Interreg
Ελλάδα-Κύπρος
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
SmartWater2020

ΔΕΣΜΟΙ ΑΝΑΠΤΥΞΗΣ

ΕΚΠΑΙΔΕΥΤΙΚΟ ΣΕΜΙΝΑΡΙΟ

Τεχνολογίες Επικοινωνιών και Ανάλυσης Δεδομένων για Ευφυή Δίκτυα Ύδρευσης

Ηράκλειο, 30/11/2020

Συμπίεση Δεδομένων με Χρήση Συμπιεστικής Δειγματοληψίας

Γιώργος Τζαγκαράκης, Ερευνητής Β’, ΙΤΕ-ΙΠ

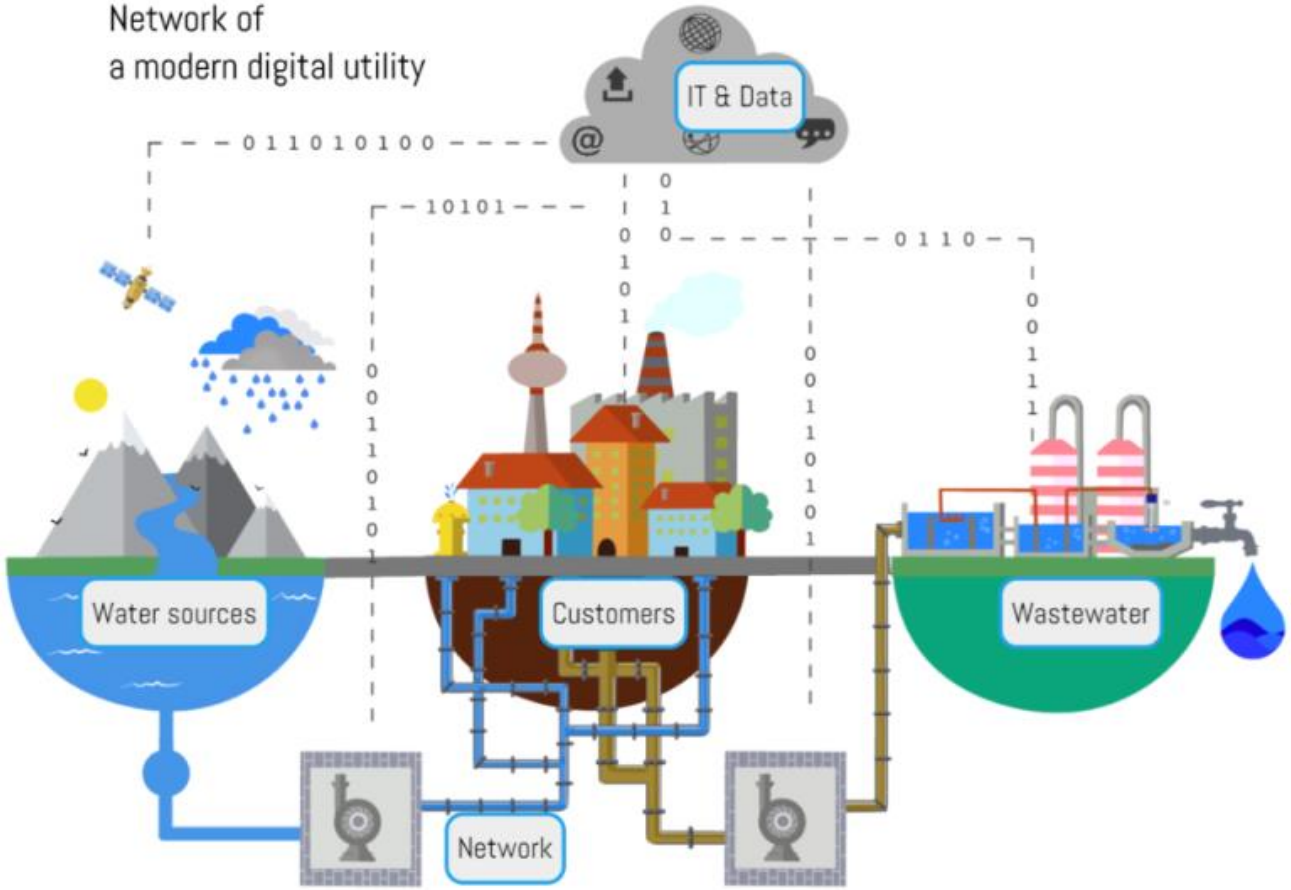
Interreg
Ελλάδα-Κύπρος
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
SmartWater2020

