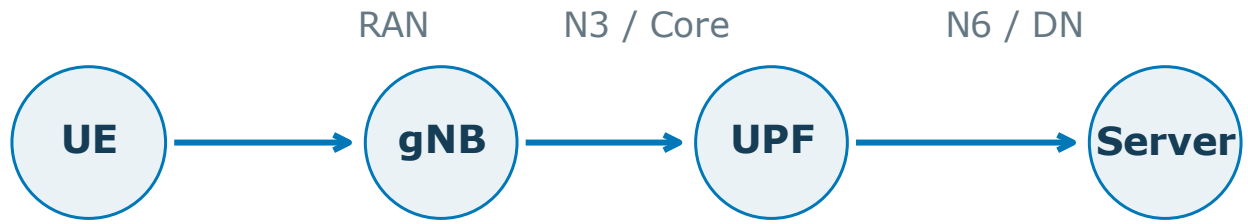
Οι απαιτήσεις QoS εξαρτώνται από την εφαρμογή: κάθε υπηρεσία έχει διαφορετική ανοχή στην καθυστέρηση, στη μεταβλητότητά της και στις απώλειες πακέτων.

| Μέγεθος | Τι μετράμε | Γιατί έχει σημασία |
|---|---|---|
| **Throughput** | Ρυθμός δεδομένων που μεταφέρονται επιτυχώς, σε συγκεκριμένο σημείο μέτρησης. | Επηρεάζει, μεταξύ άλλων, την ταχύτητα ενός download και την ποιότητα video που μπορεί να υποστηριχθεί. |
| **Latency** | Χρόνος που απαιτείται για τη μεταφορά δεδομένων από την αποστολή έως την παράδοση. | Επηρεάζει την απόκριση σε διαδραστικές εφαρμογές και την έγκαιρη λειτουργία ενός control loop. |
| **Jitter** | Μεταβολή της καθυστέρησης μεταξύ διαδοχικών πακέτων. | Μεγάλο jitter μπορεί να προκαλέσει διακοπές ή ασυνέχειες σε ήχο και video, ακόμη και όταν η μέση καθυστέρηση είναι αποδεκτή. |
| **Reliability** | Ποσοστό packets που παραδίδονται επιτυχώς, για ένα συγκεκριμένο τμήμα της διαδρομής και καθορισμένη περίοδο μέτρησης. | Είναι κρίσιμη όταν απώλειες ή καθυστερημένες παραδόσεις επηρεάζουν τη λειτουργία της εφαρμογής. |
| **Availability** | Ποσοστό χρόνου κατά το οποίο η υπηρεσία είναι διαθέσιμη, βάσει καθορισμένων κριτηρίων. | Επηρεάζει τη συνέχεια λειτουργίας, ιδιαίτερα σε υπηρεσίες που πρέπει να παραμένουν προσβάσιμες. |

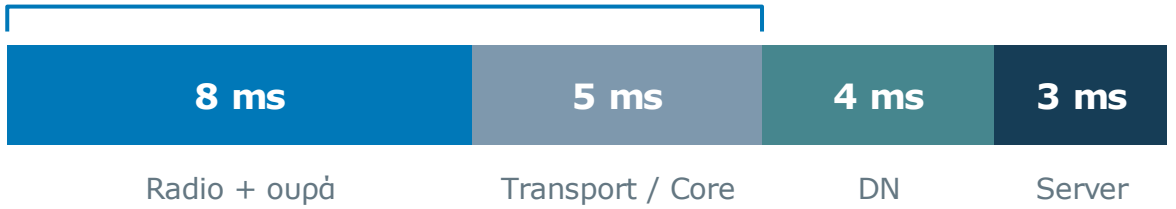
# Πού καταναλώνεται το latency budget;

Το radio είναι μόνο ένα τμήμα της end-to-end διαδρομής

- **Το Packet Delay Budget (PDB) είναι ένα όριο καθυστέρησης για τη μεταχείριση ενός QoS flow**
  - Η πραγματική καθυστέρηση είναι αυτό που μετράμε και συγκρίνουμε με το PDB.
- **Η συνολική καθυστέρηση προκύπτει από επιμέρους στάδια**: αναμονή σε ουρές, μετάδοση μέσω του RAN και του Core, μεταφορά στο Data Network και επεξεργασία στο server.
- Το PDB αφορά το τμήμα μεταξύ του UE και του σημείου τερματισμού του N6 στο UPF. Στο παράδειγμα, η καθυστέρηση σε αυτό το τμήμα είναι 8 + 5 = 13 ms.
- Η απόκριση της εφαρμογής εξαρτάται και από όσα συμβαίνουν πέρα από το UPF.




**UE-UPF: 8 + 5 = 13 ms στο παράδειγμα**




**Σύνολο παραδείγματος:** 20 ms - μετάδοση από UE έως server και επεξεργασία στο server (χωρίς την επιστροφή της απάντησης)

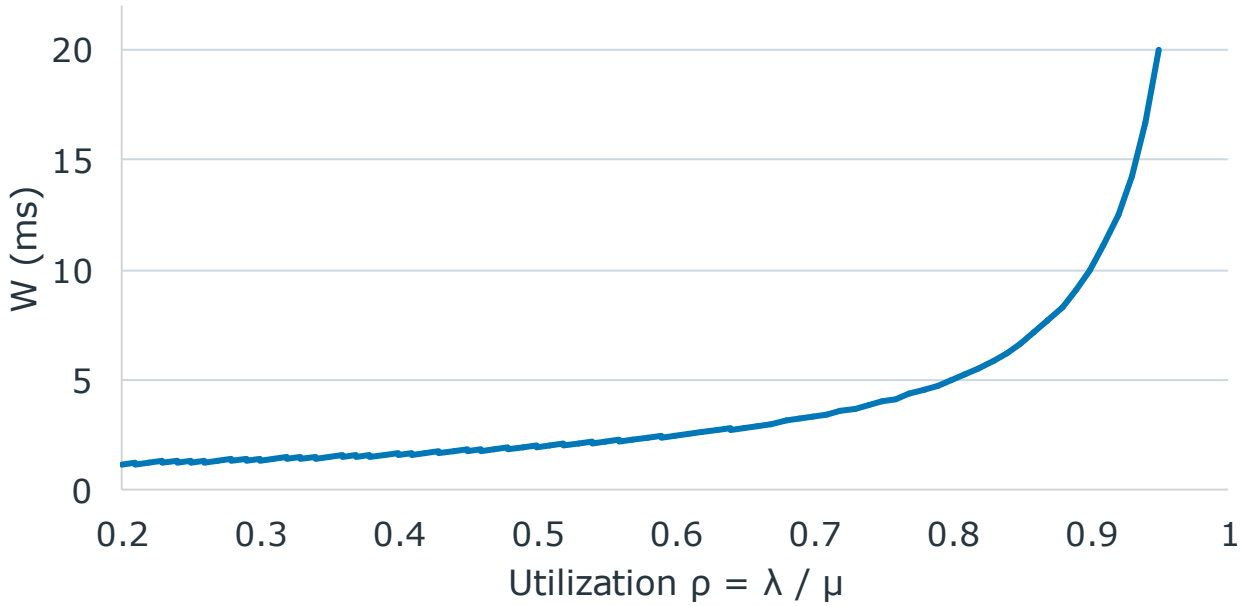
# Ουρές και καθυστέρηση υπό υψηλό load



- Όταν ο ρυθμός αφίξεων πλησιάζει τον ρυθμό εξυπηρέτησης, τα packets περιμένουν περισσότερο στην ουρά, ιδιαίτερα όταν η κίνηση είναι μεταβλητή.
- Ένα μεγαλύτερο buffer μπορεί να απορροφήσει προσωρινά ένα burst, αλλά η αναμονή των packets μπορεί να αυξήσει σημαντικά το latency.
- Αν οι αφίξεις υπερβαίνουν επίμονα την εξυπηρέτηση, η ουρά μεγαλώνει και ένα πεπερασμένο buffer τελικά οδηγεί σε απόρριψη packets.

Τα packets φτάνουν στην ουρά με μέσο ρυθμό *λ*, ενώ ο μηχανισμός εξυπηρέτησης μπορεί να εξυπηρετήσει, όταν έχει δεδομένα διαθέσιμα, κατά μέσο όρο *μ* packets ανά δευτερόλεπτο (ρυθμός εξυπηρέτησης).

**M/M/1: W = 1 / (μ − λ)**

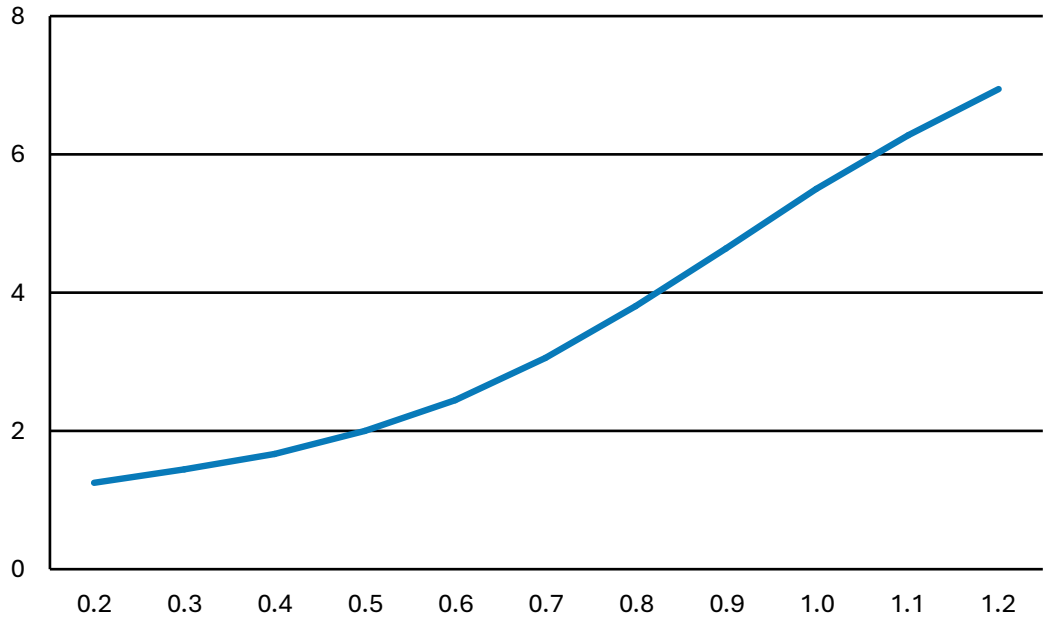



W: μέσος συνολικός χρόνος (αναμονή + εξυπηρέτηση)
μ = 1.000 packets/s, M/M/1 με ρ < 1

