

ΜΕΓΑΛΑ DATASETS ΣΤΟ ΔΙΚΤΥΟ

ΔΙΑΣΤΗΜΙΚΟ ΠΡΟΓΡΑΜΜΑ ΚΟΠΕΡΝΙΚΟΣ

Χρήστος Αργυρόπουλος
cargious@grnet.gr

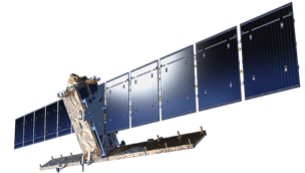
Σύνολα Δεδομένων - Πηγή

- Διαστημικό Πρόγραμμα Κοπέρνικος
 - Ευρωπαϊκός Οργανισμός Διαστήματος (ΕΟΔ)
(European Space Agency - ESA)
 - Παρατήρηση της Γης από το διάστημα
 - Υπηρεσίες παρατήρησης με όργανα σε δορυφόρους και επίγειους σταθμούς
(ευθυγράμμιση)

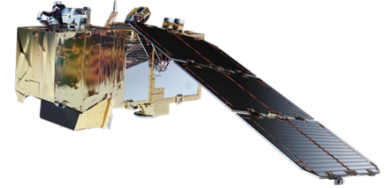
Σύνολα Δεδομένων - Όργανα

- Sentinel-1

- Παντός καιρού, ημέρας και νύχτας
- Εικόνες ραντάρ για ωκεανούς και ξηρά
- Δυνατότητα απεικόνισης με νέφωση

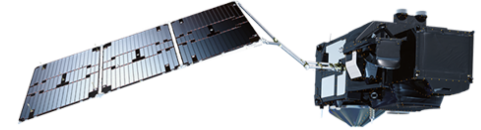


Σύνολα Δεδομένων - Όργανα



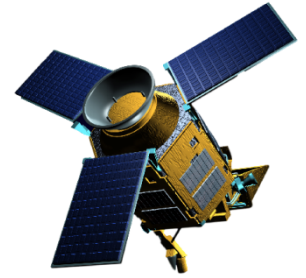
- Sentinel-2
 - Μεσαίας ανάλυσης (10μ, 20μ ή 60μ) πολυφασματικός οπτικός δορυφόρος
 - Παρατήρηση ξηράς, βλάστησης και νερού

Σύνολα Δεδομένων - Όργανα



- Sentinel-3
 - Μετρήσεις επιφανείας σε επιφανείας (300μ ανάλυση)
 - Μετρήσεις θερμοκρασίας νερού και εδάφους (0.3K ακρίβεια)
 - Μετρήσεις εξάτμισης νερού, συμπύκνωσης υδρατμών και θερμικής ακτινοβολίας της Γης

Σύνολα Δεδομένων - Όργανα

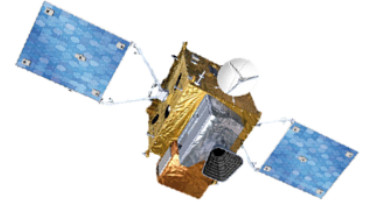


- Sentinel-5P

- Μέτρηση ατμοσφαιρικής συγκέντρωσης σημαντικών ενώσεων (O_3 , NO_2 , SO_2)
- Βελτίωση των κλιματικών μοντέλων και της πρόγνωσης καιρού
- Δορυφόρος-γέφυρα για την κάλυψη αναγκών μέχρι την εκτόξευση του Sentinel-5

Σύνολα Δεδομένων - Όργανα

- Sentinel-4 (εκτόξευση 2021)
 - Δεδομένα για την ποιότητα της ατμόσφαιρας
 - Μετρήση συγκέντρωσης επιβλαβών αερίων



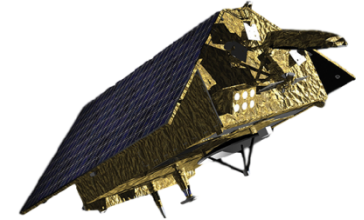
Σύνολα Δεδομένων - Όργανα



- Sentinel-5 (εκτόξευση 2021)
 - Μέτρηση ποιότητας ατμόσφαιρας και ηλιακής ακτινοβολίας
 - Μέτρηση όζοντος και κλιματολογικών συνθηκών
 - Παγκόσμια κάλυψη με μεγάλη ακρίβεια

Σύνολα Δεδομένων - Όργανα

- Sentinel-6 (εκτόξευση 2020)



- Μετρήσεις ύψους θαλάσσιων υδάτων (ακρίβεια εκατοστών)
- Παγκόσμια αποτύπωση θαλάσσιας επιφανείας σε 10 ημέρες
- Παρατήρηση θαλάσσιων ρευμάτων και θερμικών φορτίων

