

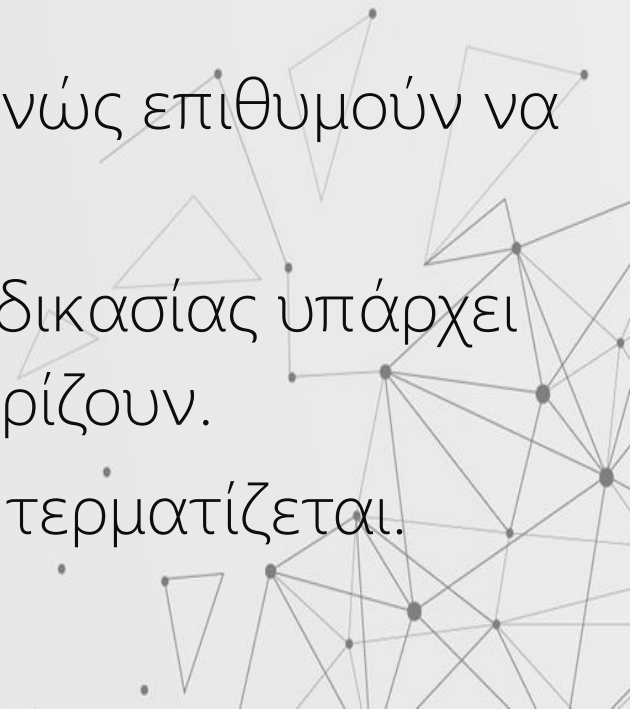
16

Κατανεμημένα Συστήματα

Αλγόριθμοι εκλογής αρχηγού
(Leader elections)

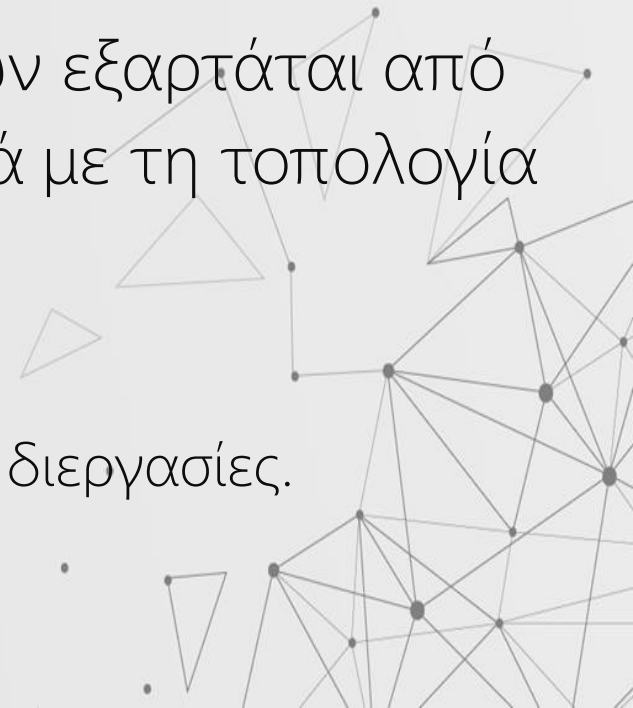
Το πρόβλημα της εκλογής αρχηγού

- Σε ένα σύστημα από N διεργασίες πρέπει να επιλεγεί μια διεργασία για να αναλάβει ένα διακεκριμένο ρόλο.
 - π.χ. ανάληψη καθηκόντων συντονιστή σε διάφορα κεντρικοποιημένα σχήματα και πρωτόκολλα συντονισμού.
- Το ζητούμενο είναι ένα σχήμα κατανεμημένου συγχρονισμού με τις εξής ιδιότητες:
- E1 (symmetry): όλες οι διεργασίες πιθανώς επιθυμούν να υποβάλλουν υποψηφιότητα.
- E2 (agreement): μετά το πέρας της διαδικασίας υπάρχει ένας μοναδικός νικητής που όλοι γνωρίζουν.
- E3 (termination): η διαδικασία εκλογής τερματίζεται.



Πλαίσιο αναφοράς και σύγκρισης

- Φόρτος στο δίκτυο:
 - συνολικός αριθμός μηνυμάτων που στέλνονται πάνω από το δίκτυο κατά τη διαδικασία εκλογής.
- Καθυστέρηση τερματισμού της διαδικασίας εκλογής.
 - καθυστέρηση αποστολής / παραλαβής μηνυμάτων που ανταλλάσσονται στο πλαίσιο μιας εκλογής.
- Η καθυστέρηση μετάδοσης μηνυμάτων εξαρτάται από τις υποθέσεις που γίνονται αναφορικά με τη τοπολογία και ικανότητες του φυσικού δικτύου.
- Υποθέτουμε:
 - αξιόπιστη επικοινωνία FIFO ανάμεσα στις διεργασίες.
 - βλάβες fail-stop (ή σύγχρονο σύστημα).



Παρατηρήσεις

- Μερικοί αλγόριθμοι δεν χειρίζονται βλάβες και πρέπει να επεκταθούν κατάλληλα για το χειρισμό τους.
- Μερικοί αλγόριθμοι υποθέτουν ότι μόνο οι διεργασίες που αρχίζουν οι ίδιες την διαδικασία ενδιαφέρονται να εκλεγούν. Έτσι, αν μια διεργασία δεν είναι starter δεν μπορεί να εκλεγεί.
- Η εκλογή αρχηγού σε ένα κατανεμημένο σύστημα μπορεί να θεωρηθεί ως ισοδύναμο πρόβλημα με τον προσδιορισμό της διεργασίας με τη μικρότερη (ή μεγαλύτερη) τιμή για μια (μεταβλητή) ιδιότητα.
 - αυτό πολλές φορές «απλοποιείται» λέγοντας ότι αναζητείται η διεργασία με το μικρότερο (ή μεγαλύτερο) αναγνωριστικό.

Αλγόριθμοι εκλογής αρχηγού

- Αλγόριθμος με δομή δέντρου.
- Αλγόριθμος FloodMax για γράφο.
- Αλγόριθμος δακτυλίου (LeLann).
- Αλγόριθμος δακτυλίου (Chang & Roberts).
- Αλγόριθμος ισχυροτέρου (Garcia-Molina).



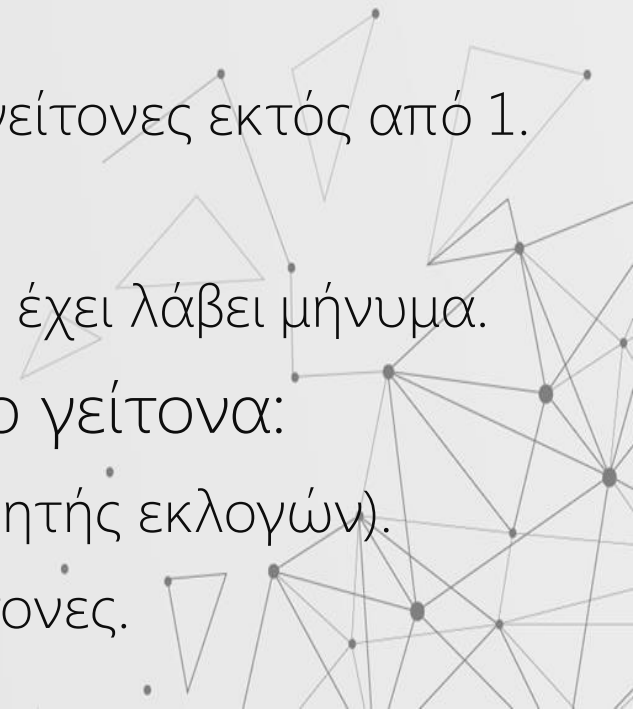
01

Αλγόριθμος με δομή
δέντρου

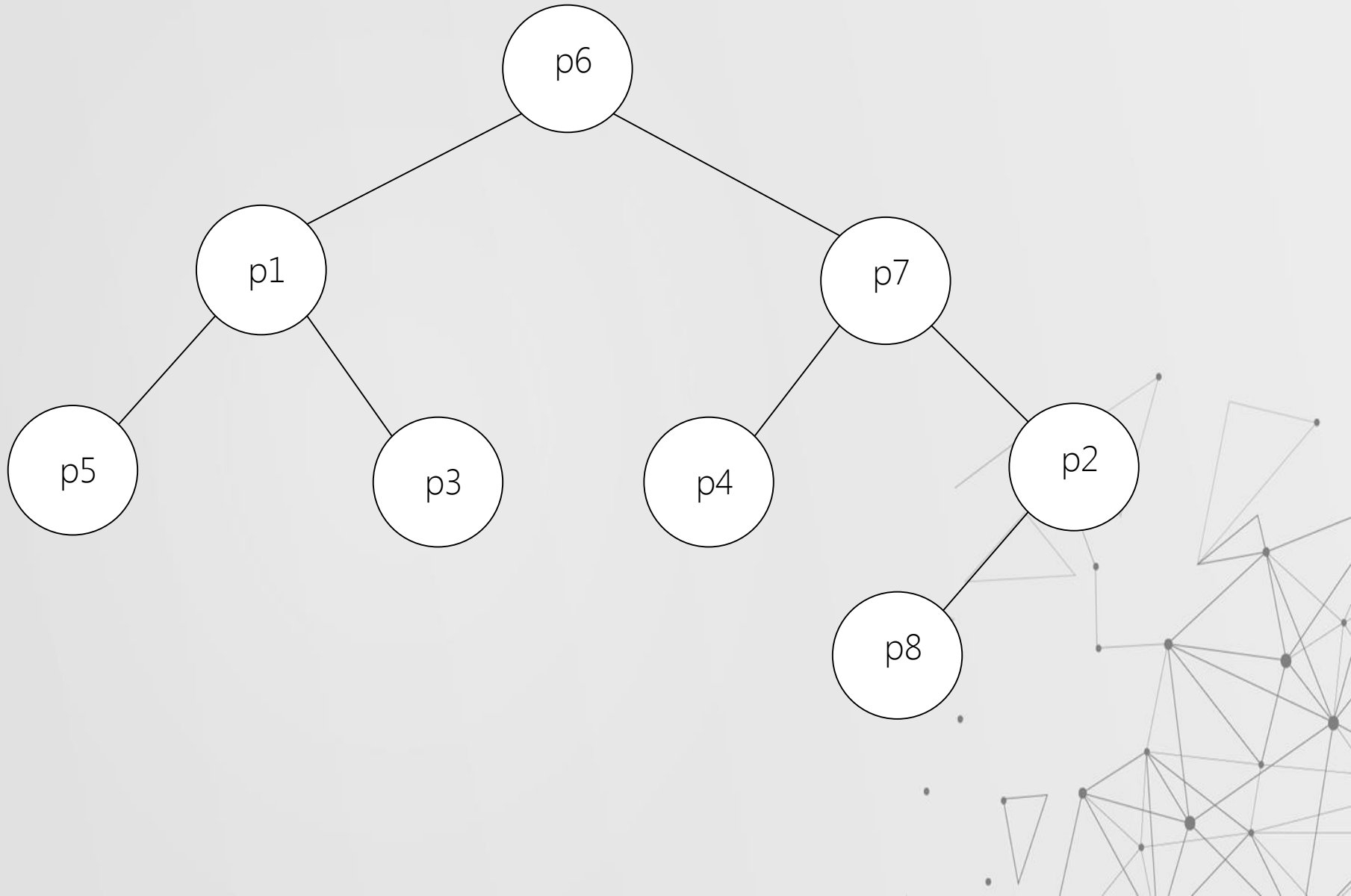


Αλγόριθμος με δομή δέντρου

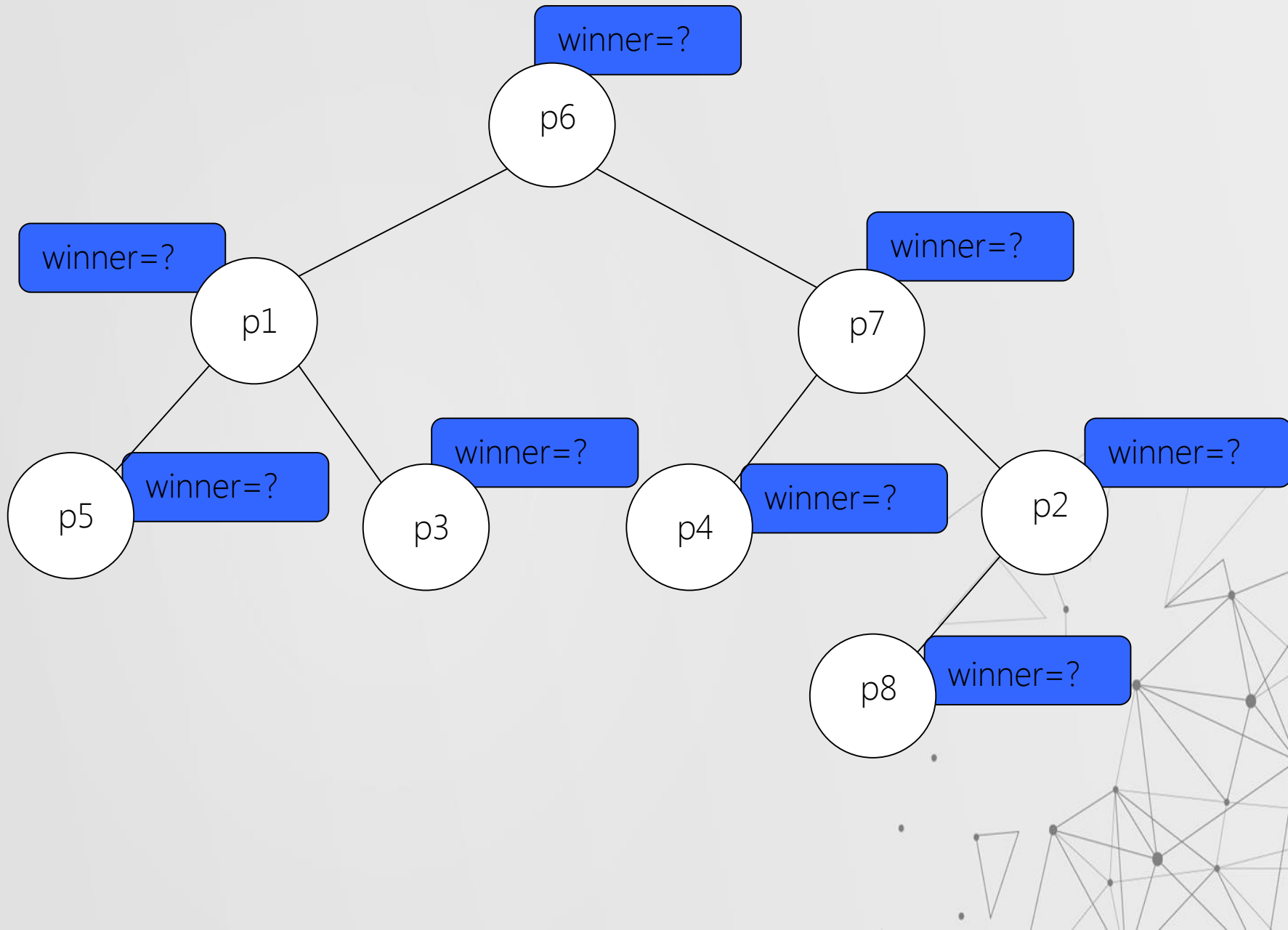
- Οι διεργασίες είναι οργανωμένες σε νοητό δέντρο.
- Η διαδικασία αρχίζει από τα φύλλα του δέντρου.
- Ο εκκινητής (starter) πρέπει να στείλει κατάλληλη ειδοποίηση (θεωρητικά αρκεί να ειδοποιηθούν μόνο οι κόμβοι-φύλλα).
- Κάθε διεργασία:
 - περιμένει για προτάσεις από όλους τους γείτονες εκτός από 1.
 - επιλέγει το μικρότερο αναγνωριστικό.
 - το στέλνει στο γείτονα από τον οποίο δεν έχει λάβει μήνυμα.
- Όταν λάβει μήνυμα από τον τελευταίο γείτονα:
 - επιλέγει το μικρότερο αναγνωριστικό (νικητής εκλογών).
 - το στέλνει σε όλους τους υπόλοιπους γείτονες.



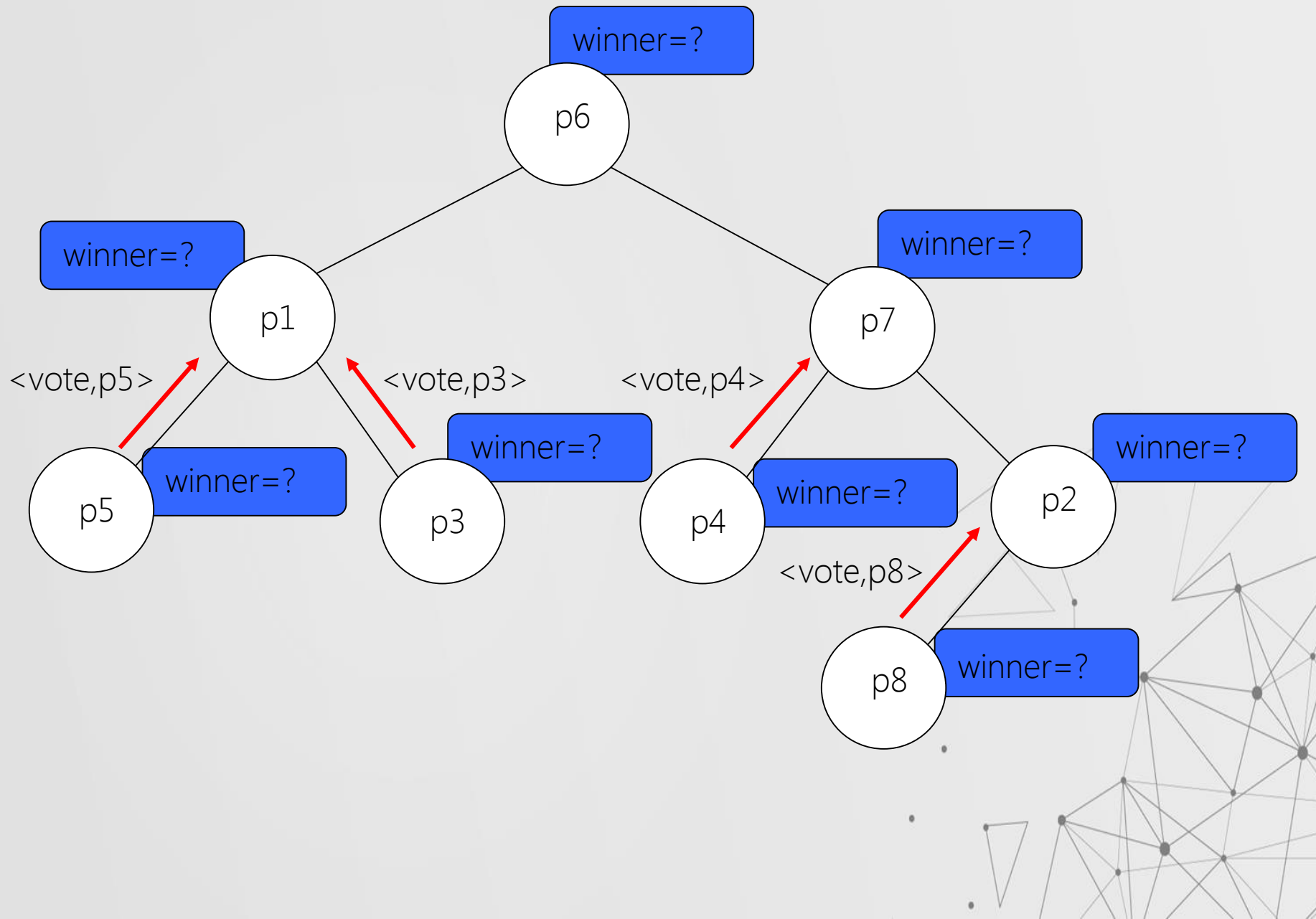
Αλγόριθμος με δομή δέντρου (1)



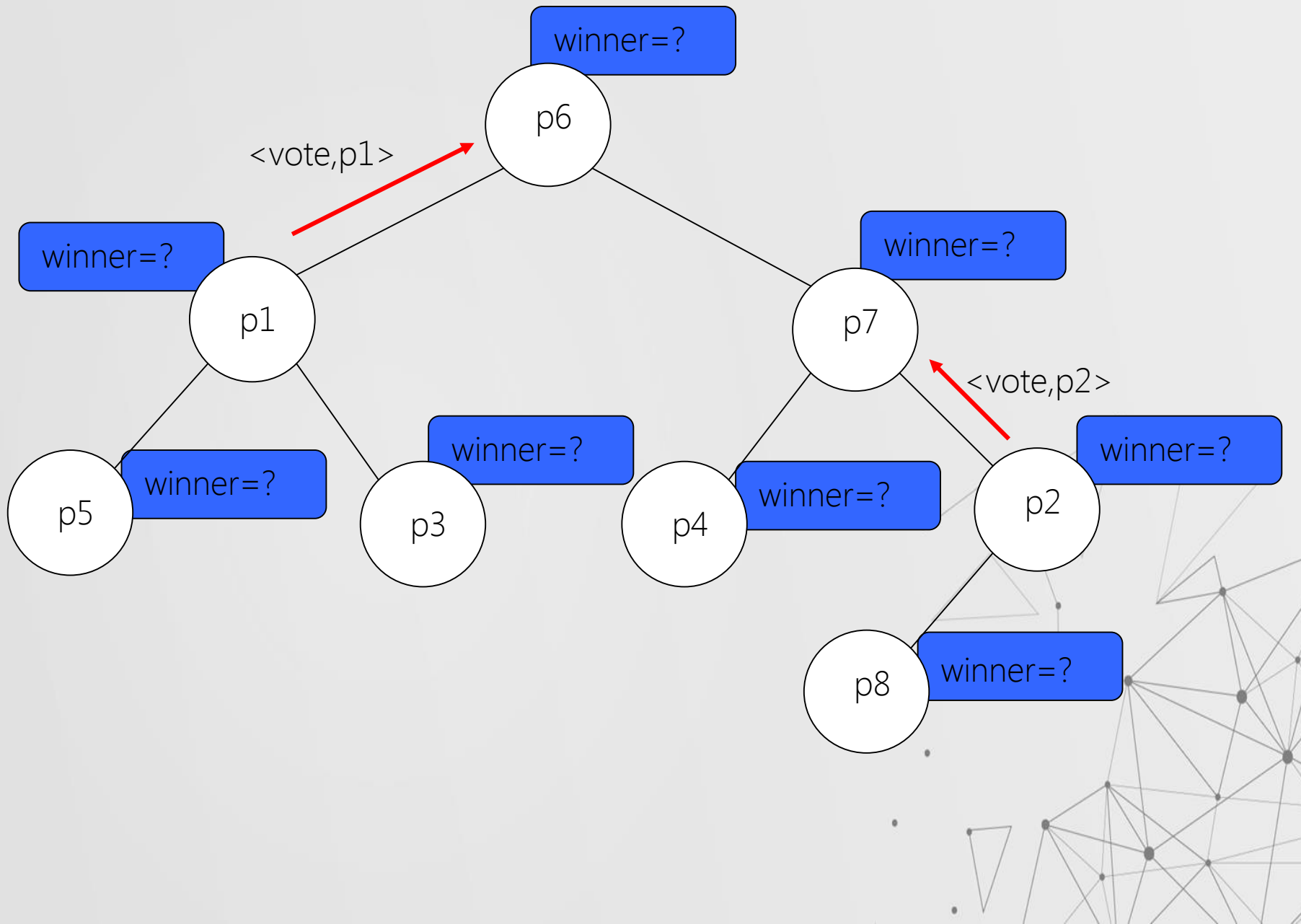
Αλγόριθμος με δομή δέντρου (2)



