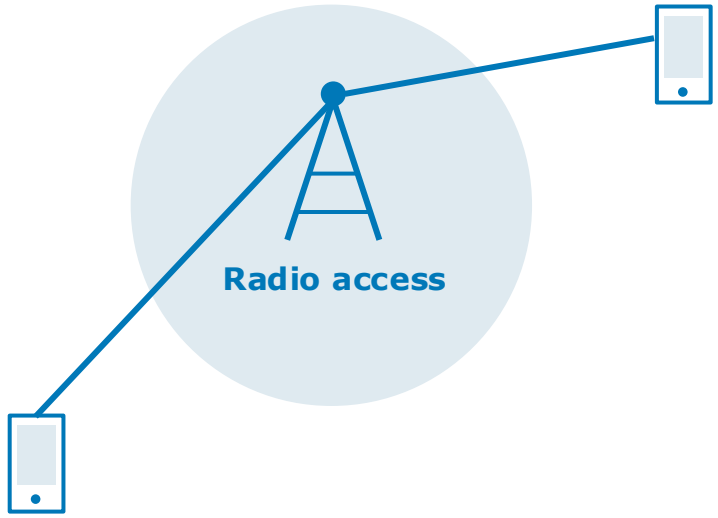
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Μ301: Παλιές και νέες προκλήσεις στα δίκτυα κινητών επικοινωνιών

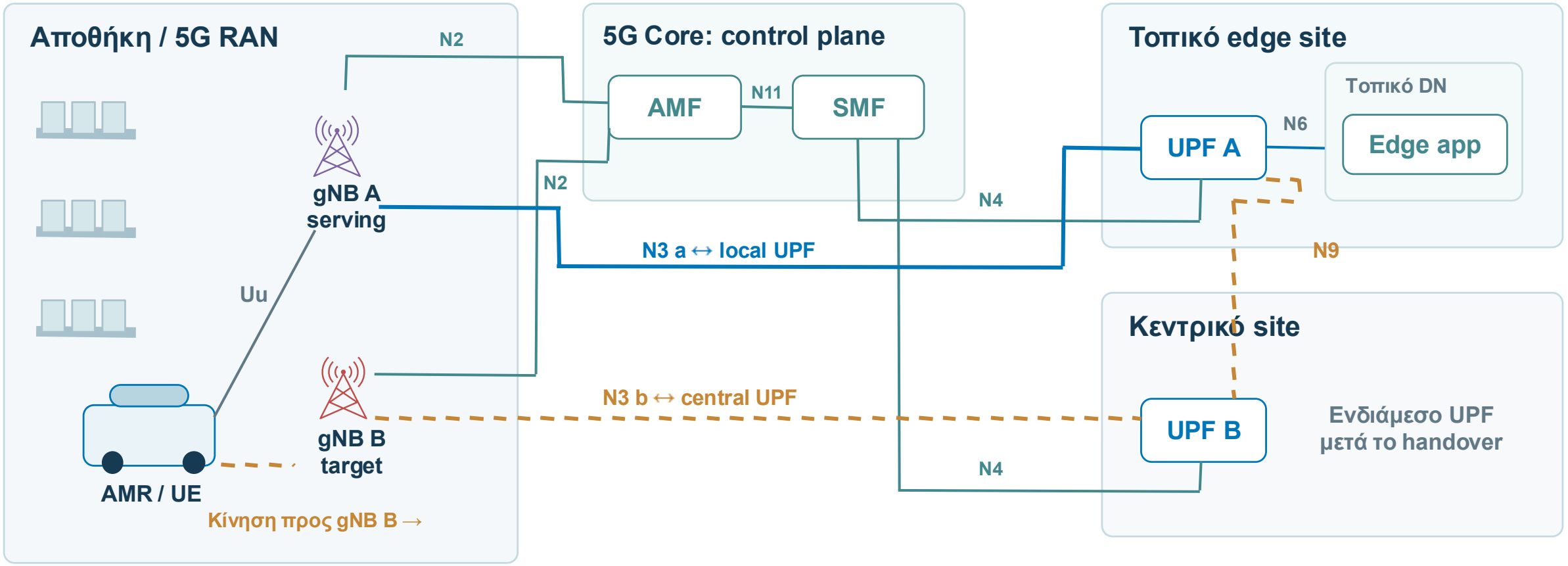
Διάλεξη #5: 5G Core - Αρχιτεκτονική και Προκλήσεις στα PDU Sessions, το UPF και τη Συνδεσιμότητα



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

Private 5G: συνδεσιμότητα και κινητικότητα

Δύο gNBs, επιλογή UPF και ενδεικτικό handover



Μετά το handover: gNB B → central UPF → local UPF → ίδια edge εφαρμογή.

Πριν: διαδρομή user plane — Μετά: N3 + N9 — Control plane (N2 / N4 / N11)

5G Core: αρχιτεκτονική και προκλήσεις

Μία ιδιωτική αποθήκη ως κοινό σενάριο για τη διάλεξη

Τέσσερις αποφάσεις σχεδιασμού

1. Τοποθέτηση και φορτίο: Πού θα τοποθετήσουμε το UPF και την εφαρμογή που επικοινωνεί με τα ρομπότ; Μπορούν να εξυπηρετήσουν την κίνηση όταν αυξάνονται τα ρομπότ ή ο όγκος των δεδομένων τους;
2. Mobility και continuity: Τι πρέπει να συμβεί όταν το ρομπότ μετακινηθεί από την κάλυψη του gNB A στην κάλυψη του gNB B, ώστε να συνεχιστεί η υπηρεσία του;
3. Αστοχία και αποκατάσταση: Τι γίνεται αν το τοπικό UPF σταματήσει να λειτουργεί ή αν χαθεί η σύνδεση προς αυτό;
4. Κοινό bottleneck και isolation: Τι συμβαίνει όταν διαφορετικές υπηρεσίες μοιράζονται έναν πόρο και η μία επηρεάζει την άλλη;

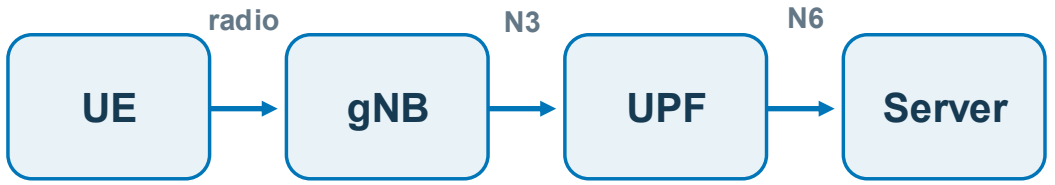
Στόχος: τεκμηριωμένες αποφάσεις για latency, κινητικότητα και συνέχεια υπηρεσίας.

Από τη διαχείριση radio πόρων στο core

To end-to-end αποτέλεσμα εξαρτάται από όλη τη διαδρομή

- Στην προηγούμενη διάλεξη εξετάσαμε πώς το gNB μοιράζει radio πόρους μεταξύ χρηστών και υπηρεσιών.
- Το core αναλαμβάνει πρόσβαση, διαχείριση σύνδεσης, πολιτικές και προώθηση της κίνησης προς το δίκτυο δεδομένων.
- Ακόμη και με καλό radio, μια μακρινή διαδρομή ή καθυστέρηση σε UPF και server μπορεί να επηρεάσει την υπηρεσία.

Από το radio στην εφαρμογή



Η κατεύθυνση και οι ετικέτες δείχνουν τη σχέση μεταξύ των λειτουργιών.

Το QoS στο radio είναι μέρος της συνολικής διαδρομής.

3GPP TS 23.501 · TS 23.502

Κοινό σενάριο: ιδιωτική αποθήκη 5G SA

Παράμετρος	Διδακτική υπόθεση
Πρόσβαση	5G SA · δύο γειτονικά cells
Συσκευές	12 κινητά ρομπότ (AMRs)
Υπηρεσίες	Control, telemetry, video
Control target	RTT ≤ 30 ms, ενδεικτικά
Συνδεσιμότητα	Ένα PDU session ανά AMR

Πριν: σύνδεση στο gNB A



Μετά: σύνδεση στο gNB B και εισαγωγή central UPF

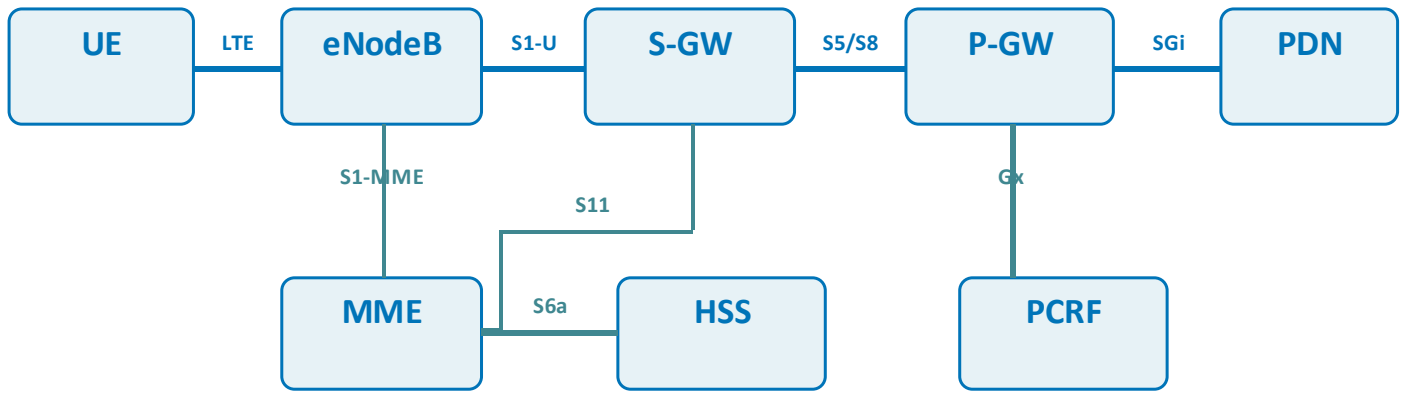


Διδακτικό παράδειγμα · οι αριθμοί αποτελούν υποθέσεις

4G: αρχιτεκτονική και βασικές λειτουργίες του EPC

Control plane και user plane σε ένα τυπικό LTE δίκτυο, χωρίς CUPS

ΑΡΧΙΤΕΚΤΟΝΙΚΗ



Τι κάνει κάθε λειτουργία

- MME: NAS σηματοδοσία, πρόσβαση και κινητικότητα. Συνεργάζεται με το HSS για στοιχεία συνδρομής.
- S-GW: προωθεί την κίνηση και λειτουργεί ως σημείο αγκύρωσης σε κινητικότητα μεταξύ eNodeBs.
- P-GW: σύνδεση με το PDN, IP allocation και επιβολή πολιτικών. Το PCRF καθορίζει πολιτικές QoS.