Κοπέρνικος - Λειτουργία

- Κέντρα ελέγχου ΕΟΔ
 - Sentinel Flight Operations Segment (FOS) --> Έλεγχος πτήσεων
 - Sentinel Payload Data Ground Segments (PDGS) --> Συλλογή/επεξεργασία δεδομένων



- Core Ground Stations (CGS) for Data Acquisition and product generation in Near Real Time
 - Matera (eGeos, IT)
 - Svalbard (K-Sat, NO)
 - Maspalomas (Inta, SP)
 - Alaska (K-Sat, USA)
- Sentinel Processing and Archiving Centers (PAC)
 - S1 (Astrium/UK, DLR/DE)
 - S2 (Astrium/UK, Indra/SP)
 - S3 (OLCI Land DLR/DE, SRAL CLS/FR, SLSTR-SYN ACRI/FR)
 - S3 (OLCI Marine EUMETSAT/DE)
- Sentinel Missions Performance Centers (MPC)
 - S1 (CLS/FR)
 - S2 (CS/FR)
 - S3 (ACRI/FR)
- Sentinel Precise Orbit Determination (POD) operated by GMV-led (Tres Cantos, Spain)
- Payload Data Management Centre (PDMC) (European Space Research Institute – ESRIN, Frascati/IT)

Μεταφορά Δεδομένων

- Ο Sentinel-1A, με ύψος τροχιάς 700χλμ., μεταφέρει δεδομένα σε επίγειους σταθμούς όταν περνά πάνω από αυτούς
- Αντίθετα οι γεωστατικοί δορυφόροι, σε ύψος 36,000 χλμ έχουν μόνιμη οπτική επαφή με τους σταθμούς βάσης τους
- Ο Sentinel-1A, με ζεύξη 36,000 χλμ και χρήση laser, συνδέθηκε με τον Alphasat (τηλεπ/κος δορυφόρος) - 1η επιτυχημένη προσπάθεια αναμετάδοσης μεγάλου όγκου δεδομένων μέσω δορυφόρου
- Άμεση αναμετάδοση των δεδομένων --> άμεσα διαθέσιμα σε εφαρμογές για ασφάλεια ναυσιπλοΐας, φυσικές καταστροφές κλπ.

Κόμβοι Δεδομένων

- **Copernicus Services Hub**
 - 200 χρήστες
 - Διαθέσιμα όλα τα δεδομένα
 - Δεδομένα κατ' απαίτηση
 - Συμφωνία στάθμης παρεχόμενης υπηρεσίας
- **International Hub** (λειτουργία από ΕΔΕΤ & Εθνικό Αστεροσκοπείο Αθηνών)
 - 4 χρήστες (National Aeronautics and Space Administration/USA, National Oceanic & Atmospheric Administration/USA, U.S. Geological Survey/USA, Geoscience/AUSTRALIA)
 - Δεδομένα 3 εβδομάδων
 - Συμφωνία στάθμης παρεχόμενης υπηρεσίας
- **Collaborative Hub** (λειτουργία από ΕΔΕΤ & Εθνικό Αστεροσκοπείο Αθηνών)
 - > 25 χρήστες (συμπεριλαμβανομένου κόμβων δεδομένων ειδικού σκοπού)
 - Δεδομένα 2-4 εβδομάδων
 - Συμφωνία στάθμης παρεχόμενης υπηρεσίας
- **Copernicus OpenAccess Hub**
 - 150.000 χρήστες
 - Διαθέσιμα όλα τα δεδομένα
- **Hub Relays, national mirror sites etc.**
 - Αύξηση διαθεσιμότητας μέσω τοπικών κόμβων
 - Δημιουργία διαφορετικών προϊόντων δεδομένων/μεταδεδομένων

Όγκος Δεδομένων

ESA (all hubs)

8x10⁶ δημοσιευμένα προϊόντα

60 PB δεδομένων έχουν διανεμηθεί σε χρήστες

Sentinel-1: 1w (σχεδόν πραγματικού χρόνου), 24h (όχι πραγματικού)

Sentinel-3: <3h (σχεδόν πραγματικού χρόνου), <1month (όχι πραγματικού)

GRNET/NOA καταναλώνει

- ~20k προϊόντα /ημέρα
- ~13 TiB /ημέρα
- Δεδομένα από κεντρικό DC διανομής
- *Εξαίρεση: Κεντρικό σημείο διανομής των δεδομένων Sentinel-5P*