Συμπίεση Δεδομένων

FORTH
INSTITUTE OF COMPUTER SCIENCE

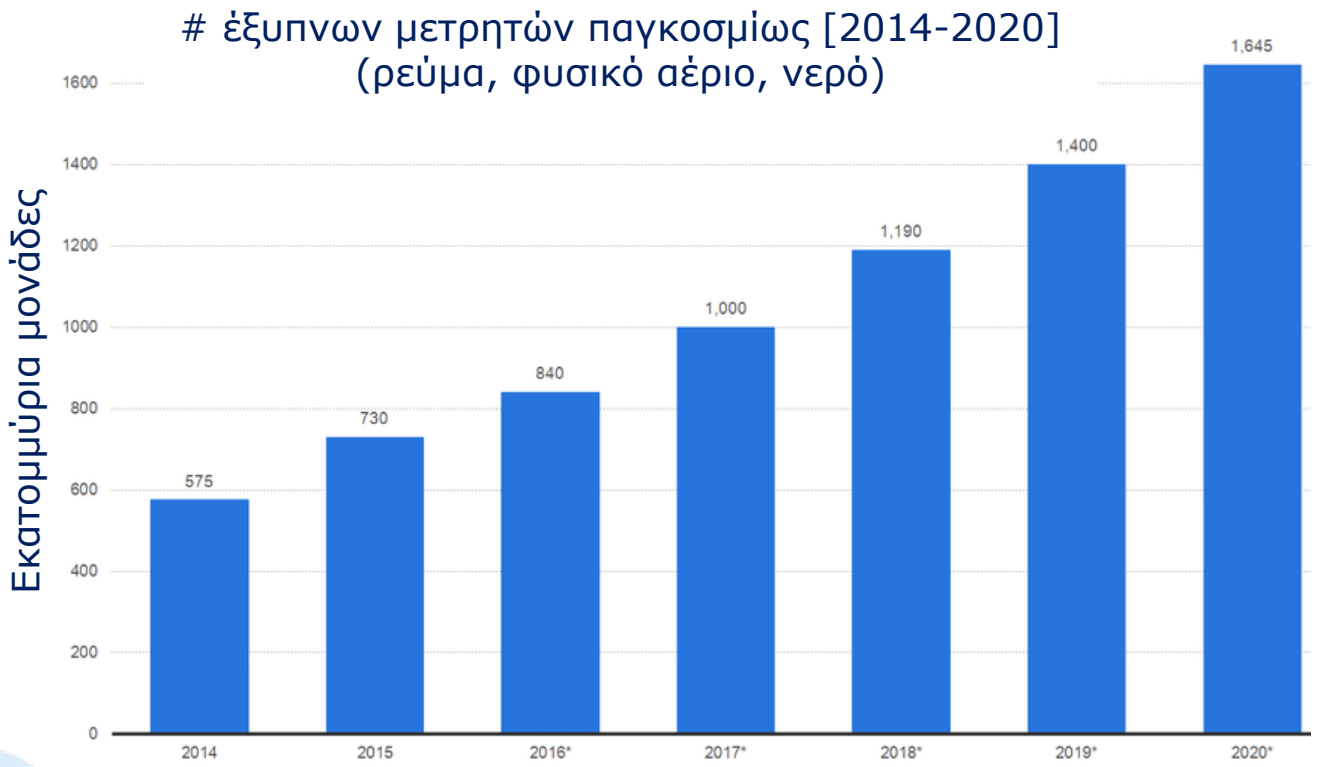
ΔΕΣΜΟΙ
ΑΝΑΠΤΥΞΗΣ

Ευφυή Δίκτυα Ύδρευσης



Μεγάλα Δεδομένα
(Big Data)

Μεγάλα Δεδομένα



Συμπίεση Δεδομένων

Συμπίεση Δεδομένων

- Είναι η τέχνη της **αναπαράστασης** της πληροφορίας σε «**συμπαγή**» μορφή

- Αυτές οι αναπαραστάσεις δημιουργούνται αναγνωρίζοντας και αξιοποιώντας τη **δομή** στα δεδομένα μας

Συμπίεση Δεδομένων - Στόχοι

- Μείωση Χώρου **Αποθήκευσης**
- Μείωση Κόστους **Τηλεμετρίας**
- Μείωση Καταναλισκόμενης **Ενέργειας**

Συμπίεση Δεδομένων - Στόχοι

- Αποθήκευση

Η συμπίεση δεδομένων μειώνει το **μέγεθος** των δεδομένων (π.χ. ενός αρχείου), μειώνοντας τον αποθηκευτικό χώρο που απαιτείται για την αποθήκευσή τους

- Μετάδοση Δεδομένων

Μειώνει το **χρόνο** και την **ενέργεια** που απαιτείται για τη μετάδοση της πληροφορίας

Σύστημα Συμπίεσης Δεδομένων

- Ένα σύστημα συμπίεσης δεδομένων αποτελείται από **δύο** βασικά μέρη

Συμπιεστής

Παίρνει τα δεδομένα και παράγει μία συμπιεσμένη αναπαράσταση

Αποσυμπιεστής

Ανακατασκευάζει τα αρχικά δεδομένα ή μία προσέγγισή τους από τη συμπιεσμένη αναπαράσταση

Μέθοδοι Συμπίεσης

- Οι μέθοδοι συμπίεσης δεδομένων χωρίζονται σε **δύο** βασικές κατηγορίες

Συμπίεση Χωρίς Απώλειες (Lossless)

- Καμία απώλεια πληροφορίας
- Τέλεια ανάκτηση των αρχικών δεδομένων

Συμπίεση Με Απώλειες (Lossy)

- Απώλεια μέρους της αρχικής πληροφορίας
- Επιτυγχάνουν υψηλούς λόγους συμπίεσης

Συμπίεση Χωρίς Απώλειες

- Κωδικοποίηση Τρέχοντος Μήκους (Run Length Encoding – RLE)

  π.χ. ΑΑΑΑΑΑΑΑΒΒΒΑΑΓΓΔΔΔΔΔ (20 σύμβολα) ➪ 8Α3Β2Α2Γ5Δ (10 σύμβολα) ➪

  (10-20)/20 = -50%

- Κωδικοποίηση Huffman
- Αλγόριθμος Lempel-Ziv (LZ77)

Συμπίεση Με Απώλειες

- JPEG (για εικόνες)
- MPEG (για βίντεο)
- MP3 (για ήχο)