Η διαδρομή χρήστη χρησιμοποιεί GTP-U στα S1-U και, στο παραπάνω παράδειγμα, S5/S8. Η σηματοδοσία ακολουθεί διαφορετικές διεπαφές.

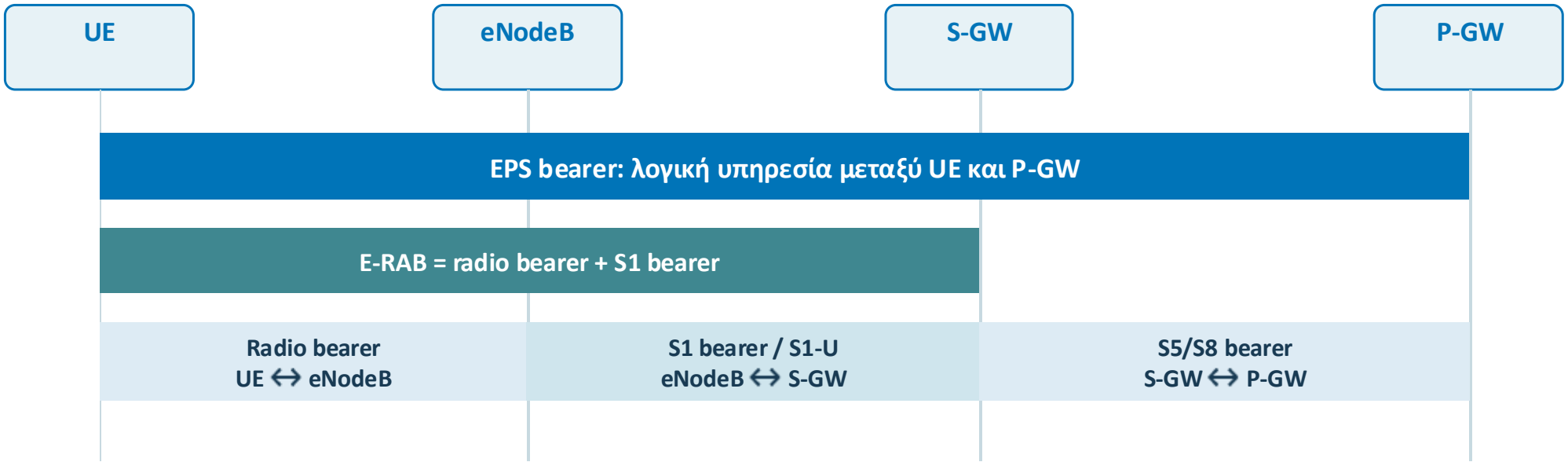
Μπλε: user plane Πράσινο: control plane Τα σχήματα είναι συνοπτικά

Το EPC παρέχει πρόσβαση, κινητικότητα και συνδεσιμότητα προς εξωτερικά δίκτυα.

3GPP TS 23.401

4G: EPS bearer και αντιστοίχιση ανά τμήμα

Μία λογική υπηρεσία QoS από το UE έως το P-GW



- Η PDN connection δημιουργεί default bearer για βασική συνδεσιμότητα. Μπορεί να προστεθεί dedicated bearer για άλλη μεταχείριση QoS.
- Τα TFTs (traffic flow templates) ταξινομούν πακέτα σε bearers. Κάθε bearer διαθέτει προφίλ QoS, π.χ. QCI και ARP.
- Στο παράδειγμα GTP-based S5/S8, οι κόμβοι αντιστοιχίζουν τις τοπικές πληροφορίες tunnel και τα TEIDs.

Η αντιστοίχιση radio bearer και tunnels συγκροτεί μια ενιαία λογική υπηρεσία QoS.

3GPP TS 23.401 §4.7

Από το EPC στο 5G Core: τι αλλάζει;

Σύγκριση λειτουργιών, συνδεσιμότητας και διαχείρισης ποιότητας υπηρεσίας

Πλευρά	4G EPC	5G Core / πρακτική συνέπεια
Control plane	MME και gateways με καθορισμένες διεπαφές.	AMF: πρόσβαση / κινητικότητα. SMF: PDU sessions. SBA υπηρεσίες μεταξύ control-plane NFs.
Συνδεσιμότητα	PDN connection με APN και EPS bearers.	PDU session με DNN και ένα ή περισσότερα QoS flows.
QoS	EPS bearer: μονάδα QoS. QCI, ARP και, όπου ισχύει, GBR.	QoS flow: μικρότερη μονάδα διαφοροποίησης QoS στο session. QFI, 5QI και ρυθμίσεις QoS.
User plane	S-GW / P-GW. **Το CUPS (Control & User Plane separation) εισήχθη στο Release 14.**	SMF ελέγχει UPF μέσω N4 / PFCP. Δυνατές διαφορετικές θέσεις και πολλαπλά UPFs.
Ευελιξία / slices	Το EPC υποστήριζε ήδη κατανεμημένο user plane και local breakout (SIPA – Selected IP Traffic Offload, LIPTO – Local IP Access). Έπειτα CUPS	Slice-aware επιλογές και πολλαπλές διαδρομές. Η απομόνωση και η απόδοση εξαρτώνται από πόρους και πολιτικές.

Παράδειγμα: ένα PDU session μπορεί να μεταφέρει control και video ως διαφορετικά QoS flows.

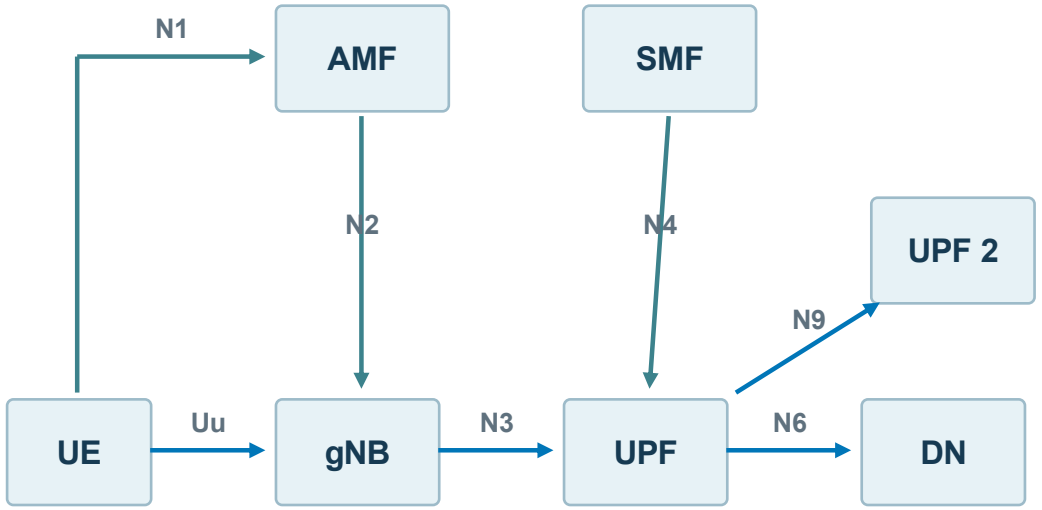
Τα 5G QoS flows δεν αντιστοιχούν υποχρεωτικά ένα προς ένα σε radio bearers.

3GPP TS 23.401 / 23.501 / 23.214

Οι διεπαφές που θα χρησιμοποιήσουμε

- N1: σηματοδοσία μεταξύ UE και AMF
- N2: σηματοδοσία μεταξύ gNB και AMF
- N3: μεταφορά user-plane packets μεταξύ gNB και UPF
- N4: έλεγχος UPF από το SMF
- N6: σύνδεση UPF με data network
- N9: user-plane διασύνδεση μεταξύ UPFs, όταν υπάρχουν περισσότερα του ενός.

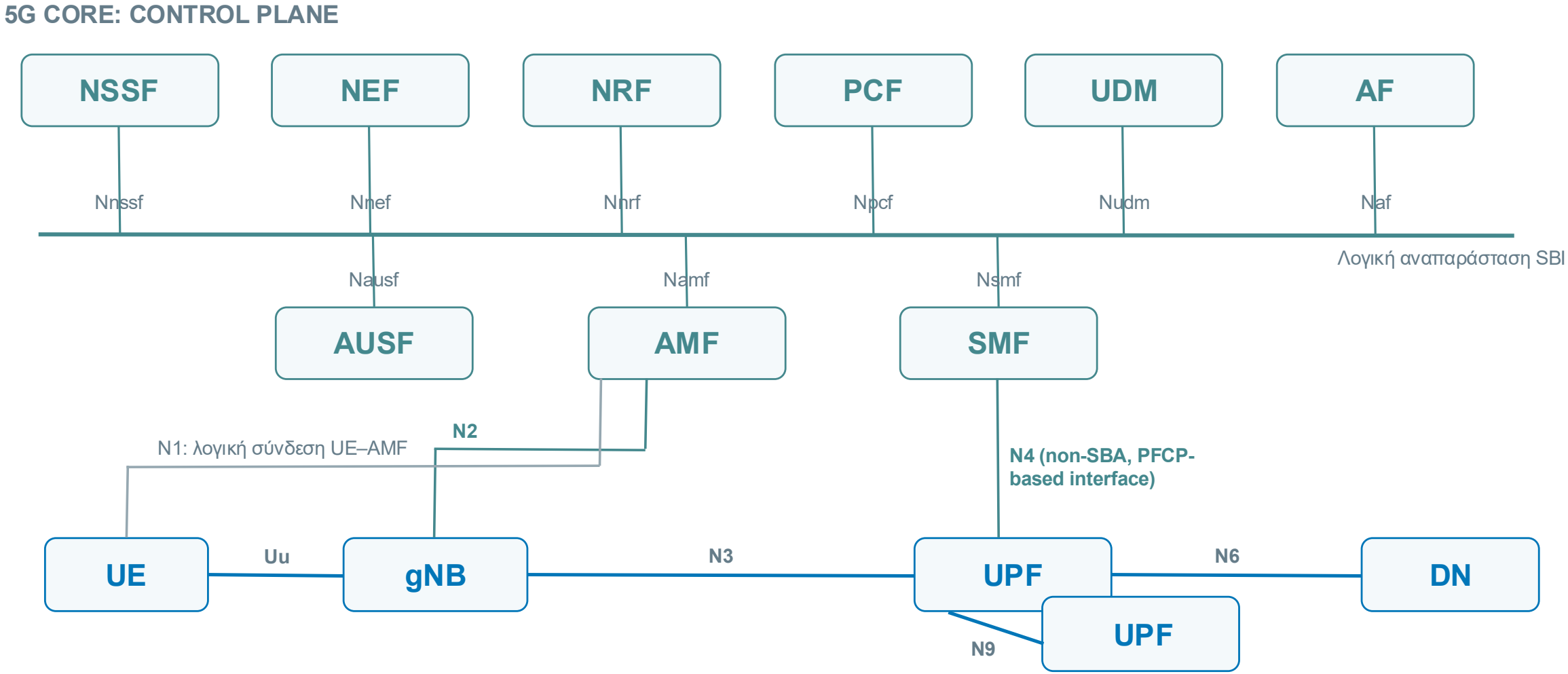
Πού τερματίζει κάθε διεπαφή;