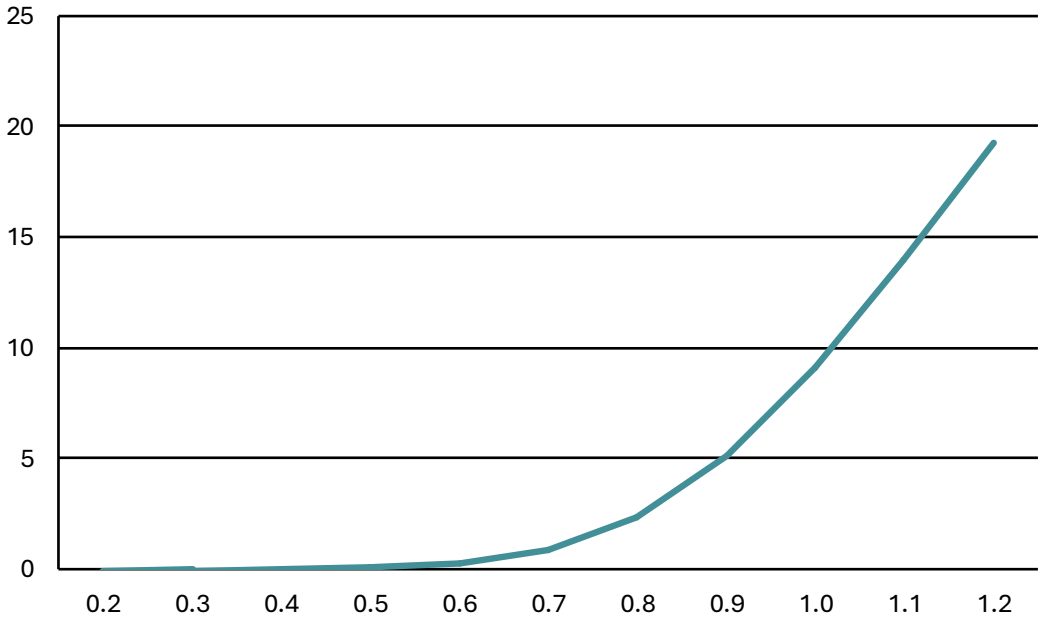
# Πεπερασμένο buffer: καθυστέρηση και απώλειες

**M/M/1/K**: όταν το buffer γεμίσει, ένα νέο packet απορρίπτεται

## Μέσος χρόνος w στο σύστημα (ms)



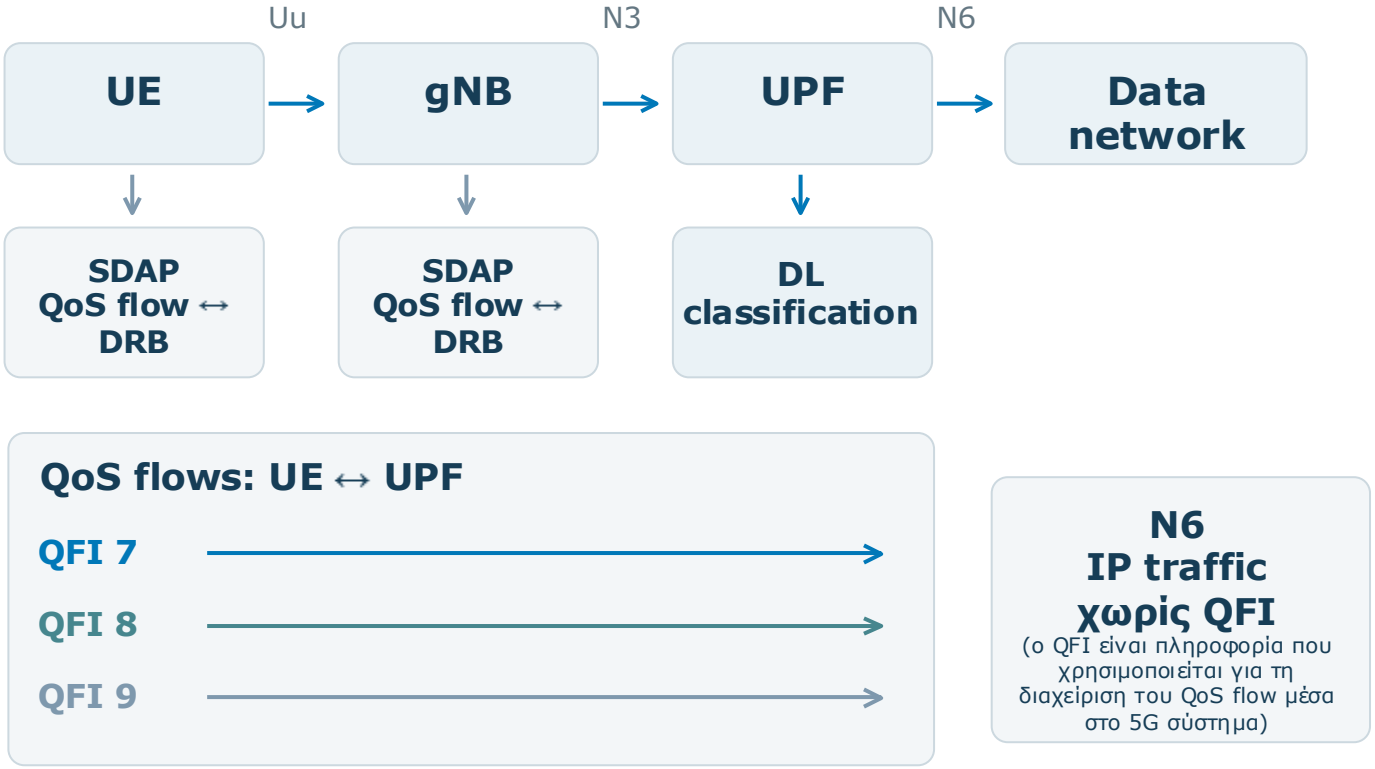

## Απορρίψεις packets (%)



# Το 5G QoS model στη διαδρομή των δεδομένων

QoS flows μέσα στο PDU session, radio bearers στο RAN

- Όπως είδαμε, **ένα PDU Session είναι η λογική σύνδεση που δημιουργεί το 5G δίκτυο μεταξύ ενός UE και ενός Data Network (DN)**.
- Ένα **PDU session μπορεί να περιλαμβάνει πολλαπλά QoS flows***.
- Το **QFI (QoS Flow Identifier) αναγνωρίζει το QoS flow** στο συγκεκριμένο PDU session και επιτρέπει τη συσχέτιση των packets με αυτό.
- Στο UE και στο gNB, το SDAP αντιστοιχίζει τα QoS flows σε Data Radio Bearers (DRBs) για τη μεταφορά τους μέσω του RAN.



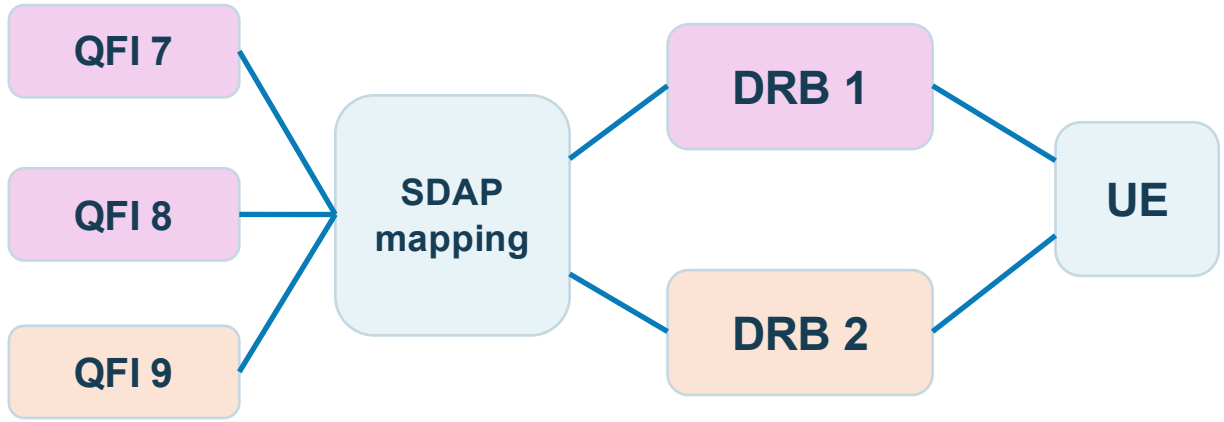

***Ένα UE μπορεί επίσης να έχει περισσότερα PDU Sessions**, προς διαφορετικά Data Networks (ή και προς το ίδιο). Ξεχωριστά sessions είναι χρήσιμα όταν απαιτείται διαφορετικό πλαίσιο συνδεσιμότητας, για παράδειγμα πρόσβαση στο Internet και σε ιδιωτικό δίκτυο, ή διαφορετικά slices.

# Data Radio Bearer (DRB)

Η λογική μεταφορά δεδομένων χρήστη στη radio διεπαφή

- Το DRB είναι ένας λογικός φορέας μεταφοράς δεδομένων χρήστη μεταξύ UE και gNB, μέσω της radio διεπαφής Uu.
- Το SDAP εφαρμόζει την αντιστοίχιση των QoS flows στα DRBs, σύμφωνα με τους κανόνες που έχει διαμορφώσει το RAN.
- Περισσότερα από ένα QoS flows μπορούν να μοιράζονται το ίδιο DRB. Η χρήση ξεχωριστού DRB εξαρτάται από την απαιτούμενη μεταχείριση.

Ενδεικτικό mapping στο ίδιο PDU Session




QFI 7 + QFI 8 → DRB 1 | QFI 9 → DRB 2

**Το DRB δεν είναι αποκλειστική λωρίδα συχνοτήτων: οι radio πόροι κατανέμονται δυναμικά.**



3GPP TS 38.300 · TS 23.501

# Τι εκφράζει το 5QI

## Χαρακτηριστικά QoS που καθοδηγούν την εξυπηρέτηση ενός flow

- Το 5QI είναι αναγνωριστικό που συνδέεται με χαρακτηριστικά QoS για τη μεταχείριση ενός QoS flow.
- Τα χαρακτηριστικά περιλαμβάνουν
  - resource type (GBR, non-GBR, delay-critical GBR)
  - priority level (βλ. επόμενο slide)
  - Packet Delay Budget και
  - Packet Error Rate.
- Το 5QI δεν ορίζει από μόνο του ρυθμό σε Mbps. Για GBR flows, το Guaranteed Flow Bit Rate (GFBR) και ο Maximum FBR καθορίζονται χωριστά.

Παράδειγμα standardized 5QI

**5QI 1** — **Προκαθορισμένα χαρακτηριστικά QoS**

| | |
|---|---|
| **Resource type** | **GBR** |
| **Priority level** | **20 (default)** |
| **Delay budget** | **100 ms** |
| **Packet error rate** | **$10^{-2}$** |

Ενδεικτική χρήση: conversational voice



3GPP TS 38.300 · TS 23.501

# Παραδείγματα standardized 5QI

Απόσπασμα του 3GPP TS 23.501 Table 5.7.4-1, Release 18



| 5QI | Resource type | Priority | PDB | PER | Παράδειγμα υπηρεσίας |
|---|---|---|---|---|---|
| 1 | GBR | 20 | 100 ms | $10^{-2}$ | Conversational voice |
| 2 | GBR | 40 | 150 ms | $10^{-3}$ | Conversational video |
| 5 | Non-GBR | 10 | 100 ms | $10^{-6}$ | IMS signalling (σημαντικά μηνύματα για τη λειτουργία μια φωνητική κλήσης) |
| 9 | Non-GBR | 90 | 300 ms | $10^{-6}$ | Buffered video / TCP |
| 82 | Delay-critical GBR | 19 | 10 ms | $10^{-4}$ | Discrete automation |

**Priority level: μικρότερη τιμή σημαίνει υψηλότερη προτεραιότητα.**

# GBR, Non-GBR και Delay-critical GBR

Η κατηγορία πόρων αλλάζει τι πρέπει να υποστηρίξει το δίκτυο

| Resource type | Χαρακτηριστικά | Πρόκληση RRM |
|---|---|---|
| GBR | Guaranteed Flow Bit Rate / Maximum Flow Bit Rate | Αποδοχή και εξυπηρέτηση δεσμευμένων απαιτήσεων |
| Non-GBR | Χωρίς ανά-flow GFBR, AMBR* | Εξυπηρέτηση βάσει QoS priority και διαθέσιμων πόρων |
| Delay-critical GBR | Αυστηρό Packet Delay Budget + Maximum Data Burst Volume | Έγκαιρη παράδοση επιτρεπόμενων bursts |