GRNET/NOA παρέχει

- IntHub
 - 36k προϊόντα - 13 TiB /ημέρα
- ColHub
 - 51k προϊόντα - 17 TiB /ημέρα
- DiasHub
 - 61k προϊόντα - 15 TiB/ημέρα
- Μέσο μέγεθος προϊόντων
 - 700 MB (15 MiB - 1.7GiB)

Σκοπός

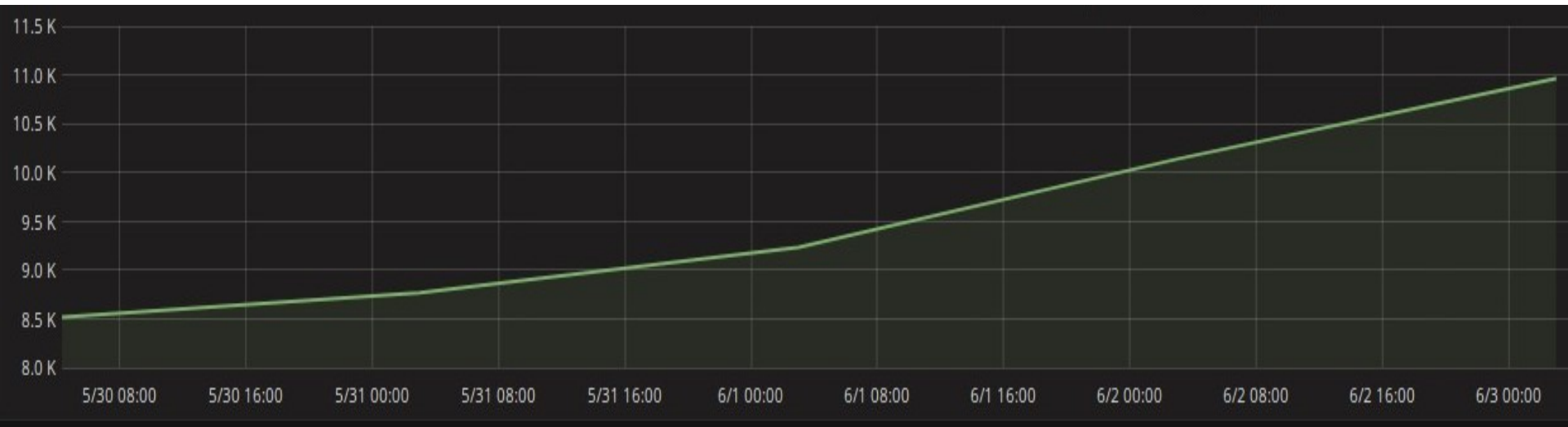
“Μεταφορά των δορυφορικών δεδομένων από το κεντρικό σημείο
διαμοιρασμού στο Κέντρο Δεδομένων ΕΔΕΤ”

Ή

“Μια συνεχή ροή μεγάλων αρχείων (TCP ροές) που πρέπει να
παραδοθούν το συντομότερο μεταξύ σημείων που απέχουν
χιλιάδες χιλιόμετρα”

