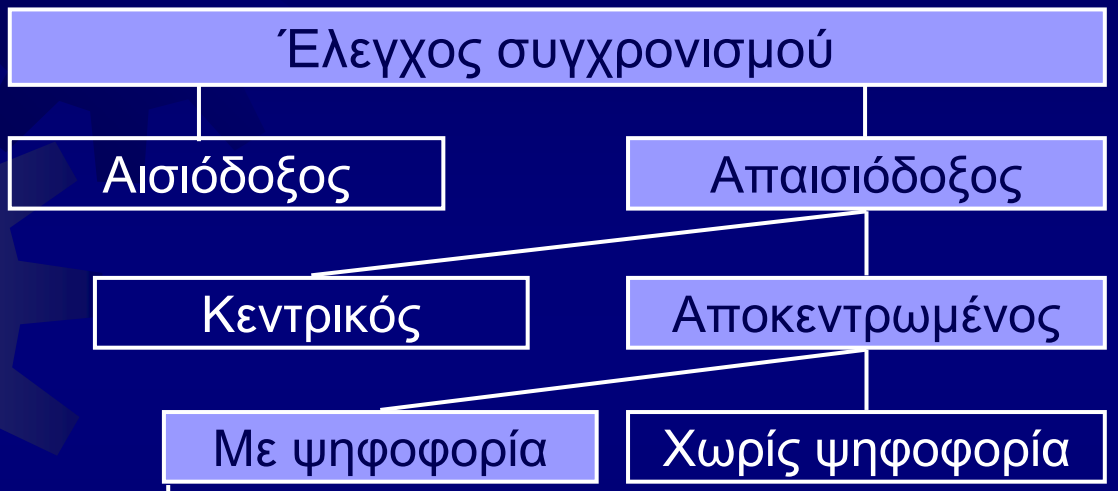
Αποκεντρωμένος έλεγχος με ψηφοφορία



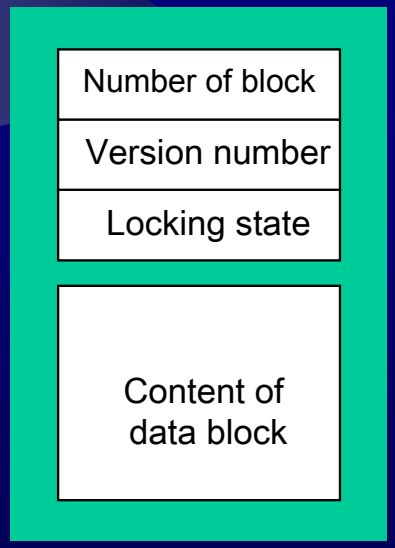
- ***Συναίνεση πλειοψηφίας***
- ***Σταθμισμένη ψηφοφορία***
- ***Εγγραφή από όλους – ανάγνωση από καθέναν***
- ***Ψηφοφορία με μάρτυρες***
- ***Κ.α.***

Μηχανισμοί ψηφοφορίας

- ανήκουν στον απαισιόδοξο έλεγχο συγχρονισμού
- έχουν υψηλές υπολογιστικές απαιτήσεις για την υλοποίηση των πρωτοκόλλων ψηφοφορίας
  - αντισταθμίζονται από την υψηλή διαθεσιμότητα πρόσβασης
- ιδιαίτερα χρήσιμοι στις περιπτώσεις μεγάλου εύρους πρόσβασης, δηλαδή, τμημάτων δεδομένων μεγέθους τουλάχιστον μερικών kilobytes
- προέλευση των μηχανισμών ψηφοφορίας εντοπίζεται στα κατανεμημένα συστήματα αρχειοθέτησης και βάσεων δεδομένων
  - Π.χ. VAX clusters
- Εφαρμογή σε κατανεμημένα με αντίγραφα μοντέλα

Data blocks

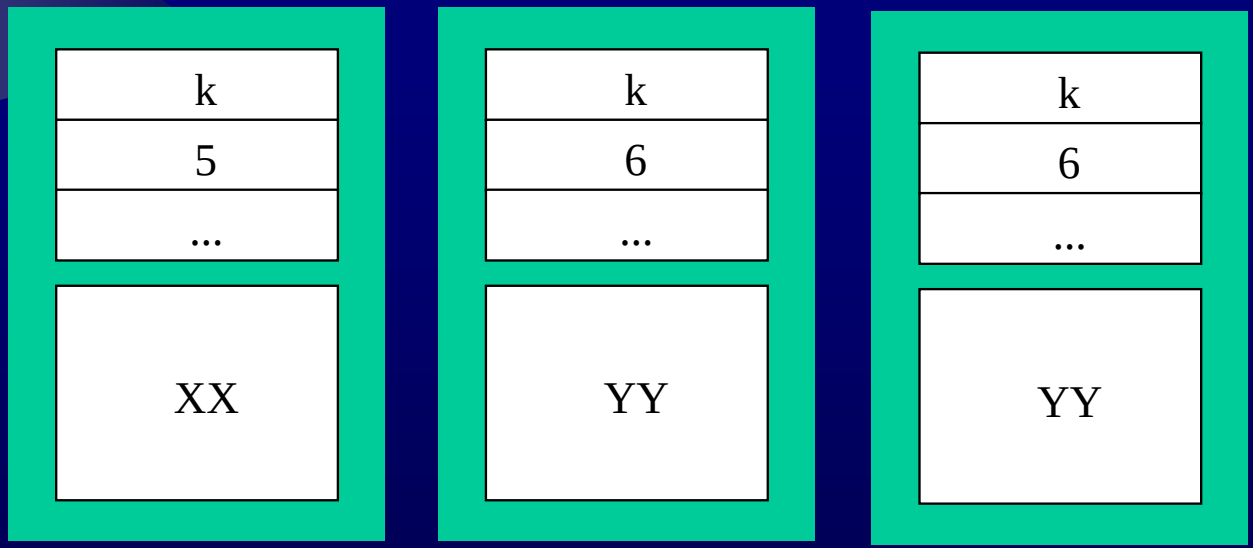
- Υποθέτουμε οτι το κλείδωμα γίνεται στο επίπεδο του data block
- Διακρίνουμε σε κλείδωμα ανάγνωσης και κλείδωμα εγγραφής
- Κάθε data block έχει αριθμό αντιγράφου (version number)
  - Στην απλή περίπτωση είναι ακέραιος αριθμός
  - Μπορεί να είναι timestamp
- Το locking state δεν μας ενδιαφέρει στο παράδειγμα