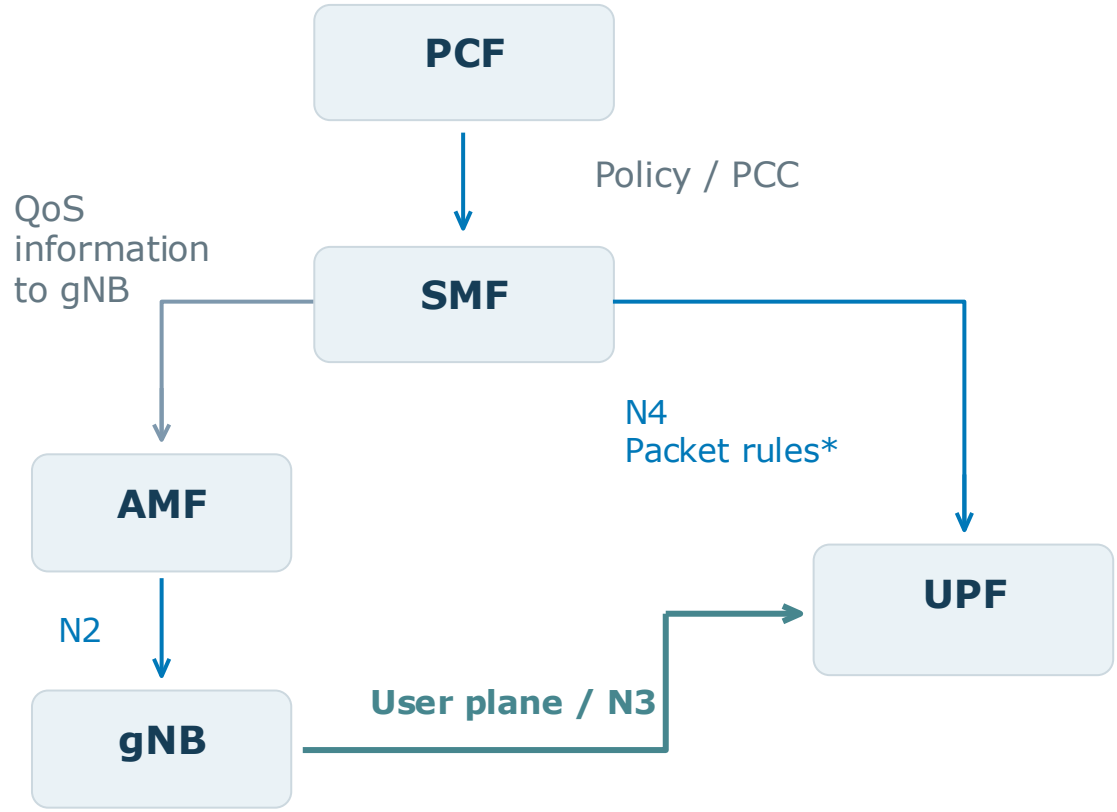
*Το AMBR (Aggregate Maximum Bit Rate) είναι ο μέγιστος αθροιστικός ρυθμός δεδομένων που επιτρέπεται για τα Non-GBR QoS flows ενός συγκεκριμένου PDU Session ή ενός UE, ανάλογα με το ποιο AMBR εφαρμόζεται.

**Guaranteed bit rate δεν σημαίνει αποκλειστικά PRBs σε κάθε slot.**

# Ποιος διαμορφώνει την εξυπηρέτηση QoS;

Policy στο Core και εφαρμογή κανόνων στο user plane και στο RAN

- **PCF, Policy Control Function:** παρέχει **πολιτικές** που καθορίζουν πώς πρέπει να αντιμετωπίζεται η κίνηση της υπηρεσίας, με βάση τη συνδρομή και τους κανόνες του operator. (Προσοχή! Ο scheduler αποφασίζει τη δυναμική κατανομή των radio πόρων)
- **SMF, Session Management Function:** διαχειρίζεται το PDU Session, διαμορφώνει τα QoS flows και εγκαθιστά τους σχετικούς κανόνες στο UPF*. Παρέχει επίσης στο RAN, μέσω του AMF, τις πληροφορίες QoS που χρειάζεται.
- **UPF, User Plane Function:** ταξινομεί τα downlink packets στα κατάλληλα QoS flows και εφαρμόζει κανόνες προώθησης και περιορισμού ρυθμού.
- **gNB:** αξιοποιεί τις πληροφορίες QoS για να διαμορφώσει τα radio bearers και να κατανείμει τους διαθέσιμους radio πόρους μέσω του scheduler.




* Το SMF εγκαθιστά στο UPF κανόνες που του λένε πώς να αναγνωρίζει και πώς να χειρίζεται τα packets του PDU Session. Packet detection rules, Forwarding action rules, QoS enforcement rules, Usage reporting rules

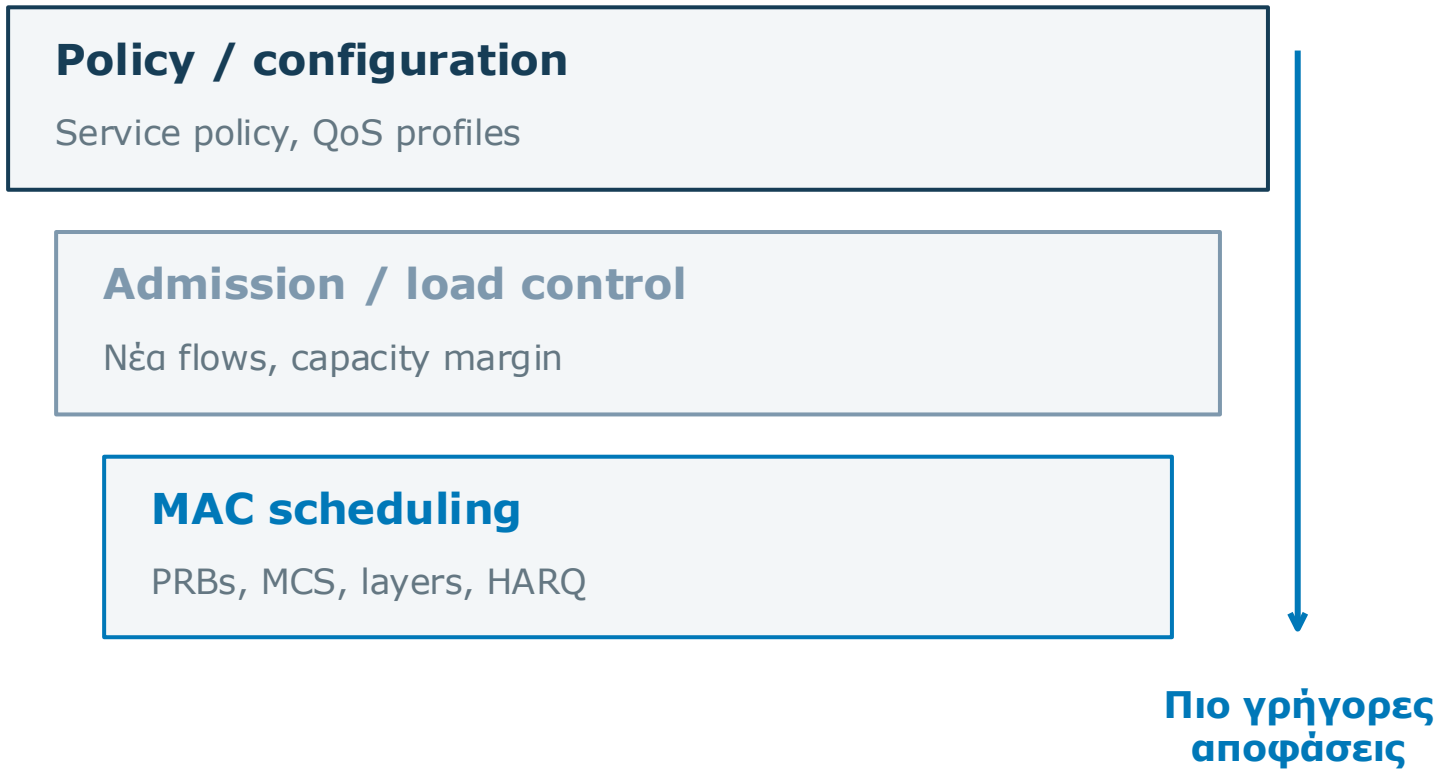


3GPP TS 23.501 · TS 38.300

# Το RRM λειτουργεί σε πολλαπλές κλίμακες

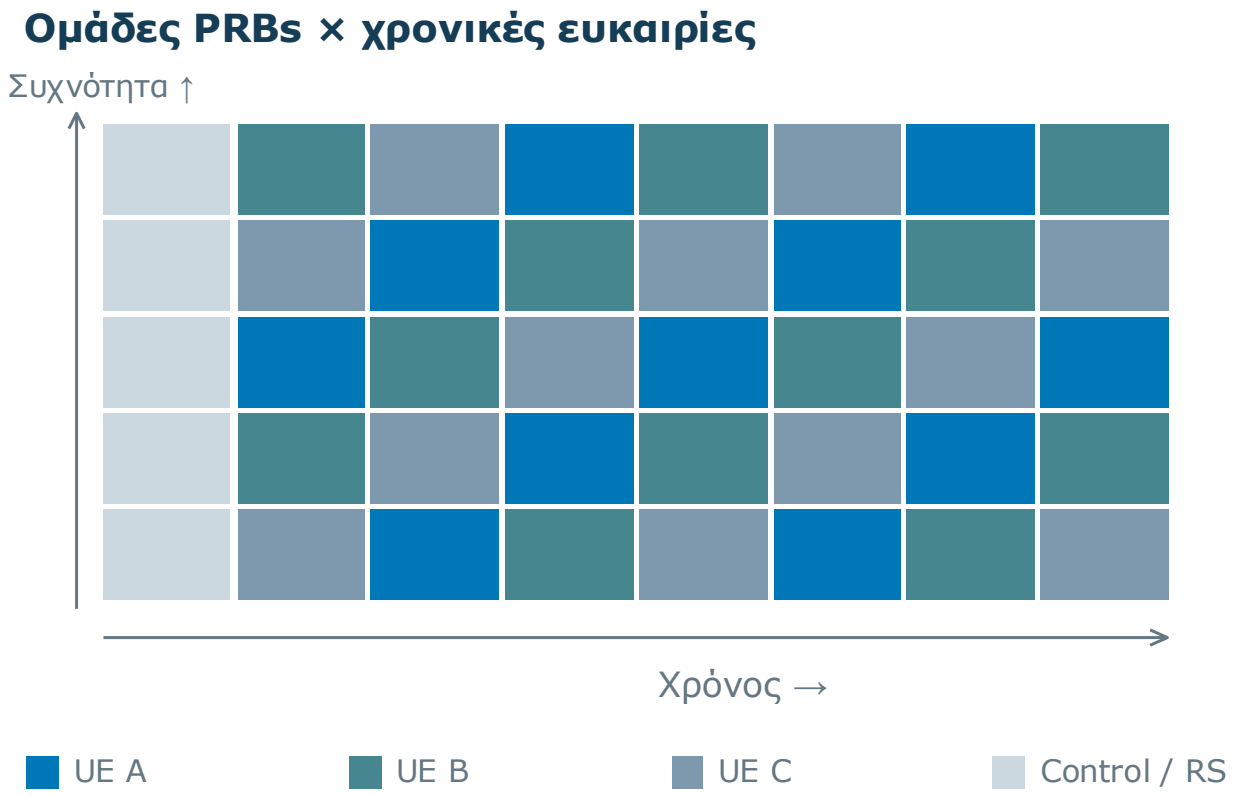
Αποφάσεις με διαφορετική χρονική κλίμακα

- **Policy / configuration**: οι γενικές πολιτικές και τα QoS profiles καθορίζουν τις απαιτήσεις των υπηρεσιών και το πλαίσιο μέσα στο οποίο λειτουργεί το δίκτυο.
- **Admission / load control**: το δίκτυο εξετάζει αν μπορεί να αποδεχθεί νέες απαιτήσεις QoS, λαμβάνοντας υπόψη τους διαθέσιμους πόρους και το υφιστάμενο load.
- **MAC scheduling**: το gNB λαμβάνει γρήγορες, επαναλαμβανόμενες αποφάσεις για την κατανομή radio πόρων, με βάση το QoS, την κίνηση στις ουρές και τις συνθήκες καναλιού.