Εφαρμογές Συμπίεσης

Συμπιεστική Δειγματοληψία

Συμπιεστική Δειγματοληψία

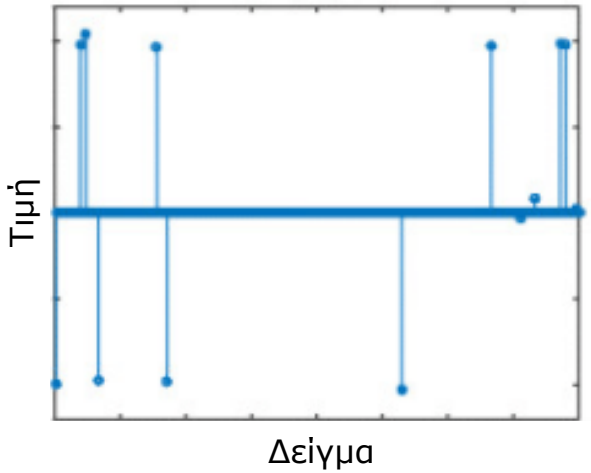
Συμπιεστική Δειγματοληψία
(Compressive Sensing)

- Ταυτόχρονη λήψη και συμπίεση των δεδομένων
- Εξαιρετικά απλή διαδικασία (συμπίεση) στα άκρα του δικτύου
- Το υπολογιστικό βάρος (αποσυμπίεση) πέφτει στο κέντρο τηλεελέγχου
- Προσφέρει ένα εγγενή μηχανισμό κρυπτογράφησης των δεδομένων

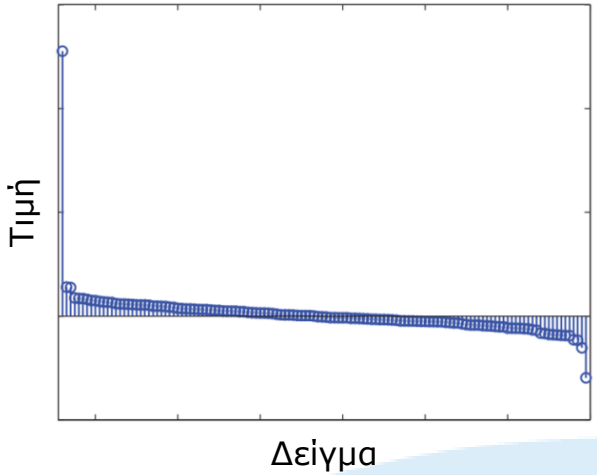
Συμπιεστική Δειγματοληψία

- Είναι μία τεχνική επεξεργασίας σήματος για την ακριβή ανακατασκευή ενός **αραιού** ή **συμπιεστικού** σήματος από ένα μικρό πλήθος **τυχαίων μετρήσεων**

Αραιό σήμα



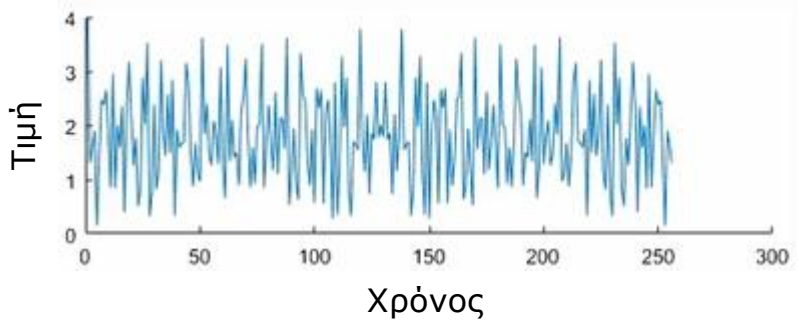
Συμπιεστικό σήμα



Συμπιεστική Δειγματοληψία

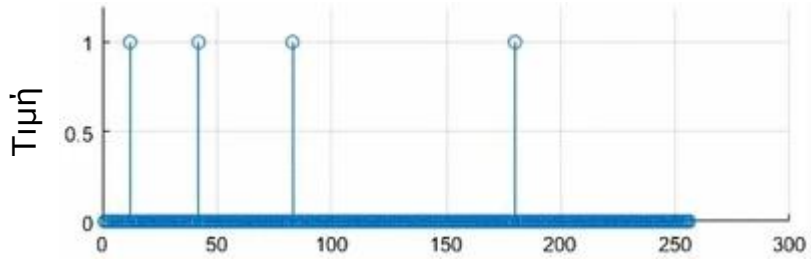
- Ένα σήμα ενδεχομένως να μην είναι από τη φύση του αραιό. Μπορούμε όμως να το αραιοποιήσουμε με ένα κατάλληλο **μετασχηματισμό**