3GPP TS 23.501 · TS 23.502

Η μεγάλη εικόνα του 5G System

Service-based control plane και διαδρομή δεδομένων



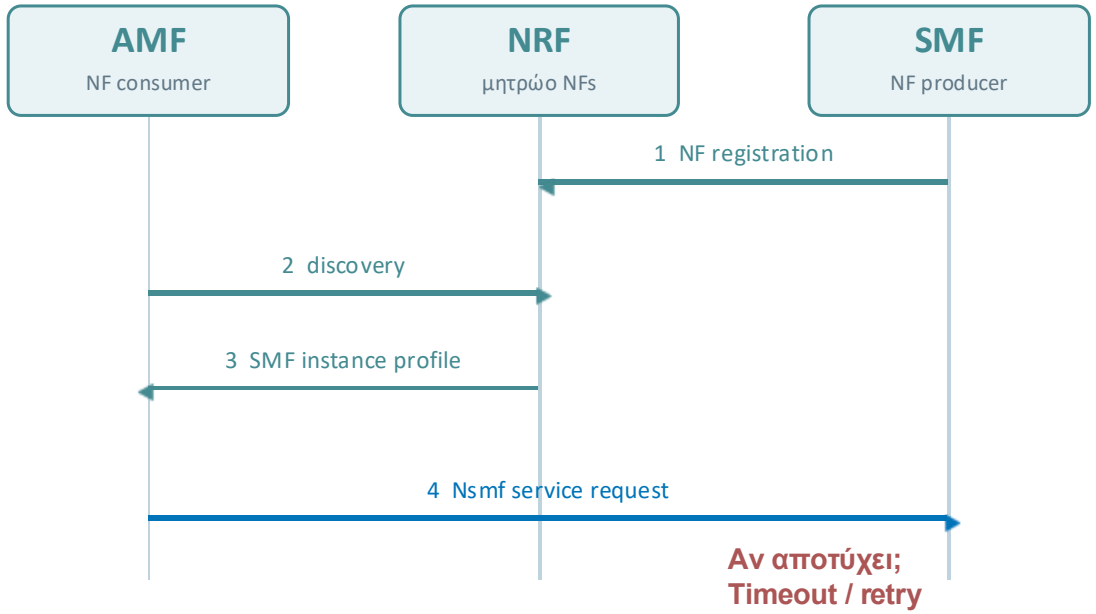
Οι διεπαφές SBI παριστάνονται λογικά. Η N4 συνδέει SMF και UPF.

3GPP TS 23.501

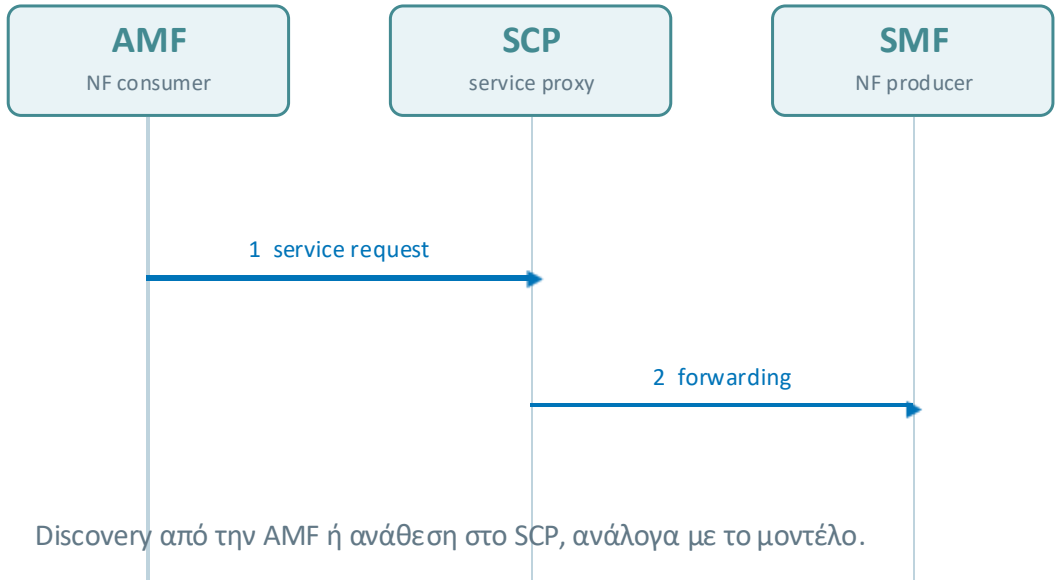
SBI: πώς επικοινωνούν οι network functions;

Η γραμμή SBA είναι λογική απεικόνιση υπηρεσιών, όχι φυσικό bus

A. Άμεση επικοινωνία με discovery



B. Έμμεση επικοινωνία μέσω SCP (Service Communication Proxy)

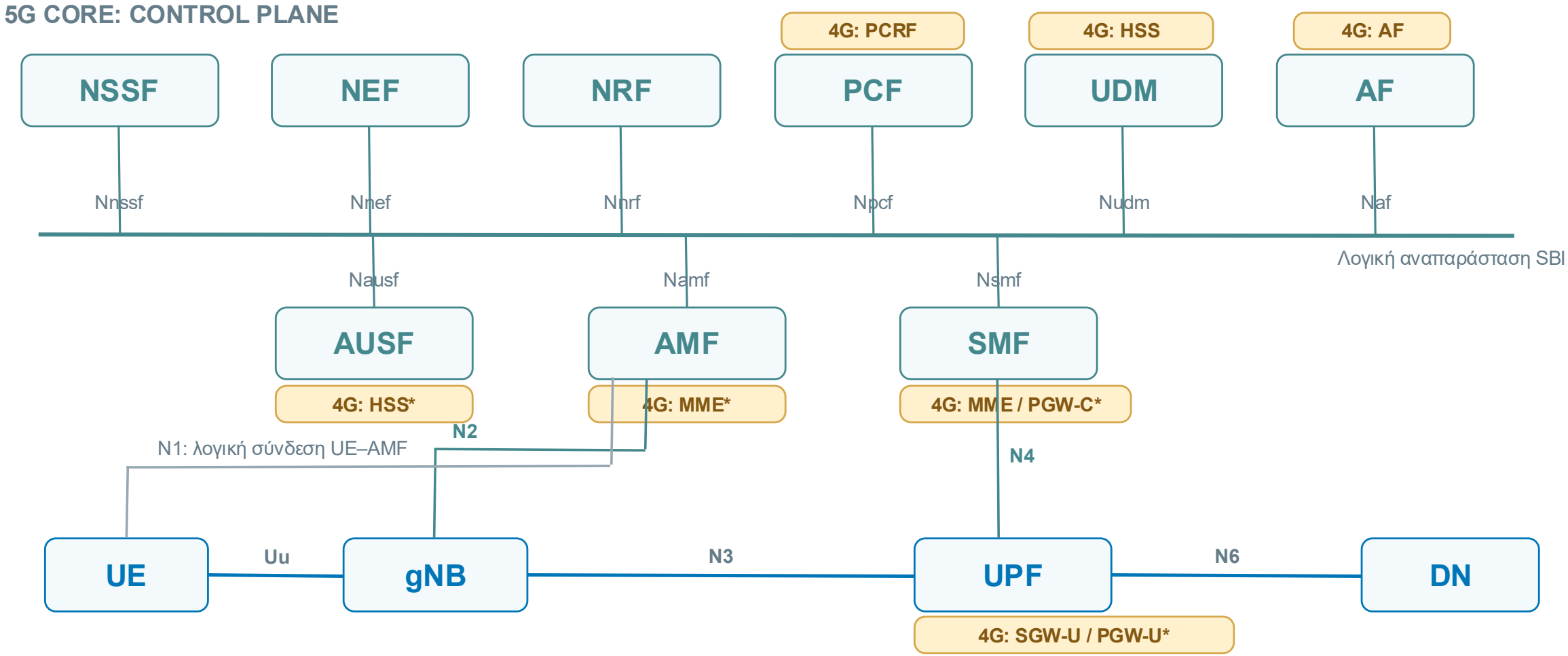


- Το NRF καταχωρίζει NF profiles και επιστρέφει διαθέσιμες υπηρεσίες / endpoints.
- Το SCP υποστηρίζει δρομολόγηση και, ανάλογα με το μοντέλο και την υλοποίηση, discovery και επιλογή producer (ποιο συγκεκριμένο instance της network function που προσφέρει την υπηρεσία θα εξυπηρετήσει ένα αίτημα).
- SBI: HTTP/2 APIs και κοινό πλαίσιο ασφάλειας (π.χ. TLS).
- Το N4 / PFCP είναι non-SBA.

3GPP TS 23.501 Annex E / TS 29.500

5G Core και 4G EPC: λειτουργική αντιστοίχιση

Ενδεικτική σύγκριση λειτουργιών, όχι ταύτιση κόμβων



* Επιμέρους συναφείς λειτουργίες. SGW-C/U και PGW-C/U προϋποθέτουν EPC με CUPS.

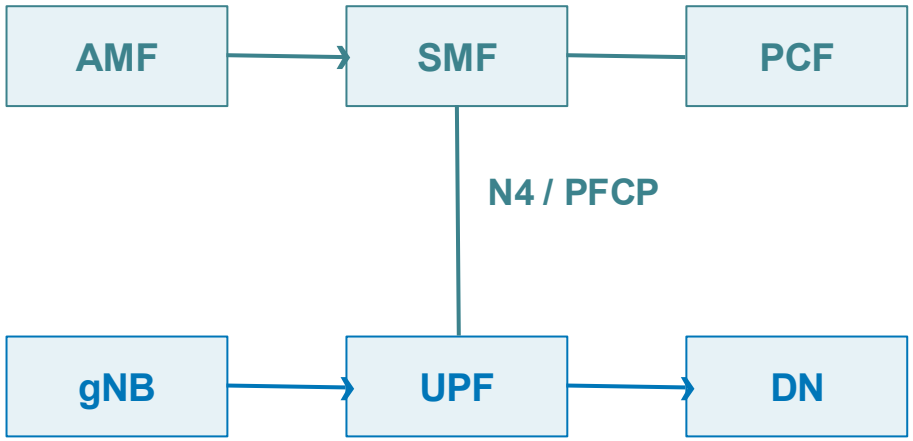
3GPP TS 23.501

Ποιος αποφασίζει και ποιος προωθεί;

Οι λειτουργίες συνδέονται με ένα συγκεκριμένο PDU session

Λειτουργία	Ρόλος στο σενάριο
AMF	Πρόσβαση και κινητικότητα
SMF	Session, επιλογή UPF, N4 rules
UPF	Ταξινόμηση και προώθηση
PCF	Πολιτικές session / QoS
UDM / AUSF	Συνδρομή / authentication