data block

Παράδειγμα χρήσης version numbers

- Ο χρήστης μπορεί να συμπεράνει, συγκρίνοντας τα version numbers, οτι το YY είναι το πλέον πρόσφατο data block
- Υπόθεση: Data black είναι συνεπές όταν η πλειοψηφία των κόμβων κατέχει ενημερωμένο αντίγραφο
- Ο χρήστης μπορεί να συμπεράνει οτι το YY είναι το πλέον πρόσφατο data block ακόμη και εάν μόνο 2 από τους 3 κόμβους του έχουν απαντήσει



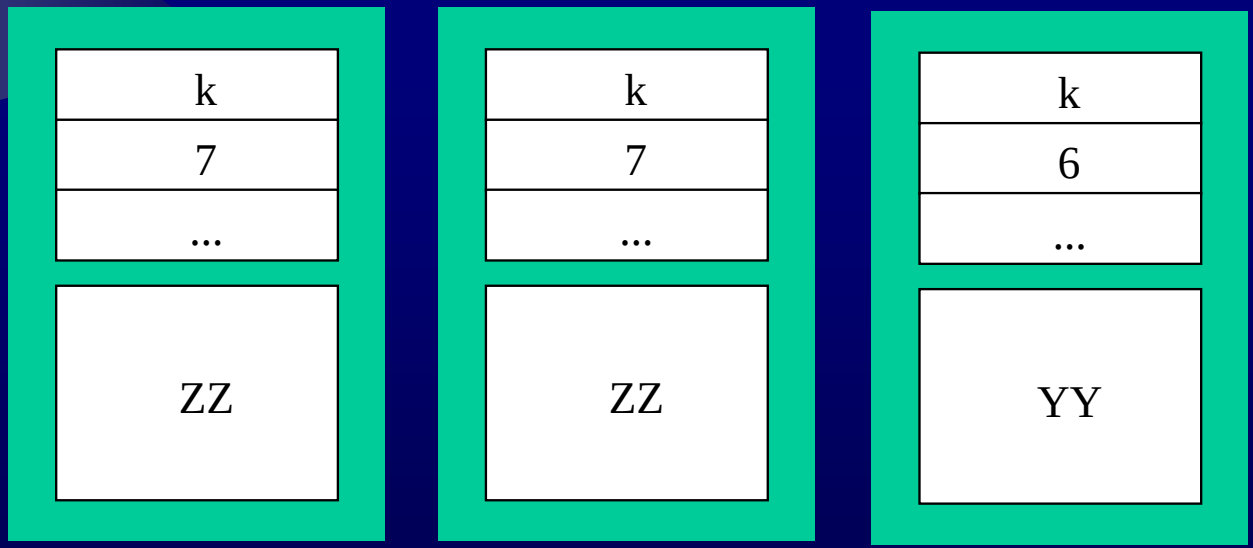
Τμήμα δεδομένων k στον κόμβο n_1

Τμήμα δεδομένων k στον κόμβο n_2

Τμήμα δεδομένων k στον κόμβο n_3

Παράδειγμα (συνέχεια)

- Χρήστης ζητεί πρόσβαση εγγραφής στο data block
- Πλειοψηφία κόμβων του «δείχνουν» το Ver. No. 6
- Χρήστης ενημερώνει data block και αλλάζει τον αριθμό αντιγράφου σε No. 7
- Κατά την ενημέρωση ο n3 crashes
- Υπάρχει συνέπεια στο data block;
  - Ναι, διότι η πλειοψηφία των κόμβων έχουν ενημερωμένο αντίγραφο
- Τι γίνεται εαν ο n1 ή n2 crashes στη συνέχεια;
  - Πάλι κάποιος κόμβος στη πλειοψηφία που θα του απαντήσει (n1 ή n2) έχει ενημερωμένο αντίγραφο



Τμήμα δεδομένων k στον κόμβο n_1

Τμήμα δεδομένων k στον κόμβο n_2

Τμήμα δεδομένων k στον κόμβο n_3

Ορισμοί (1/2)