Αρχικό σήμα



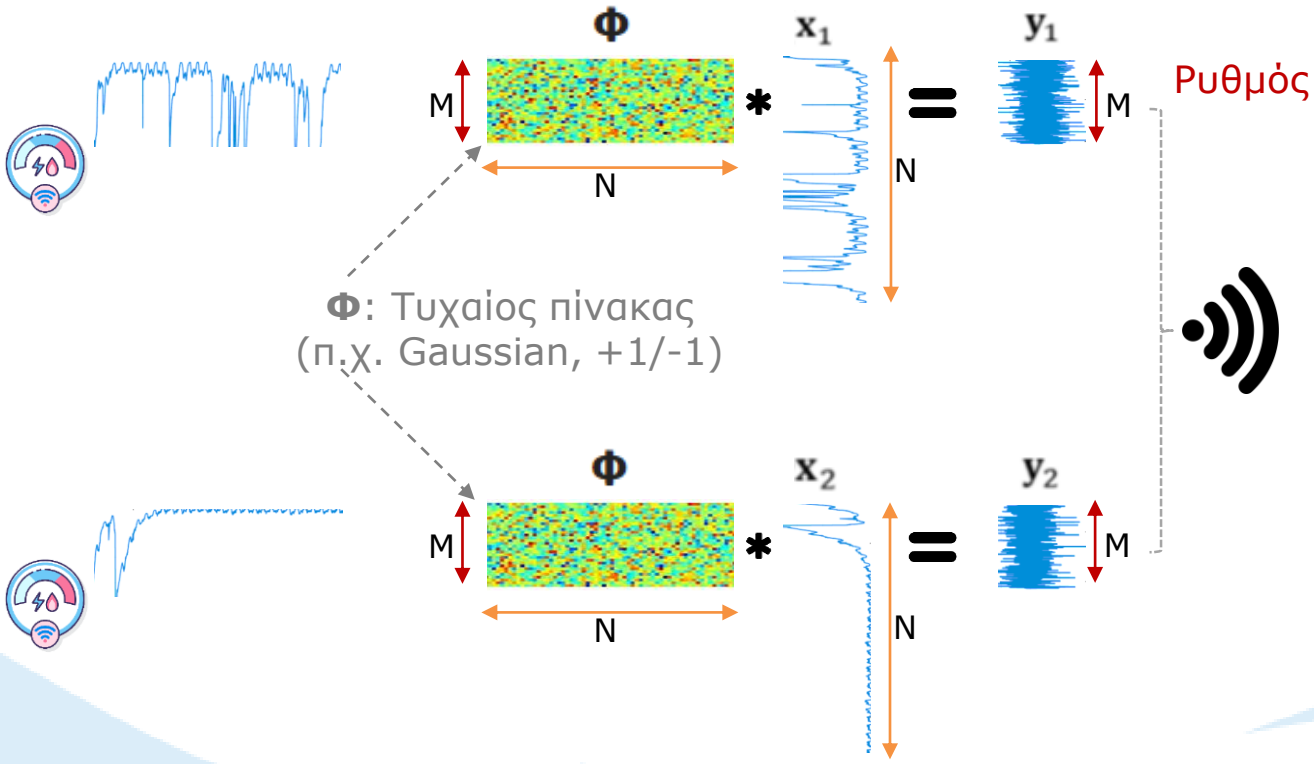
Μετασχηματισμός Fourier

Σήμα στο πεδίο συχνότητας



Συμπιεστική Δειγματοληψία

- Η συμπίεση των δεδομένων πραγματοποιείται πάνω στον αισθητήρα ή τον έξυπνο μετρητή, ενώ η αποσυμπίεση στο κέντρο τηλεελέγχου



Ρυθμός Δειγματοληψίας: $\mathrm{SR} = \frac{M}{N}$

Κέντρο ελέγχου

Ανακατασκευή

$$\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_1$$

$$\text{s.t.} \quad \|y - \Phi(\Psi\alpha)\|_2 < \delta$$

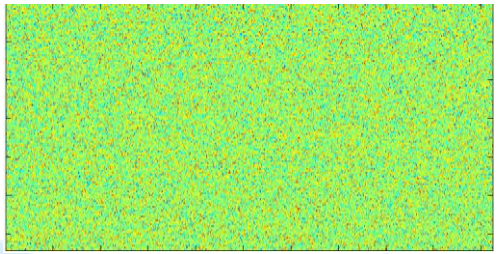
$$\hat{x} = \Psi^{-1}\hat{\alpha}$$

Ψ: Πίνακας μετασχηματισμού
(π.χ. Fourier)

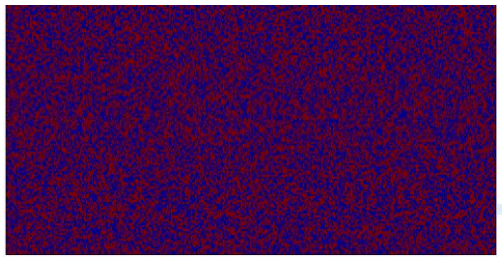
Συμπιεστής

- Ο τυχαίος πίνακας **Φ** πρέπει να πληροί συγκεκριμένες μαθηματικές ιδιότητες, για να εγγυηθεί την ανακατασκευή του αρχικού σήματος με μεγάλη ακρίβεια

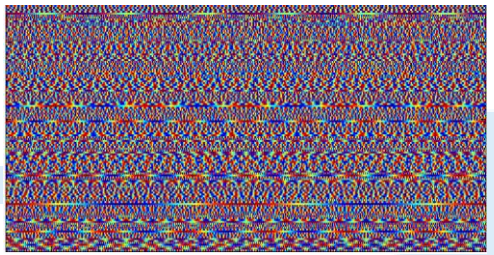
- Παραδείγματα κατάλληλων πινάκων **Φ**: (https://sites.google.com/site/igorcarron2/cs#measurement)
  - Gaussian κατανομή
  - Bernoulli κατανομή (+1/-1)
  - Συντελεστές Fourier σε Μ τυχαίες συχνότητες



Gaussian



Bernoulli



Fourier

Αποσυμπιεστής

- Ο πίνακας μετασχηματισμού **Ψ** πρέπει επίσης να πληροί συγκεκριμένες προϋποθέσεις (π.χ. να είναι ασυνάρτητος με τον **Φ**), για να εγγυηθεί την ανακατασκευή του αρχικού σήματος με μεγάλη ακρίβεια