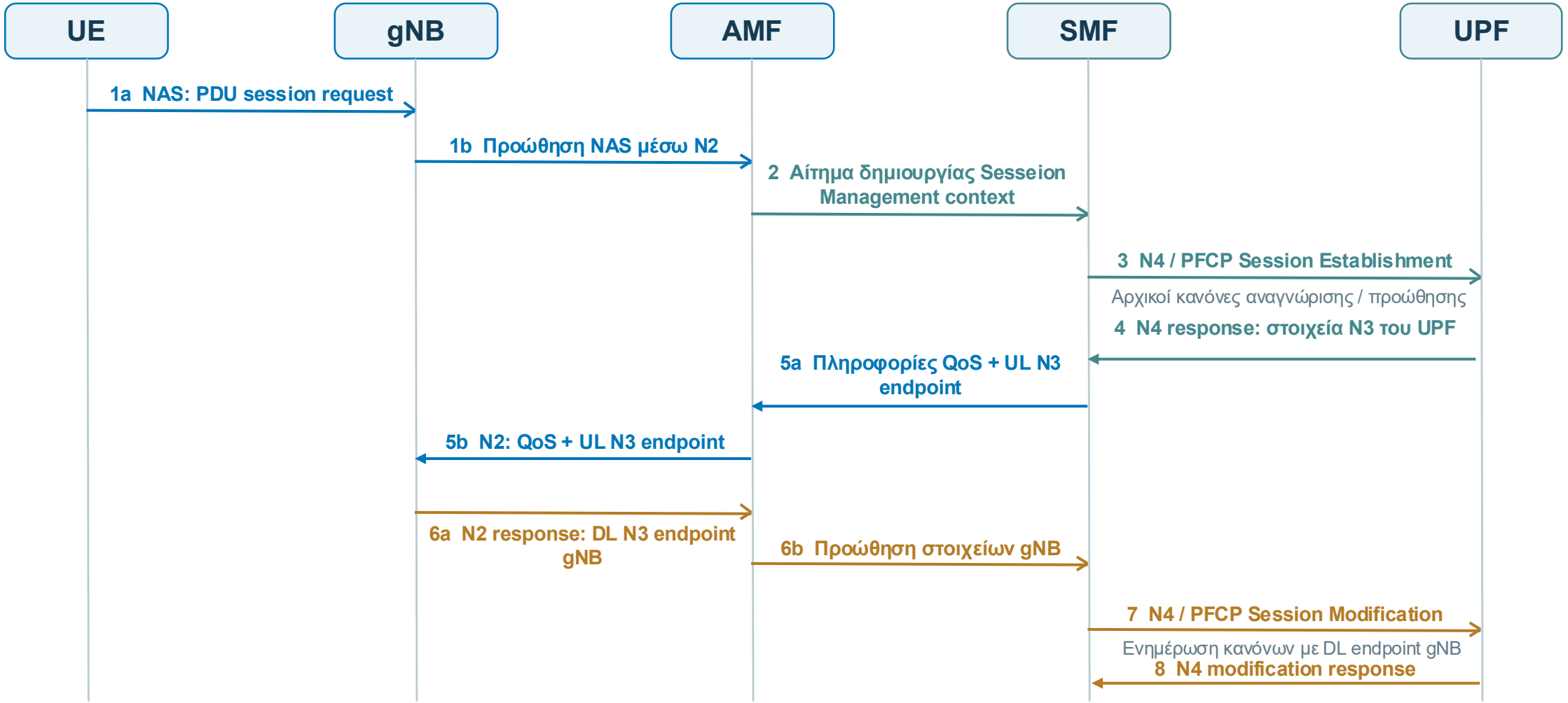
Σηματοδοσία και δεδομένα



Πράσινο: control plane · μπλε: user plane

3GPP TS 23.501 · TS 23.502

Από το αίτημα στη διαδρομή δεδομένων



Πρώτα εγκαθίστανται οι αρχικοί κανόνες. Μετά την απάντηση του gNB, συμπληρώνεται το downlink N3.

PFCP: πώς ελέγχει το SMF το UPF;

- Το SMF στέλνει στο UPF κανόνες μέσω PFCP (Packet Forwarding Control Protocol στην N4 διεπαφή)
- Το UPF επιστρέφει απόκριση για την εγκατάσταση ή τροποποίησή τους.
- Η συσχέτιση αφορά συγκεκριμένο session. Οι κανόνες περιγράφουν πώς αναγνωρίζονται, προωθούνται, περιορίζονται ή μετρώνται τα packets.



Τέσσερις οικογένειες κανόνων στο UPF

Η εγκατάσταση εξειδικεύεται για την κίνηση κάθε session

Κανόνας	Ερώτημα που απαντά
PDR - Packet Detection Rule	Ποιο packet και ποια κίνηση αντιστοιχούν;
FAR - Forwarding Action Rule	Προωθείται, απορρίπτεται ή γίνεται buffering;
QER - QoS Enforcement Rule	Επιτρέπεται η διέλευση και με ποιο όριο ρυθμού;
URR - Usage Reporting Rule	Ποια χρήση μετράται και πότε αναφέρεται;
Σχέση	Το PDR μπορεί να συνδέεται με FAR, QER και URR

Οι κανόνες συνδέουν την αναγνώριση ενός packet με ενέργειες στο user plane.

Παράδειγμα: ένα downlink packet στο UPF

Συνθετικοί κανόνες για την εντολή ελέγχου του ρομπότ

Κανόνας	Ενδεικτική ρύθμιση
PDR	Προς UE 10.1.0.8 και UDP θύρα προορισμού 5000
FAR	Προώθηση προς gNB μέσω N3 tunnel
QER	DL gate: open, QFI 7, ενδεικτικό ανώτατο όριο 1 Mbps για τη συγκεκριμένη ροή ελέγχου
URR	Μέτρηση των downlink bytes του control flow και αναφορά στο SMF με τη λήξη του session.

Το gNB / UPF tunnel endpoint και το TEID είναι στοιχεία δρομολόγησης μέσα στο δίκτυο.

Η ταυτότητα του QoS flow και η διαδρομή του packet προκύπτουν από διαμορφωμένους κανόνες.

Πορεία downlink packet: ποιο στοιχείο αλλάζει;

Από την εφαρμογή έως το gNB μέσα από τους κανόνες του UPF

Σημείο	Τι συμβαίνει στο packet
N6 → UPF	Το UPF λαμβάνει IP packet με προορισμό το UE
PDR στο UPF	Αναγνώριση session / υπηρεσίας και QoS flow
FAR στο UPF	Δημιουργία GTP-U header για το N3
N3 → gNB	Outer IP, TEID και QFI συνοδεύουν το user packet
gNB → UE	Αποενθυλάκωση και μεταφορά μέσω RAN

Η IP της εφαρμογής και το outer tunnel header εξυπηρετούν διαφορετικούς σκοπούς.

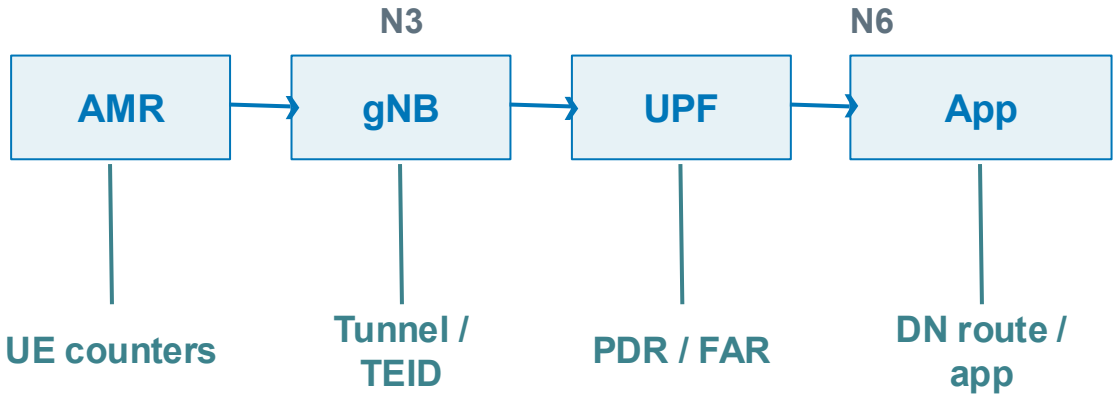
Ανάλυση προβλήματος: το session υπάρχει αλλά δεν περνούν packets

Ενδεικτική περίπτωση στο private 5G δίκτυο της αποθήκης

- Το AMR ολοκληρώνει registration και PDU session establishment, αλλά η εφαρμογή στην αποθήκη δεν λαμβάνει την τηλεμετρία του.

- Πιθανές αιτίες:
  - λανθασμένος κανόνας προώθησης,
  - μη προσβάσιμο endpoint N3,
  - πρόβλημα δρομολόγησης στο N6 ή
  - μη διαθέσιμη εφαρμογή

- Πού θα κάναμε μετρήσεις ώστε να εντοπίσουμε το πρώτο σημείο στο οποίο χάνεται το packet;
  - ➢ Επόμενο slide



Ξεχωρίζουμε την επιτυχή σηματοδοσία από τη σωστή λειτουργία της διαδρομής δεδομένων.