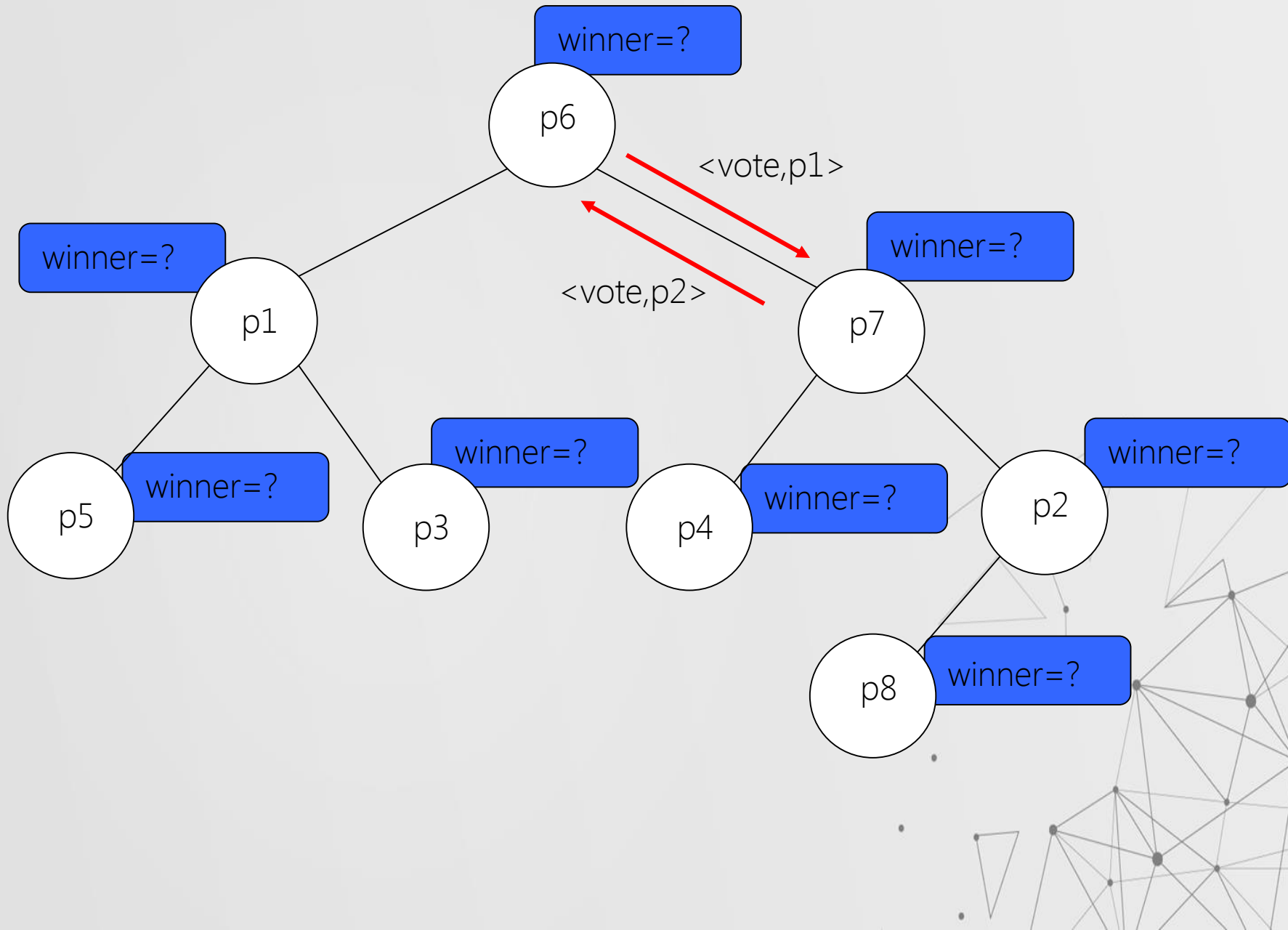
Αλγόριθμος με δομή δέντρου (3)



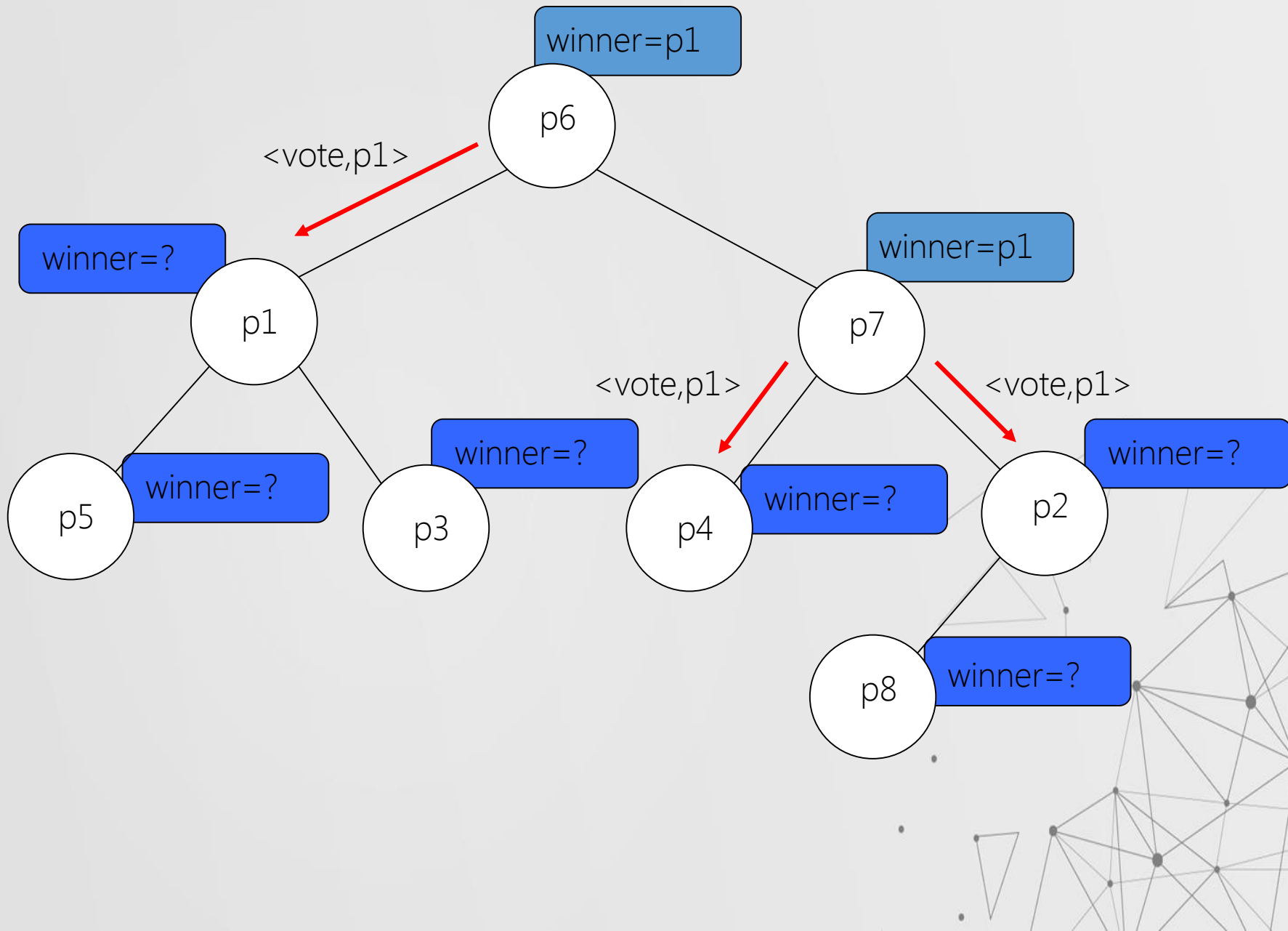
Αλγόριθμος με δομή δέντρου (4)



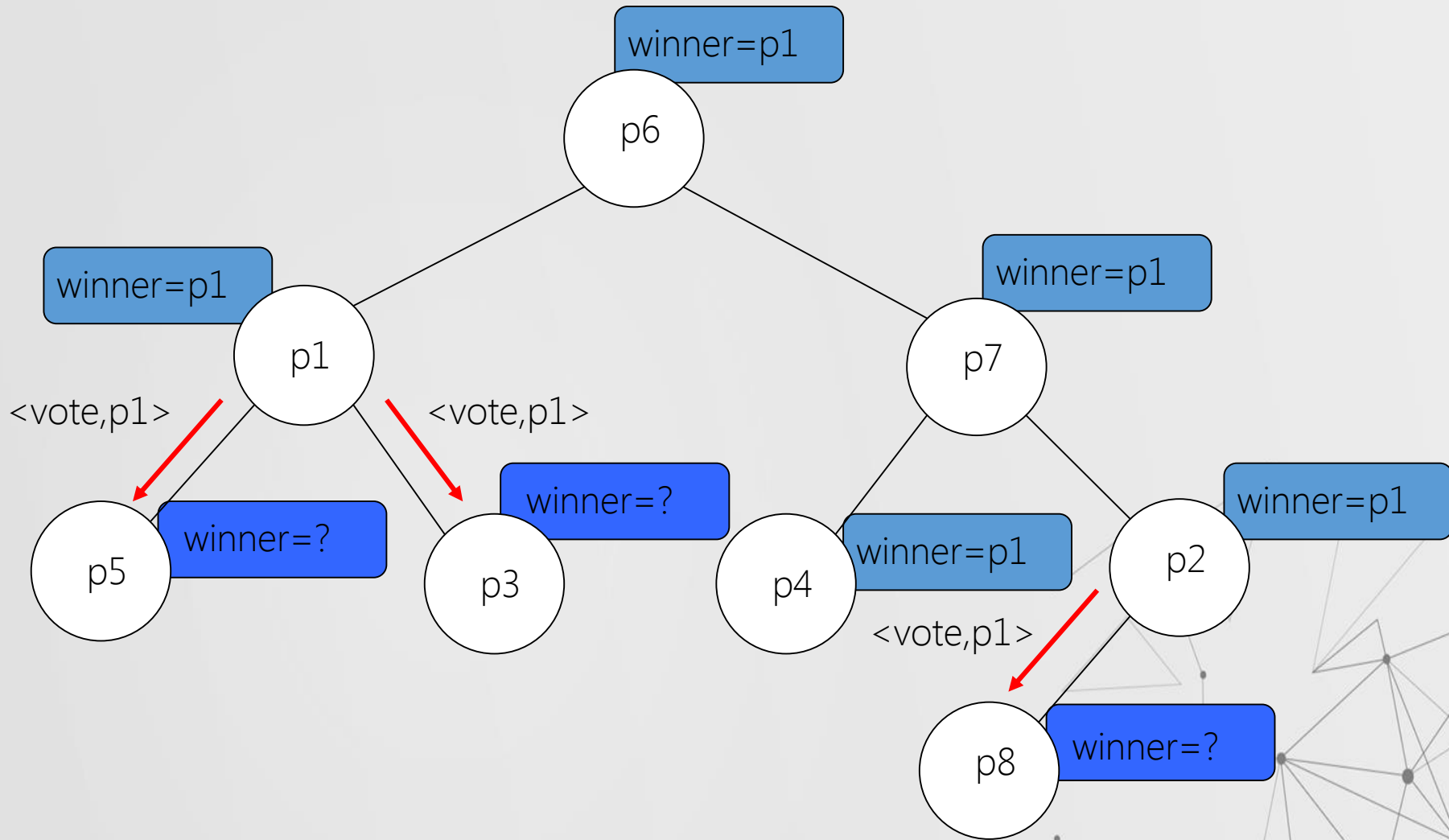
Αλγόριθμος με δομή δέντρου (5)



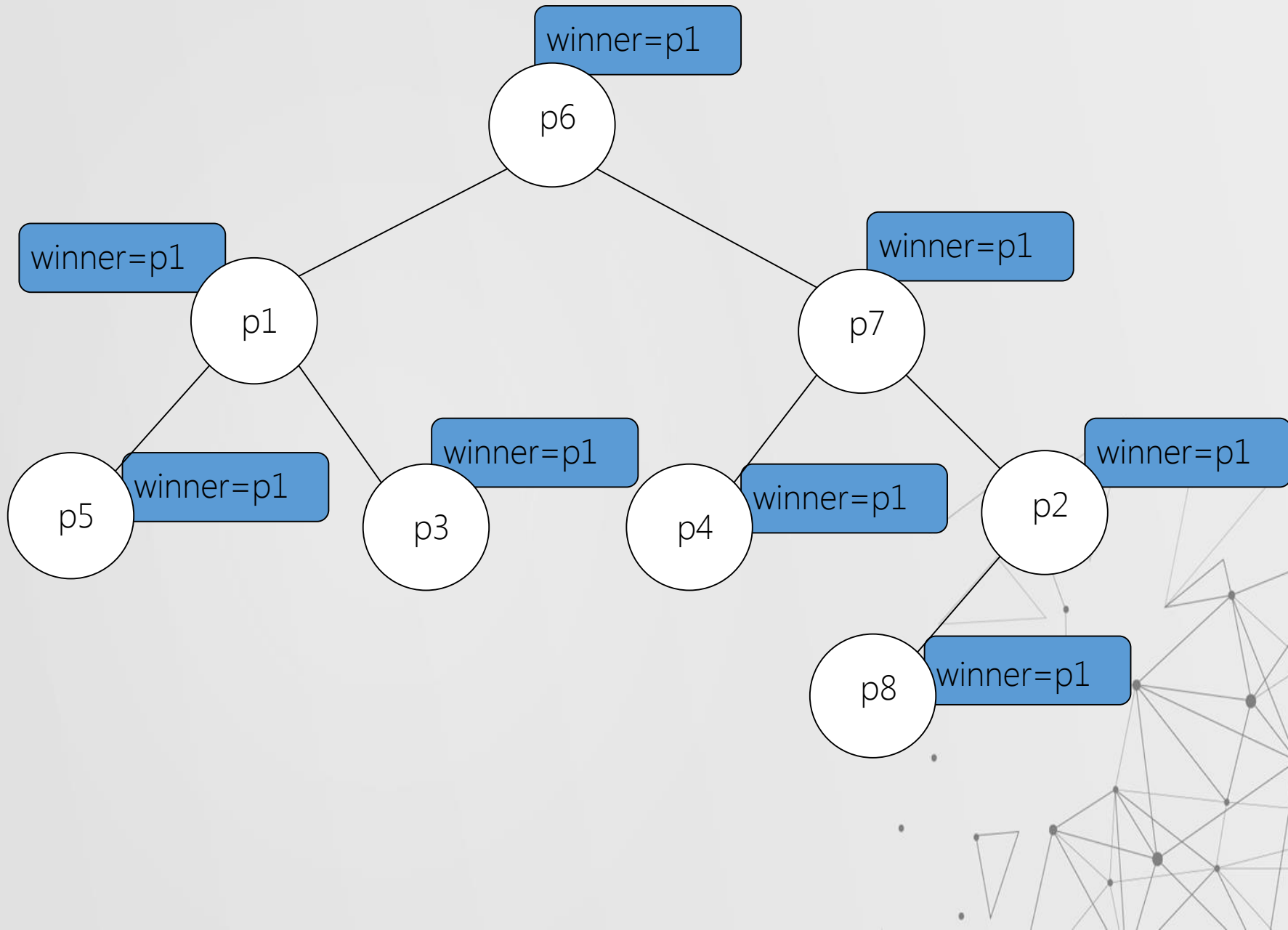
Αλγόριθμος με δομή δέντρου (6)



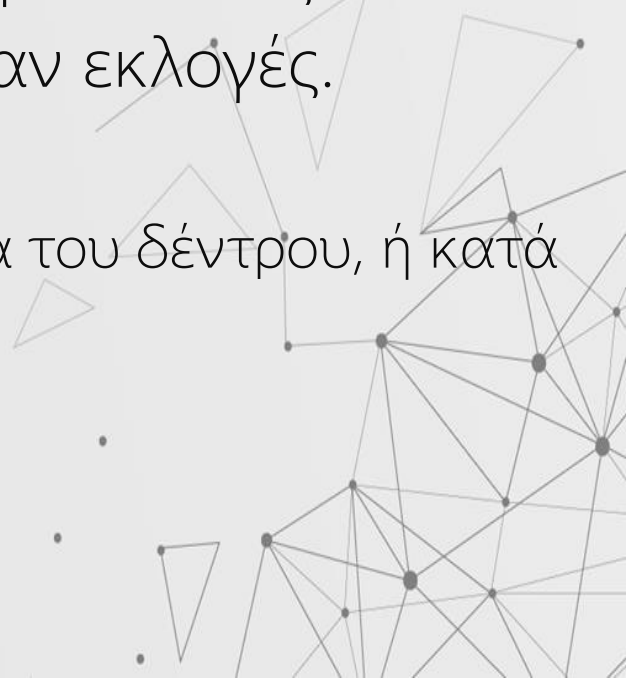
Αλγόριθμος με δομή δέντρου (7)



Αλγόριθμος με δομή δέντρου (8)



Παρατηρήσεις

- Χαρακτηριστικά απόδοσης:
 - $2(N - 1)$ μηνύματα (για N διεργασίες).
 - η καθυστέρηση αντιστοιχεί σε $h \cdot RTT$ με:
 - h το ύψος του δέντρου,
 - RTT το Round Trip Time, και
 - παράλληλη μετάδοση σε διαφορετικά κλαδιά).
 - Για να αρχίσει η διαδικασία εκλογής, πρέπει όλες οι διεργασίες να ειδοποιηθούν ότι άρχισαν εκλογές.
 - N επιπλέον μηνύματα.
 - μπορεί να σταλούν απ' ευθείας στα φύλλα του δέντρου, ή κατά μήκος των ακμών του δέντρου.
 - Βλάβες:
 - προσαρμογή του δέντρου.
 - επανεκπομπή προτάσεων.
- 

02

Αλγόριθμος FloodMax
για γράφο



Αλγόριθμος FloodMax για γράφο

- Οι διεργασίες είναι οργανωμένες σε τυχαίο γράφο με διάμετρος D (θεωρείται γνωστή).
- Σε κάθε γύρο, κάθε διεργασία:
 - στέλνει στους γείτονες της το μικρότερο αναγνωριστικό που έλαβε από τους γείτονες της (αρχικά το δικό της).
 - περιμένει να λάβει προτάσεις από όλους τους γείτονες.
- Μετά από D γύρους επικοινωνίας, κάθε διεργασία επιλέγει την διεργασία με το μικρότερο αναγνωριστικό.
- Η σειριοποίηση μηνυμάτων γίνεται με χρήση δομής FIFO.
- Χαρακτηριστικά απόδοσης:
- Πλήθος μηνυμάτων: $2ED$ (E είναι το πλήθος των ακμών του γράφου).

03

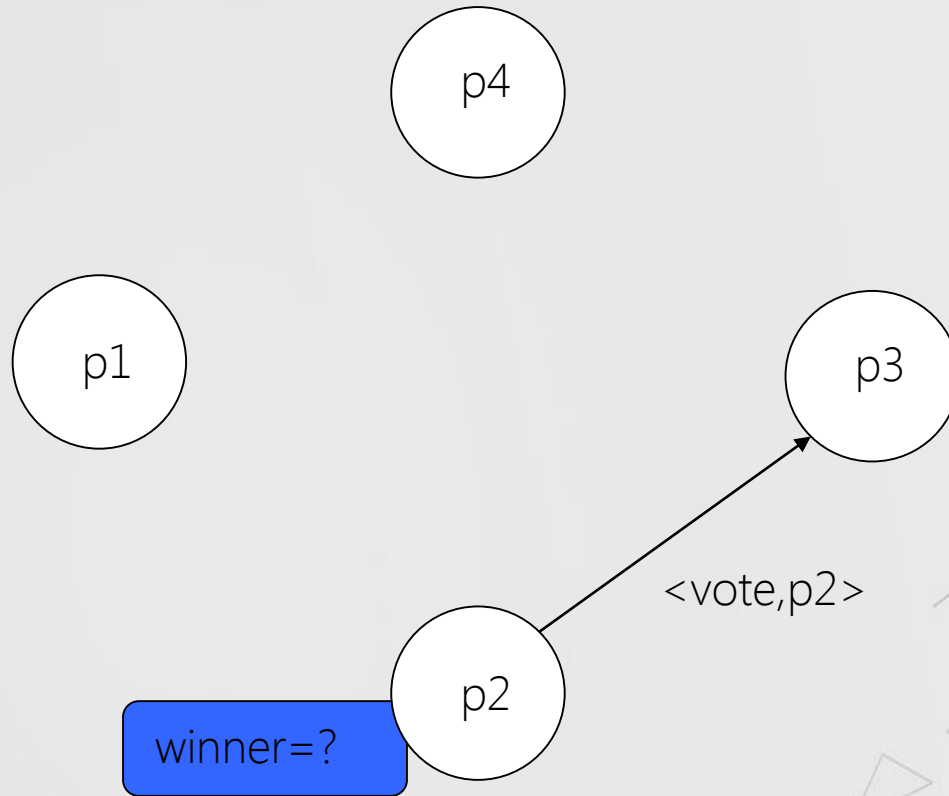
Αλγόριθμος δακτυλίου
(LeLann)



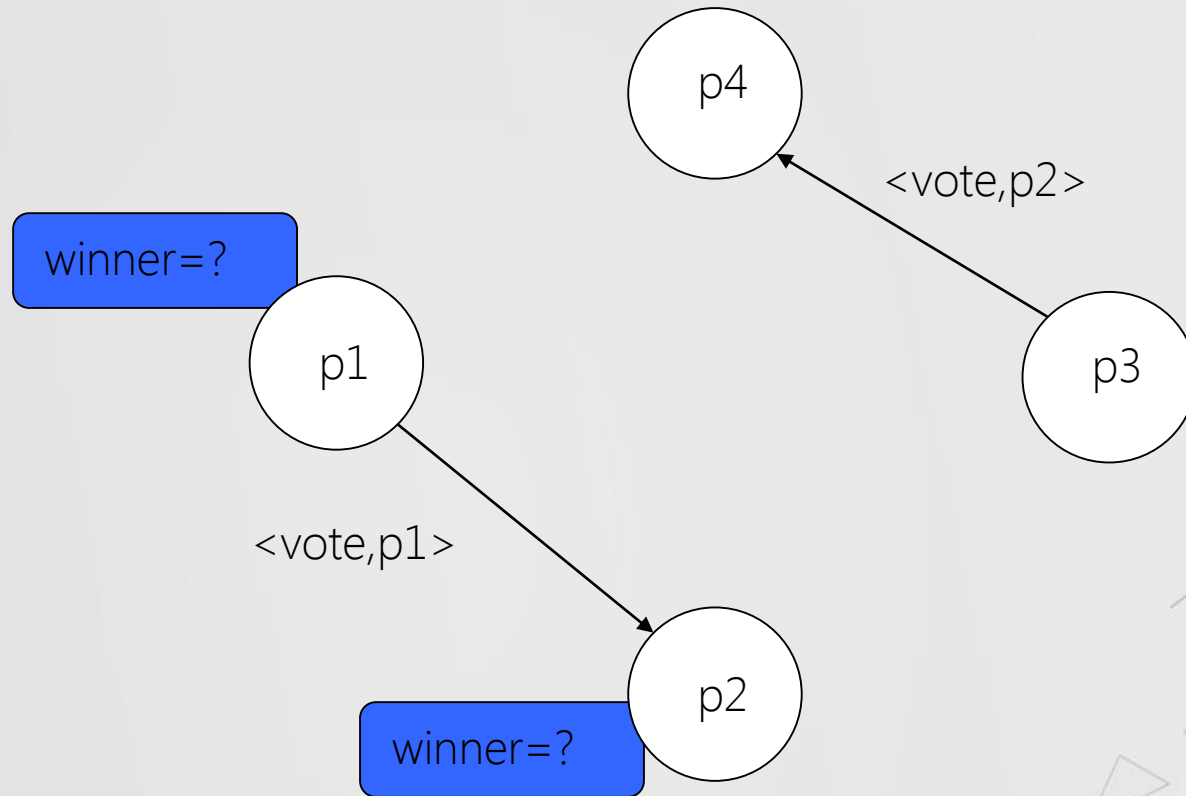
Αλγόριθμος δακτυλίου (LeLann)

- Οι διεργασίες είναι διατεταγμένες σε νοητό δακτύλιο.
- Κάθε εκκινητής στέλνει κατά μήκος του δακτυλίου ένα μήνυμα υποψηφιότητας με το δικό του αναγνωριστικό.
- Το αναγνωριστικό τοποθετείται σε λίστα.
- Όταν μια διεργασία λάβει μήνυμα υποψηφιότητας το προωθεί στην επόμενη διεργασία.
- Αν είναι εκκινητής επισυνάπτει το αναγνωριστικό του.
- Όταν ένας εκκινητής λάβει το δικό του μήνυμα υποψηφιότητας επιλέγει τη διεργασία με το μικρότερο αναγνωριστικό.
- Αν είναι ο νικητής στέλνει ανακοίνωση σε όλες τις διεργασίες.
- Απαιτούνται $N + MN$ μηνύματα για M εκκινητές.