- Παραδείγματα κατάλληλων πινάκων μετασχηματισμού **Ψ**: (https://sites.google.com/site/igorcarron2/cs#sparse)
  - Short-Time Fourier Transform
  - Wavelet Transform
  - Discrete Cosine Transform

Συμπιεστική Δειγματοληψία στην Πράξη

- **Συμπίεση στα άκρα του δικτύου:** Επιλογή παραμέτρων από το χειριστή του συστήματος

Function: CSEdge(x, seed, PhiType, samplingRatio)

Inputs:

@x: το αρχικό διάνυσμα με όλες τις μετρήσεις των αισθητήρων μήκους N

@seed: ακέραιος που καθορίζει την ακολουθία που παράγει η γεννήτρια ψευδοτυχαίων αριθμών για τη δημιουργία του πίνακα **Φ**

@PhiType: ο τύπος του πίνακα δειγματοληψίας **Φ**

@samplingRatio: το ποσοστό δειγματοληψίας των δεδομένων. Το μήκος του συμπιεσμένου διανύσματος υπολογίζεται ως M = floor(samplingRatio*N)

Output:

@y: το συμπιεσμένο διάνυσμα δεδομένων μήκους M (M≪N)

Συμπιεστική Δειγματοληψία στην Πράξη

- **Αποσυμπίεση στο κέντρο τηλεελέγχου:** Επιλογή παραμέτρων από το χειριστή του συστήματος

Function: CSDecompression(y, seed, PhiType, psi)

Inputs:

@y: το συμπιεσμένο διάνυσμα δεδομένων μήκους M

@seed: ακέραιος που καθορίζει την ακολουθία που παράγει η γεννήτρια ψευδοτυχαίων αριθμών για τη δημιουργία του πίνακα **Φ**

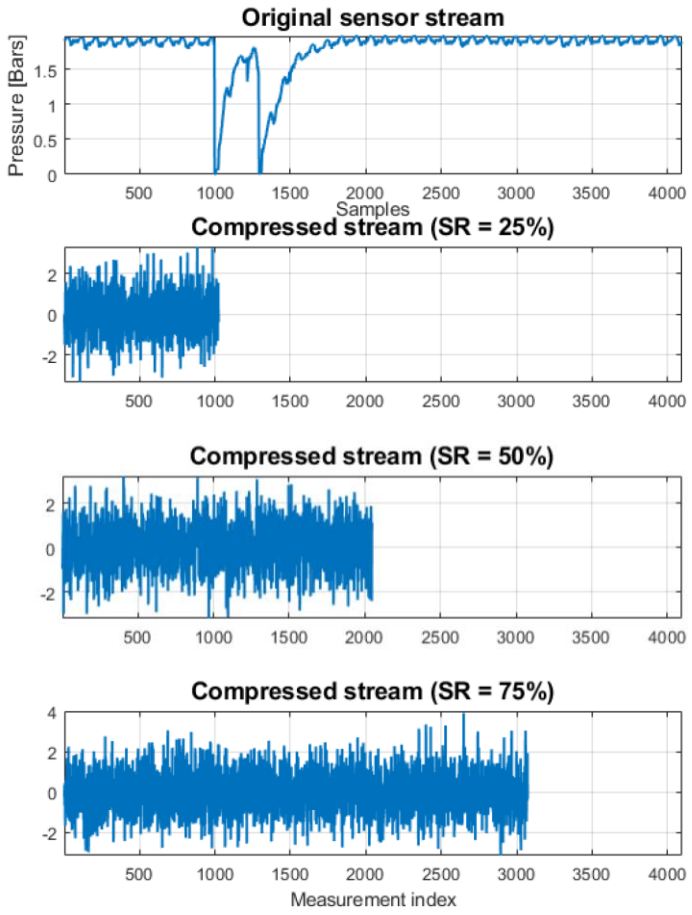
@PhiType: ο τύπος του πίνακα δειγματοληψίας **Φ**

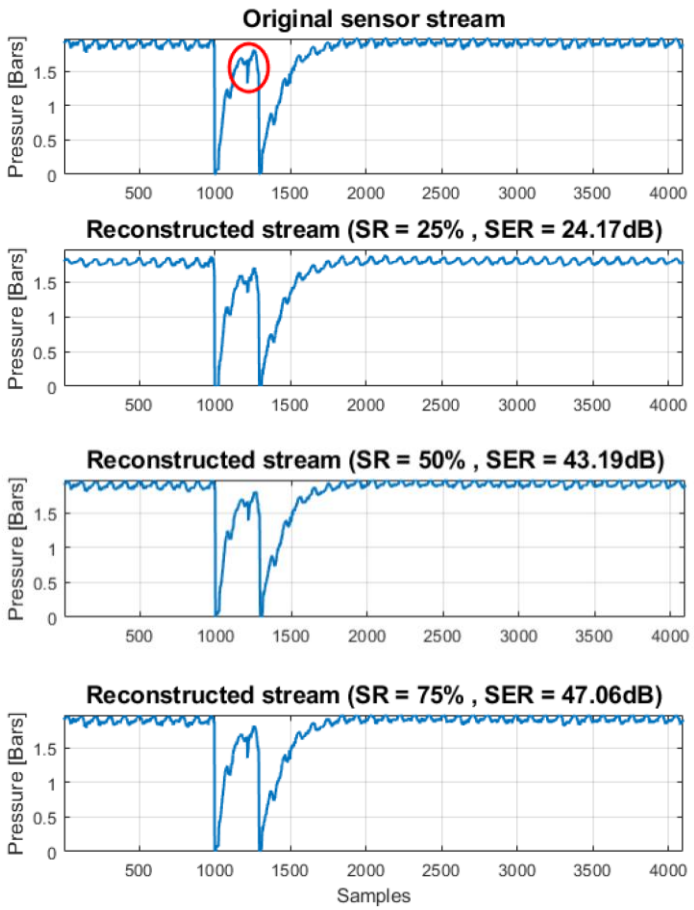
@psi: δομή για τη δημιουργία του τελεστή αραιοποίησης