Λύση: εντοπισμός απώλειας στη διαδρομή uplink

Ακολουθούμε την τηλεμετρία του AMR από την πηγή έως την εφαρμογή

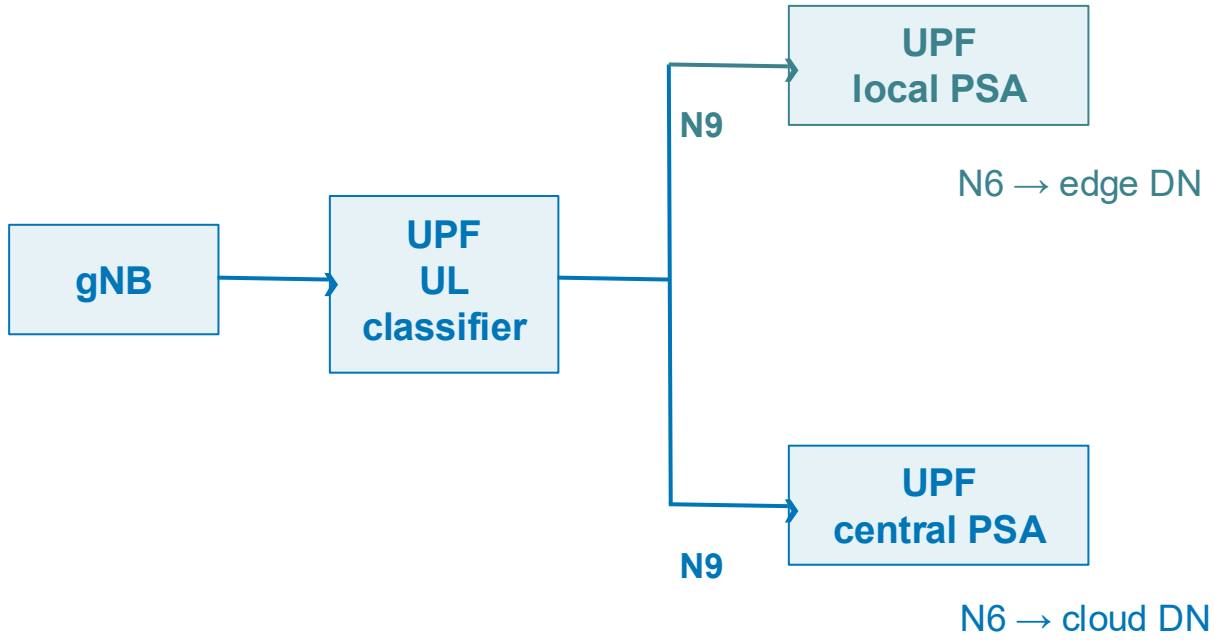
Σημείο ελέγχου	Εύρημα / τι συγκρίνουμε	Τι ελέγχουμε στη συνέχεια
1 AMR / gNB	Το AMR παράγει και στέλνει UDP πακέτα; Τα λαμβάνει το gNB;	Ελέγχουμε αν η εφαρμογή στέλνει στη σωστή IP / port και αν το radio bearer και οι uplink counters δείχνουν μετάδοση προς το gNB.
2 N3: gNB → UPF	Μετράμε GTP-U στην έξοδο gNB και στην είσοδο UPF για το ίδιο flow.	Αν φεύγει από το gNB αλλά δεν φτάνει στο UPF, ελέγχουμε δρομολόγηση και firewall για GTP-U / UDP 2152. Στο UPF ελέγχουμε αν το UL TEID αντιστοιχεί σε PDR.
3 UPF: N3 → N6	Έφτασε στο UPF; Βγαίνει το inner IP packet στο N6;	Αν το UPF λαμβάνει αλλά δεν προωθεί, ελέγχουμε αν το PDR αναγνωρίζει το πακέτο, αν το FAR ορίζει προώθηση και αν το QER gate είναι ανοικτό. Εξετάζουμε και τα drops.
4 N6 / εφαρμογή	Βγαίνει από το UPF; Φτάνει στο host και στο process της εφαρμογής;	Αν το πακέτο εξέρχεται από το UPF, ελέγχουμε τη διαδρομή και το firewall προς τον server και αν η εφαρμογή ακούει στη σωστή IP / port. Συμβουλευόμαστε και τα logs.

Παράδειγμα: N3 είσοδος στο UPF ✓, N6 έξοδος ✗ → ελέγχουμε πρώτα PDR / FAR και UPF drops.

Ίδιο flow και χρονικό παράθυρο: εντοπίζουμε το πρώτο τμήμα χωρίς επιβεβαιωμένη λήψη.

Πότε χρειάζονται περισσότερα UPFs;

- Διαφορετικές λειτουργίες μπορούν να ανατεθούν σε διαφορετικά σημεία
- Ένα UPF μπορεί να λειτουργεί ως uplink **classifier**, διαχωρίζοντας κίνηση του ίδιου PDU session **προς διαφορετικές διαδρομές.**
- Στο παράδειγμα, ένα local PDU Session Anchor εξυπηρετεί την εφαρμογή της αποθήκης, ενώ ένα άλλο anchor συνδέει την κίνηση με το κεντρικό data network (cloud DN).
- Το SMF διαχειρίζεται τους κανόνες στα εμπλεκόμενα UPFs.



Η προσθήκη UPF αλλάζει τη διαδρομή και εισάγει επιπλέον κατάσταση προς διαχείριση.

3GPP TS 23.501 §5.6.4

Local breakout χωρίς δεύτερο PDU session

Εφαρμογή στο παράδειγμα του private 5G της αποθήκης

Ζητούμενο	Τεχνική επιλογή προς συζήτηση
Κοινό session	UL classifier επιλέγει διαδρομή ανά flow
Local τηλεμετρία	UPF local PDU Session Anchor (PSA) → εφαρμογή αποθήκης
Internet traffic	Κεντρικό PSA → εξωτερικό δίκτυο
Τι πρέπει να ρυθμιστεί;	Φίλτρα, κανόνες forwarding, N9 / N6 και routing
Εναλλακτική	Δεύτερο PDU session, αν χρειάζεται άλλο πλαίσιο συνδεσιμότητας

Uplink packet από το ρομπότ	Φίλτρο στον UL classifier	Διαδρομή
Τηλεμετρία προς 10.20.0.10*	Προορισμός: εφαρμογή αποθήκης	Local PSA → edge εφαρμογή
Αίτημα προς 203.0.113.20*	Δεν ταιριάζει στο τοπικό φίλτρο	Central PSA → εξωτερικό δίκτυο

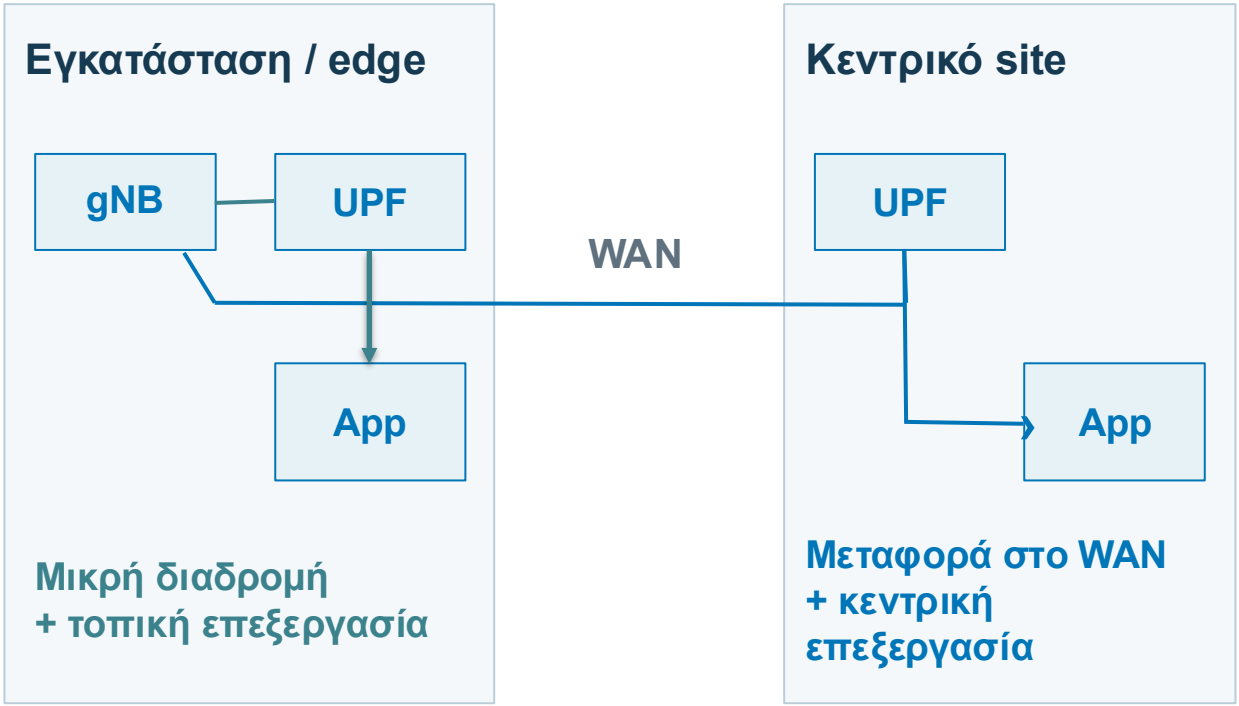
**Οι διευθύνσεις είναι ενδεικτικές· η πρόσβαση και η επιστροφή της κίνησης προϋποθέτουν κατάλληλη διαμόρφωση του data network και του routing.*

UPF και εφαρμογή: ποια διαδρομή ακολουθεί το packet;

To local breakout ωφελεί όταν το UPF και η εφαρμογή είναι κατάλληλα τοποθετημένα

Η τοποθέτηση του UPF καθορίζει πού η κίνηση εξέρχεται από το 5G Core. Η τοποθέτηση της εφαρμογής καθορίζει πόση διαδρομή απομένει μέχρι τον προορισμό.

UPF	Εφαρμογή	Τι σημαίνει για τη διαδρομή
Τοπικό	Τοπική	Τα δεδομένα μπορούν να παραμείνουν κοντά στην αποθήκη.
Τοπικό	Απομακρυσμένη	Η κίνηση βγαίνει τοπικά από το 5G Core, αλλά στη συνέχεια διασχίζει το WAN προς την εφαρμογή.
Κεντρικό	Κεντρική	Η κίνηση διασχίζει το δίκτυο μεταφοράς πριν φτάσει στο UPF και στην εφαρμογή.



Τοπικό UPF με απομακρυσμένο server δεν εξαλείφει το WAN latency.

UPF placement: τι συγκρίνουμε;

Ίδια εφαρμογή, διαφορετικό transport και διαφορετική υπολογιστική ισχύς

- Επιλογή Α: local UPF και edge application server. Μικρότερη απόσταση, **αλλά περιορισμένη διαθέσιμη υπολογιστική ισχύς.**
- Επιλογή Β: central UPF και cloud server. **Μεγαλύτερο σταθερό RTT, αλλά server με υψηλότερο μέσο ρυθμό εξυπηρέτησης.**

A Τοπικό UPF + edge εφαρμογή



RTT εκτός server: 8 ms

Ρυθμός εξυπηρέτησης **μ = 100 req/s**

Μέσος χρόνος εξυπηρέτησης: **10 ms**

B Κεντρικό UPF + cloud εφαρμογή



RTT εκτός server: 20 ms

Ρυθμός εξυπηρέτησης **μ = 1000 req/s**

Μέσος χρόνος εξυπηρέτησης: **1 ms**

Application RTT = RTT εκτός server + αναμονή στην ουρά + χρόνος εξυπηρέτησης

Άσκηση: πότε το edge παύει να "κερδίζει";

Μοντέλο M/M/1 για τον χρόνο στο application server

Application RTT = RTT εκτός server + αναμονή στην ουρά + χρόνος εξυπηρέτησης

Μέσο application RTT (υποθέτοντας μηδενική αναμονή στις ουρές

$$D_{edge} = 8 + 1000 / (100 - \lambda)$$
$$D_{central} = 20 + 1000 / (1000 - \lambda)$$

λ: requests/s | Χρόνοι: ms
Σύγκριση για λ < 100 req/s και τις υποθέσεις M/M/1

Ρυθμός άφιξης λ	Edge	Central	Μικρότερο μέσο RTT
10 requests/s	19,1 ms	21,0 ms	Edge
20 requests/s	20,5 ms	21,0 ms	Edge
23,2 requests/s	≈ 21,0 ms	≈ 21,0 ms	Όμοιο
30 requests/s	22,3 ms	21,0 ms	Cloud
60 requests/s	33,0 ms	21,1 ms	Cloud

- Πότε (για τι φόρτο) θα επιλέγαμε το edge-based deployment;
- Αρκεί ο μέσος όρος για να κρίνουμε αν τηρείται το όριο των 30 ms;

Η μικρότερη δικτυακή καθυστέρηση αντισταθμίζεται από την επιβάρυνση του server.