- **σε ένα κατανεμημένο σύστημα, πρέπει να διασφαλίσουμε οτι δύο αντίγραφα του ίδιου block δεδομένων δεν εγγράφονται ταυτόχρονα ή δεν διαβάζονται και εγγράφονται ταυτόχρονα**
  - **στρατηγική πολλαπλών-αναγνωστών-μοναδικού-συγγραφέα**
- **ψηφοφορία (votum)** για ένα αίτημα πρόσβασης σε κάποιο λογικό τμήμα δεδομένων ορίζεται ως:
- το άθροισμα των ψήφων από ένα σύνολο κόμβων
- πρόκειται για τους κόμβους που:
  - είναι προσβάσιμοι μέσω του δικτύου
  - βρίσκονται σε λειτουργία τη στιγμή της πρόσβασης
  - οι οποίοι δεν έχουν κλειδώσει αντίγραφα των τμημάτων φυσικών δεδομένων
  - δεν έχουν καμία άλλη αντίρρηση που θα απαγόρευε την πρόσβαση
    - π.χ. άρνηση της πρόσβασης λόγω ανεπαρκών δικαιωμάτων πρόσβασης

Ορισμοί (2/2)

- ψηφοφορία θεωρείται επιτυχής εάν το άθροισμα των ψήφων από τους κόμβους που ψήφισαν θετικά για το αίτημα πρόσβασης είναι ίσο ή μεγαλύτερο ενός κατώτερου ορίου QU
- κατώτερο όριο ονομάζεται **απαρτία (quorum)**
- κατώτερο όριο QU που θα επιλεχθεί, πρέπει να υποστηρίζει:
  - τη στρατηγική πολλαπλών-αναγνωστών-μοναδικού-συγγραφέα
  - να διασφαλίζει ότι μεταξύ των κόμβων που έχουν επιτρέψει την πρόσβαση υπάρχει τουλάχιστον ένας κόμβος με το πλέον πρόσφατο φυσικό αντίγραφο του επιθυμητού τμήματος δεδομένων

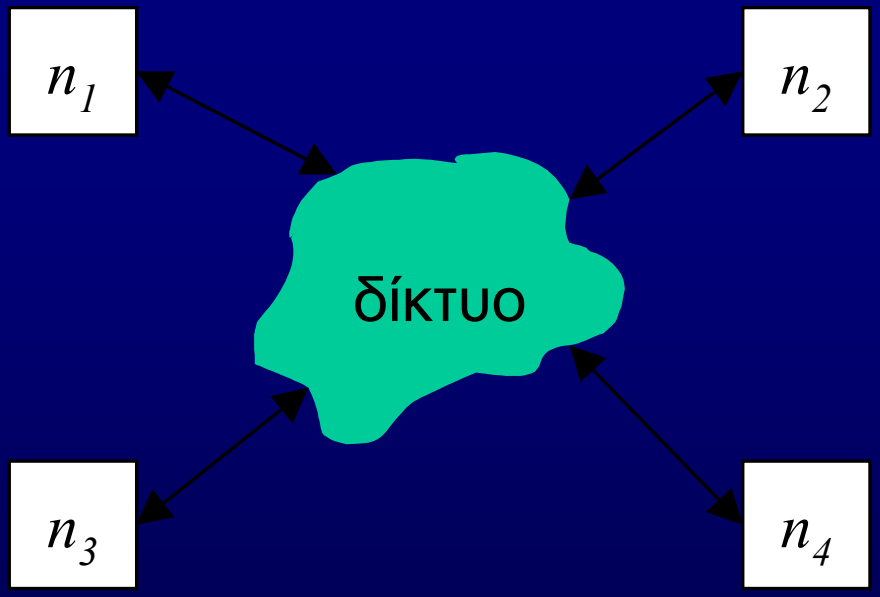
Συναίνεση πλειοψηφίας (majority consensus-MC) (1/2)

- Κάθε κόμβος που διαχειρίζεται ένα αντίγραφο έχει δικαίωμα ψήφου. Στο μηχανισμό MC μια ψηφοφορία θεωρείται επιτυχής εάν τουλάχιστον η πλειοψηφία των κόμβων με δικαίωμα ψήφου, ψήφισαν θετικά στο αίτημα πρόσβασης
- Το κατώτερο όριο QU, η απαρτία, υπολογίζεται ως εξής:

$$QU = \begin{cases} \frac{n}{2} + 1 & n\ \acute{\alpha}\rho\tau\iota\omicron\varsigma \\ \frac{n+1}{2} & n\ \pi\varepsilon\rho\iota\tau\tau\acute{\omicron}\varsigma \end{cases}$$

Συναίνεση πλειοψηφίας (majority consensus-MC) (2/2)

- Παράδειγμα (πλήρως συνδεδεμένο δίκτυο)
- Στο πλήρως συνδεδεμένο δίκτυο του σχήματος οι κόμβοι n1,...,n4 κατέχουν από ένα αντίγραφο όλων των τμημάτων δεδομένων
- Είναι προφανές ότι για να θεωρηθεί επιτυχής μία ψηφοφορία απαιτούνται τουλάχιστον τρεις ψήφοι (δηλαδή QU=3)



Σταθμισμένη ψηφοφορία (weighted voting) (1/3)

- Δίνει σε κάθε κόμβο όχι μόνο δικαίωμα ψήφου αλλά και συντελεστή βαρύτητας στην ψήφο του
- Ο συντελεστής βαρύτητας μπορεί να είναι διαφορετικός για κάθε κόμβο με δικαίωμα ψήφου
- Με τη σταθμισμένη ψηφοφορία, μπορούμε να ευνοήσουμε ορισμένους κόμβους ή ακόμη και να εγκαταστήσουμε κόμβους με μηδενική ψήφο
- Έστω w(n) ο συντελεστής βαρύτητας της ψήφου του κόμβου n, με:

$$w(n) \in \{0,1,2,...\}, \quad \forall\, n \in N$$

- Το άθροισμα όλων των συντελεστών βαρύτητας είναι W, το οποίο ορίζεται ως εξής:

$$W = \sum_{n \in N} w(n)$$

Σταθμισμένη ψηφοφορία (weighted voting) (2/3)