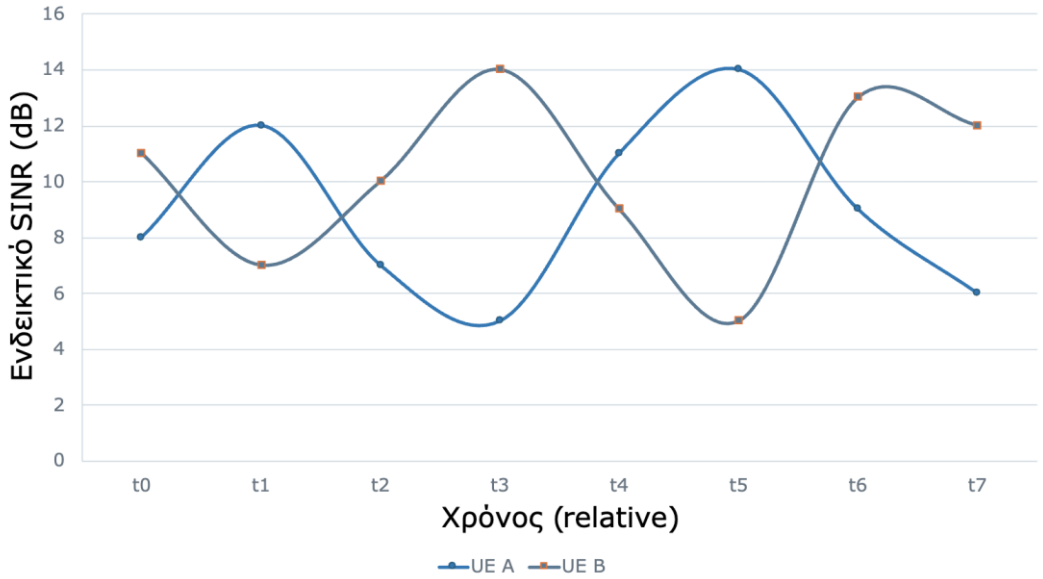
# Scheduler (υπενθύμιση)

- Ο scheduler εκχωρεί πόρους σε χρόνο και συχνότητα, δηλαδή PRBs και διαθέσιμα OFDM symbols, για τις μεταδόσεις των UEs.
- Το επιλεγμένο MCS και ο αριθμός spatial layers επηρεάζουν πόσα bits μπορούν να μεταφερθούν με τους εκχωρημένους πόρους.



# Channel-dependent scheduling

Το radio channel και η ουρά δεδομένων αλλάζουν με τον χρόνο




| CQI | Modulation | Code rate | Αποδοτικότητα[1] |
|---|---|---|---|
| 1 | QPSK | 78/1024 ≈ 0,076 | 0,15 |
| 4 | QPSK | 308/1024 ≈ 0,301 | 0,60 |
| 7 | 16-QAM | 378/1024 ≈ 0,369 | 1,48 |
| 9 | 16-QAM | 616/1024 ≈ 0,602 | 2,41 |
| 10 | 64-QAM | 466/1024 ≈ 0,455 | 2,73 |
| 12 | 64-QAM | 666/1024 ≈ 0,650 | 3,90 |
| 15 | 64-QAM | 948/1024 ≈ 0,926 | 5,55 |

## Channel quality

CQI / CSI βοηθούν στην εκτίμηση της αποδοτικότητας της μετάδοσης στους διαθέσιμους πόρους.

## Traffic state

Buffer occupancy και χρόνος αναμονής δείχνουν πόσα δεδομένα εκκρεμούν και ποια deadlines πλησιάζουν.

## Service policy

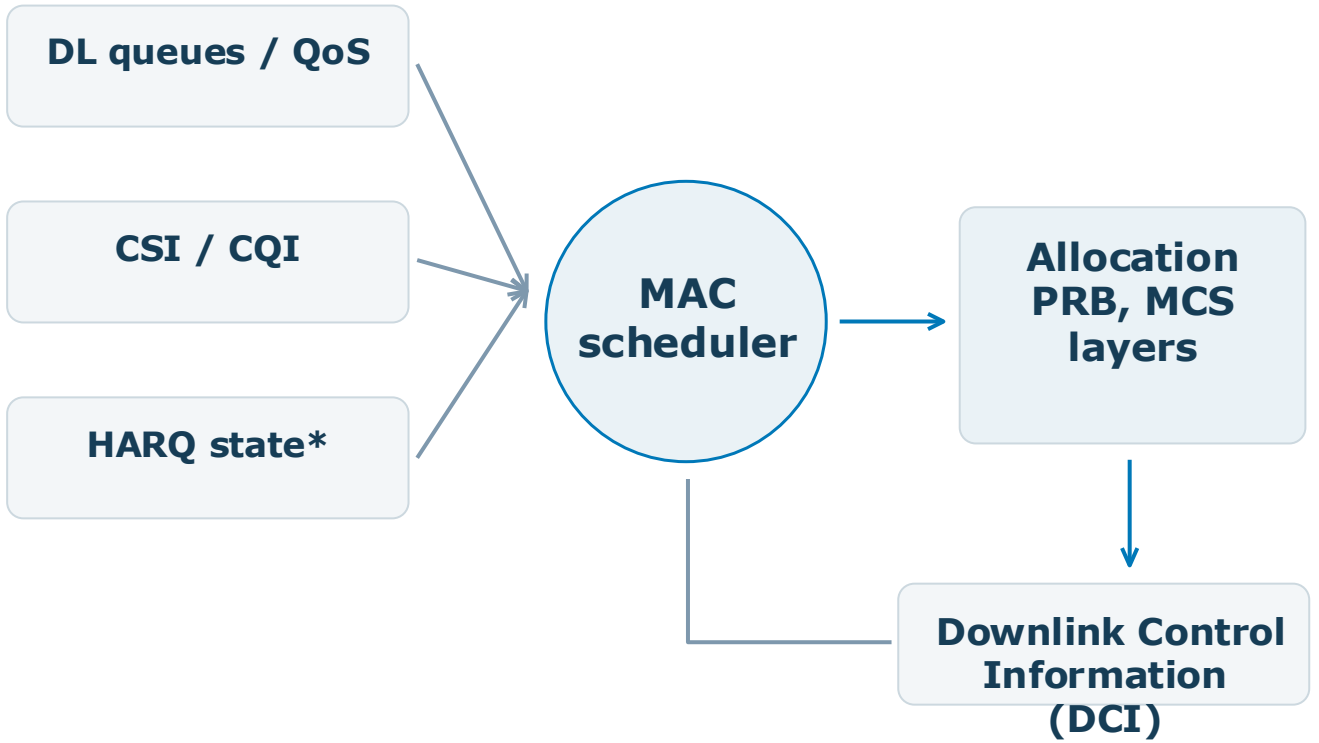
QoS priority και fairness αλλάζουν την επιλογή. Ο scheduler μπορεί να εξυπηρετήσει ένα επείγον UE με χαμηλότερο CQI.

[1] Ονομαστικά **information bits ανά data RE, ανά layer**. Υπολογίζεται ως: bits του modulation × code rate. Το **code rate** δείχνει τι ποσοστό των bits που μεταδίδονται αντιστοιχεί σε πληροφορία, αφού προστεθεί η κωδικοποίηση προστασίας από λάθη.

# Downlink scheduling στο gNB

Συνδυασμός channel state, ουρών και QoS

- Το gNB γνωρίζει τα δεδομένα που περιμένουν στις downlink ουρές και τα χαρακτηριστικά QoS των flows.
- Τα CSI/CQI reports και το HARQ feedback βοηθούν το gNB να εκτιμήσει τις radio συνθήκες και την ανάγκη επαναμετάδοσης.
- Ο scheduler επιλέγει UE, πόρους, MCS και ενδεχομένως spatial layers.
- Η εκχώρηση σηματοδοτείται στο UE μέσω downlink control information (DCI).




*Το HARQ state είναι η πληροφορία που διατηρεί το gNB για την κατάσταση των πρόσφατων μεταδόσεων προς ένα UE: ποιες έχουν επιβεβαιωθεί, ποιες περιμένουν απάντηση και ποιες χρειάζονται ενδεχομένως επαναμετάδοση.

# Ενδεικτικά scheduling policies: απόδοση και fairness

Η πολιτική ορίζει ποια συμπεριφορά επιδιώκει ο scheduler

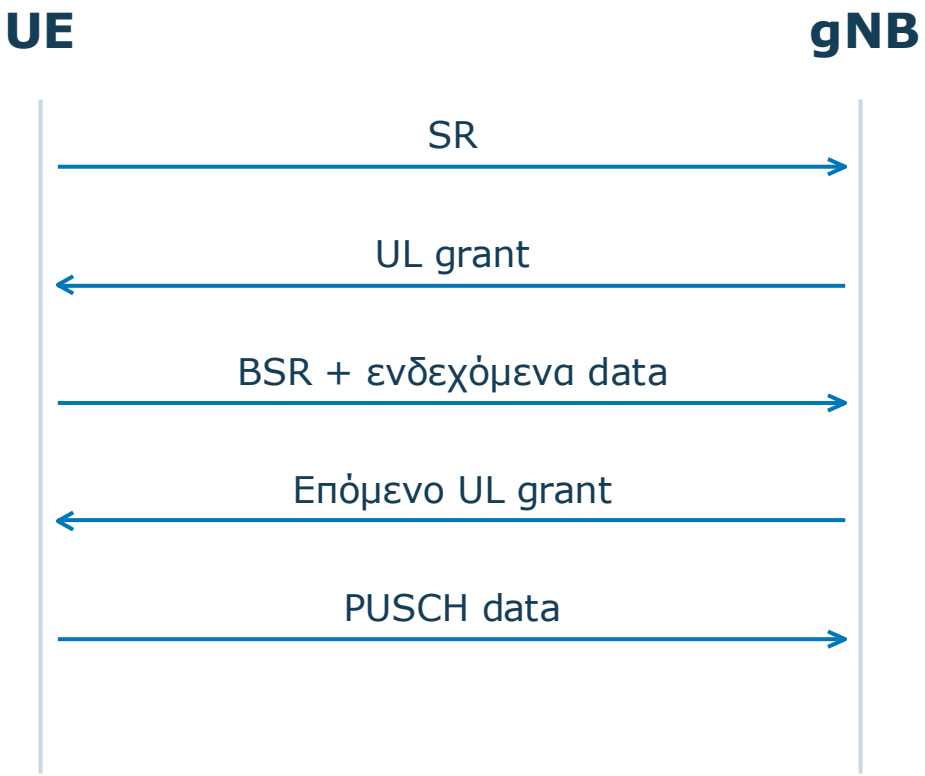
| Policy | Κανόνας επιλογής | Πλεονέκτημα | Συμβιβασμός |
|---|---|---|---|
| Round Robin | Εναλλαγή ενεργών UEs με σειρά | Απλή και προβλέψιμη πρόσβαση σε resources | Δεν αξιοποιεί πάντα τις καλές στιγμές του channel |
| Max-CQI | Προτίμηση στο καλύτερο instantaneous channel | Μεγάλο potential sum throughput | Κίνδυνος συστηματικής υποεξυπηρέτησης edge UEs |
| Proportional Fair | Σύγκριση instant rate με past average rate | Συνδυάζει channel opportunities και σχετική fairness | Δεν εγγυάται από μόνο του deadlines ή ίσους ρυθμούς |