Τι μετράμε για να εντοπίσουμε το bottleneck;

Η ίδια τιμή RTT μπορεί να προκύψει από διαφορετικές αιτίες

Σημείο	Μέτρηση / τι αναζητούμε
UE ↔ app	RTT, p95/p99, deadline-miss ratio
RAN	Radio retransmissions, queue / HOL delay
UPF: N3 → N6	Forwarding delay, packet rate, drops
Transport	RTT ανά τμήμα, loss, link utilization
Server	Queue time, processing time, CPU / load

Προσοχή: για one-way delay μεταξύ συσκευών απαιτείται συγχρονισμός ρολογιών. Τα p95 των τμημάτων δεν αθροίζονται γενικά (βλ. επόμενο slide).

Τι μετράμε για να εντοπίσουμε το bottleneck;

Αιτήματα	Χρόνος στο Α	Χρόνος στο Β	Συνολικός χρόνος
1-90	1 ms	1 ms	2 ms
91-95	10 ms	1 ms	11 ms
96-100	1 ms	10 ms	11 ms

- Με τον τρόπο υπολογισμού που χρησιμοποιήσαμε, το **p95 είναι 1 ms για καθένα από τα δύο τμήματα**.
- Όταν όμως κοιτάξουμε τον *συνολικό χρόνο*, μόνο τα αιτήματα 1–90 ολοκληρώθηκαν σε 2 ms.
- Τα υπόλοιπα 10 χρειάστηκαν 11 ms, οπότε το **συνολικό p95 είναι 11 ms**, όχι 1+1=2 ms.

Ας πούμε ότι για το ίδιο αίτημα μετράμε δύο διαδοχικά τμήματα:

$$D_{\text{συνολικό}} = D_A + D_B$$

Η σχέση αυτή ισχύει **για κάθε μεμονωμένο αίτημα**. Δεν ισχύει όμως γενικά ότι:

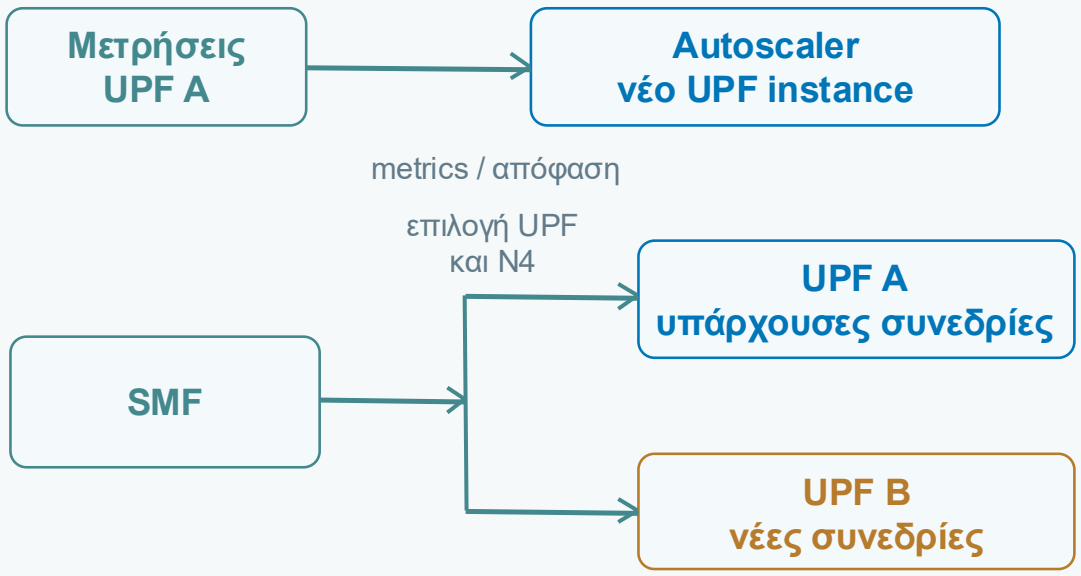
$$\underbrace{p95(D_A + D_B)}_{\text{p95 συνολικής καθυστέρησης}} = \underbrace{p95(D_A) + p95(D_B)}_{\text{άθροισμα επιμέρους p95}}$$

Κλιμάκωση UPF: scale out

Από την ανίχνευση συμφόρησης στη διάθεση επιπλέον χωρητικότητας

- Η αύξηση των packets/s μπορεί να δημιουργήσει ουρές και **να αυξήσει το forwarding delay στο UPF.**
- Μετράμε packet rate, p95/p99 delay, drops και χρήση πόρων, ώστε να εντοπίσουμε το bottleneck.
- Μια πλατφόρμα ενορχήστρωσης (π.χ., Kubernetes) μπορεί να παρακολουθεί μετρήσεις φορτίου και να δημιουργεί επιπλέον UPF instances.
- Η ανάθεση συνεδριών στα νέα instances απαιτεί συντονισμό με το 5G Core (SMF). Το SMF πρέπει να αναγνωρίσει το νέο UPF ως διαθέσιμο, να το επιλέξει για κατάλληλα PDU sessions και να εγκαταστήσει τους κανόνες τους μέσω N4/PFCP.
  - Τα υπάρχοντα sessions παραμένουν στο UPF A

Ενδεικτικό scale-out

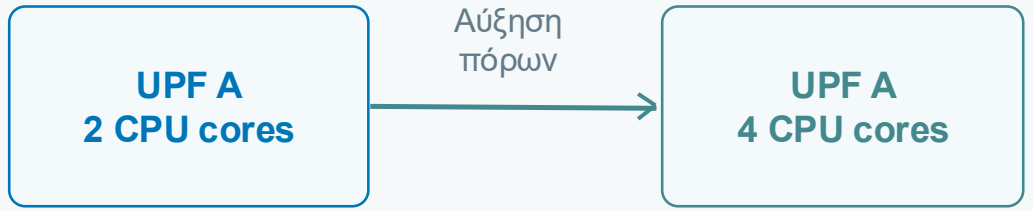


Κλιμάκωση UPF: scale up

Περισσότεροι υπολογιστικοί πόροι στο ίδιο UPF instance

- Αυξάνουμε τους πόρους του UPF A, π.χ. από 2 σε 4 CPU cores, όταν η επεξεργασία αποτελεί το bottleneck.
- Τα PDU sessions μπορούν να παραμείνουν στο ίδιο UPF, εφόσον διατηρούνται η λειτουργία και η κατάστασή του.
- Το κέρδος εξαρτάται από διαθέσιμους πόρους στον host και από τη δυνατότητα του UPF να αξιοποιήσει την επιπλέον CPU.
- Η αλλαγή μπορεί να απαιτεί επανεκκίνηση, ανάλογα με την πλατφόρμα και την υλοποίηση.

Ενδεικτικό παράδειγμα



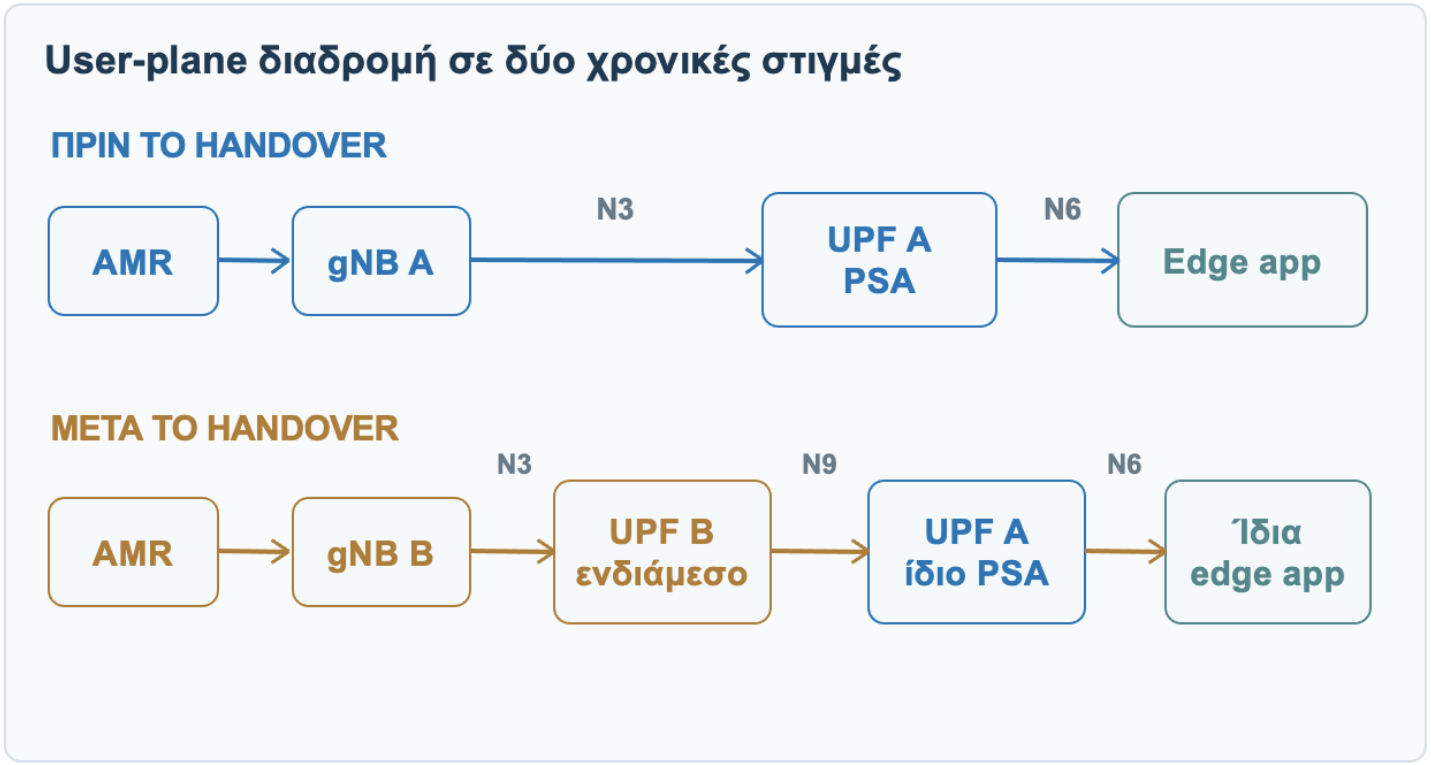
Ίδιο instance και sessions, εφόσον η αλλαγή γίνει χωρίς αντικατάσταση.