- Ο μηχανισμός συναίνεσης πλειοψηφίας (MC) δεν κάνει τη διάκριση μεταξύ ανάγνωσης και εγγραφής στα προνόμια πρόσβασης
- Με τη χρήση του μηχανισμού σταθμισμένης ψηφοφορίας μπορεί εύκολα να τεθεί ένα διαφορετικό όριο πρόσβασης όσον αφορά στην εγγραφή της πληροφορίας, το οποίο καλείται απαρτία εγγραφής (write quorum, QUw)
  - Η πρόσβαση για εγγραφή της πληροφορίας επιτρέπεται όταν η ψηφοφορία έχει θετικό αποτέλεσμα, δηλαδή, όταν το άθροισμα της βαρύτητας των ψήφων του συνόλου των κόμβων που ψήφισαν για το αίτημα πρόσβασης είναι ίσο ή μεγαλύτερο από την απαρτία εγγραφής
- Στην περίπτωση της πρόσβασης για ανάγνωση της πληροφορίας, το αποτέλεσμα της ψηφοφορίας θεωρείται θετικό όταν έχει καλυφθεί η απαρτία ανάγνωσης (read quorum, QUr)

Σταθμισμένη ψηφοφορία (weighted voting) (3/3)

- Τα όρια QUw και QUr μπορεί να είναι διαφορετικά αλλά πρέπει να πληρούν τις ακόλουθες δύο συνθήκες:
  - Για να εξασφαλιστεί οτι δεν μπορούν δύο ενέργειες εγγραφής να πραγματοποιηθούν ταυτόχρονα στο ίδιο data block $2 \times QUw > W$
  - Για να εξασφαλίζεται ανάγνωση από ενημερωμένο αντίγραφο $QUr + QUw > W$
    - ensures that a data item is not read and written by two transactions concurrently
    - read quorums always intersect with write quorums. This will ensure that read results always reflect the result of the most recent write (because the read quorum will include at least one replica that was involved in the most recent write).
- Και οι δύο προϋποθέσεις υποστηρίζουν τη στρατηγική πολλαπλών-αναγνωστών-μοναδικού-συγγραφέα και εγγυούνται τη συνέπεια μεταξύ των αντιγράφων των τμημάτων των δεδομένων
- Στην περίπτωση που η απαρτία ανάγνωσης ισούται με την απαρτία εγγραφής, τότε:

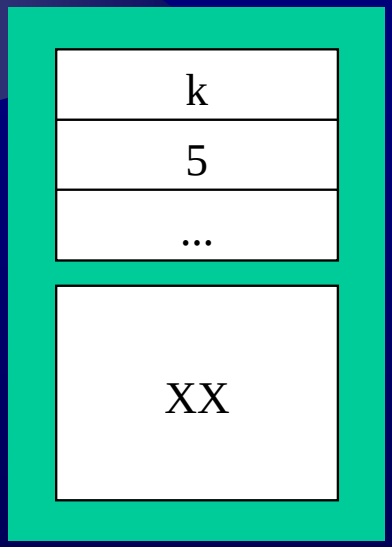
$$QUr = QUw = \begin{cases} \frac{W}{2} + 1 & , W \ \acute{\alpha}\rho\tau\iota\circ\varsigma \\ \frac{W+1}{2} & , W \ \pi\varepsilon\rho\iota\tau\acute{\circ}\varsigma \end{cases}$$

Ψηφοφορία με μάρτυρες (voting with witnesses) (1/5)

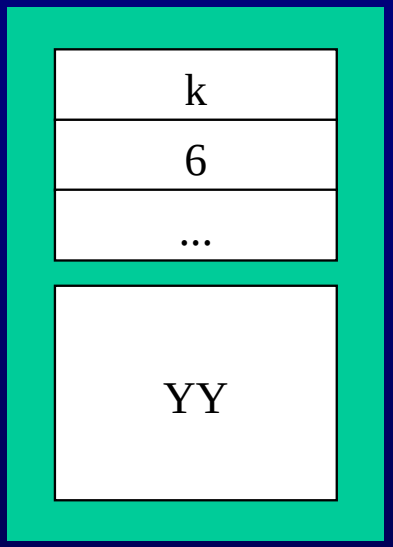
- Μάρτυρας είναι ένας κόμβος με δικαίωμα ψήφου ο οποίος δεν επεξεργάζεται το αντίγραφο του τμήματος δεδομένων για το οποίο ψηφίζει
  - απλά γνωρίζει την ελάχιστη πληροφορία που απαιτείται για τη διαδικασία της ψηφοφορίας
    - την κατάσταση κλειδώματος
    - την έκδοση του τμήματος δεδομένων
    - το συντελεστή βαρύτητας της ψήφου του
- ένας κόμβος-μάρτυρας απαιτεί λιγότερο αποθηκευτικό χώρο από ένα κόμβο που κατέχει ένα πλήρες αντίγραφο
- Καθώς ο μάρτυρας δεν κατέχει το 'περιεχόμενο' των δεδομένων δεν μπορεί να υπάρξει πραγματική πρόσβαση στα δεδομένα από τον κόμβο αυτό. Ποιο είναι επομένως το πλεονέκτημα που προσφέρει;
  - Όσον αφορά στη διαθεσιμότητα, ένας μάρτυρας μπορεί να επιφέρει σχεδόν τις ίδιες βελτιώσεις με τον κόμβο που κατέχει το πλήρες αντίγραφο, όπως φαίνεται από το ακόλουθο παράδειγμα