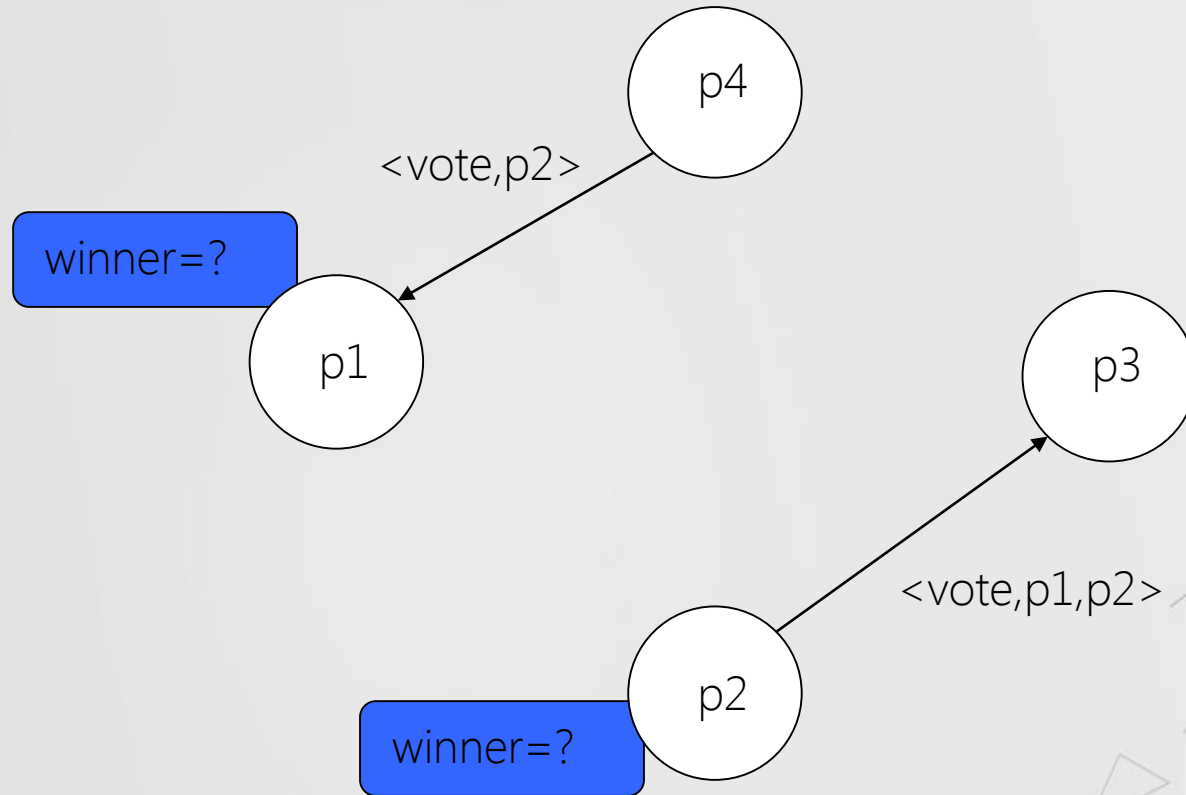
Αλγόριθμος δακτυλίου (LeLann, 1)



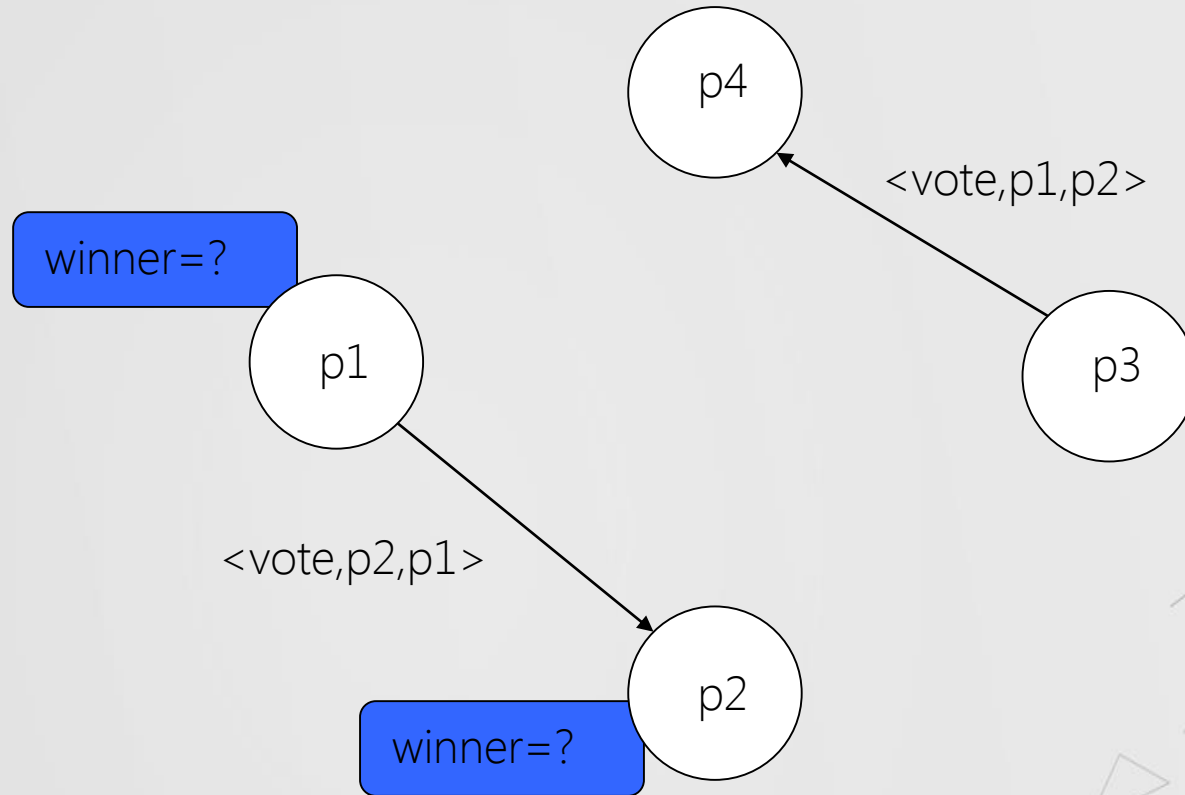
Αλγόριθμος δακτυλίου (LeLann, 2)



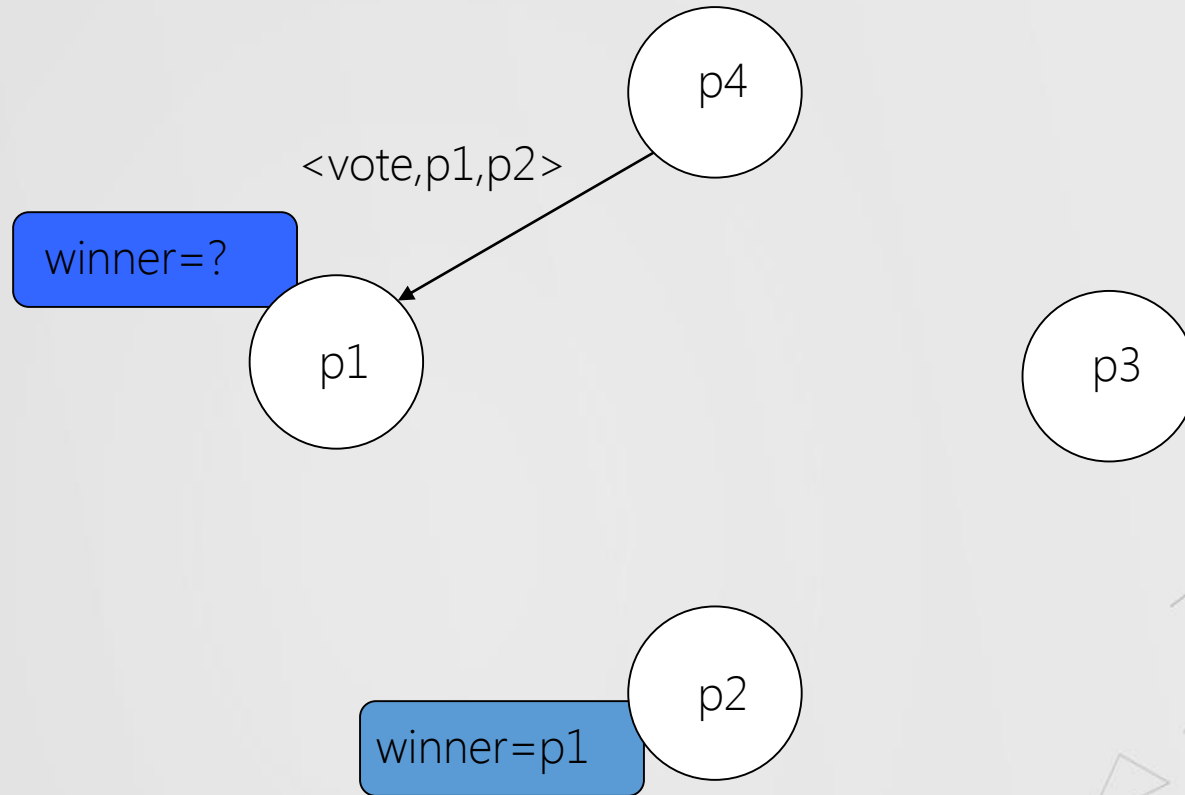
Αλγόριθμος δακτυλίου (LeLann, 3)



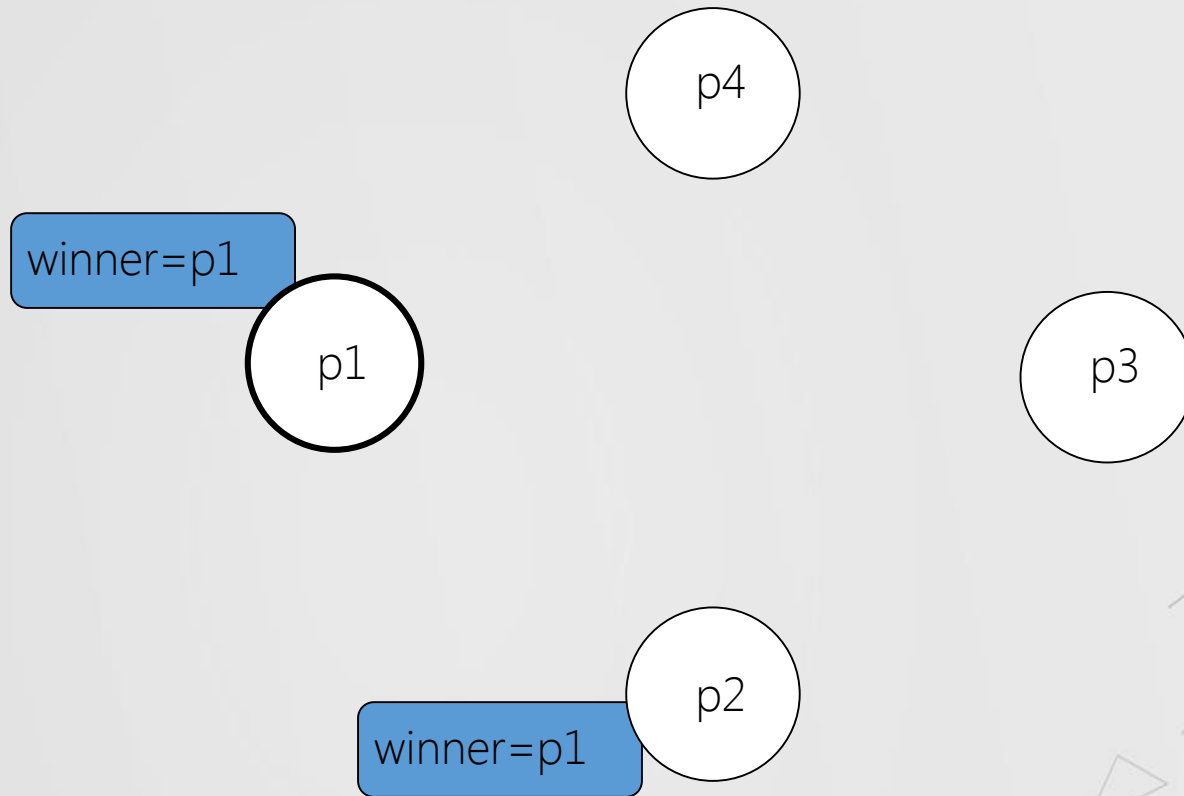
Αλγόριθμος δακτυλίου (LeLann, 4)



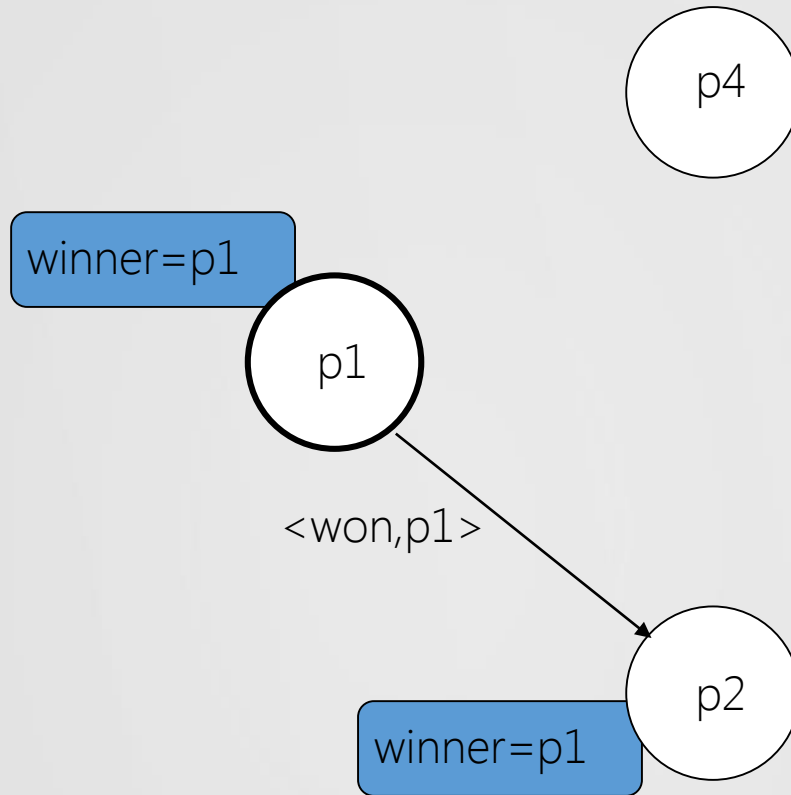
Αλγόριθμος δακτυλίου (LeLann, 5)



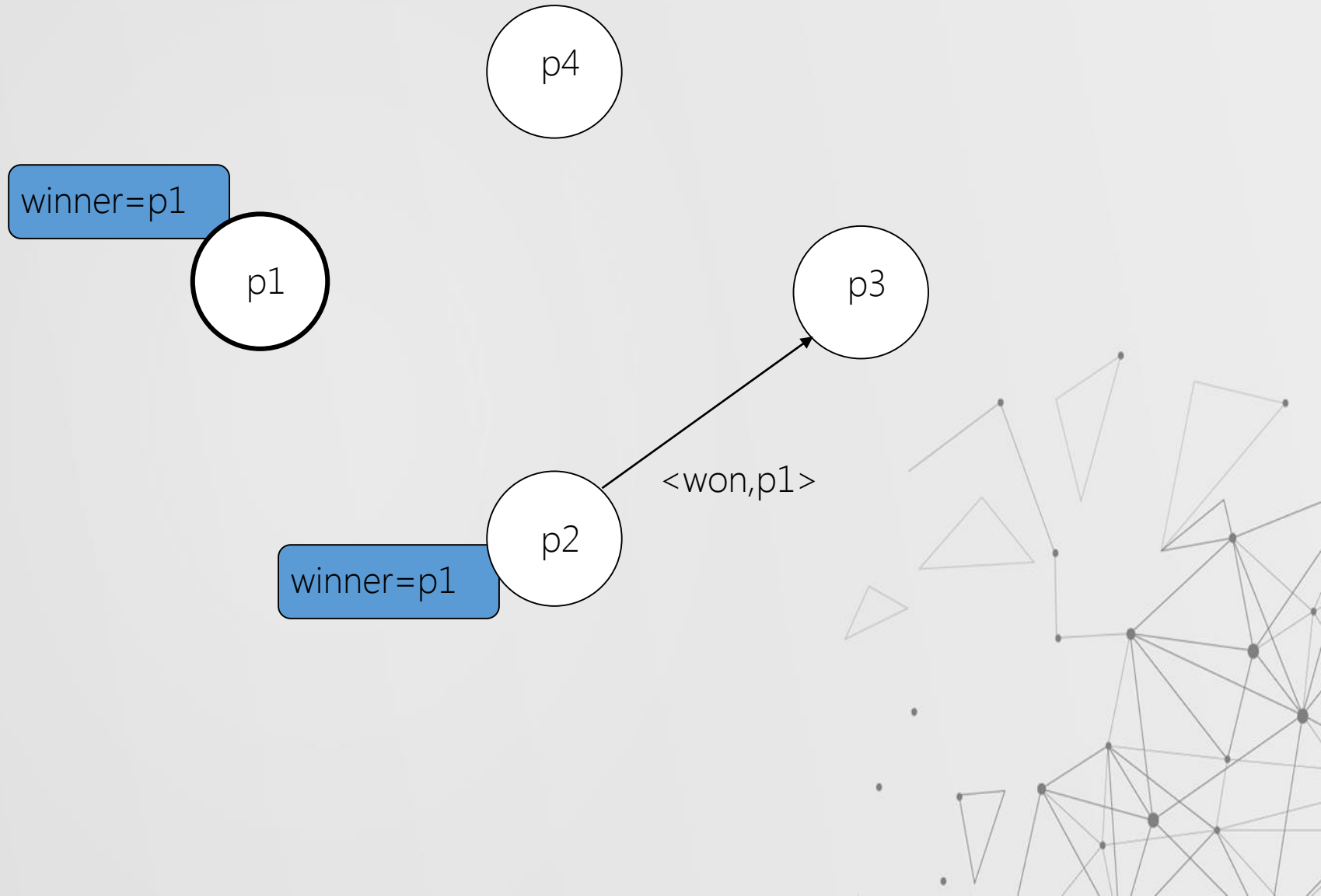
Αλγόριθμος δακτυλίου (LeLann, 6)



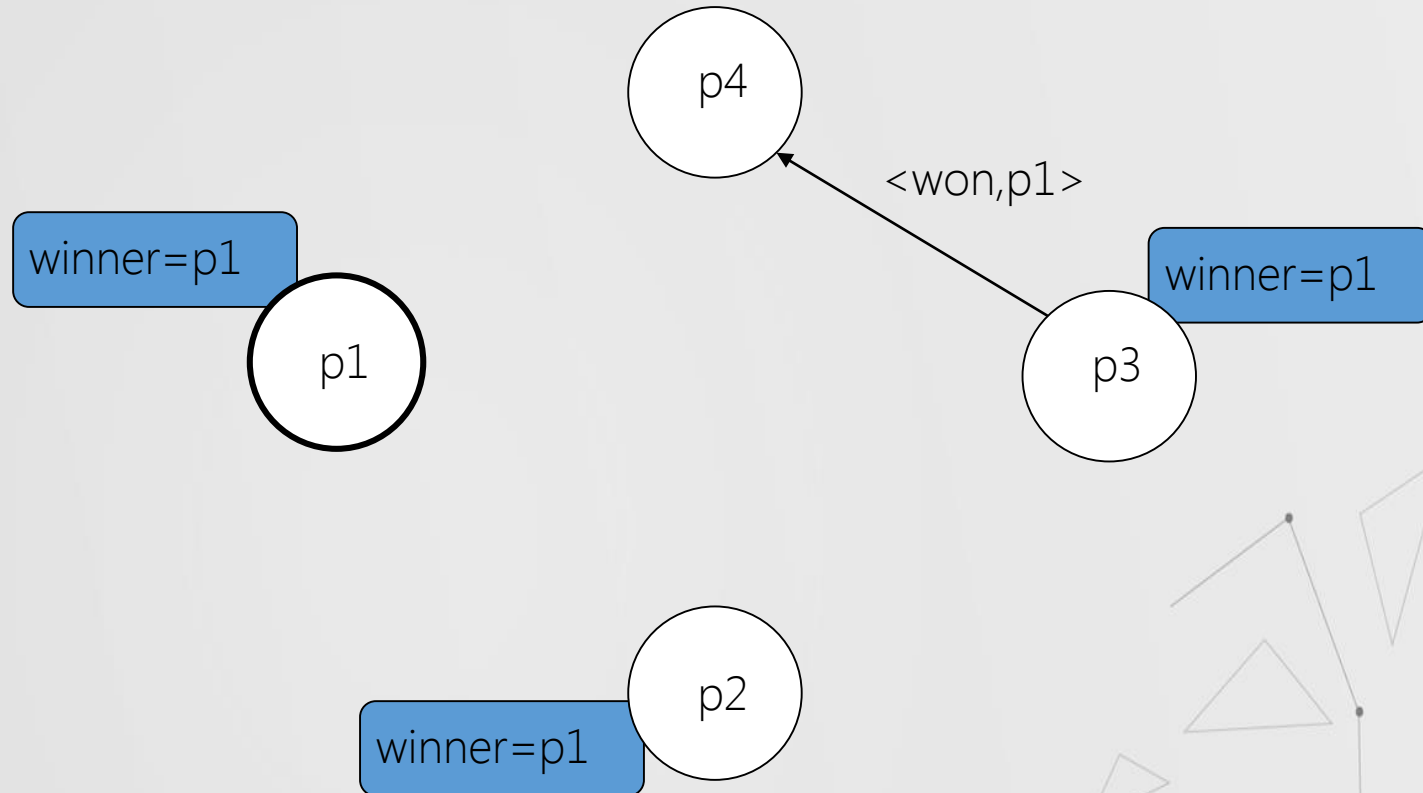
Αλγόριθμος δακτυλίου (LeLann, 7)