Πότε επιλέγω scale up και πότε scale out;

Scale up: περισσότεροι πόροι στο ίδιο UPF.
Scale out: επιπλέον UPF instance(s) και επιλογή sessions από το SMF.

Mobility και edge: αλλάζουμε διαδρομή ή και anchor;

Το AMR μετακινείται από το gNB A στο gNB B και συνεχίζει να χρησιμοποιεί την ίδια edge εφαρμογή.



- Το handover αλλάζει το gNB.
- Στο συγκεκριμένο παράδειγμα, εισάγεται και ένα ενδιάμεσο UPF B, ενώ **το UPF A παραμένει το PDU Session Anchor.**
- Η εφαρμογή παραμένει στον ίδιο edge server. Επομένως συγκρίνουμε κυρίως την επίδραση της νέας διαδρομής και των radio συνθηκών.
- Η πρόσθετη διαδρομή μέσω UPF B μπορεί να αυξήσει το RTT. Το πραγματικό αποτέλεσμα εξαρτάται από το transport, το φορτίο και την επεξεργασία στα UPFs.
- Επόμενη απόφαση: αξίζει μια άλλη διαδρομή ή η αλλαγή PSA; Η αλλαγή PSA έχει πρόσθετες συνέπειες για τη συνέχεια της συνεδρίας.

Άσκηση: αξίζει η μετάβαση στο edge (βλ. προηγούμενο slide);

Σύγκριση μέσου RTT και κόστους μετάβασης στο ίδιο σενάριο

Υποθέσεις

• Αρχικά χρησιμοποιείται cloud server. Η εναλλακτική είναι local UPF και edge server.

• Το AMR παράγει 20 requests/s επί 10 s. Είναι το συνολικό φορτίο στον επιλεγμένο server.

• Η μετάβαση γίνεται στην αρχή και διαρκεί 0,3 s. Τα αιτήματα που φτάνουν τότε χάνονται.

• Υποθέτουμε ομοιόμορφες αφίξεις και τις προηγούμενες παραμέτρους του μοντέλου M/M/1.

Υπολογισμός για λ = 20 requests/s	
Μέσο RTT cloud	20 + 1000 / (1000 − 20) ≈ 21,02 ms
Μέσο RTT edge	8 + 1000 / (100 − 20) = 20,50 ms
Μέσο όφελος	≈ 0,52 ms ανά ολοκληρωμένο αίτημα
Απώλειες στη μετάβαση	20 × 0,3 ≈ 6 αιτήματα
Αθροιστικό όφελος	194 × 0,52 ≈ 101 ms συνολικά

μ_{edge} = 100 και μ_{cloud} = 1000 requests/s.
Οι τιμές RTT είναι μέσοι χρόνοι μόνιμης κατάστασης.

Πώς θα τεκμηριώναμε την απόφαση στην πράξη;

Η μεταφορά μπορεί να βελτιώσει τη διαδρομή αλλά να διακόψει την υπηρεσία

Πληροφορία	Ερώτημα
Παραμονή στην περιοχή	Θα μείνει αρκετά ώστε να ωφεληθεί;
Disruption / packets lost	Πόσο συχνά παραβιάζονται deadlines;
SSC / IP continuity	Χρειάζεται νέα IP ή νέα σύνδεση εφαρμογής;
Μετρήσεις μετά την αλλαγή	p95/p99 RTT, packet loss, recovery time
Πολιτική απόφασης	Προτεραιότητα σε continuity ή latency;

Δεν συγκρίνουμε ms συνολικού οφέλους και χαμένα requests χωρίς κοινή συνάρτηση κόστους.

Network slicing: γιατί δημιουργούμε slices;

Διαφορετικές υπηρεσίες με διαφορετικές απαιτήσεις πάνω σε κοινή υποδομή

Ένα network slice είναι ένα λογικό δίκτυο με συγκεκριμένες λειτουργίες και χαρακτηριστικά, διαμορφωμένο για μια κατηγορία υπηρεσιών.

• Στην αποθήκη, ο έλεγχος των AMRs απαιτεί χαμηλή καθυστέρηση, ενώ το video μεγαλύτερο throughput.

• Διαφοροποιούμε τη διαχείριση και την κατανομή πόρων ανά υπηρεσία, αξιοποιώντας κοινή υποδομή.

• Στόχος: κατάλληλη εξυπηρέτηση και ελεγχόμενη αλληλεπίδραση μεταξύ υπηρεσιών.

Από τις απαιτήσεις στην υποδομή

Slice A | Control AMR
Χαμηλή καθυστέρηση και αξιοπιστία

Slice B | Video
Υψηλότερος ρυθμός μεταφοράς

Λειτουργίες και πολιτικές ανά slice
Κοινές ή αποκλειστικές network functions

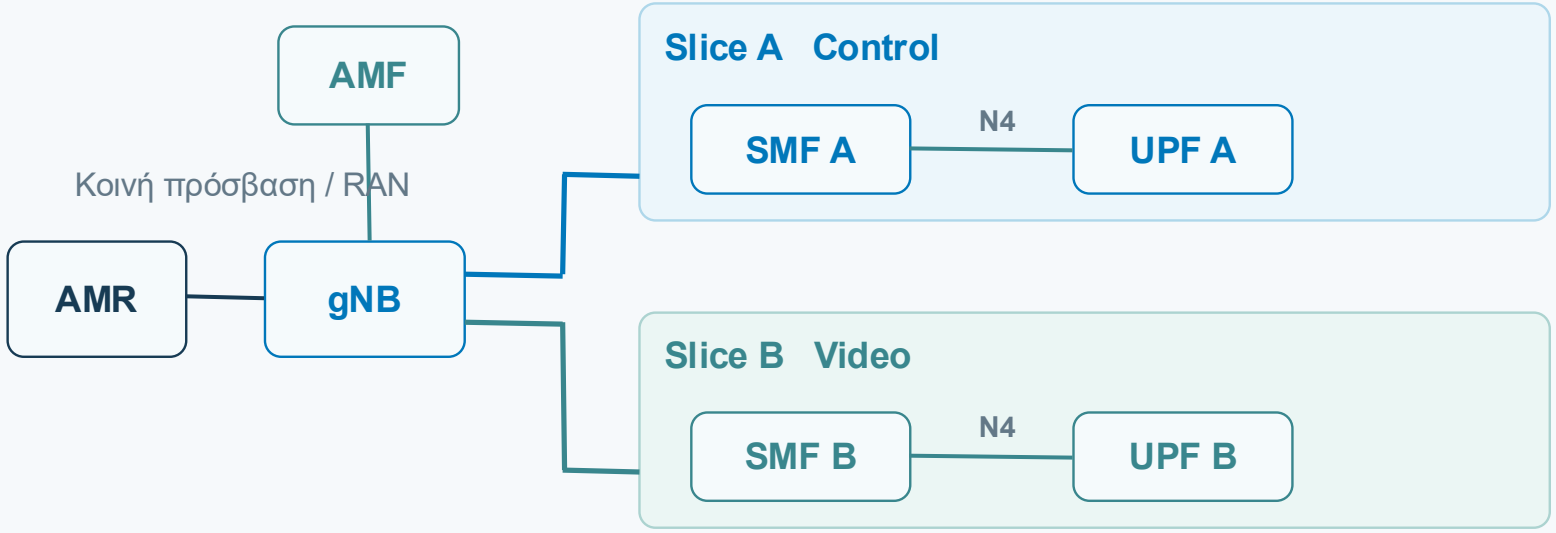
Κοινή φυσική υποδομή
RAN • transport • compute

Το slicing επιτρέπει διαφοροποίηση υπηρεσιών. Η απομόνωση εξαρτάται από τη διαχείριση πόρων.

Αρχιτεκτονική slicing: κοινές και ειδικές λειτουργίες

Ένα AMR χρησιμοποιεί δύο PDU sessions, ένα για κάθε υπηρεσία

Απλοποιημένη συνδεσιμότητα και ενδεικτικός διαχωρισμός



Γραμμές ανά slice: λογική απεικόνιση των δύο sessions

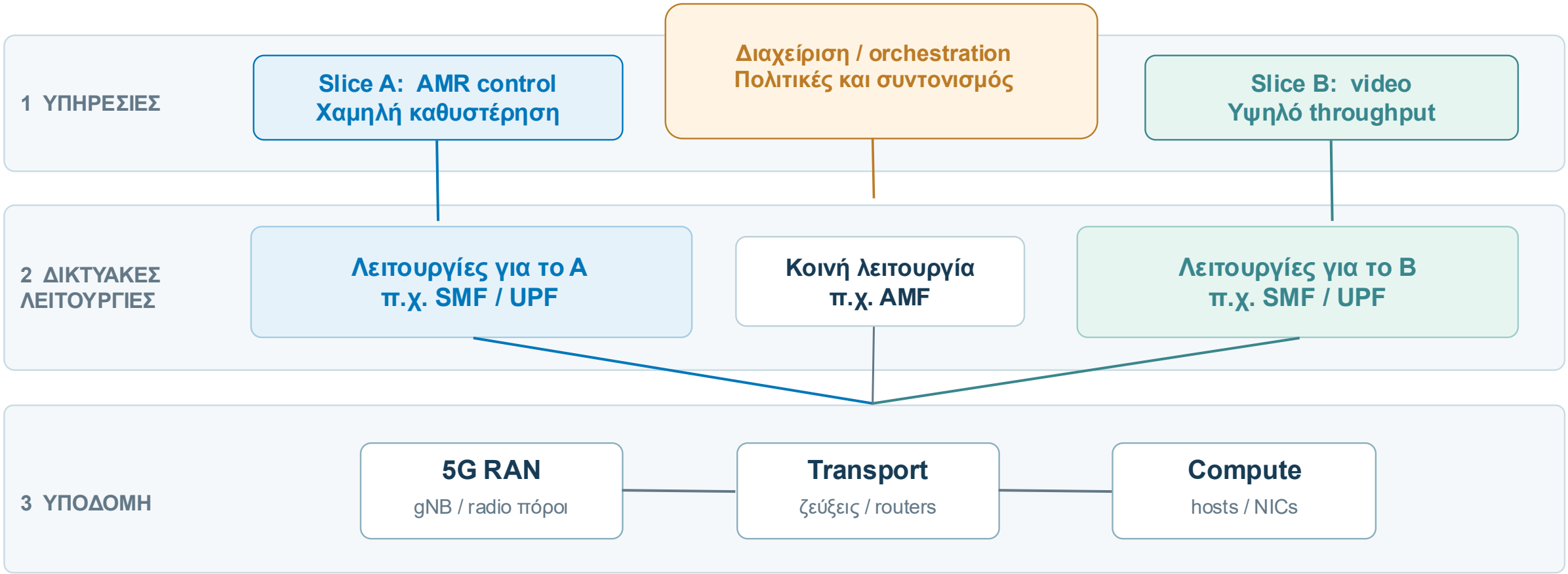
Πώς το διαβάζουμε

- PDU session 1 συνδέεται με το Slice A, session 2 με το Slice B.
- S-NSSAI: SST και προαιρετικά SD. Αναγνωρίζει το slice, χωρίς να εγγυάται πόρους.
- Στο παράδειγμα τα AMF / gNB είναι κοινά, ενώ κάθε slice χρησιμοποιεί ξεχωριστά SMF / UPF instances.