Ψηφοφορία με μάρτυρες (voting with witnesses) (2/5)

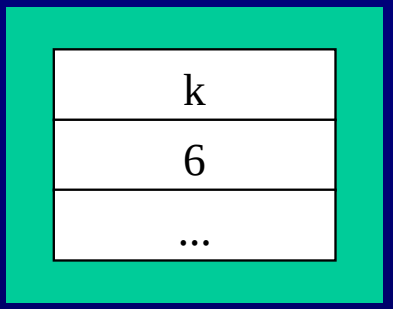
- Παράδειγμα:
  - κόμβοι n1 και n2 κατέχουν πλήρη αντίγραφα του τμήματος δεδομένων k
  - κόμβος n1 κατέχει την έκδοση 5 των δεδομένων με περιεχόμενο 'XX'
  - κόμβος n2 κατέχει την έκδοση 6 του ίδιου τμήματος δεδομένων με περιεχόμενο 'YY'
  - κόμβος n3 είναι μάρτυρας
    - γνωρίζει μόνο ότι η τελευταία έκδοση του τμήματος δεδομένων k είναι η 6
    - Δεν γνωρίζει απολύτως τίποτε για το περιεχόμενο των δεδομένων



Τμήμα δεδομένων k στον κόμβο n_1



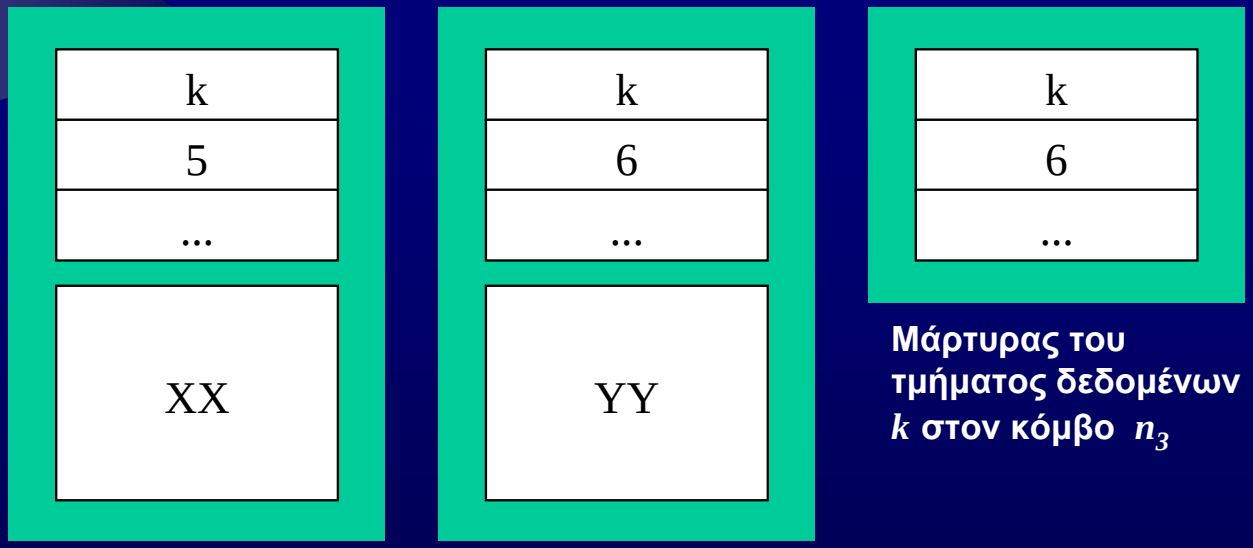
Τμήμα δεδομένων k στον κόμβο n_2



Μάρτυρας του τμήματος δεδομένων k στον κόμβο n_3

Ψηφοφορία με μάρτυρες (voting with witnesses) (3/5)

- Η ψηφοφορία είναι επιτυχής εάν:
  - και οι τρεις κόμβοι ή
  - ο μάρτυρας και ο κόμβος n2 ή
  - οι κόμβοι n1 και n2
- ψηφίσουν για το αίτημα πρόσβασης
- Το πλέον πρόσφατο περιεχόμενο των δεδομένων 'YY' μπορεί να προσδιοριστεί από τον αριθμό έκδοσης. Στην περίπτωση αιτήματος ανάγνωσης, η ανάγνωση θα γίνει προφανώς από το αντίγραφο του κόμβου n2