Output:

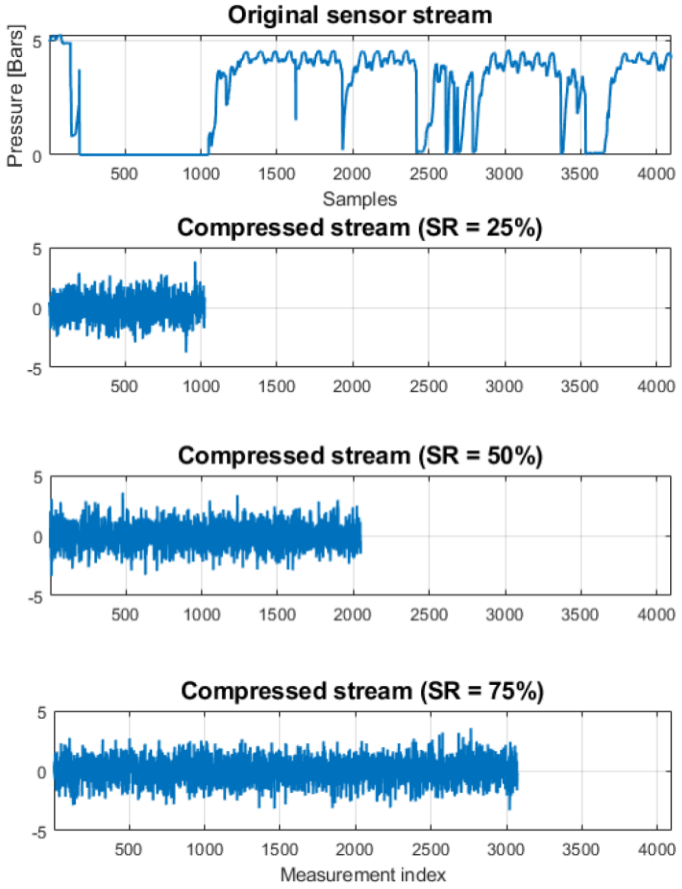
@x: το αποσυμπιεσμένο διάνυσμα δεδομένων μήκους N

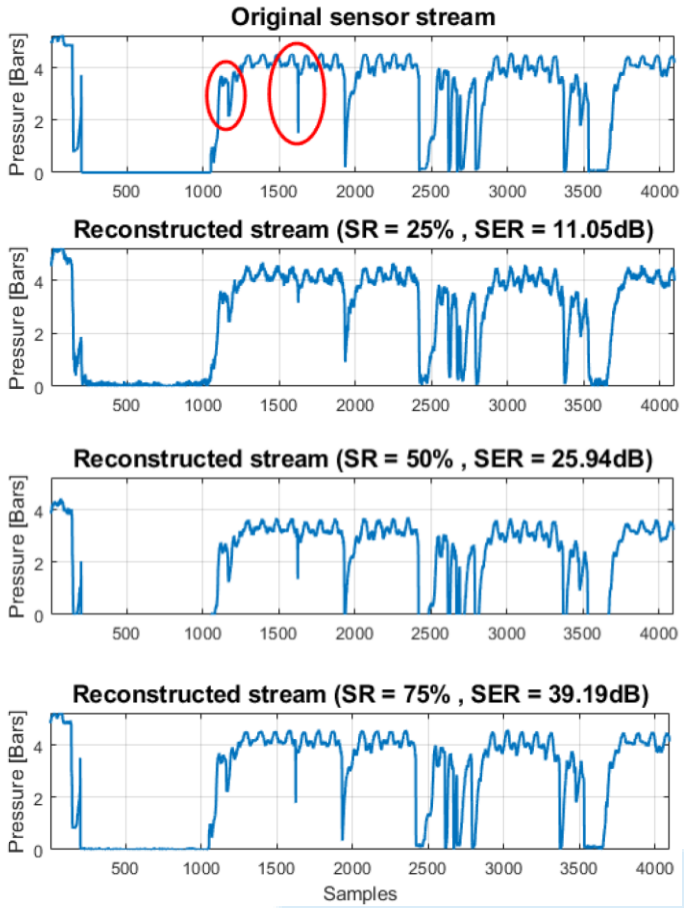
Δεδομένα ΔΕΥΑΜ





Δεδομένα ΔΕΥΑΜ





Interreg
Ελλάδα-Κύπρος
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
SmartWater2020

Συμπιεστική Δειγματοληψία κ' Κρυπτογράφηση

FORTH
INSTITUTE OF COMPUTER SCIENCE

ΔΕΣΜΟΙ
ΑΝΑΠΤΥΞΗΣ

Κρυπτογράφηση Δεδομένων

- Η συμπιεστική δειγματοληψία προσφέρει ένα **εγγενή μηχανισμό** (ασθενούς) **κρυπτογράφησης** λόγω του τυχαίου πίνακα **Φ**
- Δε μπορεί να εγγυηθεί τέλειο απόρρητο, αλλά **υπολογιστικό απόρρητο** των δεδομένων
- Υπολογιστικά ανέφικτο ένας «εισβολέας» να ανακατασκευάσει το αρχικό σήμα χωρίς να γνωρίζει το σύστημα τυχαίας δειγματοληψίας