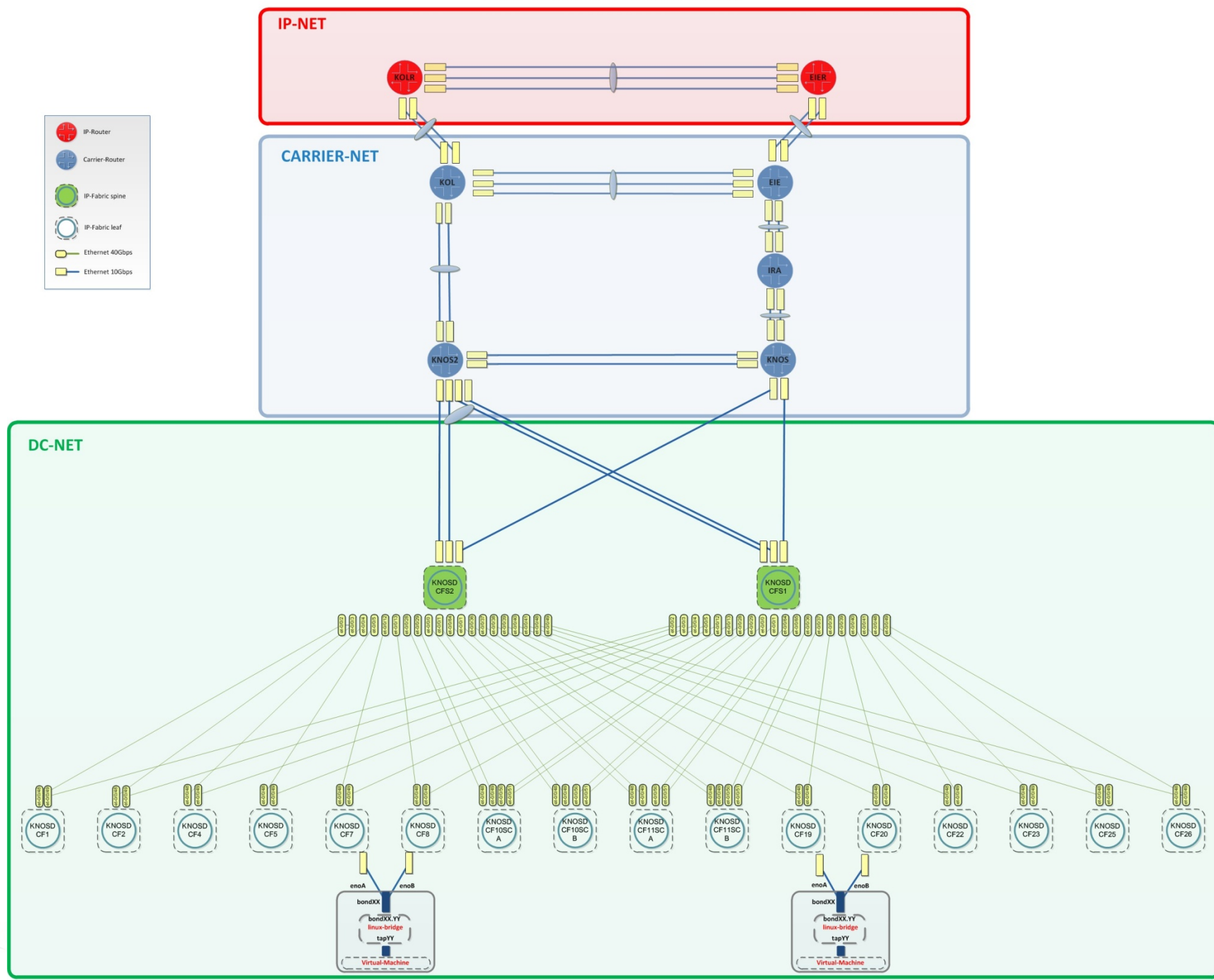
Το πρόβλημα

Ουρά αρχείων $\rightarrow \infty$



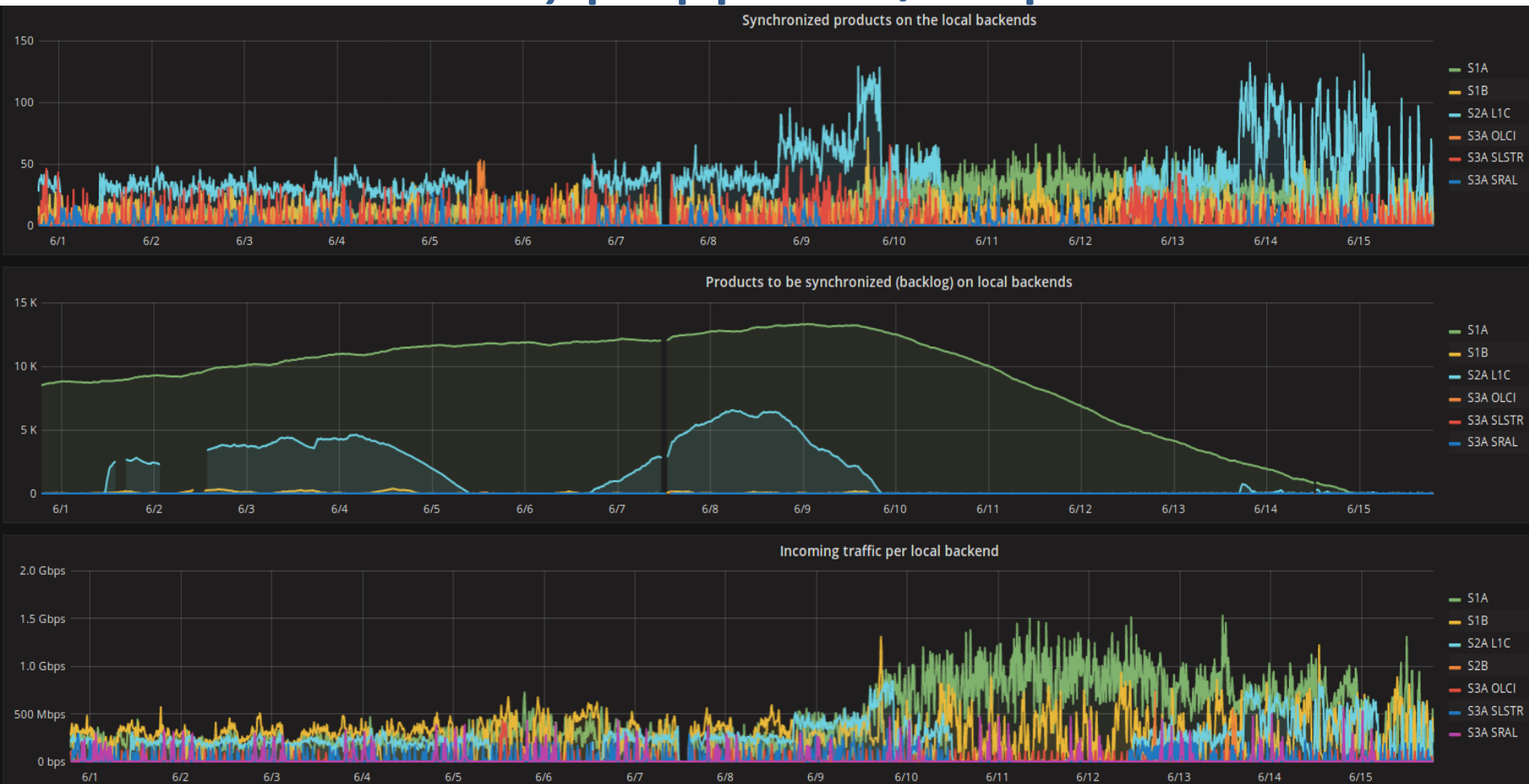
Πολύπλοκο οικοσύστημα

- DC με spine-leaf collapsed CLOS τοπολογία
- EVPN-VXLAN για L2 σύνδεση εξυπηρετητών & δρομολογητών
- Spine μεταγωγείς δρουν ως δρομολογητές
- Carrier δίκτυο (MPLS/IS-IS) παρέχει L2VPNs για σύνδεση DC-σε-IP
- Carrier συνδέσεις πάνω από οπτικό δίκτυο DWDM
- Εφεδρείες σε πολλαπλά επίπεδα WAN/DC
- Πολλαπλές οδεύσεις μεταξύ των εξυπηρετητών
- Εξυπηρετητές με οδεύσεις προς 2 μεταγωγείς (1x2 ή 2x2)
- Εικονικές μηχανές πάνω στους εξυπηρετητές
- Ganetti/KVM στοίβα εικονικοποίησης
- Ειδικό λογισμικό για την διαχείριση των προϊόντων δεδομένων
- Το κεντρικό σημείο διανομής χιλιάδες χιλιόμετρα μακριά
- Πολλαπλές διαχειριστικές περιοχές (δυσκολία συνεννόησης)



Η πυροσβεστική λύση

Αυξηση ροών/ουρά



Υπόθεση Εργασίας

*Το γίνόμενο $bandwidth * RTT$ απαιτεί μεγάλο TCP παράθυρο*

- Υπολογισμός μεγέθους καταχωρητή:
 $RTT = \text{propagation} + \text{transmission} + \text{processing} + \text{queuing}$
 $buffer(size) = bandwidth * RTT$
- Απλοποίηση λόγω μεγάλης απόστασης:
 $buffer\ size = bandwidth * propagation\ delay$
- Βασικές προ-υποθέσεις:
Δεν υπάρχει κορεσμός σε γραμμή δικτύου και απώλεια πακέτων
(no congestion, no back-off algorithm)

Συμπέρασμα: Με σταθερό RTT το $bandwidth$ είναι ανάλογο του μεγέθους καταχωρητή

Περιβάλλον Δοκιμών

- Δοκιμές μεταξύ των Κέντρων Δεδομένων σε Κνωσό και ΥΠΕΠΘ
- Χρήση εργαλείου iperf με 1 TCP σύνδεση
- Μηδενικός κορεσμός/απώλειες σε WAN/DC δίκτυα
- Χωρίς βελτιστοποίηση σε καταχωρητές:
HOST A, DC A → HOST B, DC B
1η δοκιμή: 0.0-10.0 sec 1.58 Gbits/sec
2η δοκιμή: 0.0-10.0 sec 2.00 Gbits/sec
3η δοκιμή: 0.0-10.0 sec 3.11 Gbits/sec
- >> 1Gbit/s vs. GRNET-Central_EU DC

Περιβάλλον Δοκιμών



Ο υπολογισμός για 8ms RTT με βασικό μέγεθος καταχωρητή δίνει <3.1 Gbit/sec, κοντά στη μέγιστη τιμή που παρατηρήθηκε



Παρατηρήθηκε όμως μεγάλη διακύμανση

Δεν υπήρχαν απώλειες πακέτων άρα το RTT άλλαζε ?