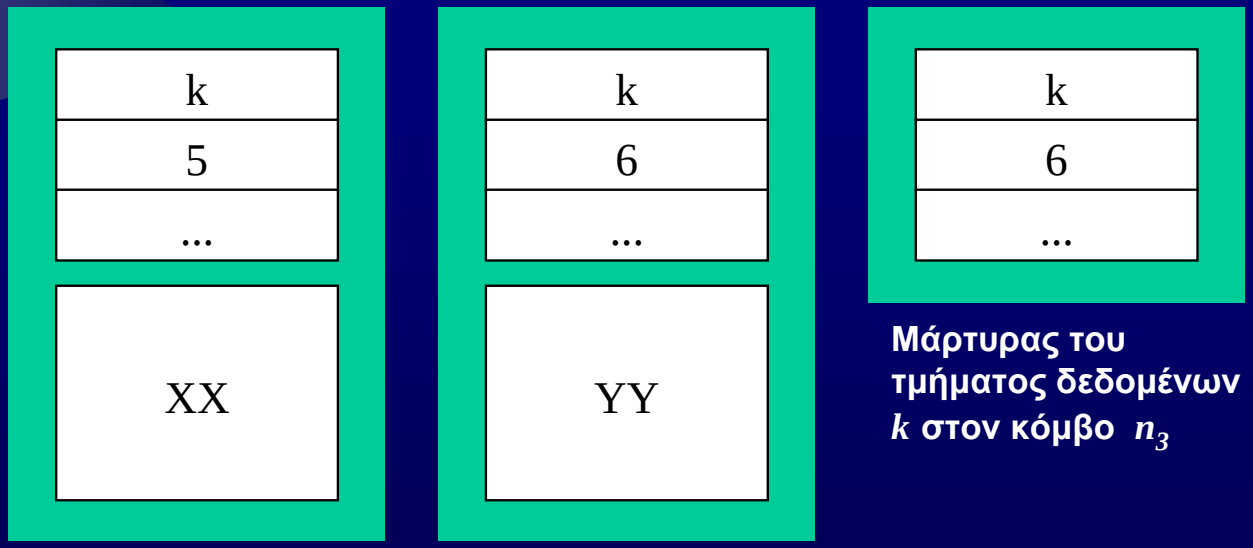
Τμήμα δεδομένων k στον κόμβο n_1

Τμήμα δεδομένων k στον κόμβο n_2

Μάρτυρας του τμήματος δεδομένων k στον κόμβο n_3

Ψηφοφορία με μάρτυρες (voting with witnesses) (4/5)

- Έστω ότι μόνο ο κόμβος n1 και ο μάρτυρας ψήφισαν για το αίτημα πρόσβασης για ανάγνωση (έστω ότι ο κόμβος n2 είναι μη διαθέσιμος)
- Παρότι η ψηφοφορία μπορεί να θεωρηθεί επιτυχής (το όριο απαρτίας ανάγνωσης έχει καλυφθεί), η θέση που αιτείται την πρόσβαση μπορεί μόνο να αντιληφθεί ότι η έκδοση 5 των δεδομένων είναι παλαιά αλλά δεν είναι δυνατή η πρόσβαση στο περιεχόμενο των δεδομένων



Τμήμα δεδομένων k στον κόμβο n_1

Τμήμα δεδομένων k στον κόμβο n_2

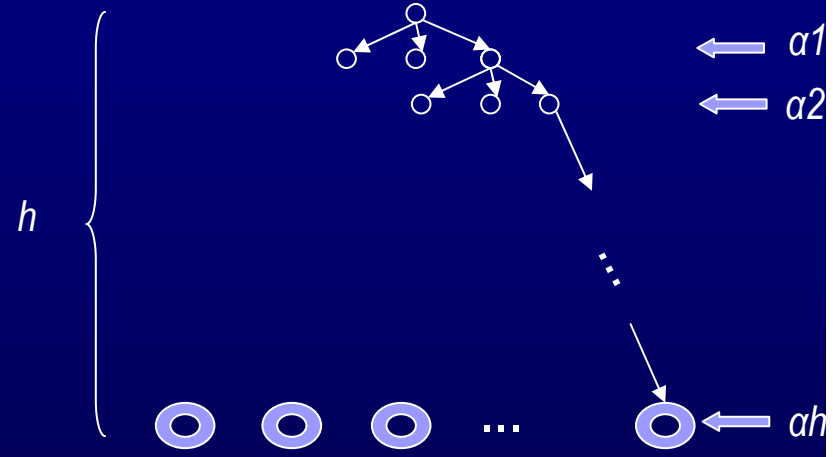
Μάρτυρας του τμήματος δεδομένων k στον κόμβο n_3

Ψηφοφορία με μάρτυρες (voting with witnesses) (5/5)

- Με άλλα λόγια, η κάλυψη της απαρτίας ανάγνωσης QUr επαρκεί για να υποστηρίξει τη στρατηγική πολλαπλών-αναγνωστών-μοναδικού-συγγραφέα και να εγγυηθεί ότι μεταξύ των κόμβων που ψηφίσαν για το αίτημα πρόσβασης υπάρχει τουλάχιστον ένας κόμβος που κατέχει το πλέον πρόσφατο αριθμό έκδοσης του φυσικού αντιγράφου του επιθυμητού λογικού τμήματος δεδομένων
- Η κάλυψη της απαρτίας ανάγνωσης δεν επαρκεί για να εγγυηθεί ότι μεταξύ των κόμβων που ψηφίσαν για το αίτημα πρόσβασης υπάρχει τουλάχιστον ένας κόμβος που κατέχει το πλέον πρόσφατο περιεχόμενο του επιθυμητού τμήματος δεδομένων
- Αυτό που απαιτείται, εκτός από την κάλυψη της απαρτίας, είναι ένα όριο που να εγγυάται την επιτυχία της πρόσβασης, εφόσον η ψηφοφορία είναι επιτυχής

Ιεραρχική ψηφοφορία (hierarchical voting) (1/6)