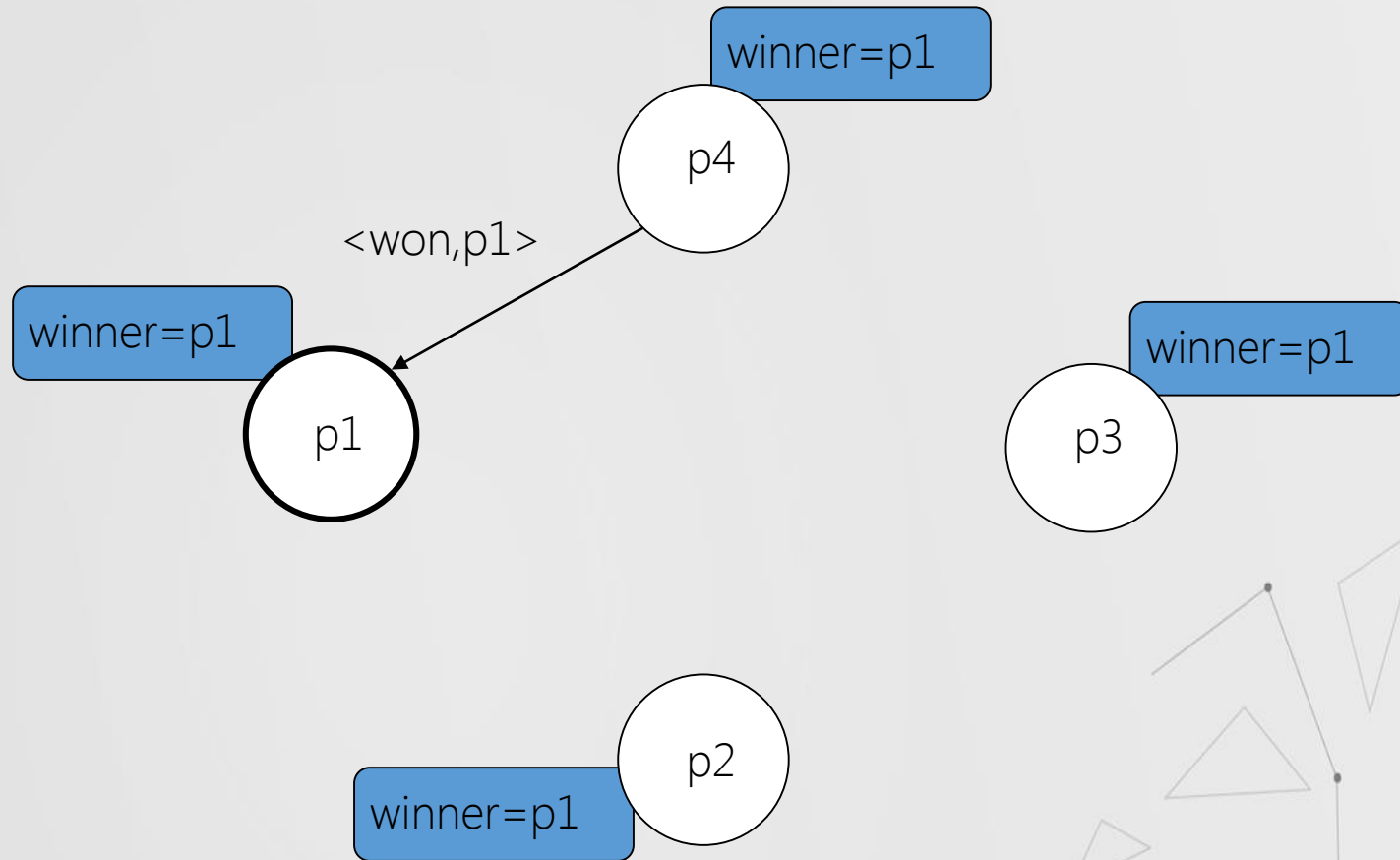
Αλγόριθμος δακτυλίου (LeLann, 8)



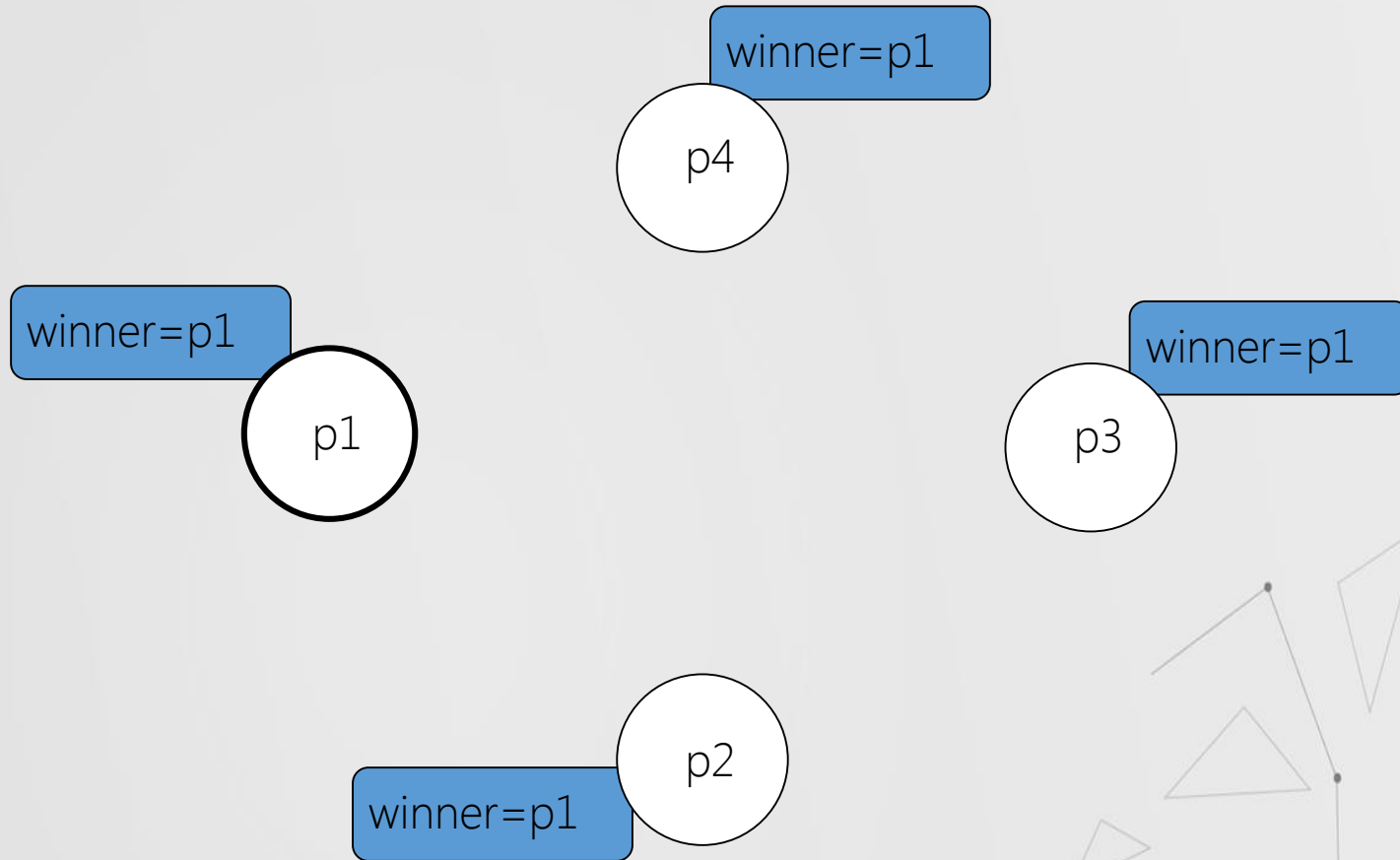
Αλγόριθμος δακτυλίου (LeLann, 9)



Αλγόριθμος δακτυλίου (LeLann, 10)



Αλγόριθμος δακτυλίου (LeLann, 11)



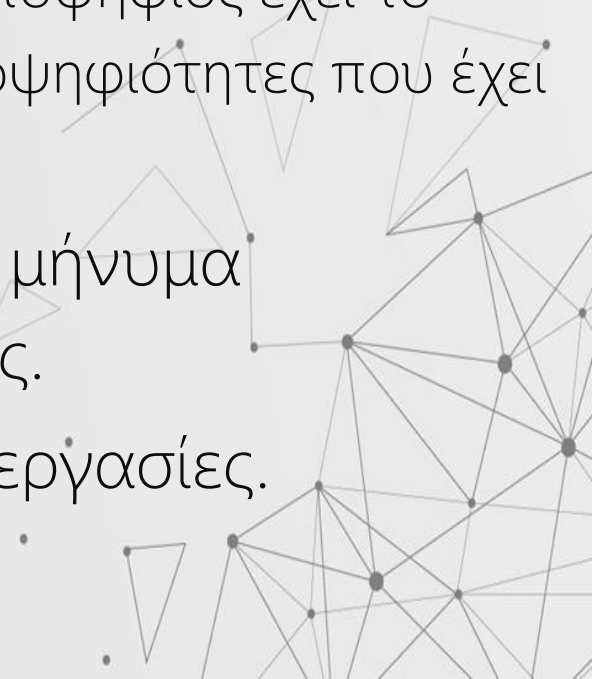
04

Αλγόριθμος δακτυλίου
(Chang & Roberts)

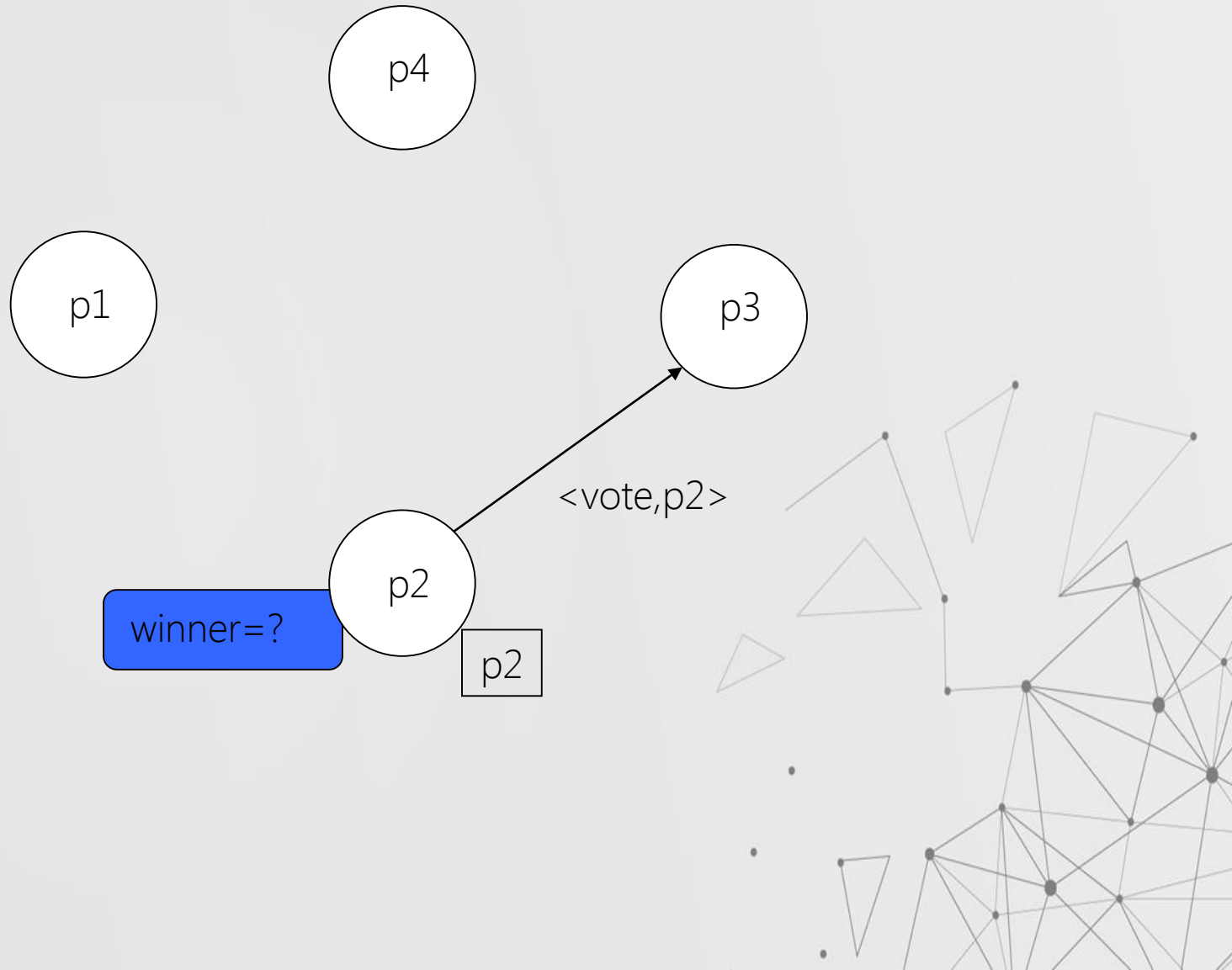


Αλγόριθμος δακτυλίου (Chang & Roberts)

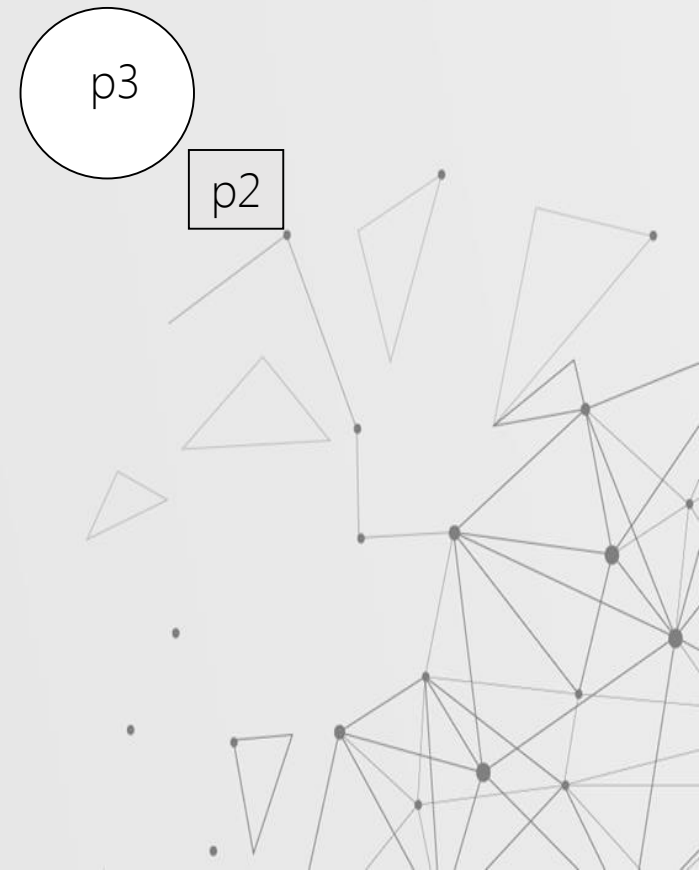
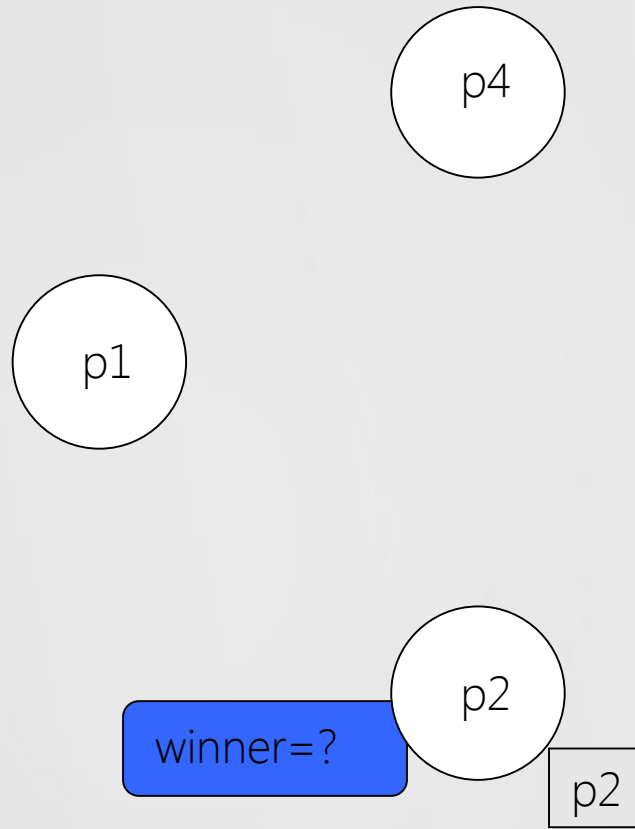
- Βελτιωμένη έκδοση του προηγούμενου αλγόριθμου.
- Κάθε εκκινητής κρατά το μικρότερο αναγνωριστικό της διεργασίας από την οποία έλαβε υποψηφιότητα.
- Όταν μια διεργασία λάβει μήνυμα υποψηφιότητας:
 - αν δεν είναι εκκινητής το προωθεί στην επόμενη διεργασία.
 - αν είναι εκκινητής το προωθεί μόνο αν ο υποψήφιος έχει το μικρότερο αναγνωριστικό από όλες τις υποψηφιότητες που έχει λάβει μέχρι στιγμής.
- Όταν ένας εκκινητής λάβει το δικό του μήνυμα υποψηφιότητας ανακηρύσσεται νικητής.
- Ο νικητής ενημερώνει τις υπόλοιπες διεργασίες.



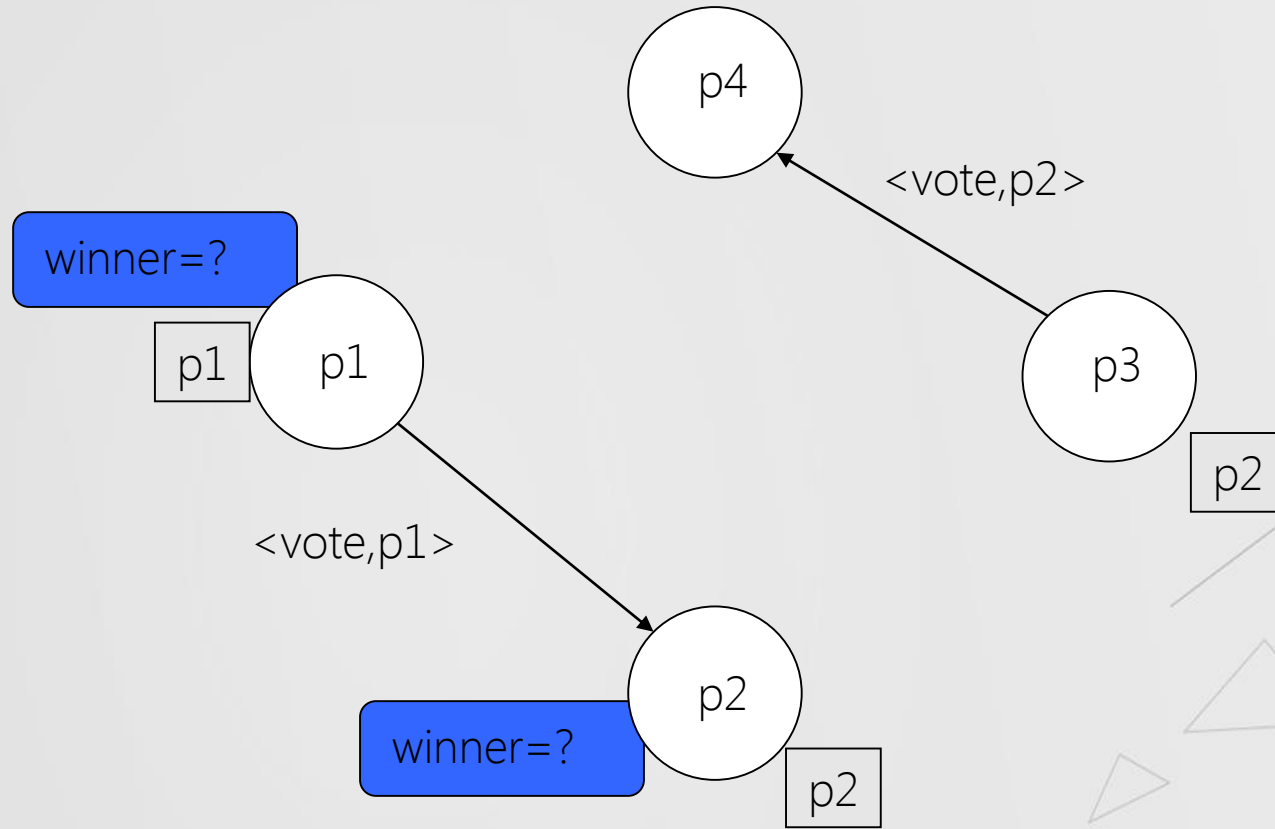
Αλγόριθμος δακτυλίου (Chang & Roberts)



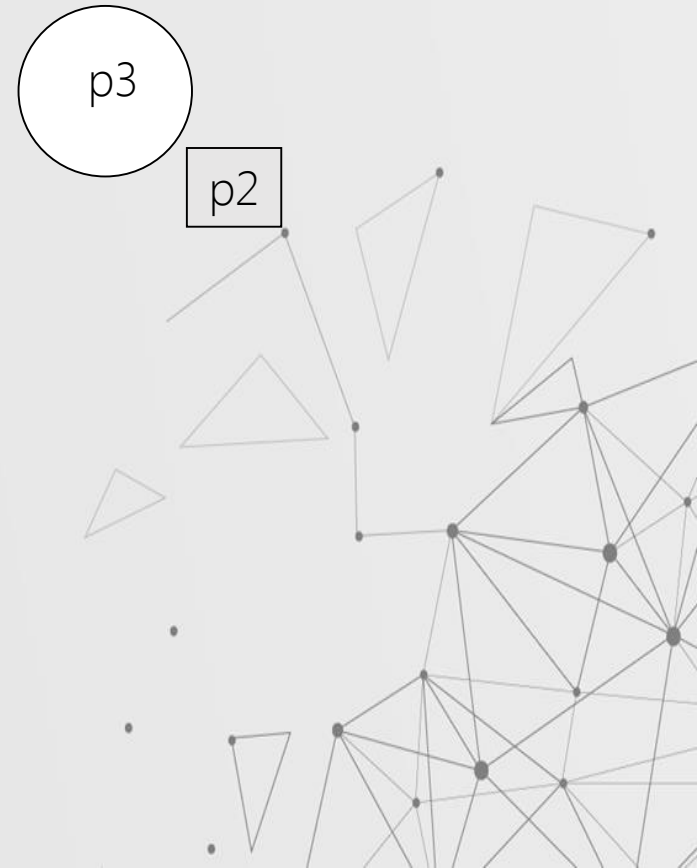
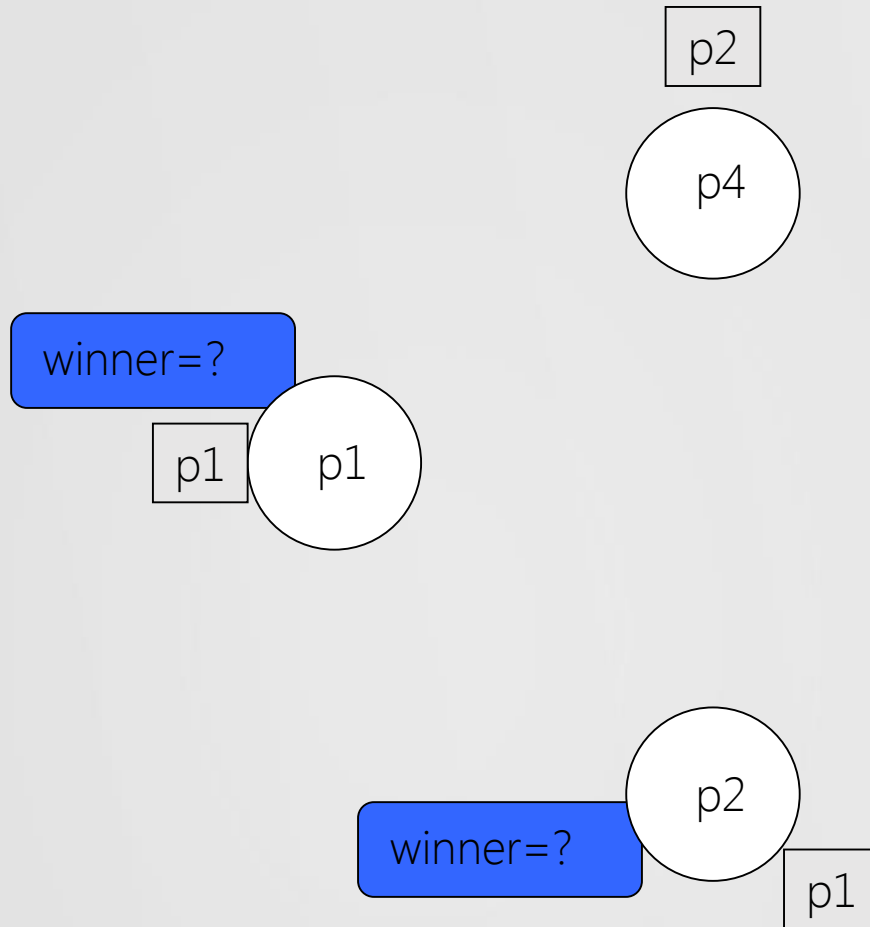
Αλγόριθμος δακτυλίου (Chang & Roberts)