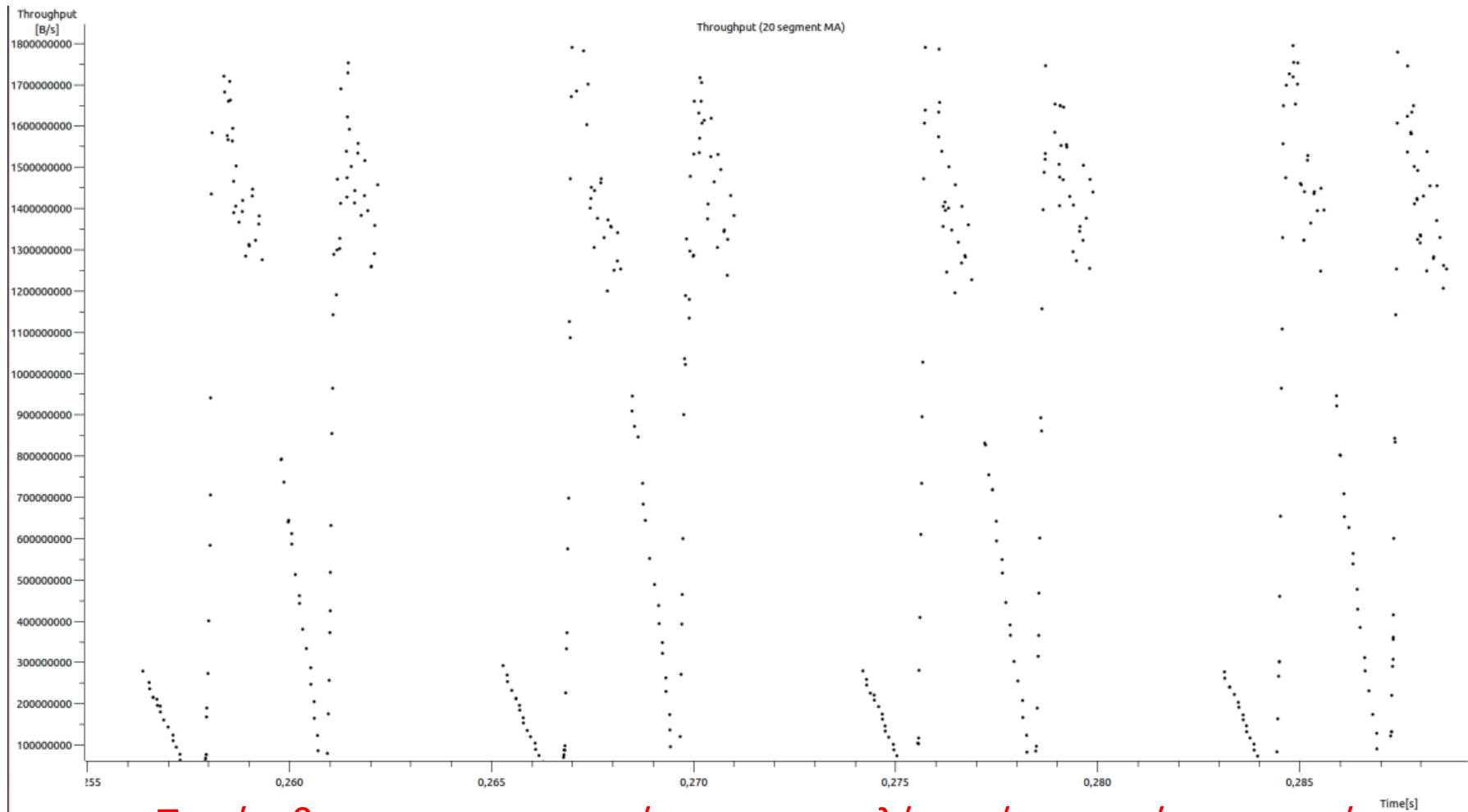
- Μετά από δοκιμές RTT ~ [8, 12]ms
- Πολλαπλά μονοπάτια >128



Η κατανόηση της πηγής διακύμανσης ήταν δύσκολη υπόθεση (DC/Carrier/IP δίκτυο)