- Το πεδίο εφαρμογής των μηχανισμών ιεραρχικής ψηφοφορίας είναι τα δίκτυα με ιδιαίτερα πολλούς κόμβους
  - Έστω ότι *n=100*
  - *Ε*φαρμόζοντας το μηχανισμό συναίνεσης πλειοψηφίας, για να είναι μια ψηφοφορία επιτυχής απαιτούνται 51 ψήφοι
  - Το γεγονός αυτό οδηγεί σε μεγάλο επικοινωνιακό φόρτο κατά τη διαδικασία ψηφοφορίας
- Οι κόμβοι που κατέχουν ένα αντίγραφο οργανώνονται λογικά σε μια δομή ανεστραμμένου δένδρου ύψους *h*
  - Η ρίζα του δένδρου βρίσκεται σε ύψος 0
  - Όλοι οι κόμβοι με αντίγραφα κατέχουν τις θέσεις των φύλλων (στο ύψος *h*)
  - Επιπλέον, έστω ότι από τη ρίζα του δένδρου ξεκινούν α_1 διακλαδώσεις και έστω ότι κάθε μία από αυτές έχει με τη σειρά της α_2 υπο-διακλαδώσεις, κλπ.
  - Δηλ., έχουμε α_h φύλλα για όλες της υπο-διακλαδώσεις σε ύψος *h-1*



Ιεραρχική ψηφοφορία (hierarchical voting) (2/6)

- Μια ψηφοφορία για πρόσβαση για ανάγνωση (ή εγγραφή) πληροφορίας θεωρείται επιτυχής σε ύψος i εάν τουλάχιστον QU^i κόμβοι με δικαίωμα ψήφου ψηφίσουν θετικά για τη συγκεκριμένο αίτημα
- Το συνολικό QU^{total} *δίνεται από τον τύπο:*

$$QU^{total} = \prod_{i=1}^{h} QU^i \quad \forall\, i = 1, \ldots, h$$

Ιεραρχική ψηφοφορία (hierarchical voting) (3/6)

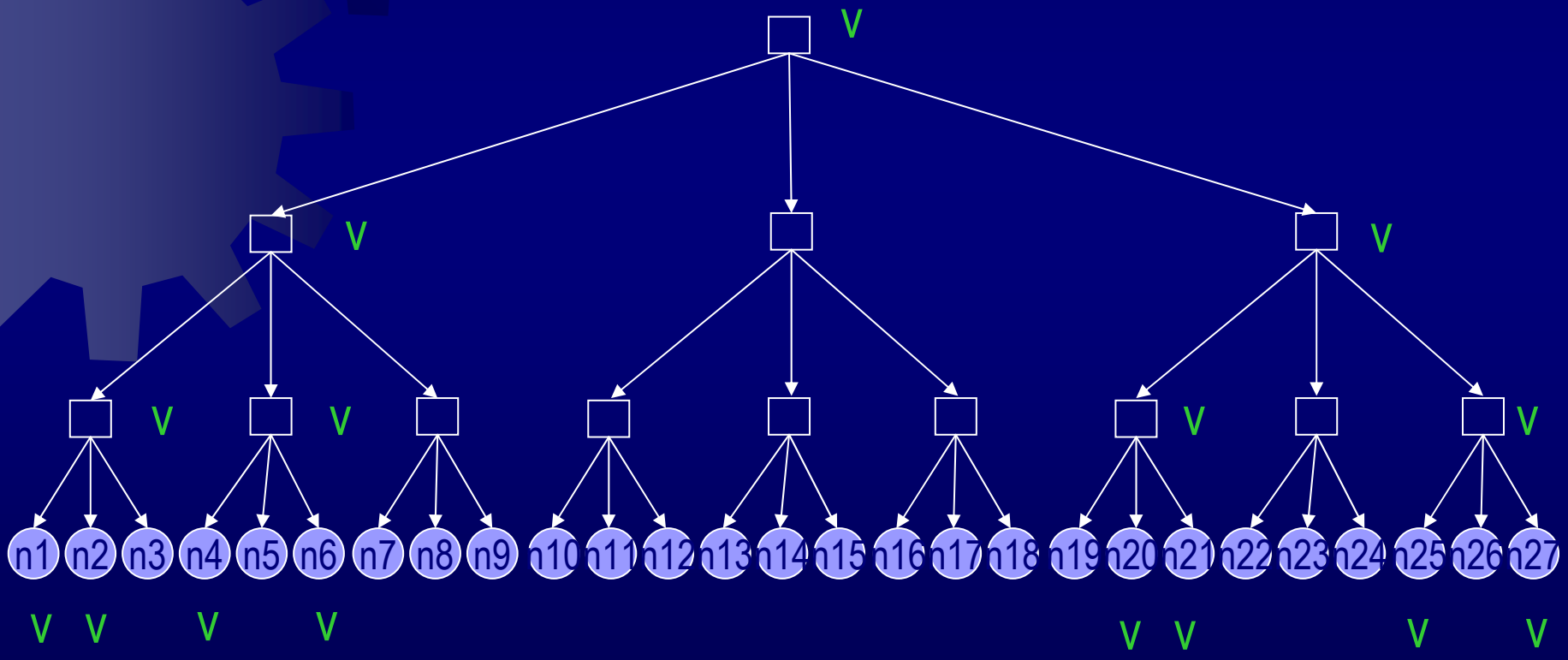
Παράδειγμα

- Έστω αριθμός αντιγράφων n=27. Οι κόμβοι με αντίγραφα (και επομένως δικαίωμα ψήφου) μπορούν να οργανωθούν σε δένδρο ύψους 3 όπου $\alpha_1=\alpha_2=\alpha_3=3$
- Εάν δώσουμε σε κάθε κόμβο με αντίγραφο βαρύτητα ψήφου ίση με 1 και εφαρμόσουμε το μηχανισμό συναίνεσης πλειοψηφίας, τότε για μια επιτυχή ψηφοφορία απαιτούνται 14 ψήφοι
- Με εφαρμογή του μηχανισμού ιεραρχικής ψηφοφορίας η κατάσταση βελτιώνεται αναφορικά με τον ελάχιστο αριθμό των κόμβων που απαιτείται για μια επιτυχή ψηφοφορία
- Σε αυτή την περίπτωση για να είναι πετυχημένη μια ψηφοφορία για πρόσβαση για ανάγνωση πληροφορίας απαιτούνται μόνο 8 ψήφοι
  - Εφόσον α_1=3, στο ύψος 1 το QU^1=2
  - Σε κάθε μία από τις υπο-διακλαδώσεις στο ύψος 1, το QU^2=2 διότι α_2=3
  - Το ίδιο συμβαίνει και στο ύψος 2
  - Επομένως, το συνολικό όριο QU^{total} είναι $2\times2\times2=8$