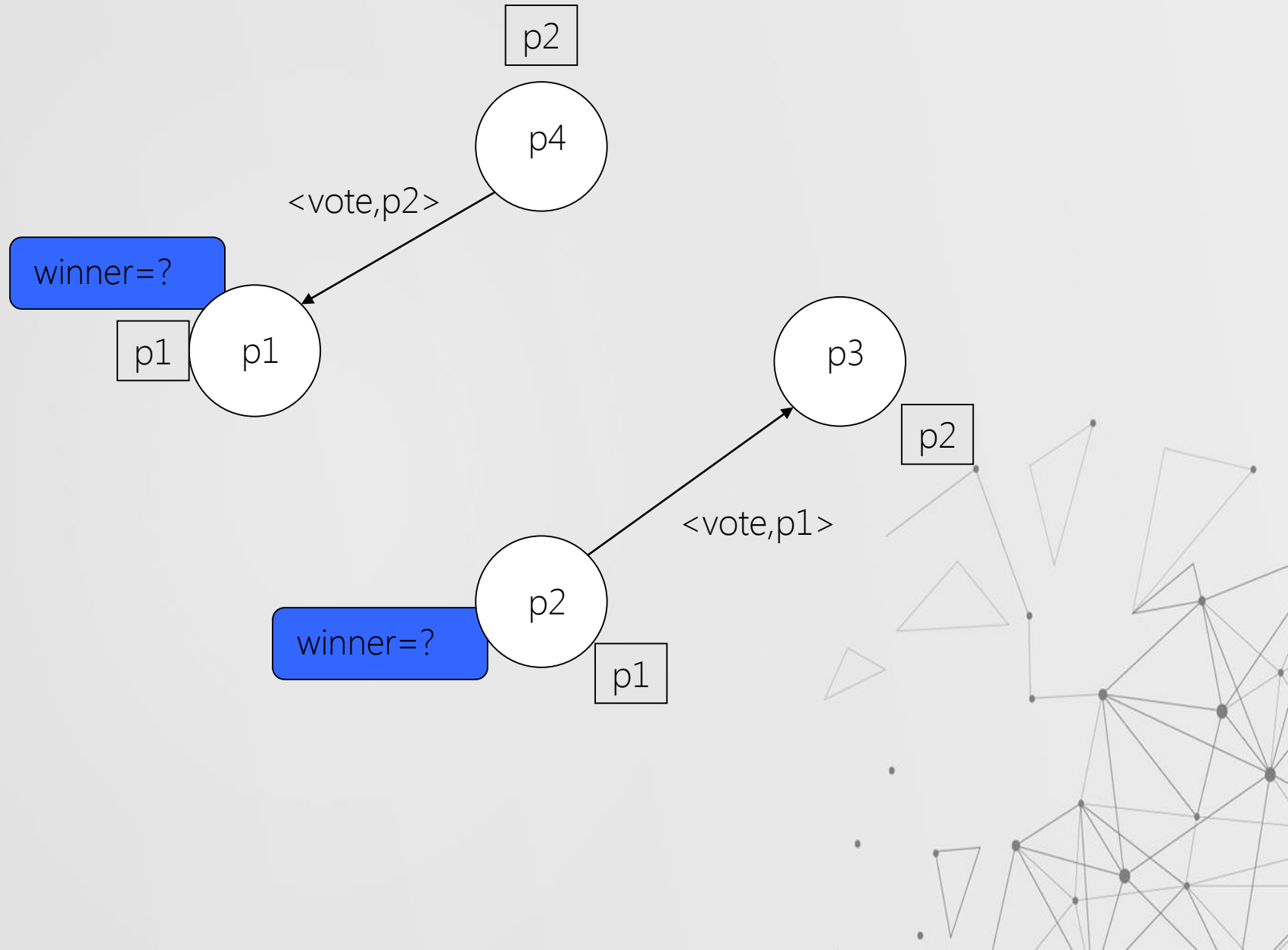
Αλγόριθμος δακτυλίου (Chang & Roberts)



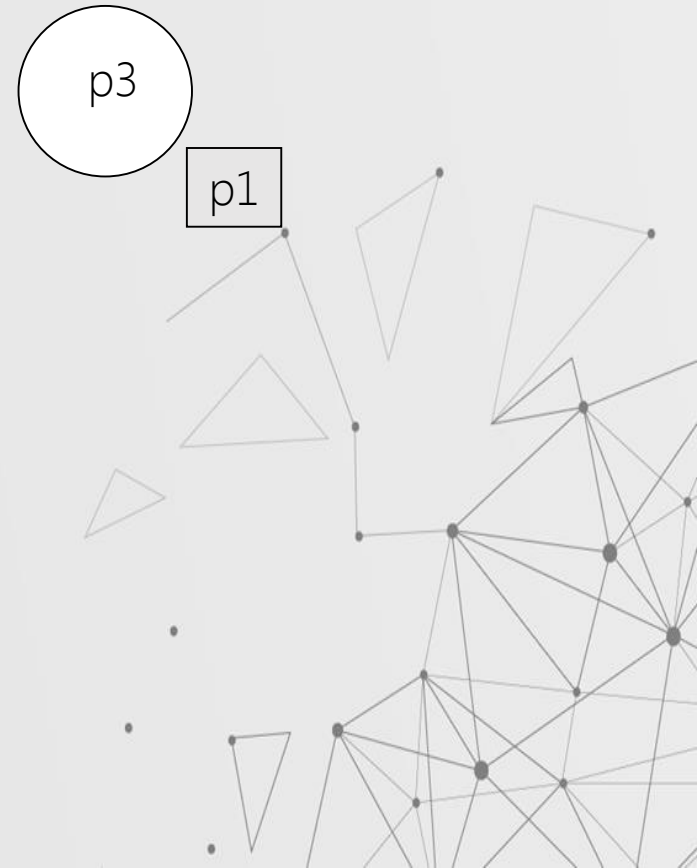
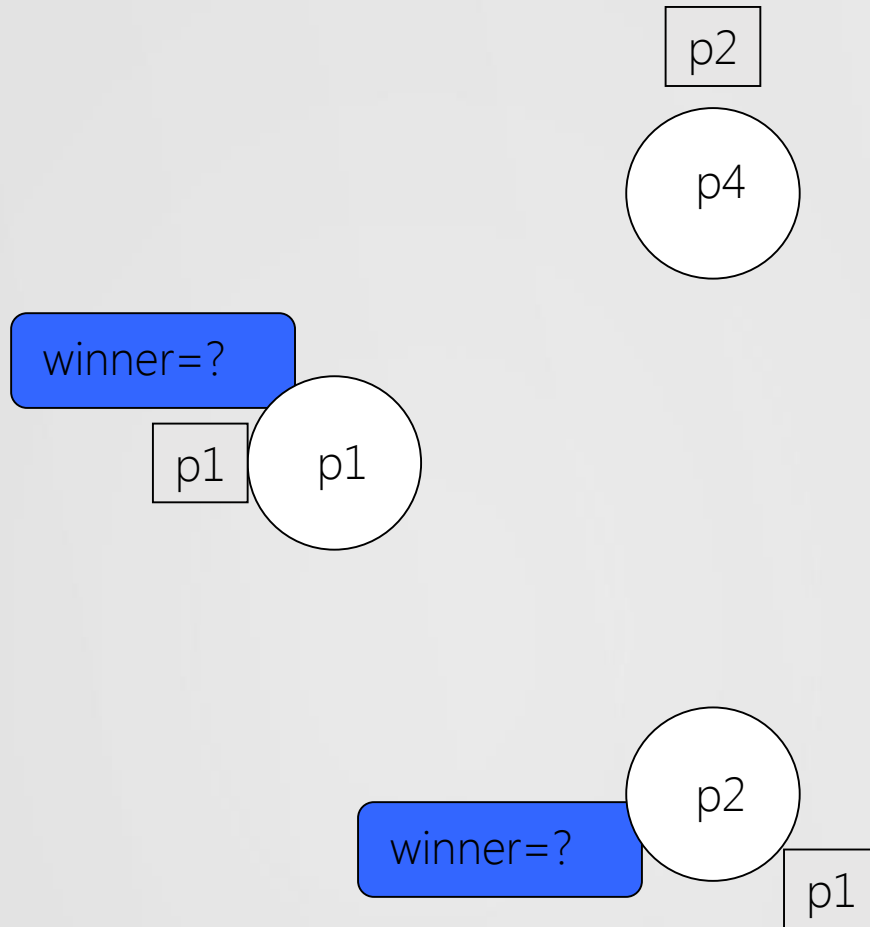
Αλγόριθμος δακτυλίου (Chang & Roberts)



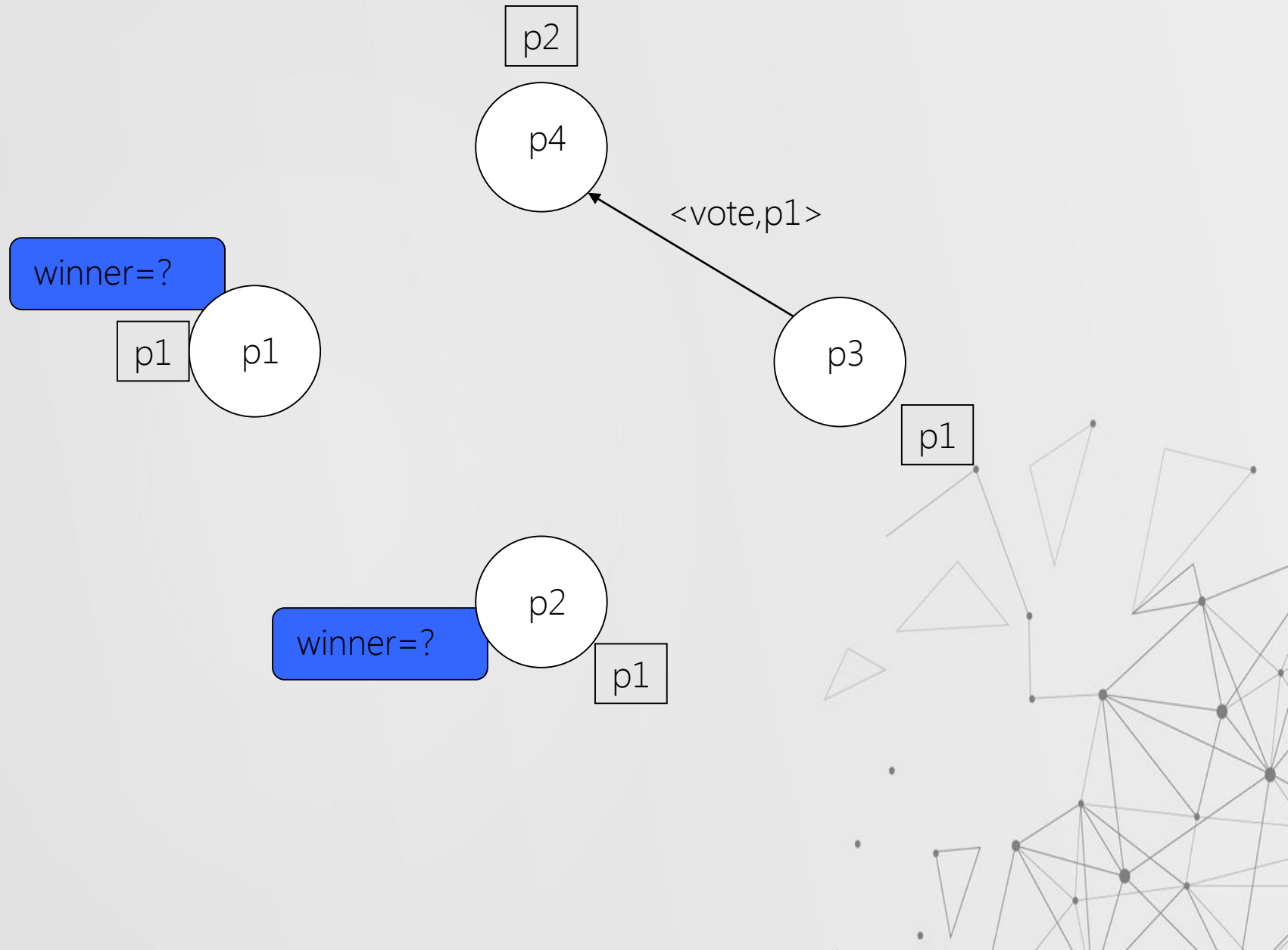
Αλγόριθμος δακτυλίου (Chang & Roberts)



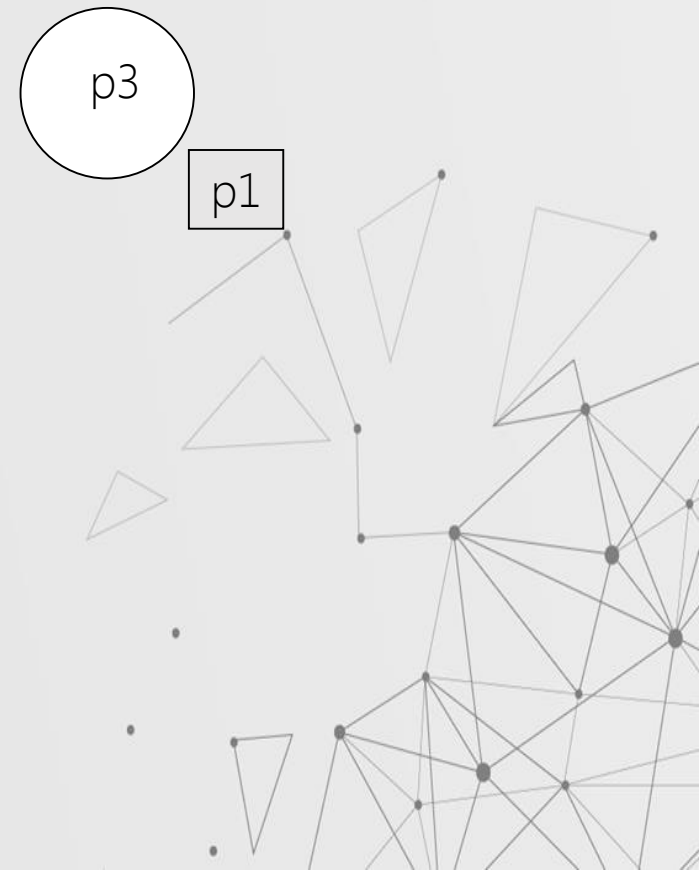
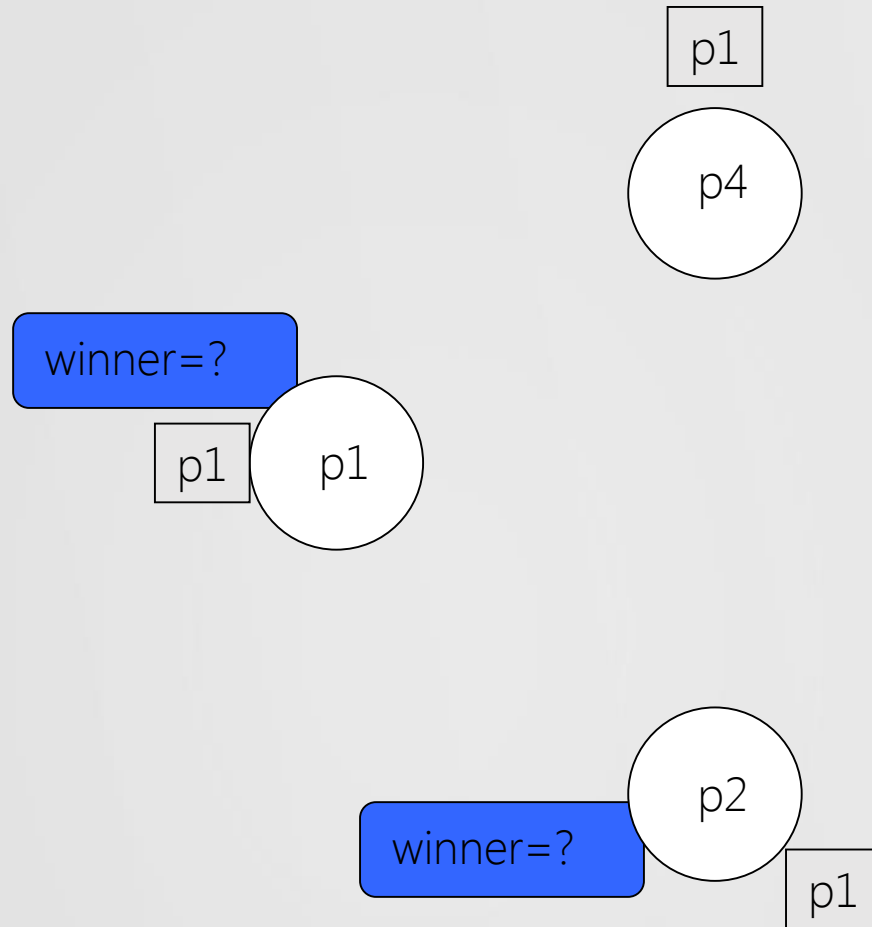
Αλγόριθμος δακτυλίου (Chang & Roberts)



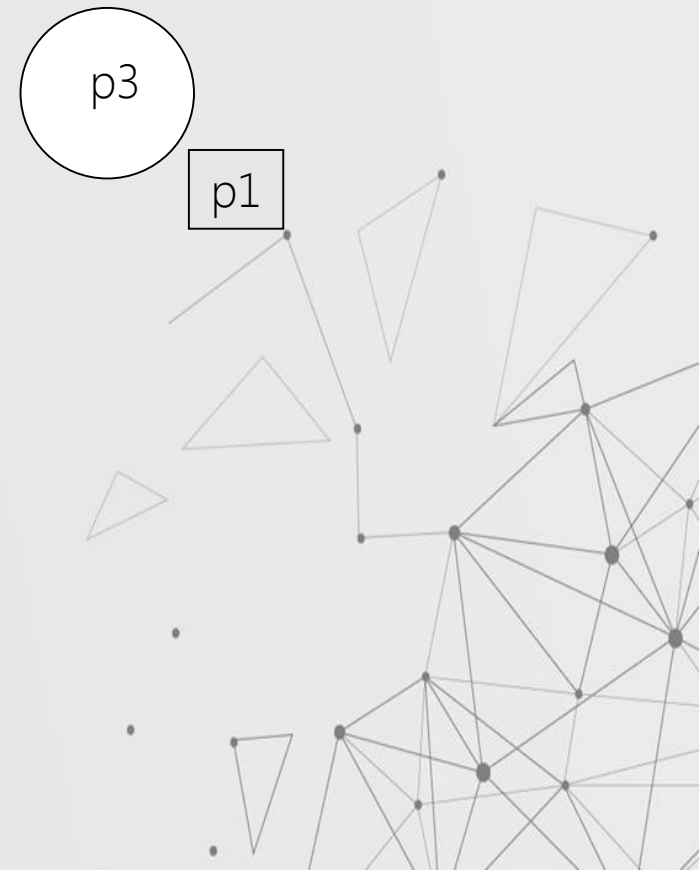
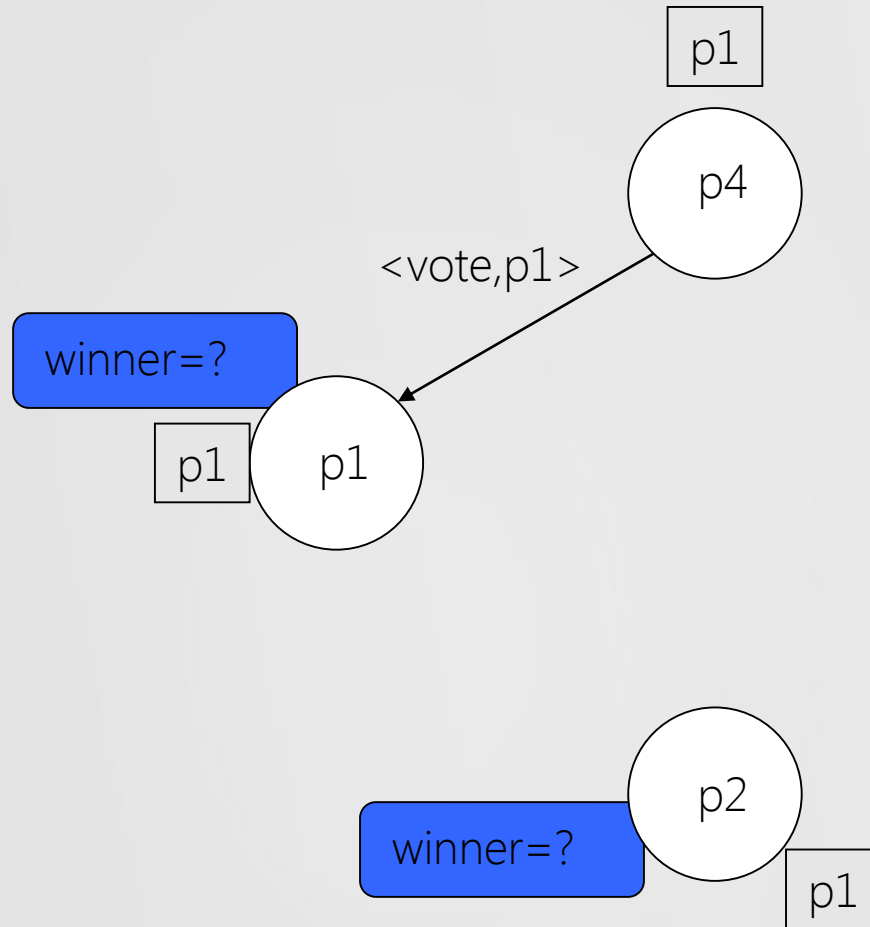
Αλγόριθμος δακτυλίου (Chang & Roberts)