- Ο λόγος της διακύμανσης ήταν μια οπτική διαδρομή που σε συνδυασμό με την κατανομή κίνησης σε πολλαπλές διαδρομές ήταν δύσκολο να εξηγηθεί
- Με αύξηση μεγέθους καταχωρητή στο δέκτη των δεδομένων παρατηρήθηκε βελτίωση, αλλά όχι η προβλεπόμενη

Περιβάλλον δοκιμών



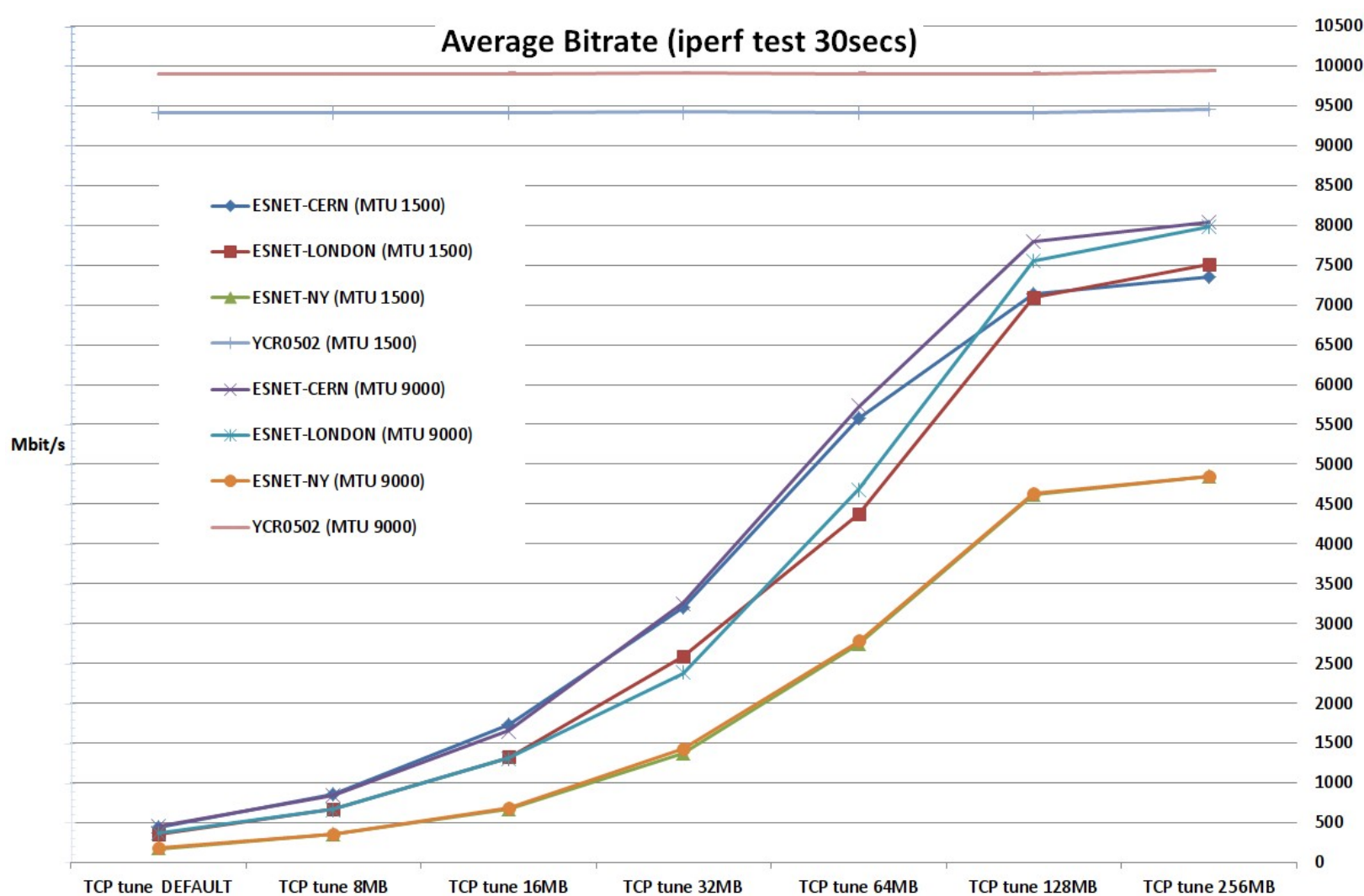
Το μέγεθος του καταχωρητή του αποστολέα πρέπει να είναι αρκετά μεγάλος για να υποστηρίξει μεγάλο αριθμό υπό μετάδοση δεδομένων