Στην απλή PF μορφή, το metric είναι r_i(t) / R_avg,i(t). **Ένα UE αποκτά προτεραιότητα όταν η τρέχουσα ευκαιρία είναι καλή σε σχέση με τον προηγούμενο μέσο ρυθμό του.**

# Uplink – Dynamic scheduling: το gNB χρειάζεται πληροφορία για τις ουρές

SR, BSR και grant στην κοινή περίπτωση dynamic scheduling

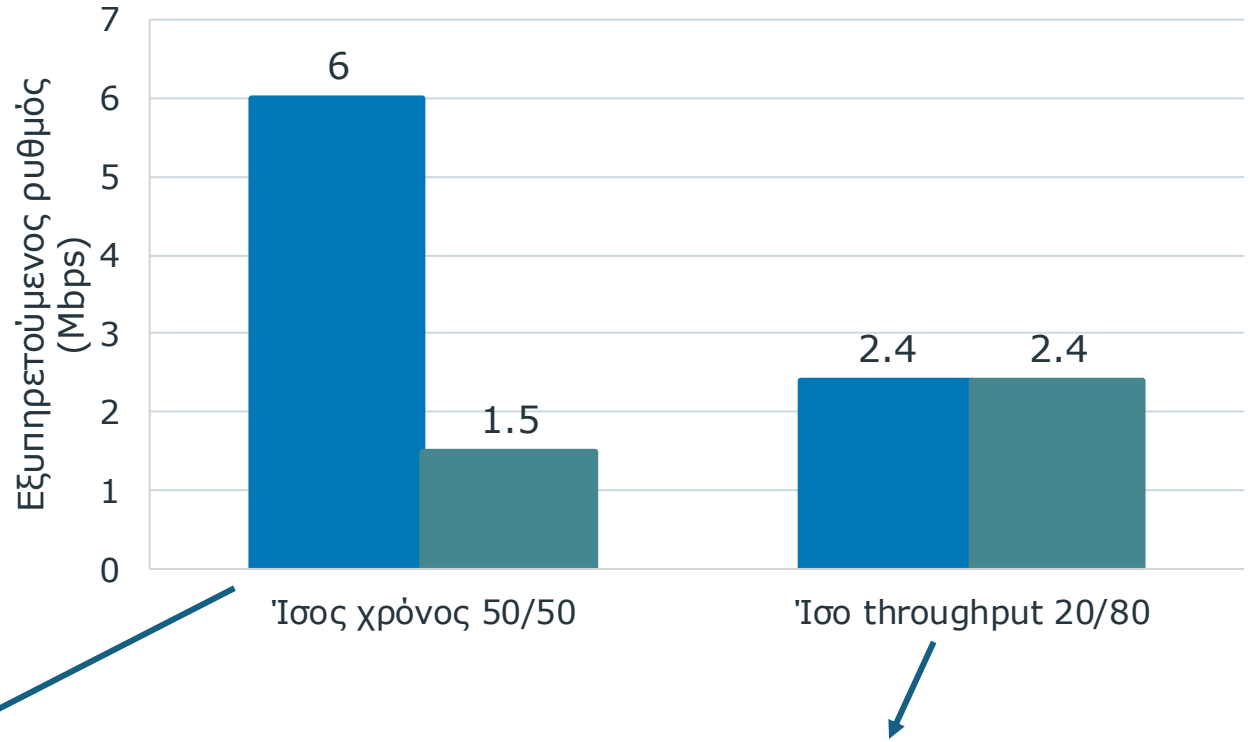
- Όταν χρειάζεται uplink πόρους και δεν διαθέτει κατάλληλο grant, το UE μπορεί να στείλει **Scheduling Request (SR).**
- Το **Buffer Status Report (BSR)** αποστέλλεται από το UE και ενημερώνει το gNB πόσα δεδομένα έχει διαθέσιμα προς αποστολή στο uplink.
- Το **Power Headroom Report (PHR)** πληροφορεί το gNB για το περιθώριο ισχύος που έχει το UE σε σχέση με την επιτρεπόμενη μέγιστη ισχύ εκπομπής του.
- Το gNB εκδίδει **UL grant**
- **Το UE επιλέγει τα δεδομένα** που θα τοποθετήσει στη μετάδοση, βάσει των κανόνων prioritization.



# Δίκαιοι πόροι ή δίκαιο throughput;

## Η επιλογή του κριτηρίου αλλάζει το αποτέλεσμα

- Ένα UE με χαμηλότερο spectral efficiency χρειάζεται συνήθως περισσότερους radio πόρους για να μεταφέρει τον ίδιο αριθμό bits.
- Ίσος χρόνος ή ίσοι πόροι δεν συνεπάγονται ίσο throughput. Η επιδίωξη ίσων ρυθμών μπορεί να μειώσει το συνολικό throughput.
- Η έννοια της δικαιοσύνης πρέπει να ορίζεται: συγκρίνουμε μερίδιο πόρων, επιτυγχανόμενο ρυθμό ή ικανοποίηση QoS;




| UE | Μερίδιο χρόνου | Υπολογισμός | Μέσο throughput |
|---|---|---|---|
| A | 50% | $12 \times 0{,}5$ | **6 Mbps** |
| B | 50% | $3 \times 0{,}5$ | **1,5 Mbps** |

**Συνολικό throughput: 6 + 1,5 = 7,5 Mbps.**

Τα δύο UEs παίρνουν **ίσο χρόνο**, αλλά δεν πετυχαίνουν ίσο ρυθμό, επειδή έχουν διαφορετικές radio συνθήκες.

| UE | Μερίδιο χρόνου | Υπολογισμός | Μέσο throughput |
|---|---|---|---|
| A | 20% | $12 \times 0{,}2$ | **2,4 Mbps** |
| B | 80% | $3 \times 0{,}8$ | **2,4 Mbps** |

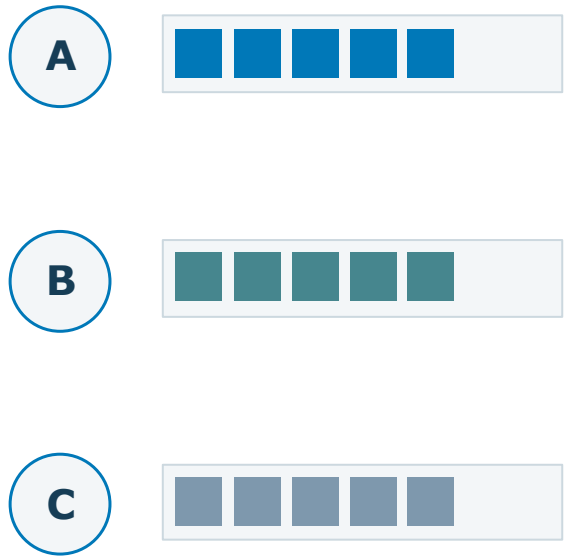
**Συνολικό throughput: 2,4 + 2,4 = 4,8 Mbps.**

Το B χρειάζεται **τέσσερις φορές περισσότερο χρόνο** από το A για να πετύχει τον ίδιο ρυθμό, επειδή στο παράδειγμα μπορεί να μεταφέρει το ένα τέταρτο των δεδομένων ανά ίση ευκαιρία εξυπηρέτησης.

# QoS-aware scheduling και ηλικία packet

Η επείγουσα ανάγκη δεν φαίνεται μόνο στο CQI

- Το **Head-of-Line (HOL) delay** είναι ο χρόνος που περιμένει στην ουρά το packet στην κεφαλή της.
- Ένα **deadline-aware** κριτήριο μπορεί να συνδυάζει τον **χρόνο** που απομένει, τις απαιτήσεις **QoS** και τον εφικτό **ρυθμό στο radio channel**.
- Το **Earliest Deadline First (EDF)** δίνει προτεραιότητα στο πλησιέστερο deadline, ενώ μέθοδοι όπως το Modified Largest Weighted Delay First (M-LWDF) συνδυάζουν καθυστέρηση και radio conditions.




HOL 9 ms / deadline 10 ms

**Υπόλοιπο: 1 ms**

HOL 20 ms / deadline 50 ms

**Υπόλοιπο: 30 ms**

Elastic • χωρίς fixed deadline

**Παρακολουθούμε το ιστορικό εξυπηρέτησης**



Andrews et al. · TS 38.300

# Άσκηση: ποιος παίρνει τους διαθέσιμους πόρους;

Έστω ότι υπάρχουν 12 μονάδες radio πόρων.

Υποθέτουμε ίδια απόδοση ανά μονάδα πόρου, ώστε να συγκρίνουμε τις αποφάσεις προτεραιότητας.

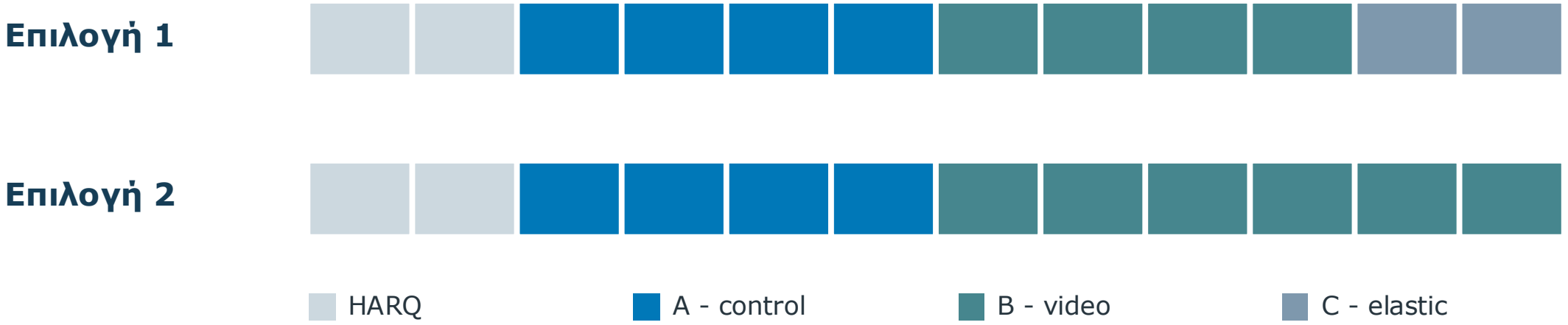
Από τις 12 μονάδες, οι 2 χρησιμοποιούνται για HARQ retransmission.

| Flow | Δεδομένα προς εξυπηρέτηση | Χρονική απαίτηση |
|---|---|---|
| A) control | 4 units | Πρέπει να φύγουν τώρα |
| B) video | 6 units | Deadline σε 5 ευκαιρίες |
| C) elastic | Συνεχές backlog | Χωρίς fixed deadline εδώ |

# Άσκηση: Μία εφικτή κατανομή και το κόστος της

Πρώτα το άμεσο deadline, μετά απόφαση για το υπόλοιπο budget

- Στην πρώτη επιλογή, 2 units δίνονται στο HARQ, 4 στο A, 4 στο B και 2 στο C. Το A ολοκληρώνεται πριν από το deadline του.
  - Το B διατηρεί backlog 2 units, άρα θα χρειαστεί εξυπηρέτηση σε επόμενη ευκαιρία πριν λήξει το deadline του.
- Η δεύτερη επιλογή ολοκληρώνει και το B, αλλά αφήνει το C χωρίς πόρους.
  - Αν αυτό επαναλαμβάνεται, το C μπορεί να υποστεί starvation.