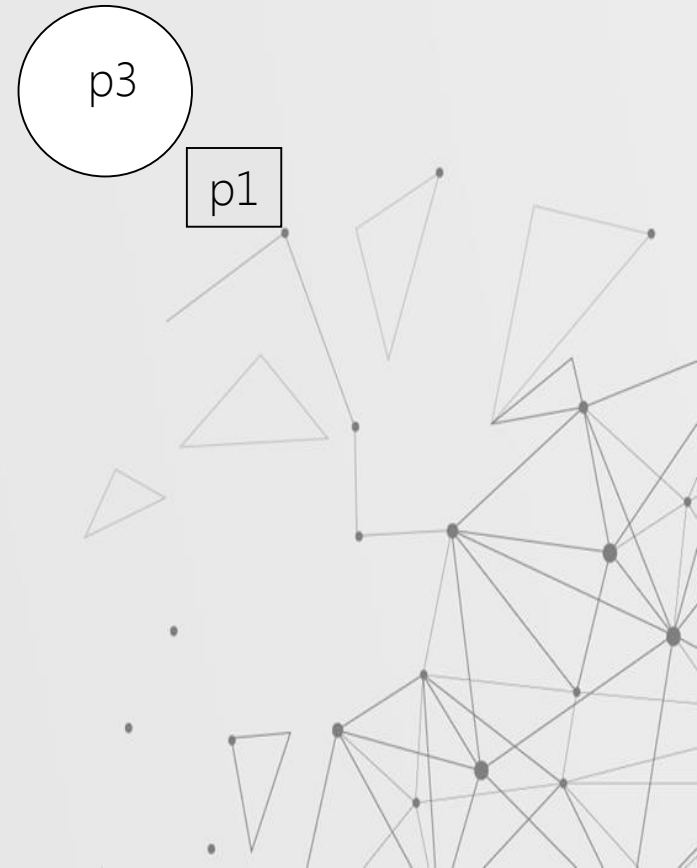
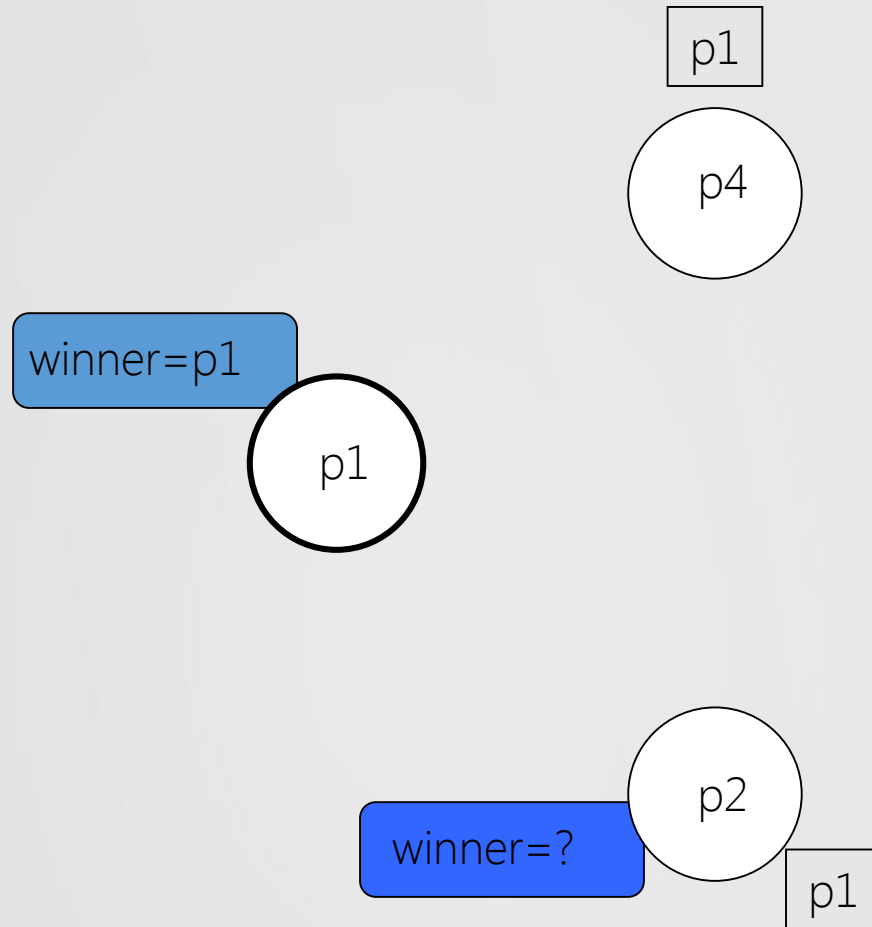
Αλγόριθμος δακτυλίου (Chang & Roberts)



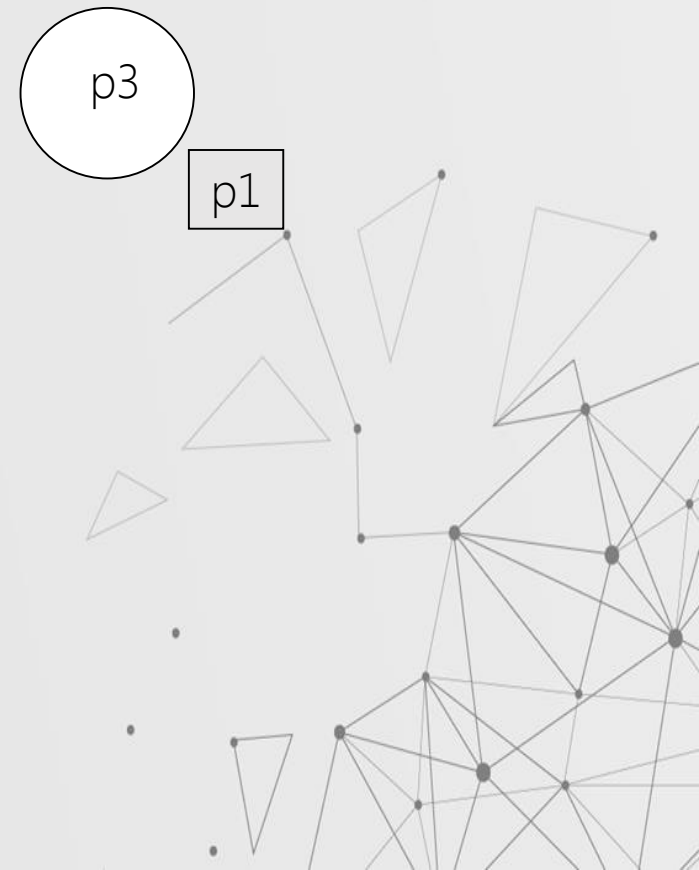
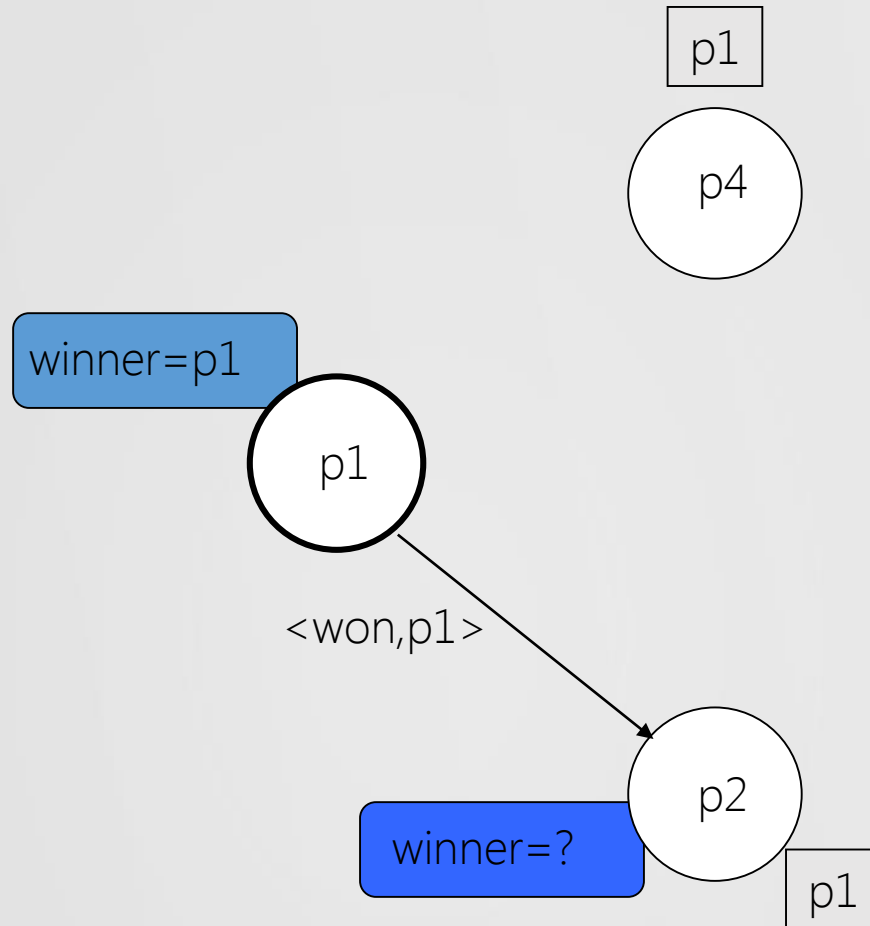
Αλγόριθμος δακτυλίου (Chang & Roberts)



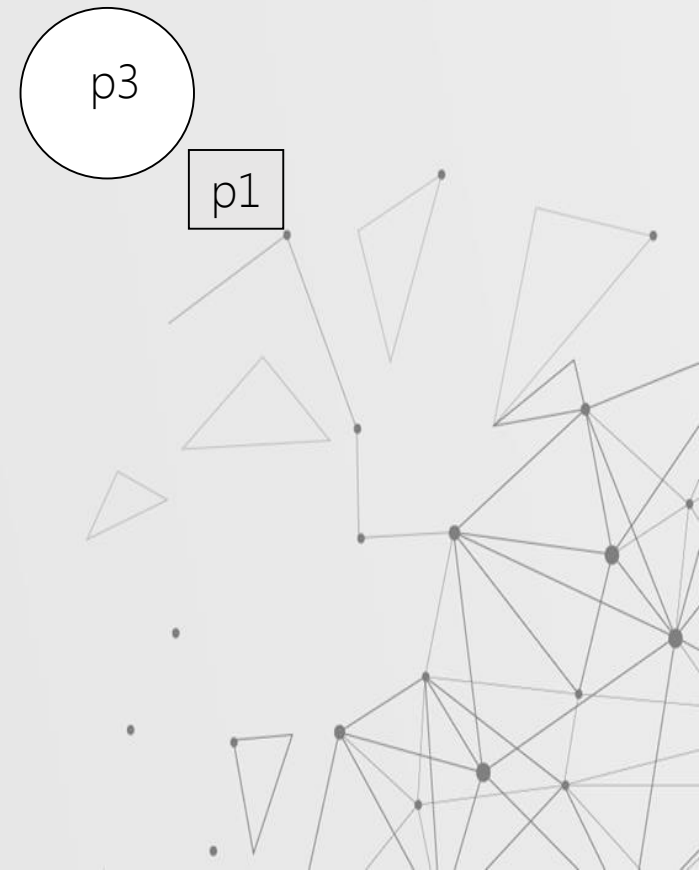
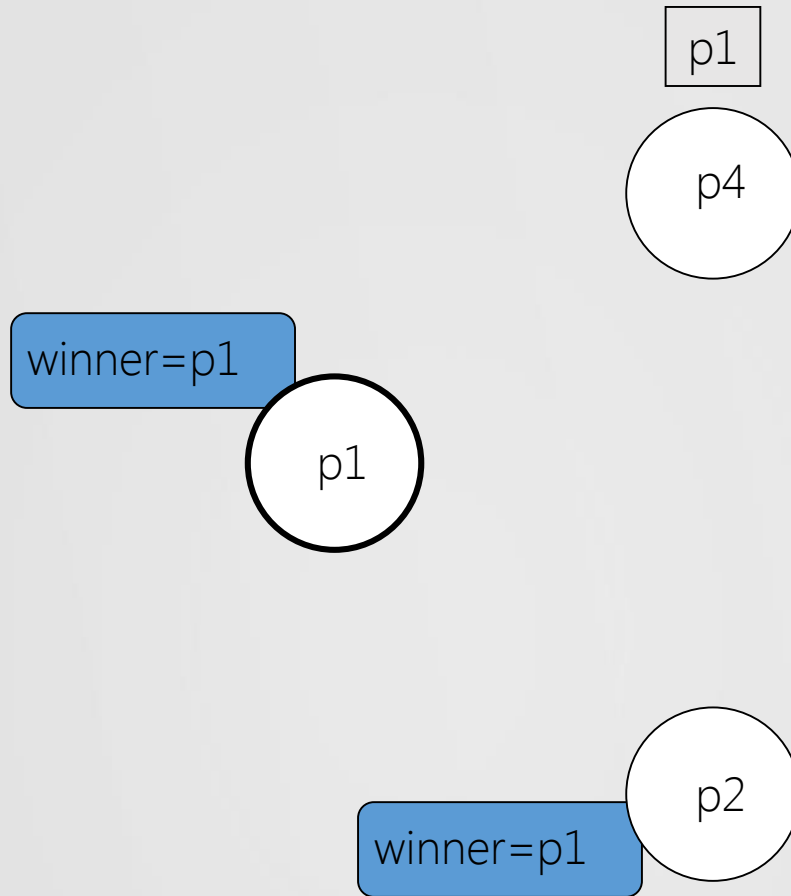
Αλγόριθμος δακτυλίου (Chang & Roberts)



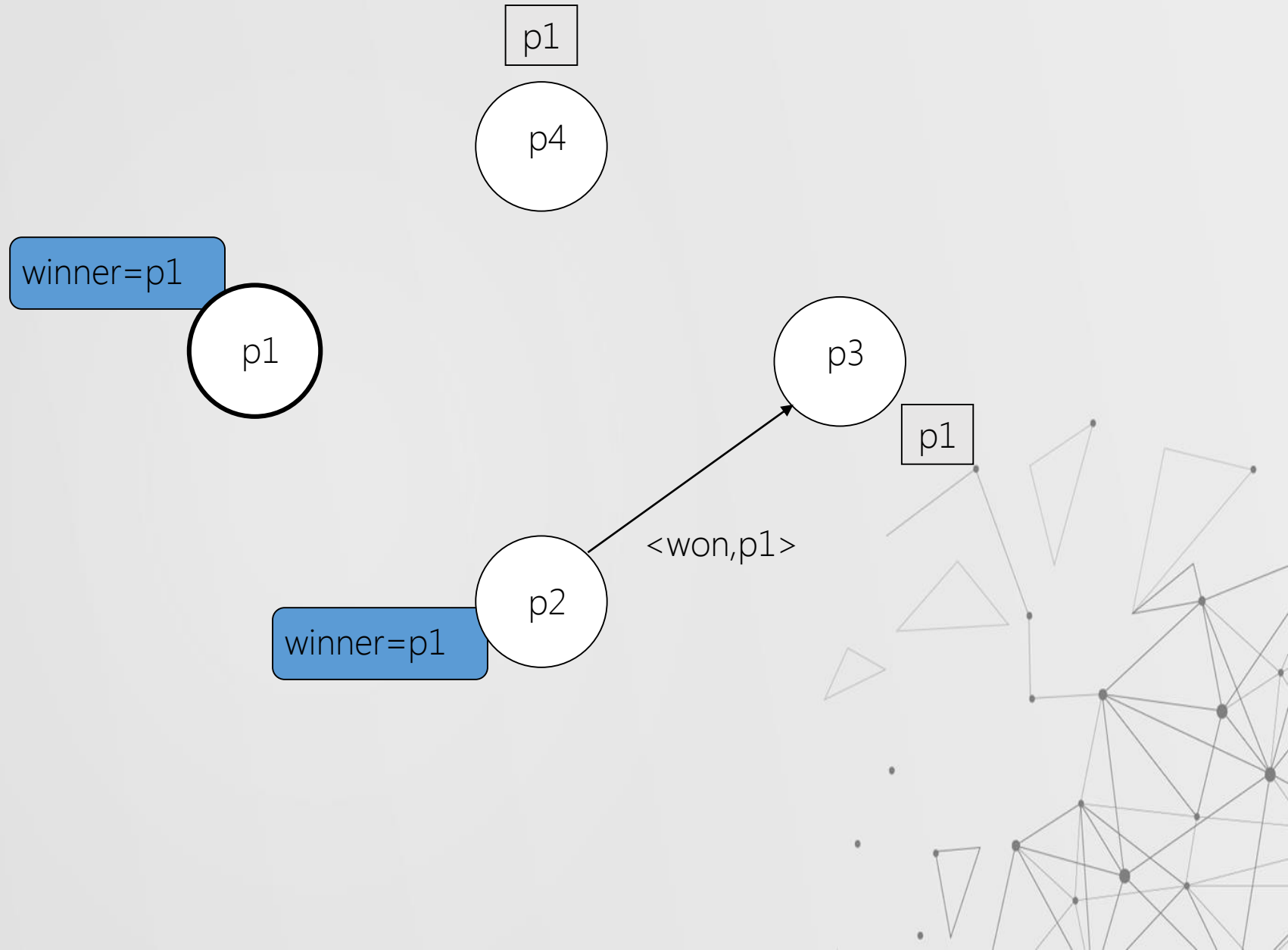
Αλγόριθμος δακτυλίου (Chang & Roberts)



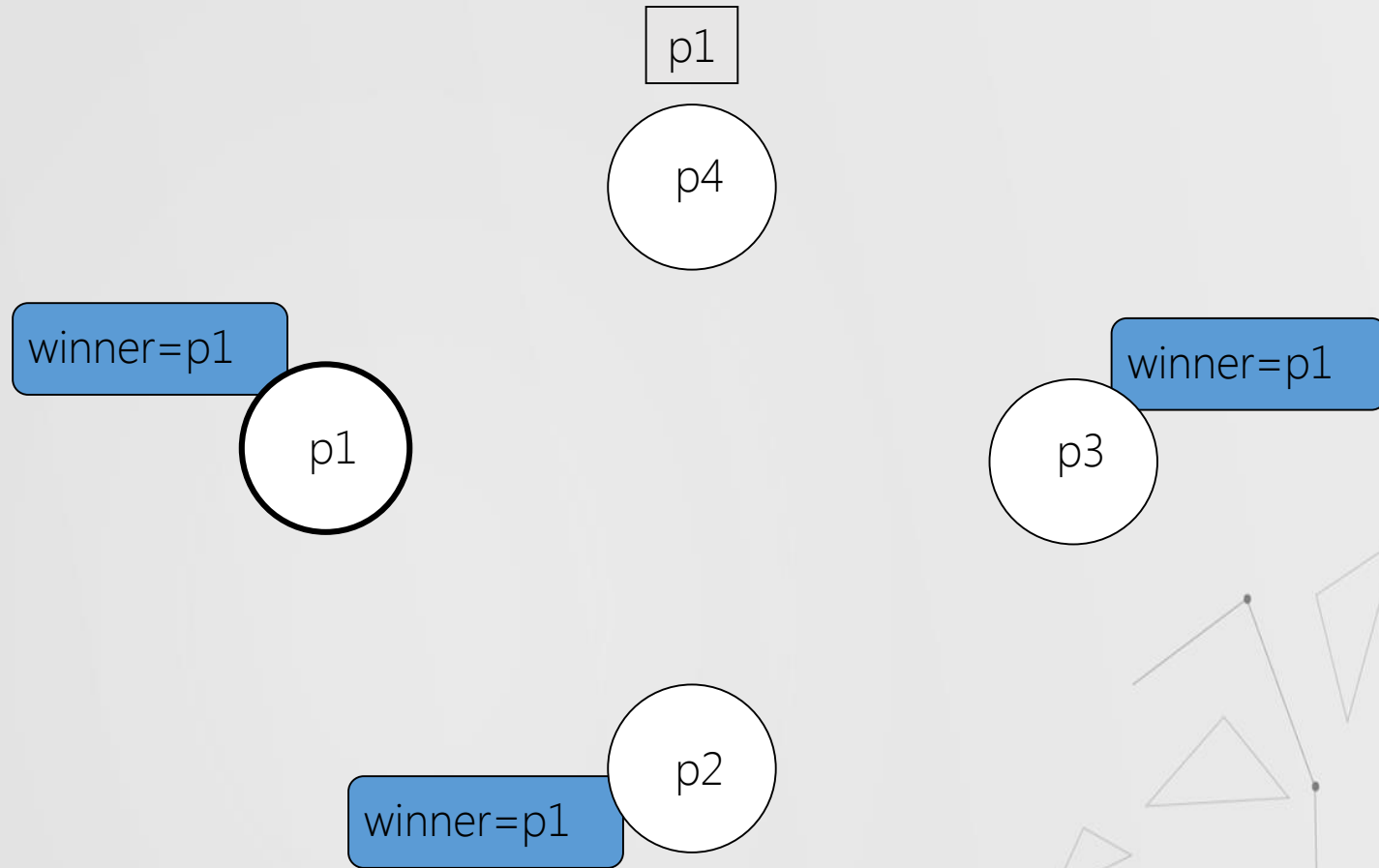
Αλγόριθμος δακτυλίου (Chang & Roberts)



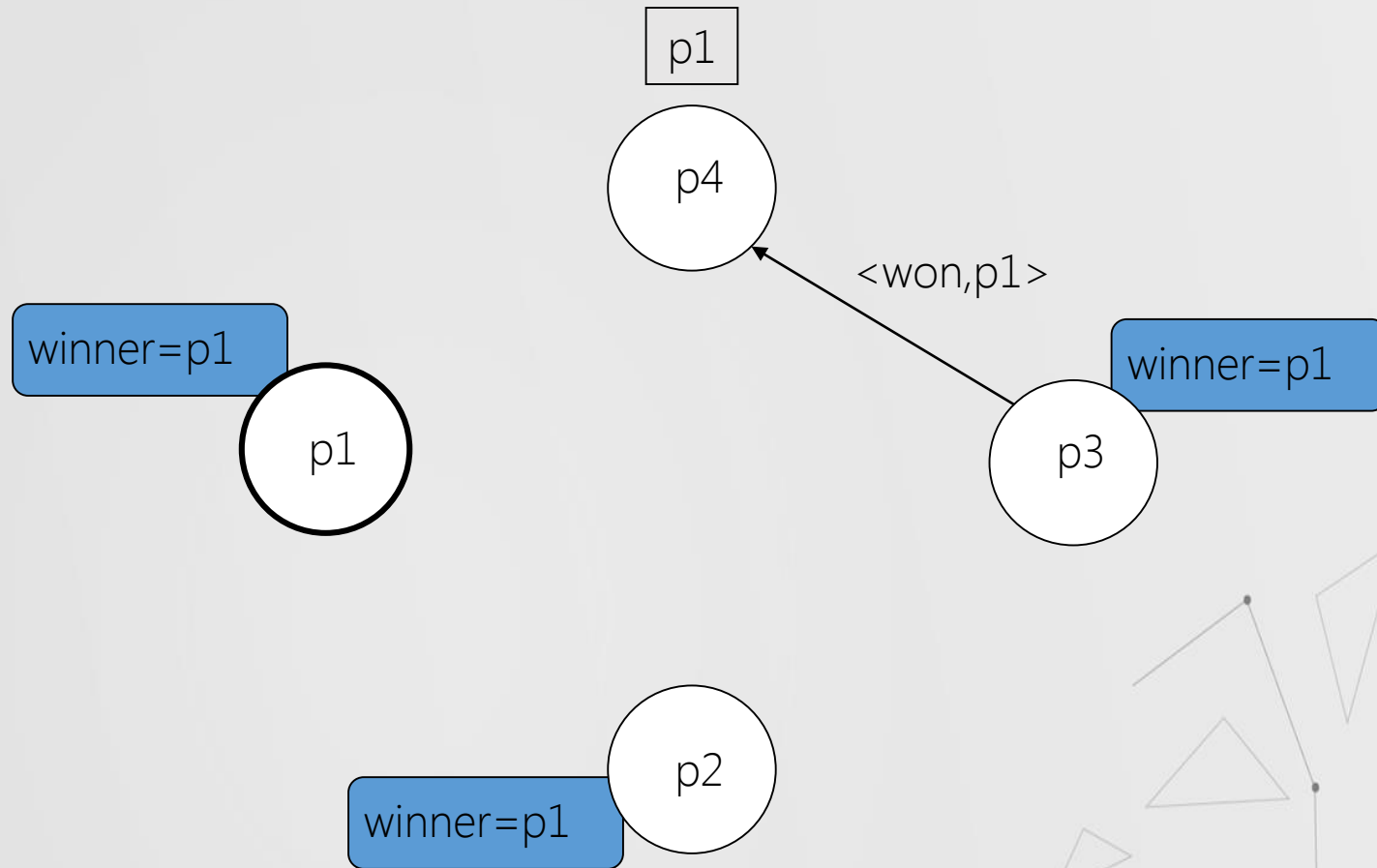
Αλγόριθμος δακτυλίου (Chang & Roberts)