Κοινό gNB, transport ή host παραμένουν πιθανά σημεία ανταγωνισμού, ακόμη και με ξεχωριστά UPFs.

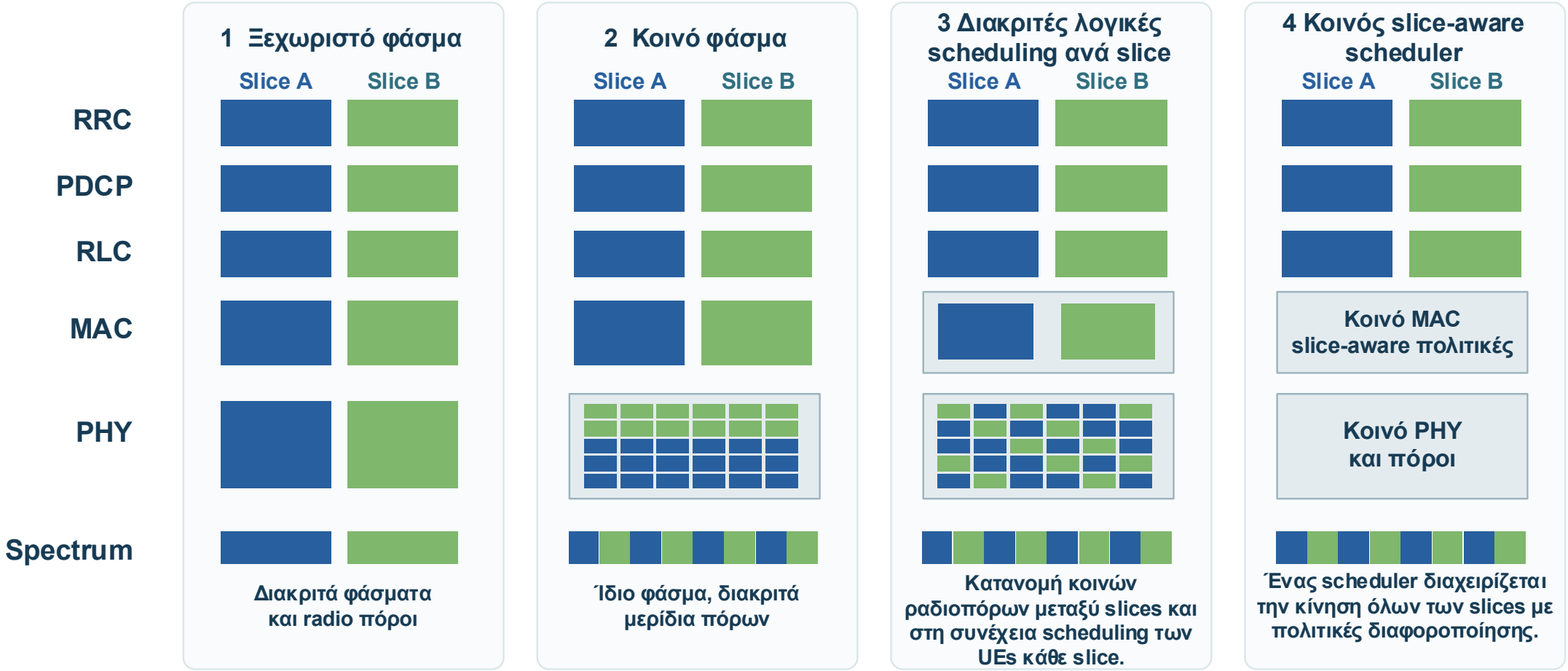
Network slicing: από τις υπηρεσίες στην υποδομή

Δύο λογικά δίκτυα, διαμορφωμένα πάνω σε κοινούς φυσικούς πόρους



RAN slicing: επιλογές κοινής χρήσης και απομόνωσης

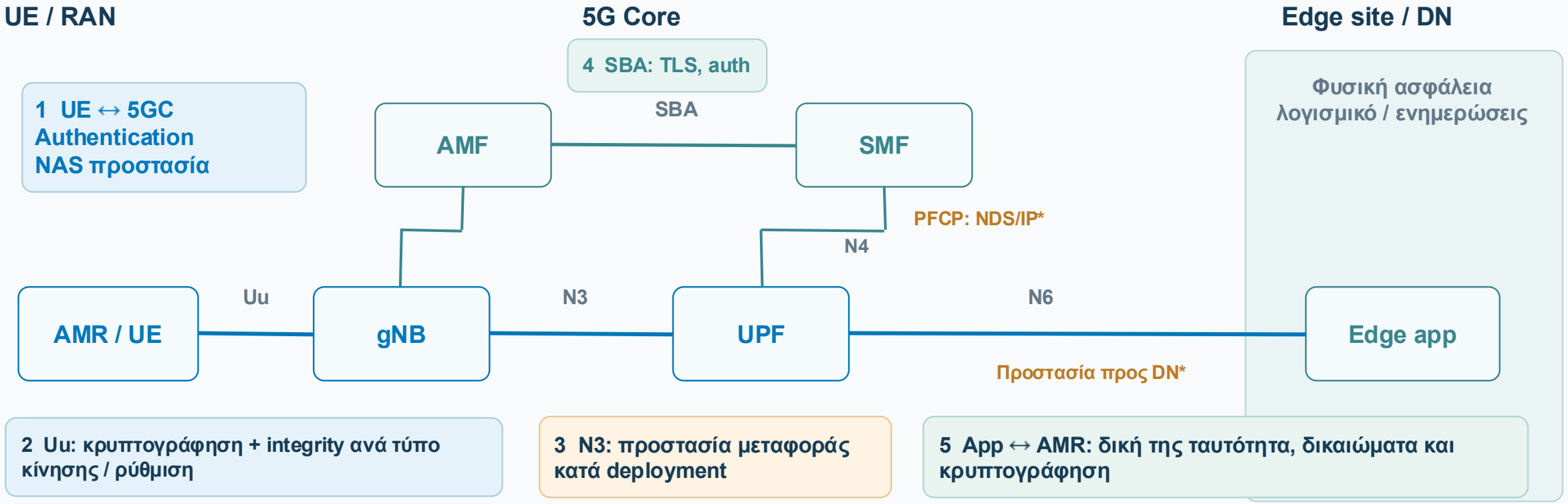
Ενδεικτικά σενάρια για δύο slices, από ξεχωριστό φάσμα έως κοινό MAC / PHY



Βασισμένο στο παλαιότερο εκπαιδευτικό υλικό: 5G System Design, Wiley

Ασφάλεια 5G: προστασία σε διαφορετικά σημεία

Παράδειγμα ιδιωτικής αποθήκης με AMR και edge εφαρμογή



* Ανά όριο εμπιστοσύνης / deployment. Λεπτομέρειες και παραδοχές στο επόμενο slide.

Η προστασία της 5G σύνδεσης και η ασφάλεια της εφαρμογής καλύπτουν διαφορετικά όρια εμπιστοσύνης.

Ασφάλεια ανά διεπαφή: τι προστατεύεται και πώς

Οι μηχανισμοί επιλέγονται με βάση το πρωτόκολλο και το όριο εμπιστοσύνης

Διεπαφή	Τι μεταφέρεται	Ενδεικτική προστασία	Περιγραφή
Uu	UE ↔ gNB RRC και user data	Access Stratum security: κρυπτογράφηση. Integrity σύμφωνα με το είδος κίνησης και τη ρύθμιση.	Η κρυπτογράφηση προστατεύει το περιεχόμενο στην ασύρματη ζεύξη. Η ακεραιότητα (integrity) επιτρέπει στον παραλήπτη να ανιχνεύει αλλοίωση των προστατευμένων μηνυμάτων.
N1 / NAS	UE ↔ AMF NAS σηματοδοσία	5G-AKA (Authentication and Key Agrement) ή EAP-AKA′ (Extensible Authentication Protocol). NAS integrity και κρυπτογράφηση.	Το authentication ελέγχει τη συνδρομή και οδηγεί στην παραγωγή κλειδιών. Τα κλειδιά NAS χρησιμοποιούνται για την προστασία της σηματοδοσίας μεταξύ UE και AMF.
SBA	AMF ↔ SMF υπηρεσίες 5G Core	TLS, ταυτοποίηση NFs και εξουσιοδότηση υπηρεσιών, σύμφωνα με το deployment.	Το TLS προστατεύει την επικοινωνία μεταξύ των άκρων της σύνδεσης. Τα πιστοποιητικά χρησιμοποιούνται για την ταυτοποίηση, ενώ οι μηχανισμοί εξουσιοδότησης ελέγχουν αν μία NF επιτρέπεται να καλέσει μια συγκεκριμένη υπηρεσία.
N2 / N3	gNB ↔ AMF / UPF NGAP / GTP-U	Προστασία ακεραιότητας, εμπιστευτικότητας και replay. IPsec ESP / IKEv2 υποστηρίζεται.	Το IPsec ESP μπορεί να προστατεύει τα μεταφερόμενα IP πακέτα. Το IKEv2 διαπραγματεύεται τις παραμέτρους ασφάλειας και τα κλειδιά. Η ενεργοποίηση και τα άκρα προστασίας καθορίζονται με βάση το deployment.
N4 / N9	SMF ↔ UPF / UPF ↔ UPF PFCP / user plane	NDS/IP για εσωτερικές non-SBA διεπαφές, εκτός αν παρέχεται ασφάλεια με άλλα μέσα.	Το NDS/IP εφαρμόζει προστασία στο επίπεδο IP, π.χ. με IPsec, ώστε να προστατεύεται η μεταφορά των μηνυμάτων PFCP στο N4 και της κίνησης στο N9. Το PFCP χρησιμοποιείται από το SMF για τη διαμόρφωση του UPF.
N6 / εφαρμογή	UPF ↔ DN AMR ↔ edge app	Transport / segmentation κατά το όριο εμπιστοσύνης. App TLS ή DTLS όπου ενδείκνυται.	Το segmentation περιορίζει ποιες οντότητες μπορούν να επικοινωνήσουν. Το TLS ή DTLS μπορεί να προστατεύει τα δεδομένα μεταξύ των άκρων της εφαρμογής, ενώ η εφαρμογή ελέγχει την ταυτότητα και τα δικαιώματα των χρηστών ή συσκευών.