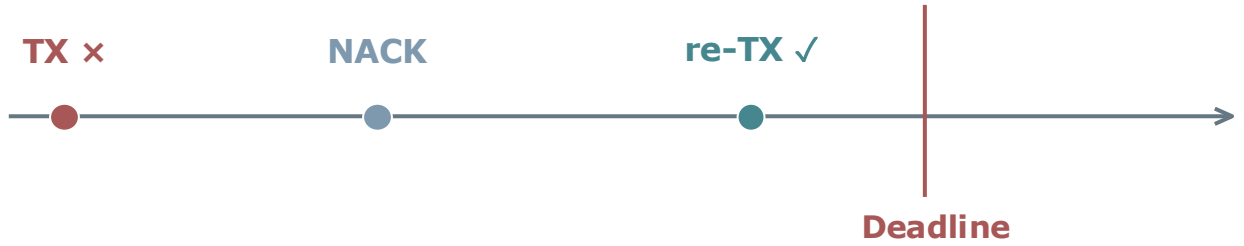

Η επιλογή εξαρτάται από το deadline του B, τη μελλοντική διαθεσιμότητα πόρων και την ανάγκη εξυπηρέτησης του C.

# Έγκαιρη και αξιόπιστη παράδοση: αποφάσεις του RAN

Οι επαναμεταδόσεις βοηθούν την παράδοση, αλλά χρειάζονται χρόνο. Latency και reliability μοιράζονται πόρους.

- Ένα πιο robust MCS μπορεί να μειώσει την πιθανότητα αποτυχίας, αλλά συχνά απαιτεί περισσότερους πόρους για το ίδιο payload
- Το HARQ επιτρέπει επαναμετάδοση μετά από αποτυχία, εφόσον υπάρχει διαθέσιμος χρόνος και δυνατότητα εκχώρησης πόρων.
- Για μια υπηρεσία με αυστηρό deadline έχει σημασία η επιτυχής παράδοση πριν από το όριο, όχι μόνο η τελική άφιξη του packet.

**Έγκαιρη επιτυχής παράδοση**




**Επιτυχής αλλά εκπρόθεσμη παράδοση**




Το δίκτυο μπορεί να ακολουθήσει διαφορετικές προσεγγίσεις:

- Να επιλέξει πιο robust πρώτη μετάδοση, ώστε να μειώσει την πιθανότητα να χρειαστεί HARQ. ➡ Περισσότεροι πόροι
- Να αφήσει αρκετό χρονικό περιθώριο για επαναμετάδοση, αν αποτύχει η πρώτη προσπάθεια. ➡ Προτεραιοποίηση scheduler
- Αν το deadline είναι τόσο κοντά ώστε η επαναμετάδοση δεν προλαβαίνει, η επιτυχία της πρώτης προσπάθειας γίνεται ακόμη πιο σημαντική.



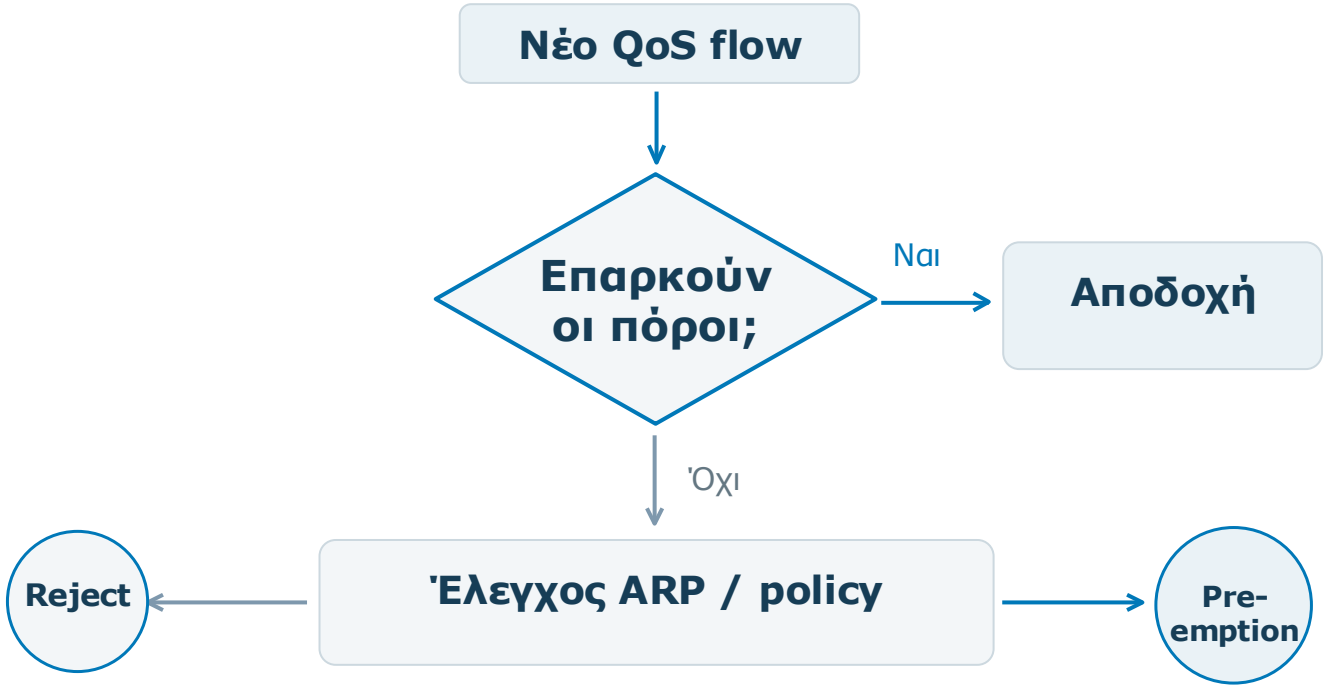
TS 38.300 · TS 38.321 · 3GPP TS 23.501

# Admission control και ARP

- Το admission control εξετάζει αν το δίκτυο μπορεί να αποδεχθεί ένα νέο QoS flow, λαμβάνοντας υπόψη και τις υπάρχουσες απαιτήσεις.
- Το ARP (Allocation and Retention Priority) είναι μια παράμετρος QoS που βοηθά το δίκτυο να αποφασίσει αν θα αποδεχθεί ή θα διατηρήσει ένα QoS flow όταν οι πόροι είναι περιορισμένοι.
- Το ARP ορίζει προτεραιότητα για allocation (1 – 15, με το 1 υψηλότερο) και retention, καθώς και αν ένα flow μπορεί να εκτοπίσει ή να εκτοπιστεί από άλλο.
- Το ARP χρησιμοποιείται σε αποφάσεις αποδοχής και διατήρησης. Δεν καθορίζει από μόνο του ποιο packet θα μεταδοθεί στο επόμενο slot.

**Allocation:** μπορεί το δίκτυο να δεχτεί ένα νέο flow ή τις νέες απαιτήσεις ενός flow;

**Retention:** αν δεν επαρκούν οι πόροι, ποια ήδη αποδεκτά flows μπορούν να διατηρηθούν;




**Η αποδοχή νέου flow είναι διαφορετική από την επιλογή packet στο επόμενο slot (βλ. επόμενο slide)**

# ARP και 5QI: δύο διαφορετικές προτεραιότητες



- Το ARP χρησιμοποιείται όταν εξετάζεται αν ένα flow μπορεί να γίνει δεκτό ή να διατηρηθεί με περιορισμένους πόρους.
- Το 5QI συνδέει το flow με χαρακτηριστικά QoS, όπως priority level, Packet Delay Budget και resource type.
- Το priority level του 5QI σχετίζεται με τη μεταχείριση της κίνησης κατά την εξυπηρέτηση. Το ARP δεν λέει από μόνο του στον MAC scheduler ποιο packet θα μεταδώσει στο επόμενο slot.
- Το ARP priority και το 5QI priority είναι διαφορετικές παράμετροι (δεν συγκρίνουμε τις αριθμητικές τους τιμές μεταξύ τους)

## ARP

**Γίνεται δεκτό ή διατηρείται το flow;**

1-15, με 1 = υψηλότερη
+ κανόνες pre-emption (εκτοπισμού άλλου flow)

## 5QI priority

**Ποια σχετική προτεραιότητα έχει στην εξυπηρέτηση;**

1-127, με 1 = υψηλότερη
+ άλλα QoS χαρακτηριστικά

**ARP: αποδοχή / διατήρηση | 5QI priority: σχετική προτεραιότητα εξυπηρέτησης**

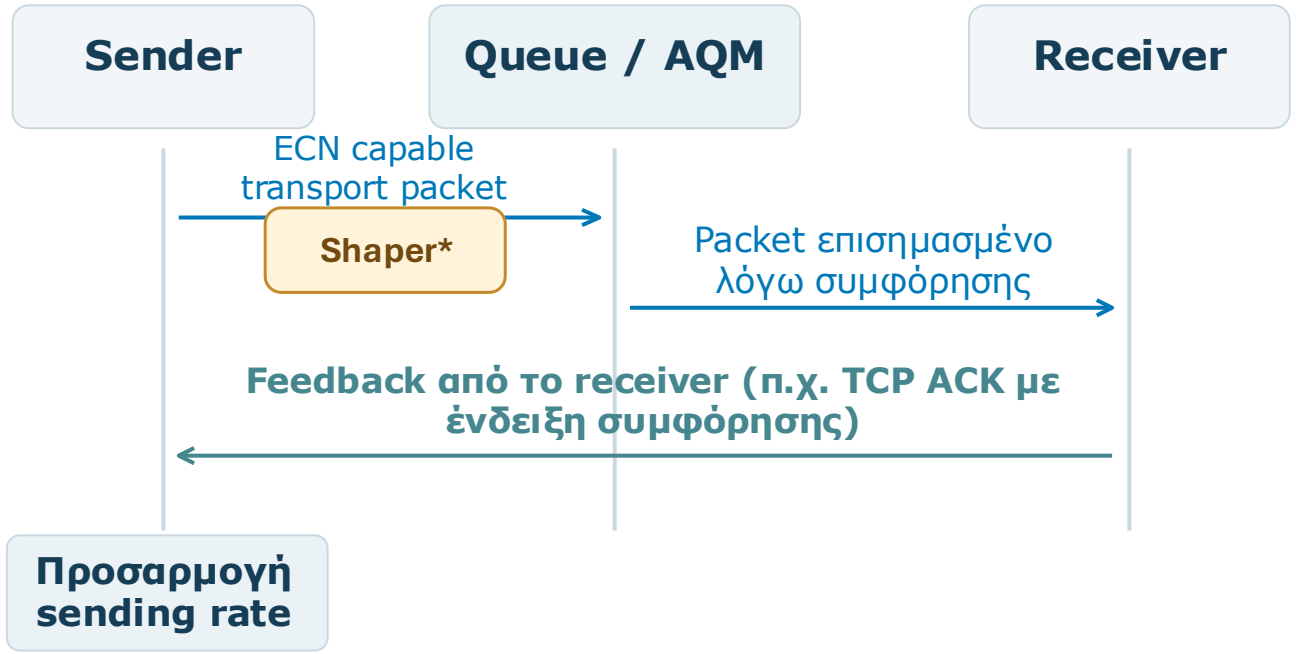


3GPP TS 23.501 §§5.7.2.2, 5.7.3.3

# Congestion control και διαχείριση ουρών: Scheduling, shaping και AQM

- Το **scheduling** επιλέγει ποια packets εξυπηρετούνται.
- Το **traffic shaping** είναι ένας μηχανισμός που ρυθμίζει τον ρυθμό με τον οποίο προωθούνται τα δεδομένα προς το επόμενο τμήμα του δικτύου.
- Το **Active Queue Management (AQM)** μπορεί να απορρίπτει ή να επισημαίνει packets πριν η ουρά γεμίσει, ώστε να περιορίσει την επίμονη συμφόρηση.
- Το **ECN (Explicit Congestion Notification)** είναι μηχανισμός που επιτρέπει στο δίκτυο να σηματοδοτήσει **συμφόρηση με επισήμανση σε packet**, αντί να βασίζεται μόνο σε απόρριψη.
- Ο sender μπορεί τότε να προσαρμόσει τον ρυθμό αποστολής.