Κρυπτογράφηση Δεδομένων

- Οι μαθηματικές ιδιότητες του τυχαίου πίνακα **Φ** διασφαλίζουν ότι η ενέργεια των τυχαίων μετρήσεων είναι σχεδόν ίδια με την ενέργεια του αρχικού σήματος

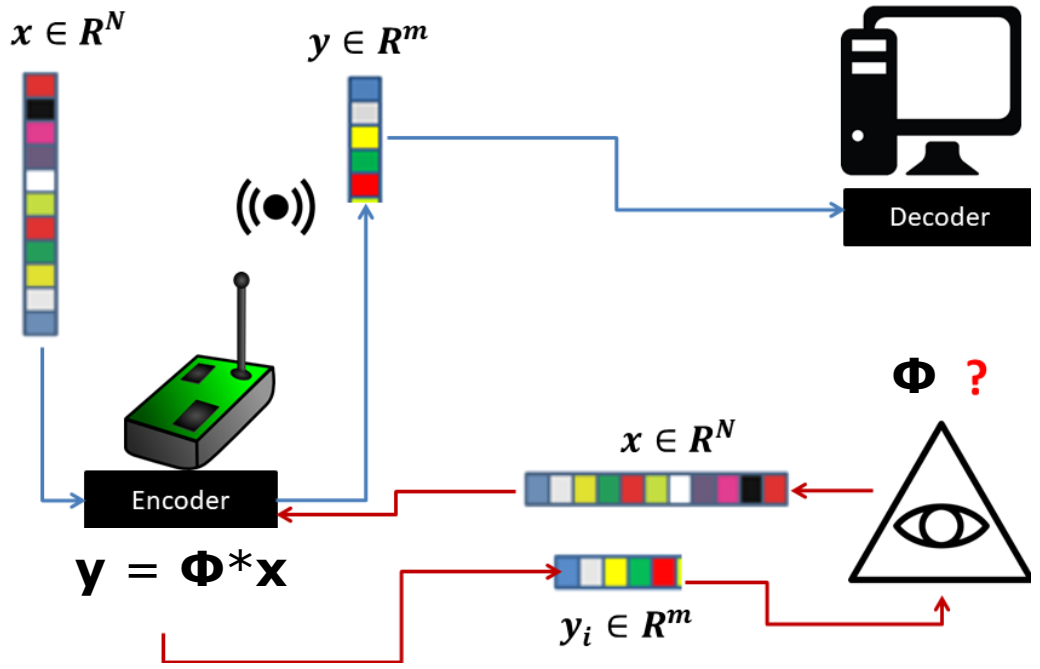
- Οι μετρήσεις που παράγονται με τυχαία δειγματοληψία είναι δυσδιάκριτες εάν τα αντίστοιχα αρχικά σήματα είναι κανονικοποιημένα στην ίδια ενέργεια

- Οι τυχαίοι πίνακες Bernoulli είναι ασφαλείς μόνο για αρκετά μεγάλο N

* Κανονικοποίηση σήματος $s_n = \frac{s}{\sqrt{\frac{\sum_{i=1}^{N} |s_i^2|}{N}}}$

Παράδειγμα Επίθεσης

- Επίθεση επιλεγμένου-απλού κειμένου (CPA)
- Σε μια CPA ο εισβολέας μπορεί (ενδεχομένως προσαρμοστικά) να ζητήσει τα κρυπτογραφημένα μηνύματα αυθαίρετων μηνυμάτων απλού κειμένου
- Αυτό προϋποθέτει αλληλεπίδραση του εισβολέα με το μηχανισμό κρυπτογράφησης