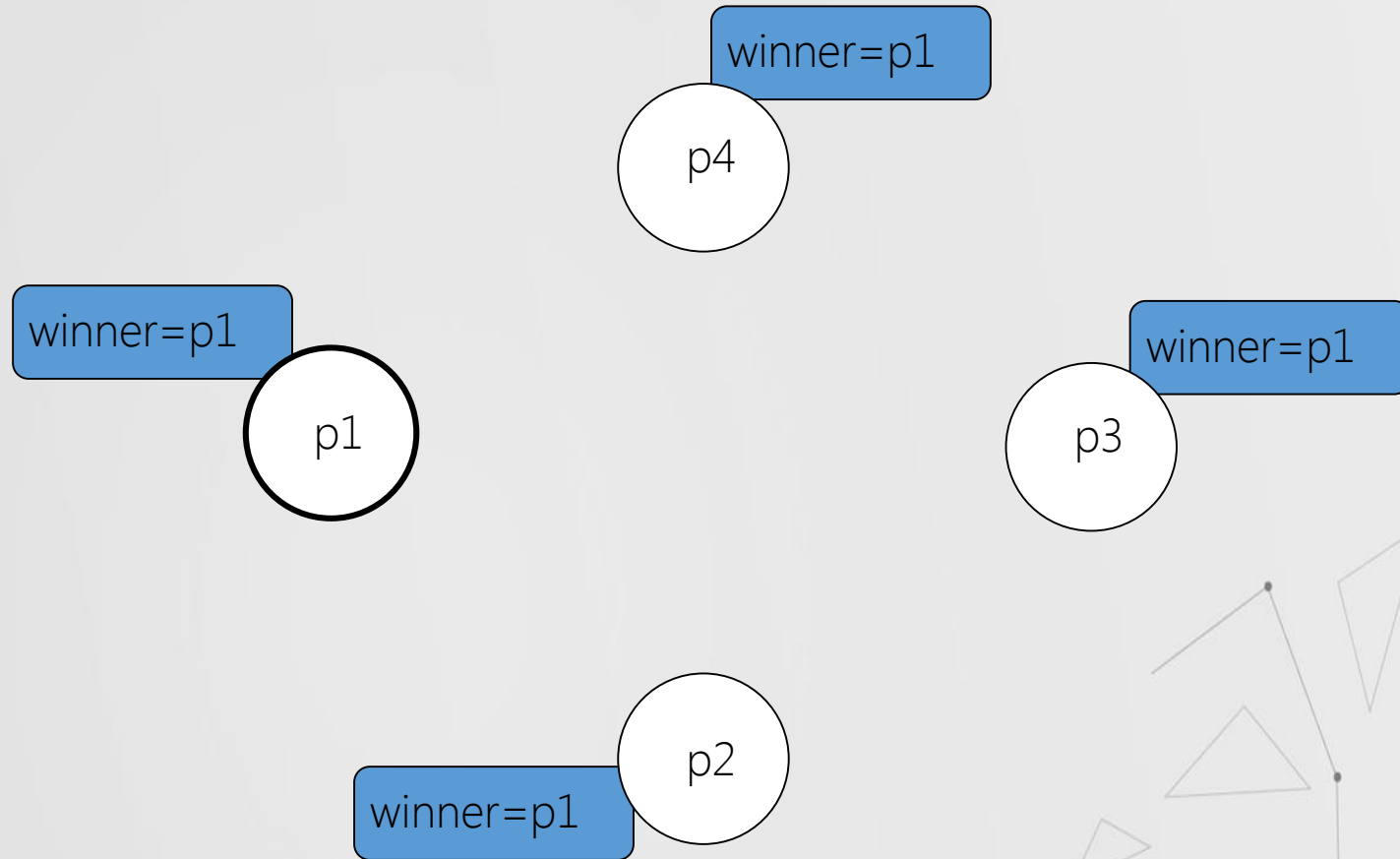
Αλγόριθμος δακτυλίου (Chang & Roberts)



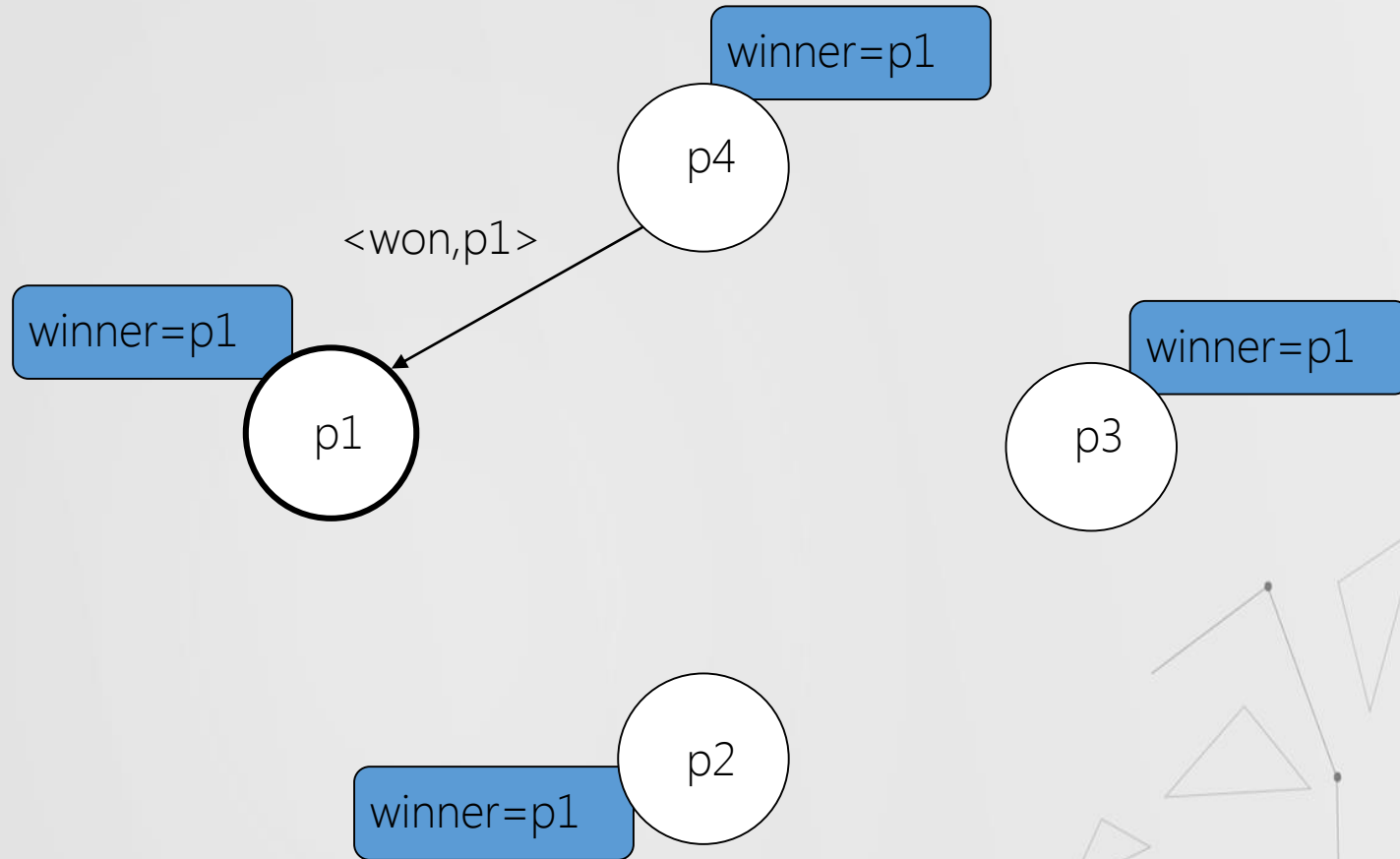
Αλγόριθμος δακτυλίου (Chang & Roberts)



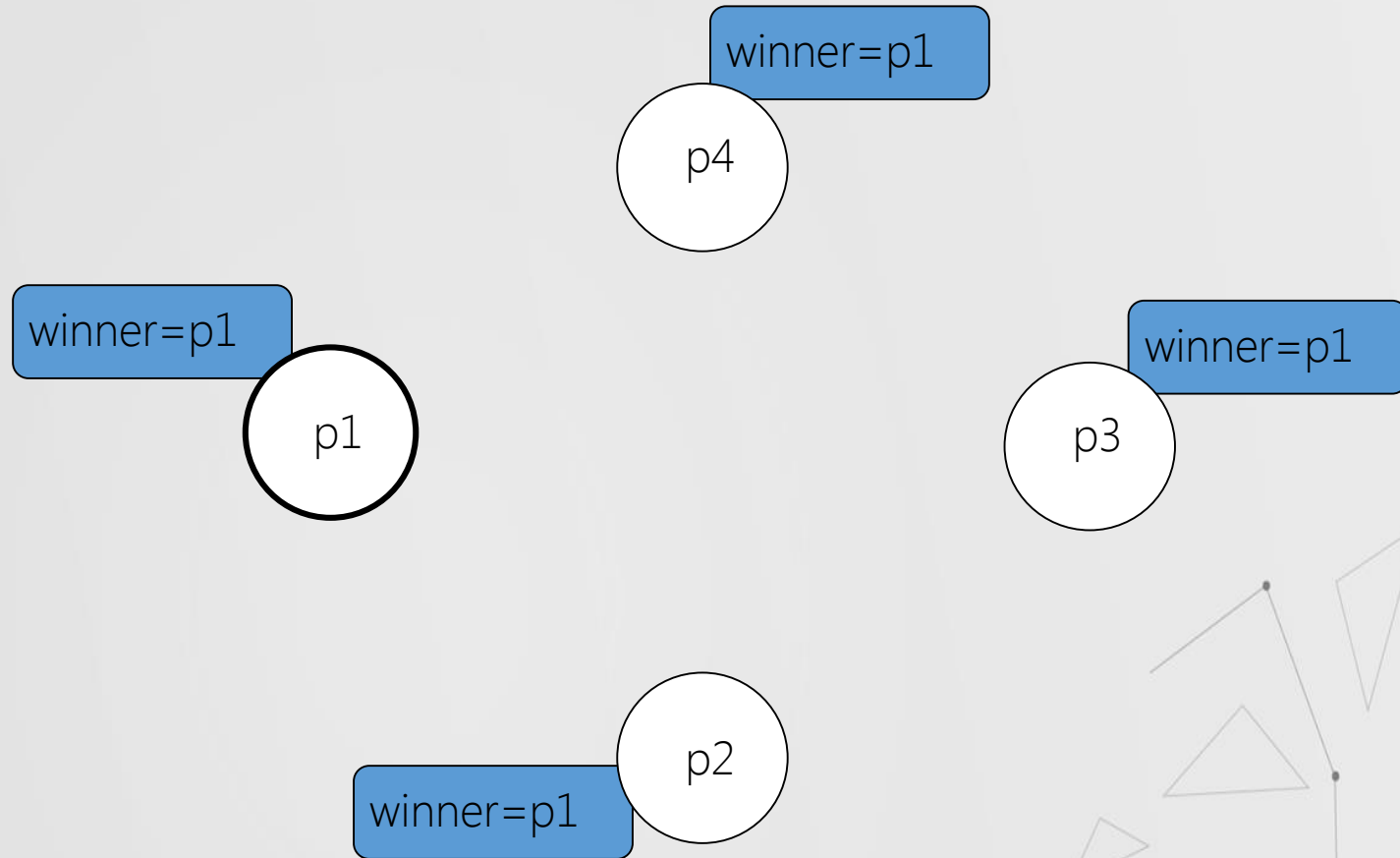
Αλγόριθμος δακτυλίου (Chang & Roberts)



Αλγόριθμος δακτυλίου (Chang & Roberts)



Αλγόριθμος δακτυλίου (Chang & Roberts)



Παρατηρήσεις

- Δημιουργούνται, κατά μέσο όρο, λιγότερα μηνύματα σε σχέση με τον αλγόριθμο του LeLann.
- Στη χειρότερη περίπτωση οι διεργασίες είναι διατεταγμένες με αύξουσα σειρά αναγνωριστικού σε σχέση με την κατεύθυνση επικοινωνίας.
 - Απαιτούνται $O(MN)$ μηνύματα για M εκκινητές ($\sim$ LeLann).
- Στην καλύτερη περίπτωση οι διεργασίες είναι διατεταγμένες με φθίνουσα σειρά αναγνωριστικού σε σχέση με την κατεύθυνση επικοινωνίας.
 - Απαιτούνται $N + M - 1 + N$ μηνύματα για M εκκινητές – $O(N)$.
- Αν μόνο μια διεργασία είναι εκκινητής απαιτούνται $2N$ μηνύματα όπως και στον αλγόριθμο του LeLann.

05

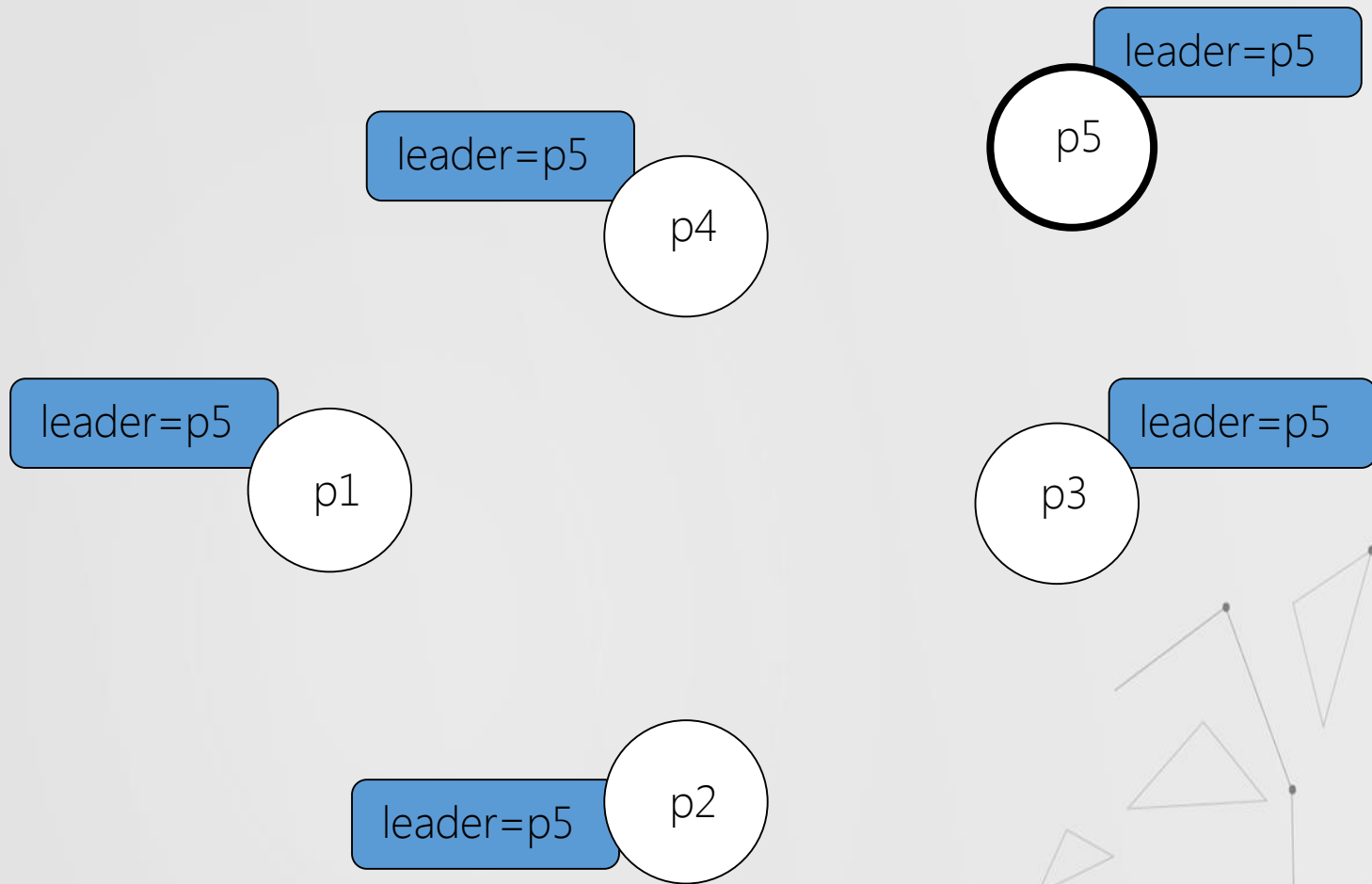
Αλγόριθμος ισχυροτέρου
(Garcia-Molina)



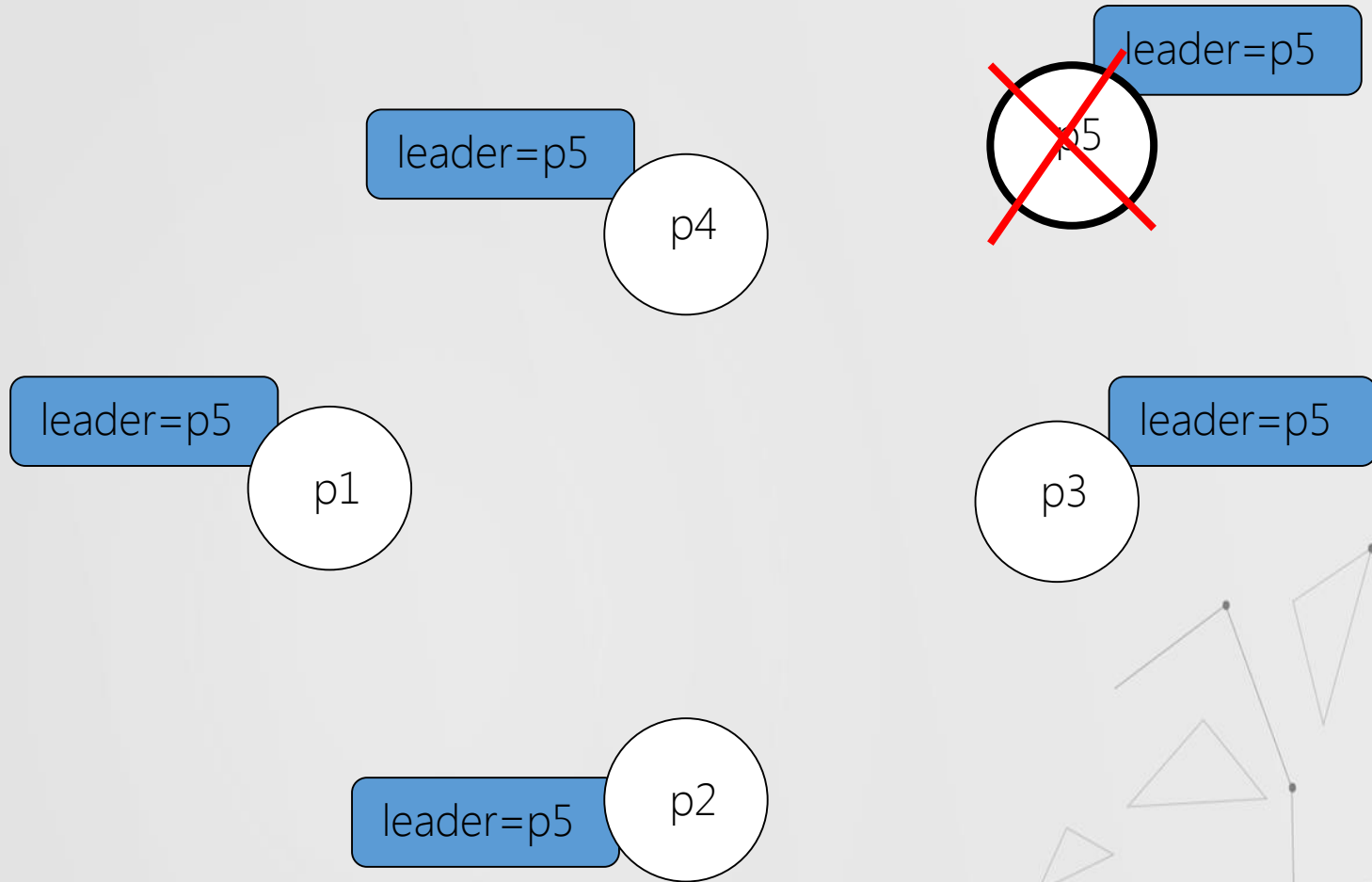
Αλγόριθμος ισχυροτέρου (Garcia-Molina)

- Υποθέτουμε σύγχρονο σύστημα.
- Όταν μια διεργασία θεωρήσει ότι ο αρχηγός έχει παρουσιάσει βλάβη εκπέμπει μήνυμα υποψηφιότητας.
- Κάθε διεργασία που λαμβάνει μήνυμα υποψηφιότητας με μικρότερο αναγνωριστικό από το δικό της, εκπέμπει δικό της μήνυμα υποψηφιότητας.
- Αν μια διεργασία δεν λάβει νέο μήνυμα υποψηφιότητας μέσα σε συγκεκριμένο χρονικό διάστημα, επιλέγει ως αρχηγό την διεργασία με το μεγαλύτερο αναγνωριστικό.
- Τερματισμός το αργότερο σε $NT_{\max}$ από την έναρξη της διαδικασίας ή σε $2T_{\max}$ αν το δίκτυο υποστηρίζει παράλληλη εκπομπή από όλες τις διεργασίες (π.χ. μέσω ξεχωριστών συνδέσεων ανάμεσα σε κάθε ζευγάρι).