**Ενδεικτική λογική ECN-capable transport**



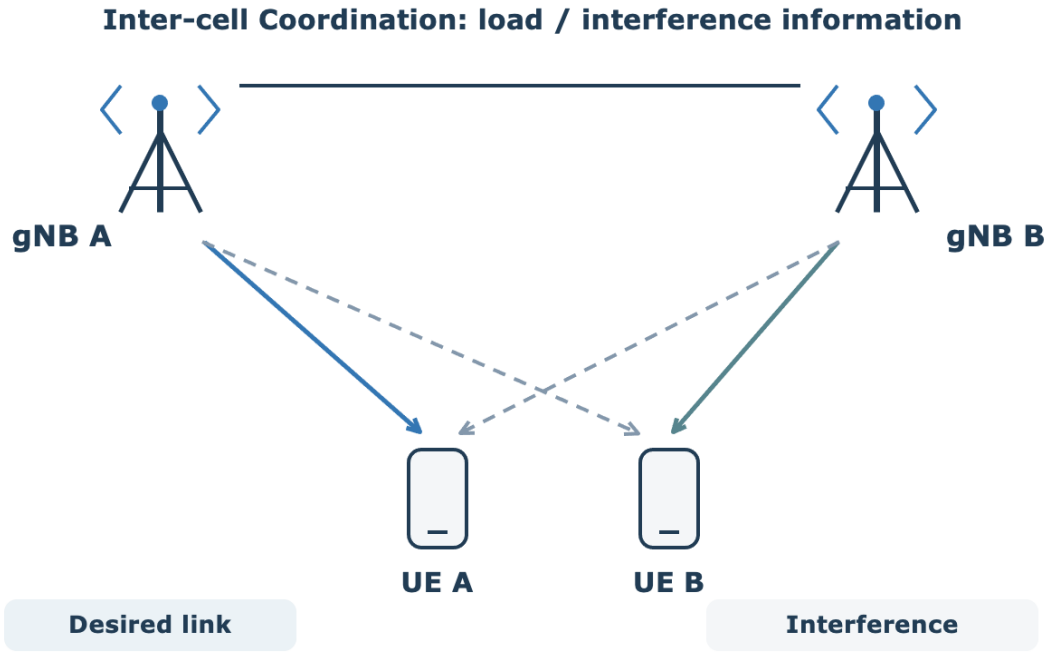

* Ένας shaper θα μπορούσε, ανάλογα με τον σχεδιασμό και την υλοποίηση, να βρίσκεται:
- Στον server ή στην έξοδό του, ώστε να ρυθμίζει την κίνηση πριν μπει στο δίκτυο.
- Σε έναν router της διαδρομής, ώστε να ελέγχει τον ρυθμό προς μια σύνδεση.
- Στο UPF, εφόσον η υλοποίηση εφαρμόζει εκεί σχετικούς μηχανισμούς ελέγχου ρυθμού.
- Σε λειτουργία της πλευράς του RAN

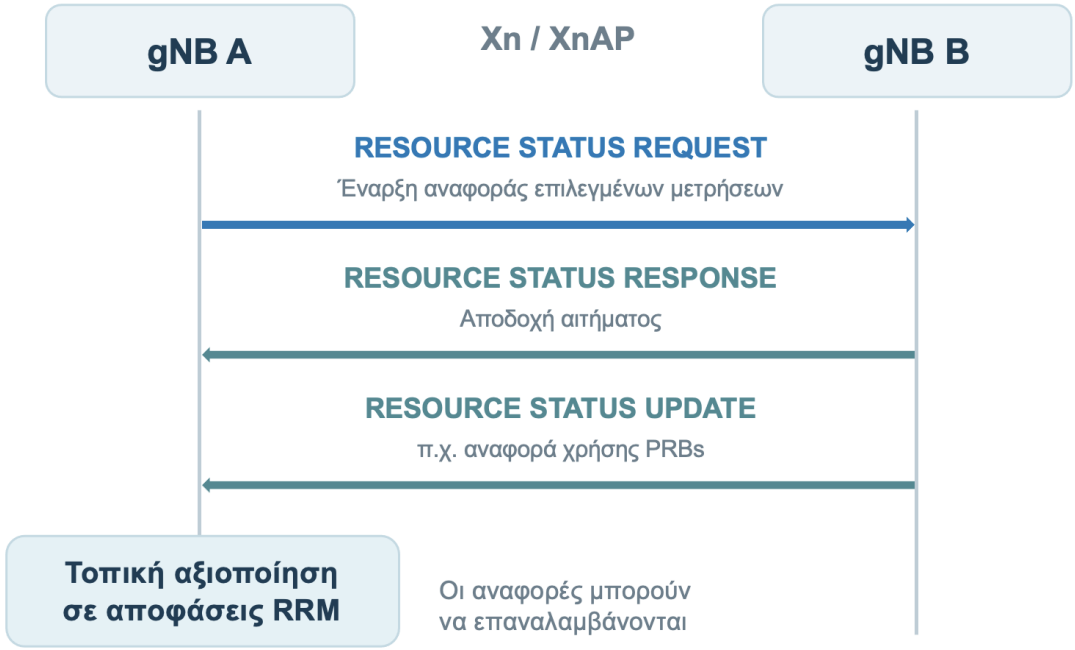


RFC 7567 / 3168 · TS 38.300

# Το scheduling ενός cell επηρεάζει τα γειτονικά

Interference coordination και ευρύτερο RRM – Resource Status Reporting μέσω Xn




- Οι ταυτόχρονες μεταδόσεις γειτονικών cells δημιουργούν παρεμβολή, επηρεάζοντας το SINR και τον ρυθμό που μπορούν να επιτύχουν τα UEs.
- Μηχανισμοί RRM μπορούν να συντονίζουν τη χρήση πόρων, την ισχύ ή τις παραμέτρους εξυπηρέτησης μεταξύ cells.
- Η ανταλλαγή πληροφοριών για load και interference βοηθά τον συντονισμό, χωρίς να απαιτείται ένας κοινός scheduler για όλα τα cells.

> **! Η βελτιστοποίηση ενός cell δεν εγγυάται το καλύτερο αποτέλεσμα για το δίκτυο.**

# Το resource allocation ως optimization

Στόχος και περιορισμοί πριν από την επιλογή αλγορίθμου

- Η βελτιστοποίηση ενός cell δεν εγγυάται το καλύτερο αποτέλεσμα για το δίκτυο.
- Ορίζουμε συνάρτηση -> objective function η οποία εκφράζει το στόχο (π.χ., συνολικό throughput, utility με διαφορετικά βάρη για UEs/traffic types, κτλ.)
- Ορίζουμε περιορισμούς (constraints) για την εύρεση πιθανών λύσεων: οι περιορισμοί μπορεί να εκφράζουν διαθέσιμους πόρους, την ισχύ και τις απαιτήσεις που το δίκτυο επιδιώκει να τηρήσει.
- Αν οι απαιτήσεις δεν μπορούν να ικανοποιηθούν ταυτόχρονα, χρειάζεται επανεξέταση του admission ή των στόχων.

## Κοινό πρόβλημα κατανομής πόρων

$$\begin{aligned}
&\underset{\{x_{ik}\}}{\text{maximize}} && \sum_{i\in\mathcal{U}} w_i R_i \\
&\text{subject to} && \sum_{i\in\mathcal{U}_c} x_{ik} \le 1, && \forall\, c\in\mathcal{C},\ k\in\mathcal{K},\\
& && x_{ik}+x_{jk}\le 1, && \forall\,(i,j)\in\mathcal{E},\ k\in\mathcal{K},\\
& && R_i \ge g_i, && \forall\, i\in\mathcal{G},\\
& && x_{ik}\in\{0,1\}, && \forall\, i\in\mathcal{U},\ k\in\mathcal{K},\\
& && R_i=\sum_{k\in\mathcal{K}} r_{ik}x_{ik}, && \forall\, i\in\mathcal{U}.
\end{aligned}$$

Η μεταβλητή απόφασης είναι $x_{ij}$ και ορίζεται $x_{ik}=1$ εάν ο πόρος $k$ εκχωρηθεί στον χρήστη $i$, ή 0 διαφορετικά.

$\mathcal{C}$: cells $\mathcal{U}$: χρήστες $\mathcal{U}c$: χρήστες του cell c $\mathcal{K}$: πόροι
$\mathcal{E}$: ζεύγη χρηστών που δεν μοιράζονται τον ίδιο πόρο
$\mathcal{G}$: χρήστες με ελάχιστο στόχο ρυθμού

# Rules, optimization και learning-based RRM

Διαφορετικά εργαλεία για το ίδιο πρόβλημα αποφάσεων

- Ορίσαμε το πρόβλημα – **πώς θα το λύσουμε;**
- Οι αποφάσεις RRM μπορούν να βασίζονται σε κανόνες, μαθηματική βελτιστοποίηση ή μάθηση από δεδομένα.
- Η επιλογή μεθόδου εξαρτάται από
  - το διαθέσιμο μοντέλο
  - τον χρόνο απόφασης
  - την ικανότητα αντιμετώπισης μεταβαλλόμενων συνθηκών
  - ή την ικανότητα μεταφοράς σε διαφορετικά περιβάλλοντα.

| Προσέγγιση | Τι χρησιμοποιεί | Πλεονέκτημα | Κρίσιμη δυσκολία |
|---|---|---|---|
| Rules / heuristics | Μετρήσεις και προκαθορισμένους κανόνες | Συνήθως γρήγορη και εύκολα ερμηνεύσιμη απόφαση | Οι κανόνες και οι παράμετροι χρειάζονται κατάλληλη ρύθμιση για διαφορετικές συνθήκες. |
| Optimization | Μαθηματικό μοντέλο, στόχο και περιορισμούς | Ρητή διατύπωση, το μοντέλο μπορεί να επιλυθεί με νέες παραμέτρους και να γενικευθεί | Ακρίβεια μοντέλου και χρόνος επίλυσης. |
| Learning-based | Δεδομένα και εκπαιδευμένο μοντέλο ή policy | Αξιοποιεί σύνθετα μοτίβα· μπορεί να δώσει γρήγορη απόφαση μετά την εκπαίδευση. | Απόδοση σε νέες συνθήκες και τήρηση περιορισμών. |

**Οι προσεγγίσεις μπορούν να συνδυαστούν:** ένα μοντέλο μάθησης μπορεί να προβλέπει την κίνηση, ενώ ένας optimizer να αποφασίζει την κατανομή.