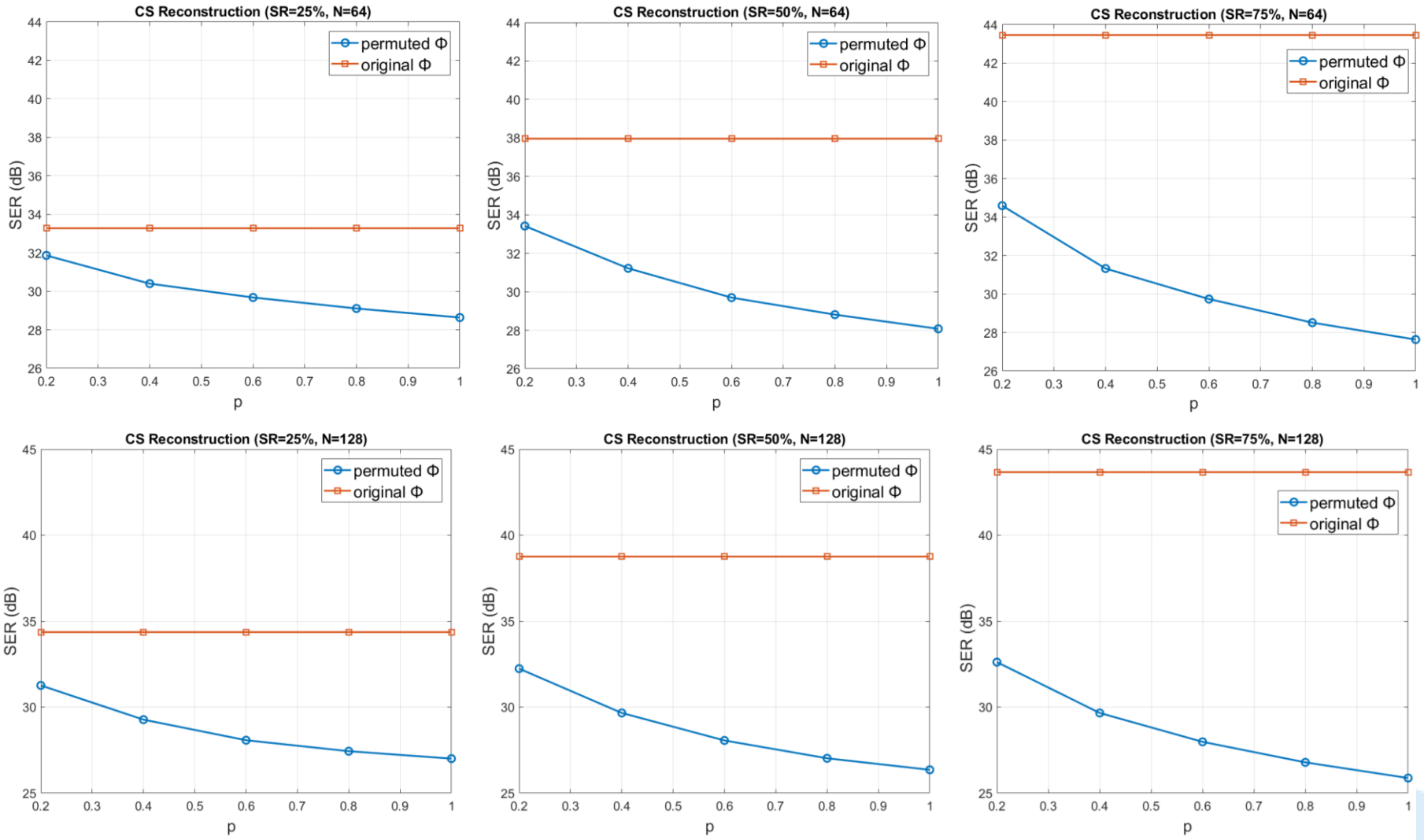
Σχεδιαστικοί Στόχοι Κρυπτογράφησης

- **Αντοχή ενάντια σε επιθέσεις τύπου CPA:** μετατρέπουμε τη ντετερμινιστική συμπεριφορά της συνάρτησης κρυπτογράφησης αφού έχουμε επιλέξει το κλειδί (seed), σε ένα σχήμα πιθανοτικής κρυπτογράφησης
- **Χειρισμός σημάτων με διαφορετική ενέργεια:** τα σήματα πρέπει να κανονικοποιηθούν κατά την κρυπτογράφηση, και η ενέργεια πρέπει να μεταδίδεται εμπιστευτικά μαζί με τις τυχαίες μετρήσεις
- **Ασφάλεια μέσω σχεδιασμού:** είναι σημαντικό να βασιστεί ο τρόπος λειτουργίας σε γνωστά και αξιόπιστα κρυπτογραφικά θεμέλια και μεθόδους κατασκευής κρυπτογραφικών συστημάτων

Δεδομένα ΔΕΥΑΜ

- Ο εισβολέας έχει γνώση του **Φ** ως προς ένα ποσοστό (p%) μεταθέσεων των γραμμών του

$$\mathrm{SER}\left(\mathbf{x},\widehat{\mathbf{x}}\right) = 10\,\log_{10}\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x}-\widehat{\mathbf{x}}\|_2^2}$$



Σύνοψη

Σύνοψη

- Οι σύγχρονες εφαρμογές σε ευφυή δίκτυα ύδρευσης καθιστούν απολύτως αναγκαία τη λειτουργία της συμπίεσης δεδομένων (μείωση κόστους τηλεμετρίας, αύξηση διάρκειας ζωής του δικτύου αισθητήρων)
- Η μέθοδος συμπιεστικής δειγματοληψίας προσφέρει ένα ισχυρό, και εύκολα υλοποιήσιμο, μηχανισμό συμπίεσης των παρατηρούμενων δεδομένων
- Η μέθοδος συμπιεστικής δειγματοληψίας παρέχει ένα εγγενή μηχανισμό κρυπτογράφησης (και αποθορυβοποίησης) των δεδομένων, χωρίς να απαιτείται χρήση πρόσθετου εξοπλισμού ή λογισμικού

FORTH
INSTITUTE OF COMPUTER SCIENCE

ΔΕΣΜΟΙ
ΑΝΑΠΤΥΞΗΣ

Συμπιεστική Δειγματοληψία στην Πράξη

- **Αποσυμπίεση στο κέντρο τηλεελέγχου:** Επιλογή παραμέτρων από το χειριστή του συστήματος

Function: CSDecompression(y, seed, PhiType, psi)

Inputs:

…

@psi: δομή για τη δημιουργία του τελεστή αραιοποίησης, η οποία περιλαμβάνει

@psi.gMax: ακέραιος που καθορίζει τη μέγιστη κλίμακα του STFT (max scale = logN + gMax)

@psi.gLevels: ακέραιος που καθορίζει την ελάχιστη κλίμακα του STFT (min scale = logN - gLevels)

@psi.tRedundancy: ακέραιος που καθορίζει τη μετατόπιση του STFT παραθύρου στο χρόνο

@psi.fRedundancy: ακέραιος που καθορίζει τη μετατόπιση του STFT παραθύρου στη συχνότητα

@psi.gWindow: τύπος παραθύρου (η παρούσα υλοποίηση υποστηρίζει την επιλογή 'iterate sine')