Ιεραρχική ψηφοφορία (hierarchical voting) (4/6)

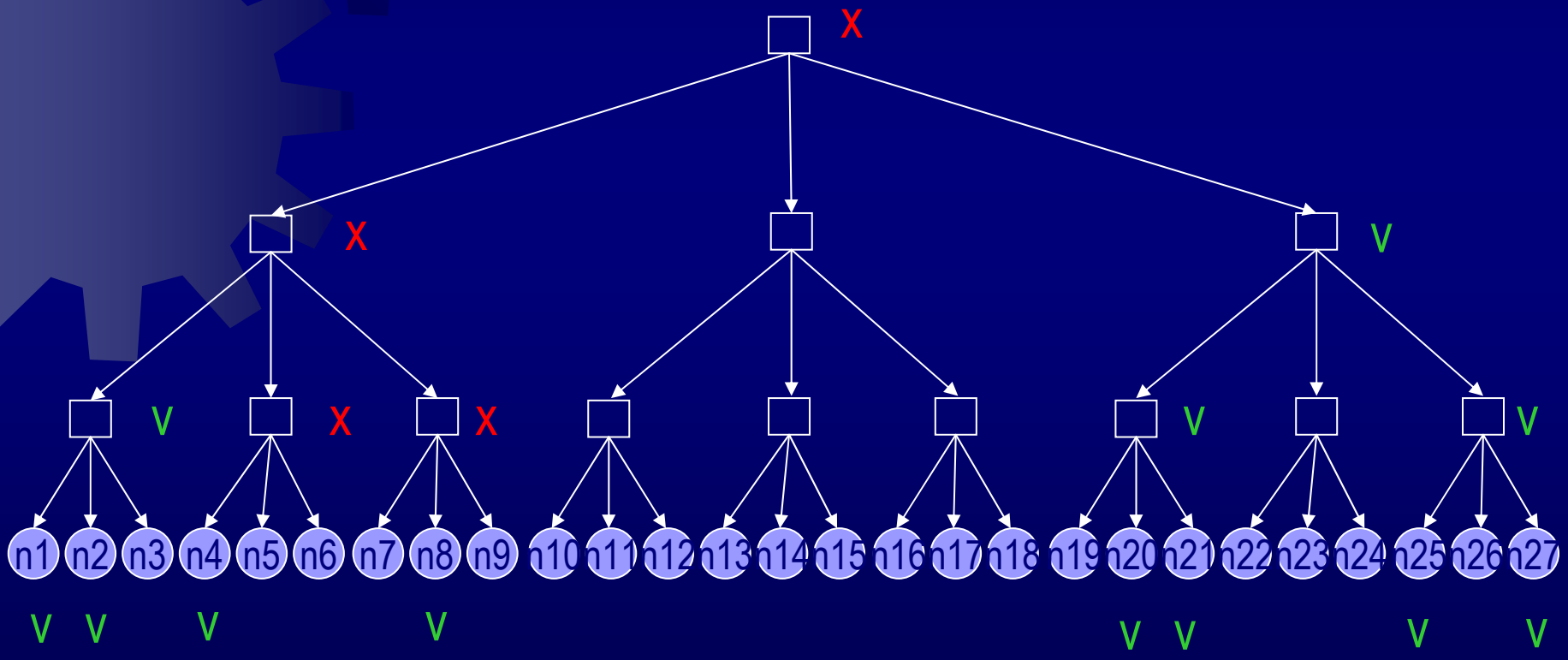
Πλεονέκτημα

- Σχηματισμός quorum από οποιαδήποτε 2 από 3 nodes σε κάθε παρακλάδι
- Δεν μπορεί να υπάρξει άλλο quorum που να μην τέμνει το προηγούμενο άρα δεν μπορούν δύο ενέργειες πρόσβασης να πραγματοποιηθούν ταυτόχρονα



Ιεραρχική ψηφοφορία (hierarchical voting) (5/6)

- ***Μειονέκτημα 1***
  - Παρόλο που ο μηχανισμός ιεραρχικής ψηφοφορίας απαιτεί λιγότερες ψήφους για μια επιτυχή ψηφοφορία, πρέπει ωστόσο αυτές να δοθούν από ένα επακριβώς ορισμένο υποσύνολο κόμβων
  - Στο παρακάτω quorum 8 nodes δεν έχουμε επιτυχημένη ψηφοφορία



Ιεραρχική ψηφοφορία (hierarchical voting) (6/6)

- ***Μειονέκτημα 2***
  - Συνεπές data block δεν εξασφαλίζει πρόσβαση στο ενημερωμένο αντίγραφο (κάτι που ισχύει π.χ. στη συναίνεση πλειοψηφίας)
  - Στο παράδειγμα, ο αριθμός είναι το version number