Δοκιμές στον πραγματικό κόσμο

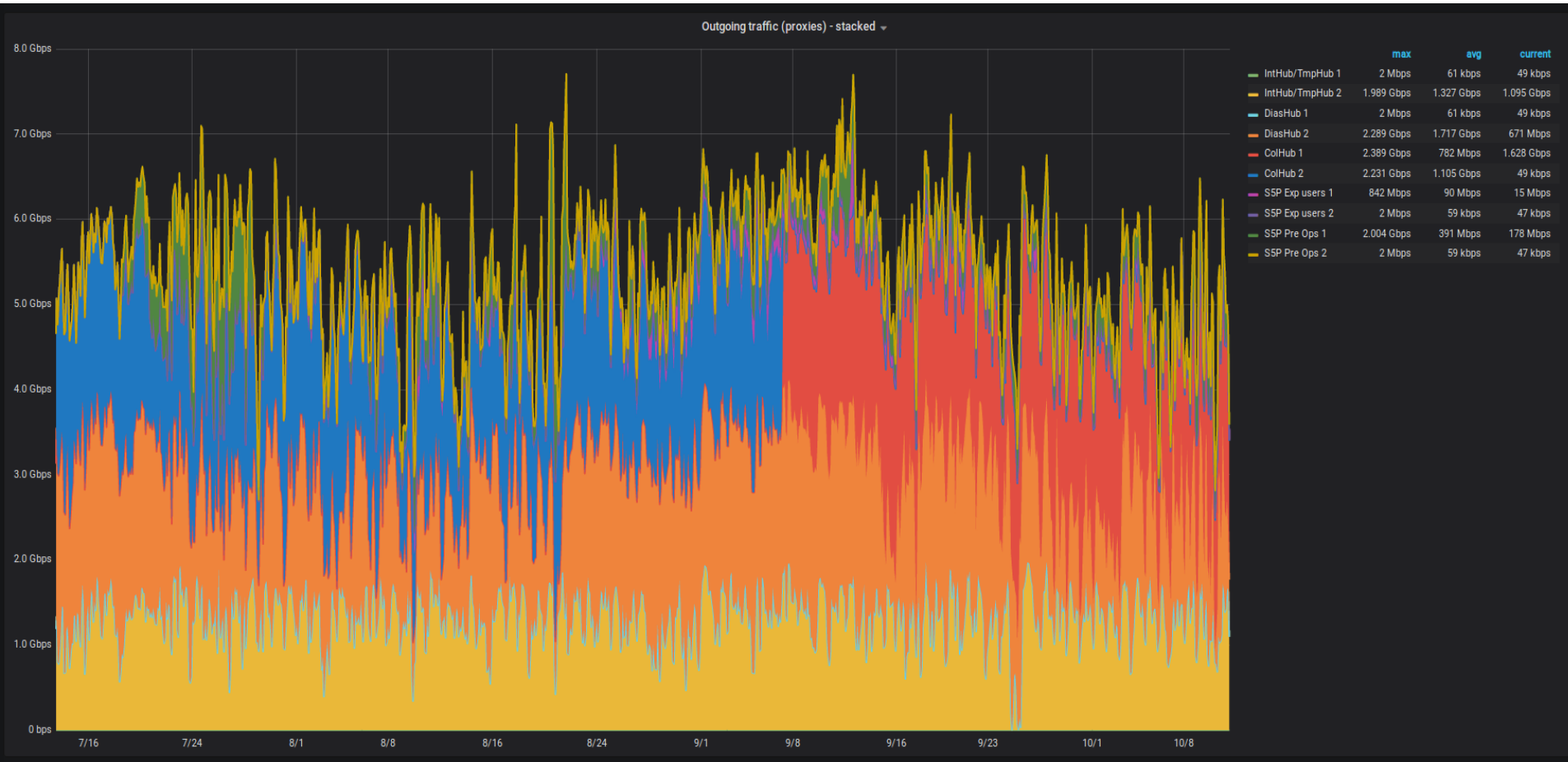
- 1) Χρήση εργαλείου μετρήσεων perfSONAR
- 2) Χρήση διακομιστών ESnet (256 Mbyte buffers & 10Gbit NICs)
- 3) Μηδενικές απώλεις και κορεσμός συνδέσεων
- 4) Πραγματική κίνηση σε συνδέσεις ΕΔΕΤ με GEANT 1-1,5Gbps και διαμοιρασμός κίνησης με L3-L4 κριτήρια

*πραγματικός κόσμος = ακαδημαϊκά δίκτυα

Πραγματικός κόσμος



Αποτελέσματα



Νέες Διαστάσεις στην Έρευνα & Εκπαίδευση