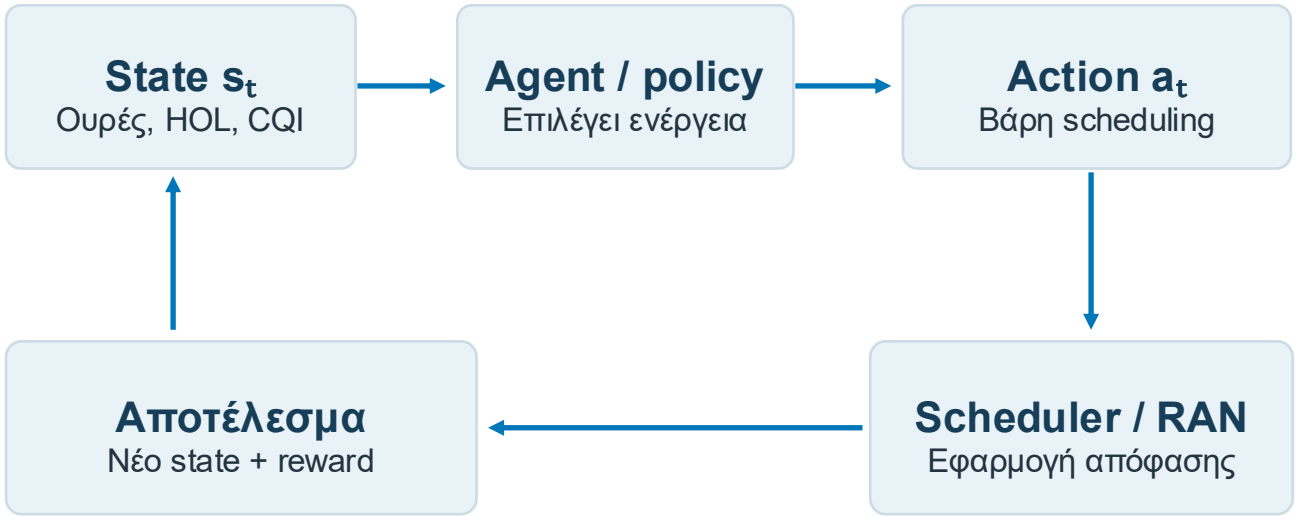


TS 38.300 · Kushner & Whiting · O-RAN / NTT DOCOMO

# Reinforcement learning για RRM

## Ο agent μαθαίνει μια policy από την αλληλεπίδραση με το δίκτυο

- **State (κατάσταση):** Οι είσοδοι της απόφασης, π.χ. μέγεθος ουρών, HOL delay και αναφορές CQI.
- **Action (ενέργεια):** Η policy επιλέγει βάρη προτεραιότητας που χρησιμοποιεί ο scheduler για την κατανομή πόρων.
- **Reward (ανταμοιβή):** Αξιολογεί το αποτέλεσμα, π.χ. throughput και παραβιάσεις deadline.
  - Στην εκπαίδευση καθοδηγεί τη βελτίωση του policy (π.χ., Σε μια πιθανοτική policy, μπορεί να αυξηθεί η πιθανότητα επιλογής μιας ενέργειας σε μια συγκεκριμένη κατάσταση)




Εκπαίδευση: το reward καθοδηγεί τη βελτίωση της policy.

Λειτουργία: η εκπαιδευμένη policy επιλέγει ενέργειες με βάση το τρέχον state.

**Αξιολογούμε KPIs (π.χ., throughput), παραβιάσεις deadline και δικαιοσύνη μεταξύ χρηστών.**

# Οι προκλήσεις που κρατάμε

QoS απαιτήσεις, πεπερασμένοι πόροι και αποφάσεις υπό μεταβλητότητα

- Το QoS περιγράφει τις απαιτήσεις εξυπηρέτησης. Το RAN τις λαμβάνει υπόψη όταν κατανέμει τους περιορισμένους radio πόρους.
- Το admission αποφασίζει ποιες απαιτήσεις γίνονται δεκτές, το scheduling κατανέμει πόρους και το congestion control αντιμετωπίζει την υπερβολική ζήτηση.
- Rules, optimization και learning συγκρίνονται με κοινούς περιορισμούς και μετρικές που δείχνουν την επίδραση στις υπηρεσίες.

1

**Απαιτήσεις υπηρεσίας**

QoS profile / 5QI

2

**Ανάθεση πόρων - Εφικτή εξυπηρέτηση**

RRM, scheduling, admission
(rule-based, heuristic, optimization, learning-driven)

3

**Μετρήσιμο αποτέλεσμα**

Delay, throughput, reliability



3GPP TS 23.501 · TS 38.300 · ETSI TS 103 982 · RFC 7567 / 3168 · Kushner & Whiting

# Εξαμηνιαία εργασία

# Ενδεικτικά σενάρια: ποιος παίρνει τους πόρους;

Private 5G SA, στοιχεία εμπνευσμένα από εγκαταστάσεις, αριθμητικές τιμές: παραδοχές



## Παράδειγμα 1. Αποθήκη με 12 AMRs

- **Δίκτυο:** 4.000 m², 2 cells, n78, TDD, 3,75 GHz / 20 MHz, local UPF και edge.
- **Control DL:** 200 bytes / 20 ms ανά AMR = 80 kbps payload· server → AMR deadline 30 ms.
- **Κίνηση UL:** telemetry 300 bytes / 100 ms ανά AMR, 4 ενεργές κάμερες × 4 Mbps.
- **Απόφαση:** προστασία DL control με μεταβαλλόμενο radio κανάλι και αυξανόμενο φορτίο.

## Παράδειγμα 2. Ορυχείο με 6 οχήματα

- **Δίκτυο:** υπόγειες στοές, 2 cells, n78, TDD, 3,5 GHz / 40 MHz, 100 κινητά UEs.
- **Video UL:** 4 οχήματα × 3 Mbps payload προς σταθμό ελέγχου.
- **Control DL:** 200 bytes / 50 ms ανά όχημα, one-way στόχος έως την εφαρμογή 50 ms.
- **Απόφαση:** πόροι UL / DL, video έναντι control, επιδείνωση καναλιού σε στοές.

# Τεχνική ανάλυση και επιλογή μηχανισμού

Ορίστε συγκεκριμένη απόφαση RRM και τεκμηριώστε τη λειτουργία της

## Χαρακτηρίστε το σύστημα

- **Περιβάλλον:** τοπολογία, κύτταρα, bandwidth, UL / DL, αριθμός UEs και κινητικότητα.
- **Traffic / QoS:** μέγεθος packet, inter-arrival, ρυθμός UL / DL, burst, endpoints και deadline.
- **Απόφαση:** PRBs, βάρη scheduler, ισχύς ή admission· επίπεδο UE, flow ή slice.
- **Χρονισμός:** πού και κάθε πότε λαμβάνεται η απόφαση; Ποιες μετρήσεις είναι διαθέσιμες;

## Περιγράψτε τον μηχανισμό

- **Είσοδοι:** ουρές / HOL, CQI, QoS και ιστορικό· ποια είναι η ηλικία των μετρήσεων;
- **Λειτουργία:** δείξτε πώς παράγεται η απόφαση με κανόνα, εξίσωση ή ψευδοκώδικα.
- **Στόχος / reward:** τι μεγιστοποιείται και ποιο trade-off κωδικοποιείται;
- **Περιορισμοί:** πόροι και QoS· ξεχωρίστε hard constraints από στόχους ή ποινές.

**Αιτιολογήστε την επιλογή heuristic, optimization, learning ή συνδυασμού τους.**

# Βιβλιογραφία, baselines και συγκρισιμότητα

Τα τρία papers στηρίζουν την απόφαση, δεν παρουσιάζονται ως ανεξάρτητες περιλήψεις



## Επιλογή και κριτική σύνθεση

- **3 papers:** ένα κεντρικό, μία συγκρίσιμη εναλλακτική και μία συμπληρωματική διάσταση. Τουλάχιστον 2 του 2025-2026
- **Σύγκριση:** μοντέλο traffic / καναλιού, είσοδοι, αποφάσεις, QoS constraints και κόστος.
- **Τεκμήρια:** ποιο σχήμα ή πείραμα στηρίζει κάθε ισχυρισμό και τι δεν αποδεικνύει;

## Σύγκριση χωρίς υλοποίηση

- **Στα papers:** αναλύστε τις baselines, τα KPIs και τις πειραματικές συνθήκες τους.
- **Δική σας πρόταση:** επιλέξτε συγκρίσιμες baselines και KPIs, εξηγήστε πώς θα ελέγχατε την εισήγηση.
- **Δίκαιη σύγκριση:** ίδιο σενάριο, traffic, radio συνθήκες, χρονικός ορίζοντας και επίπεδο απόφασης.
- **Όρια:** μην κατατάσσετε μεθόδους από ποσοστά διαφορετικών papers και διαφορετικών πειραμάτων.

**Το σχέδιο αξιολόγησης δεν απαιτεί εκτέλεση πειραμάτων ή παραγωγή αποτελεσμάτων.**

# Παραδοτέα, παράδειγμα και πηγές

Ατομική εργασία ή ομάδα δύο ατόμων

## Τι παραδίδετε

- **Πρόταση:** σενάριο, τεχνικό ερώτημα, 3 papers και η μεταξύ τους συμπληρωματικότητα.
- **Τελική εργασία:** τεχνική ανάλυση, συγκριτική σύνθεση, εισήγηση και σχέδιο αξιολόγησης.
- **Παράδειγμα:** 2 flows ή λίγοι πόροι, δείξτε υπολογιστικά πώς προκύπτει μία απόφαση.
- **Μεταβολή:** αλλάξτε μία παραδοχή, π.χ. load ή ηλικία CQI, και αναλύστε την επίδραση.
- **Παρουσίαση:** υποστηρίξτε την επιλογή και να είστε σε θέση να απαντήσετε σε τεχνικές ερωτήσεις.

## Πηγές έμπνευσης για τα σενάρια

**Siemens: private 5G και AMRs στο Karlsruhe**

references.siemens.com/en/reference/manufacturing-karlsruhe?id=35686

**Nokia: ARENA2036 και φάσμα 3,7–3,8 GHz**

nokia.com/bell-labs/collaboration-opportunities/arena2036/

**Nokia / Boldyn: Callio FutureMINE, 2025**

nokia.com/newsroom/nokia-and-boldyn-networks-power-the-future-of-mining-at-callio-pyhaejaervi-with-private-5g-network/

# Βαθμολόγηση και χρονοδιάγραμμα

Πρόταση, τελική εργασία και παρουσίαση με τεχνική συζήτηση

## Συνιστώσες βαθμού (40% στον τελικό βαθμό του μαθήματος)

### 10% Πρόταση θέματος

Σενάριο, ερώτημα και επιλογή papers.

### 40% Τελική εργασία

Τεχνική ανάλυση, κριτική σύνθεση και τεκμηρίωση της επιλογής.

### 50% Παρουσίαση και συζήτηση

Σαφήνεια, κατανόηση και απαντήσεις σε τεχνικές ερωτήσεις.

## Προθεσμία και παρουσιάσεις

**22/01/2026**

Η ολοκλήρωση της εργασίας θα γίνει πριν από την ημερομηνία εξέτασης του μαθήματος.

Οι παρουσιάσεις θα προγραμματιστούν ad hoc, κατόπιν συνεννόησης.