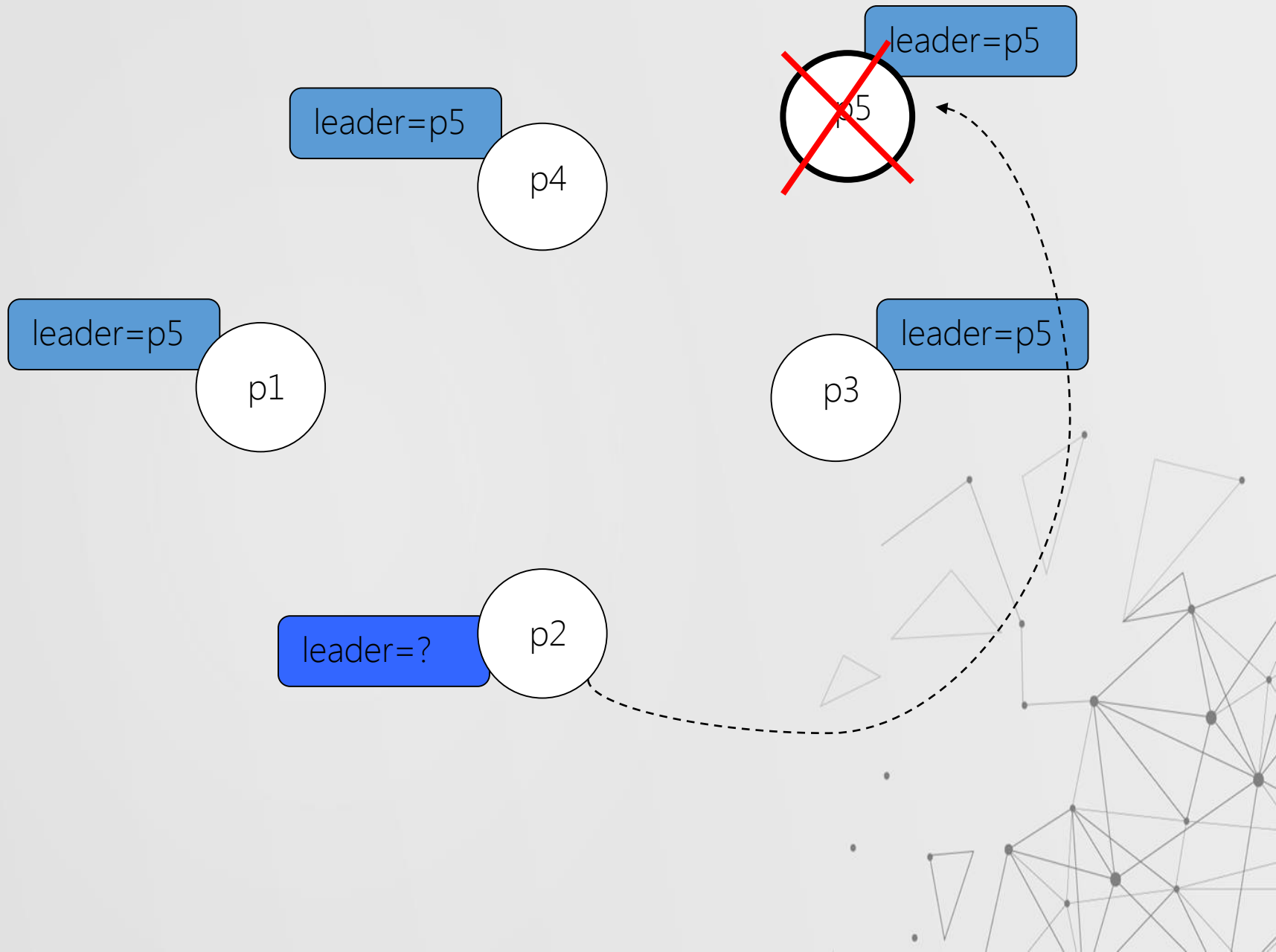
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



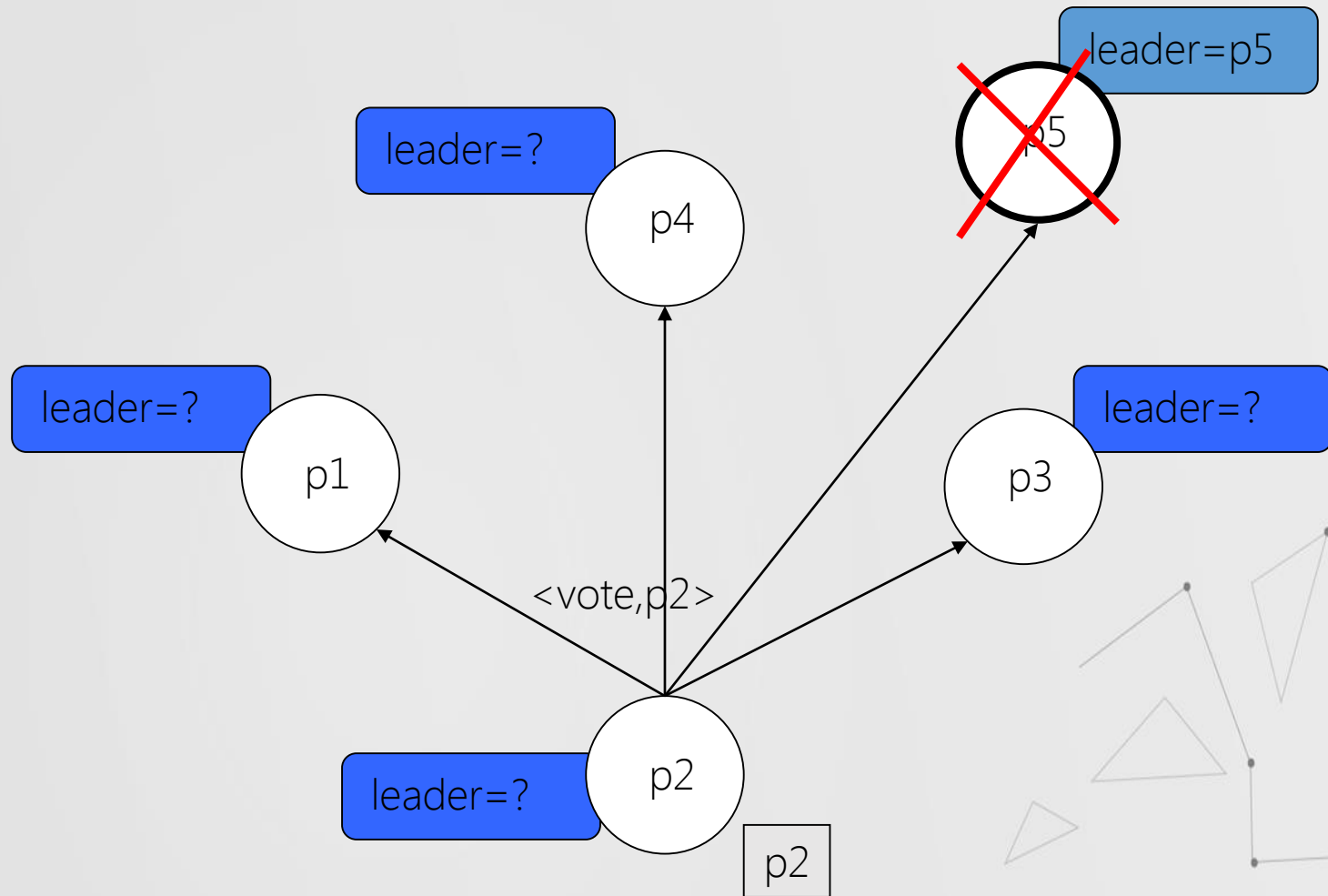
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



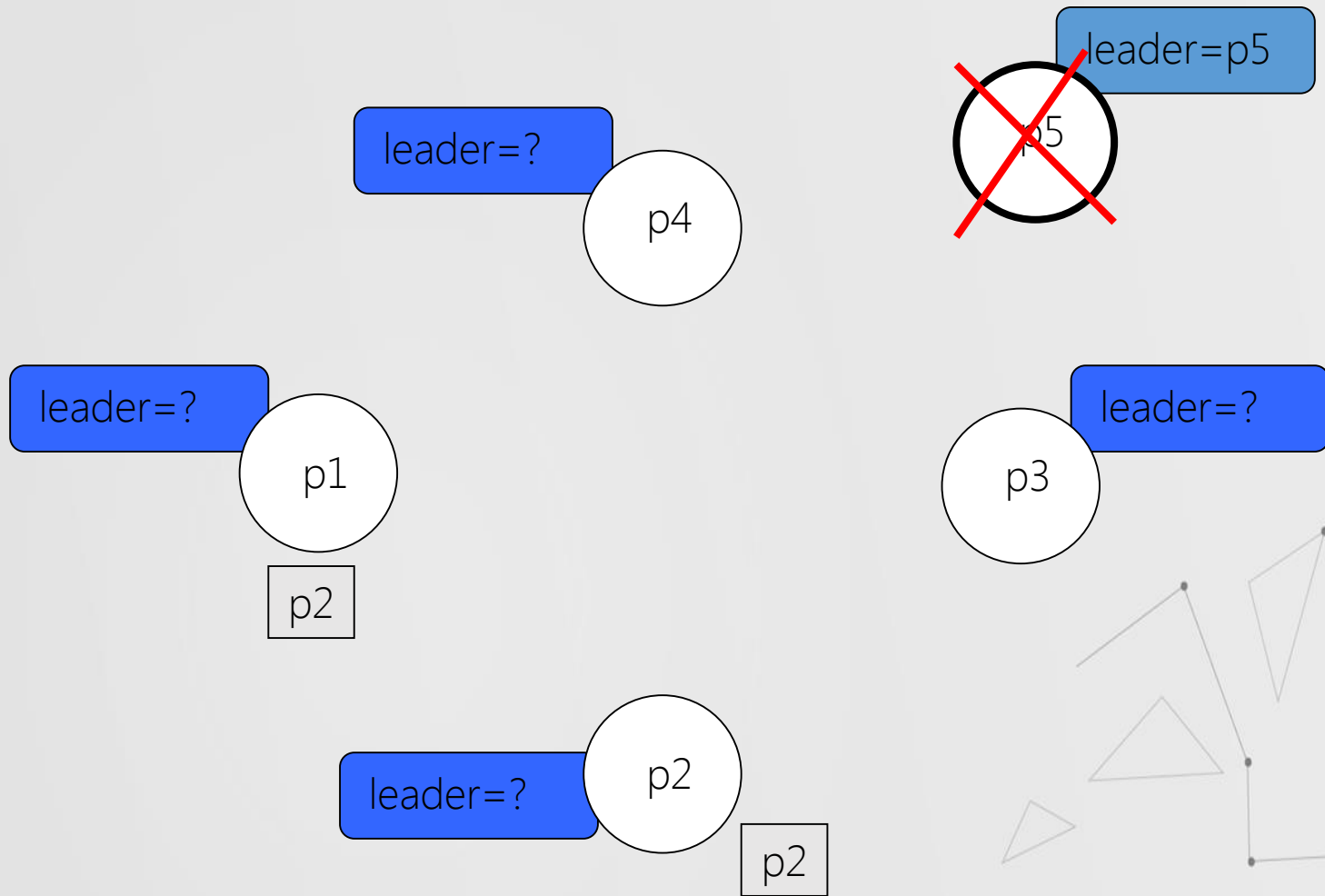
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



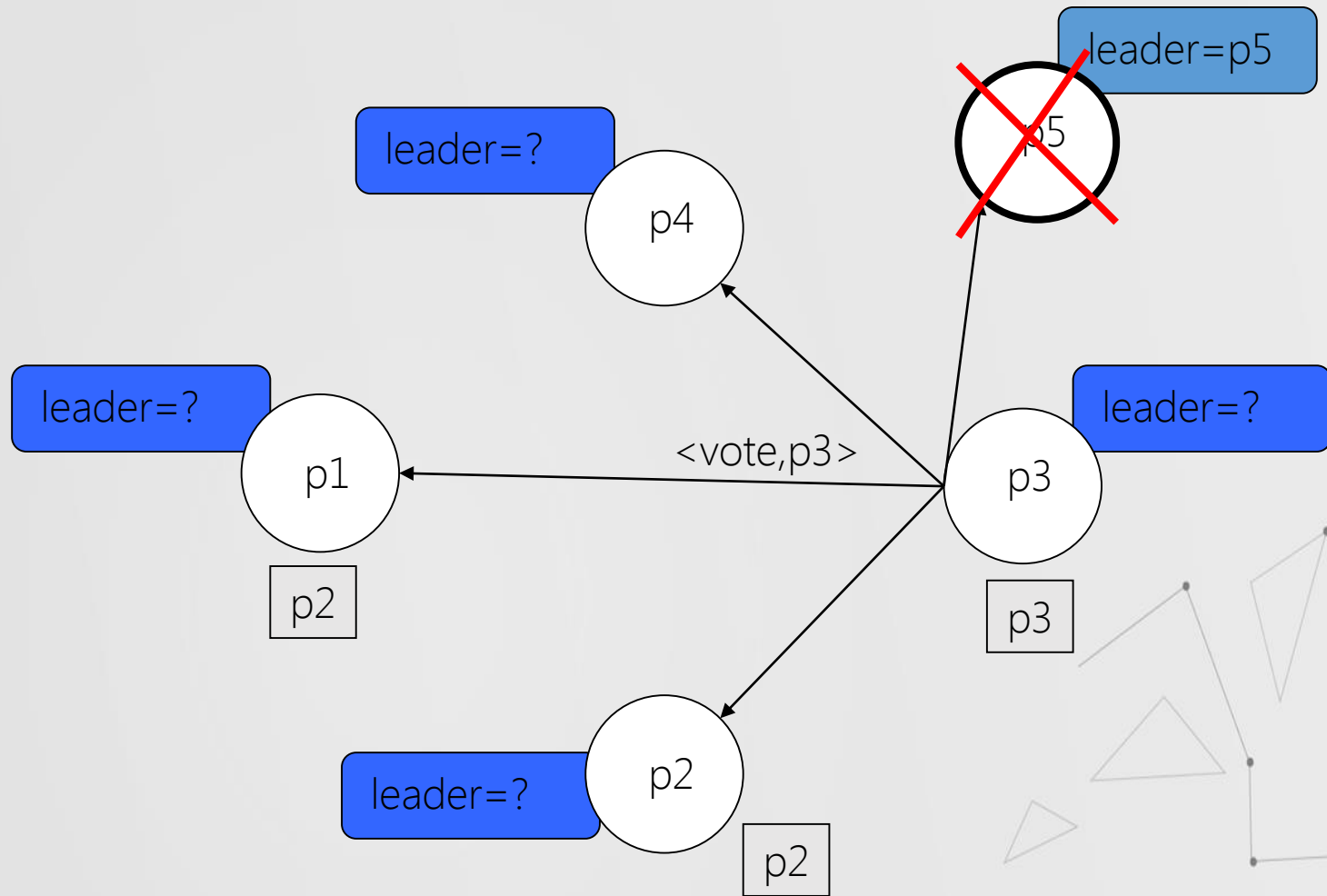
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



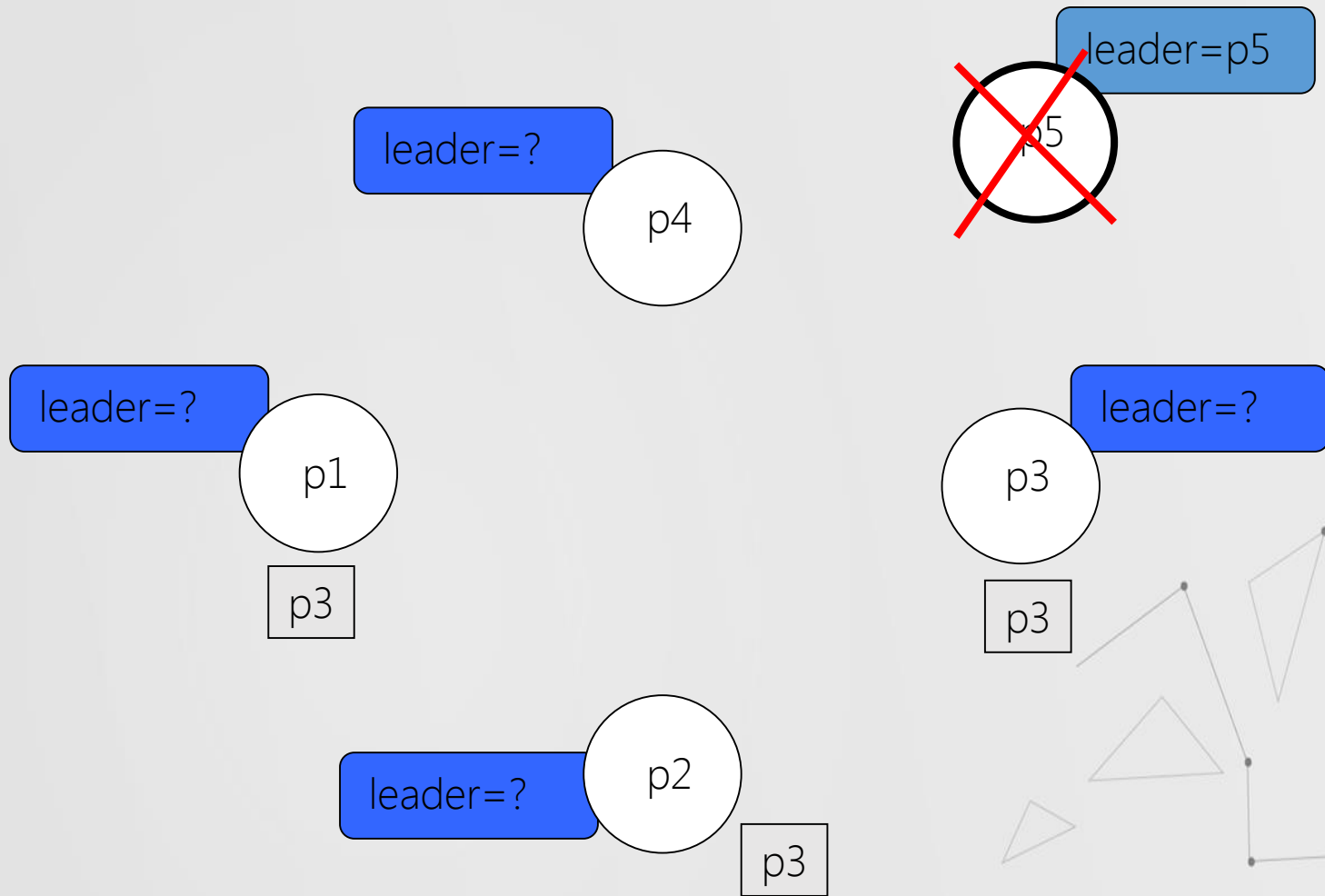
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



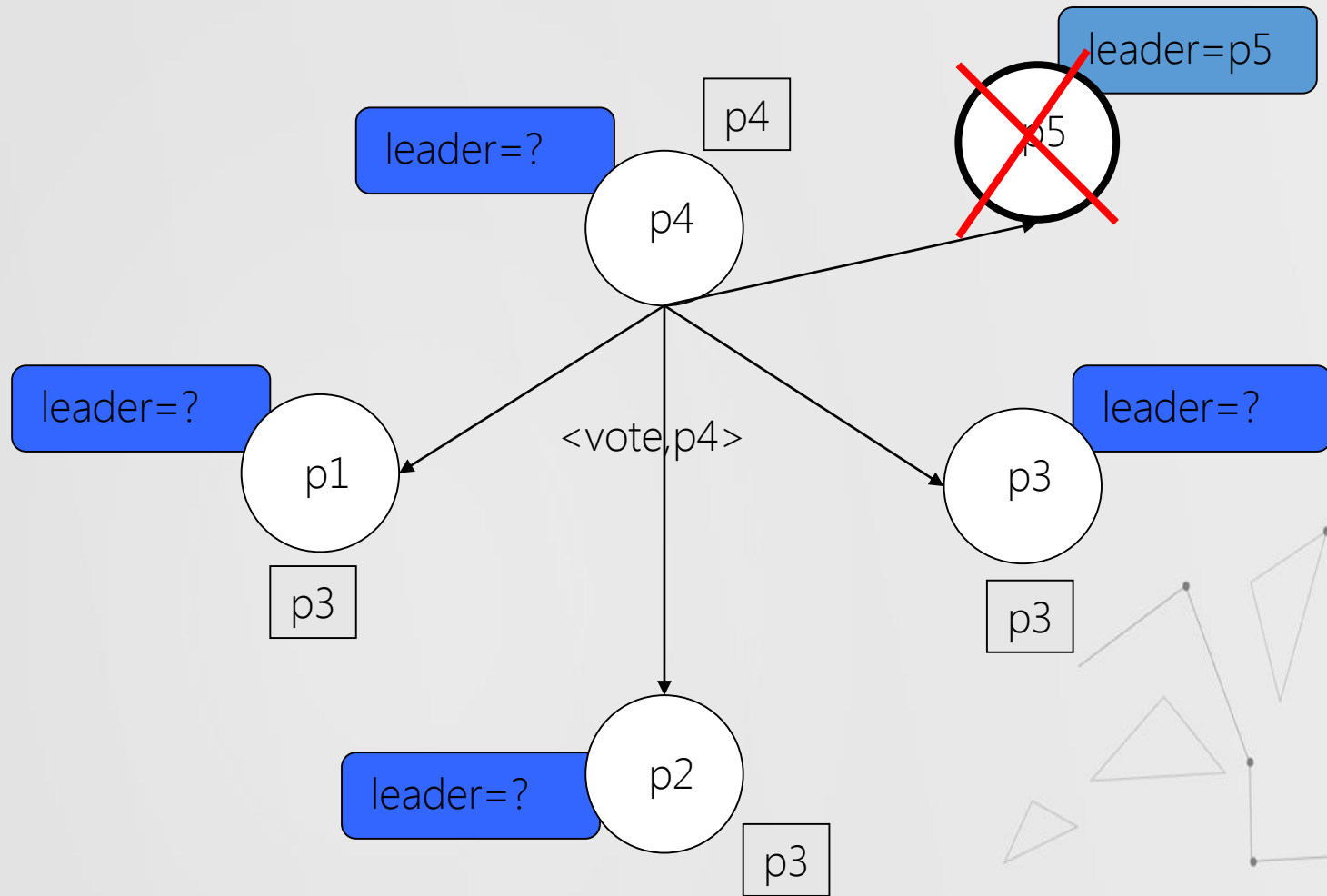
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



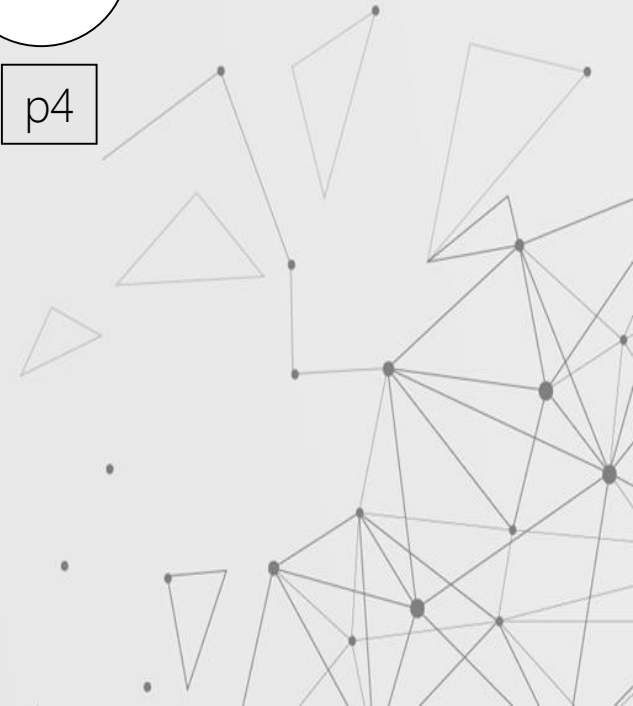
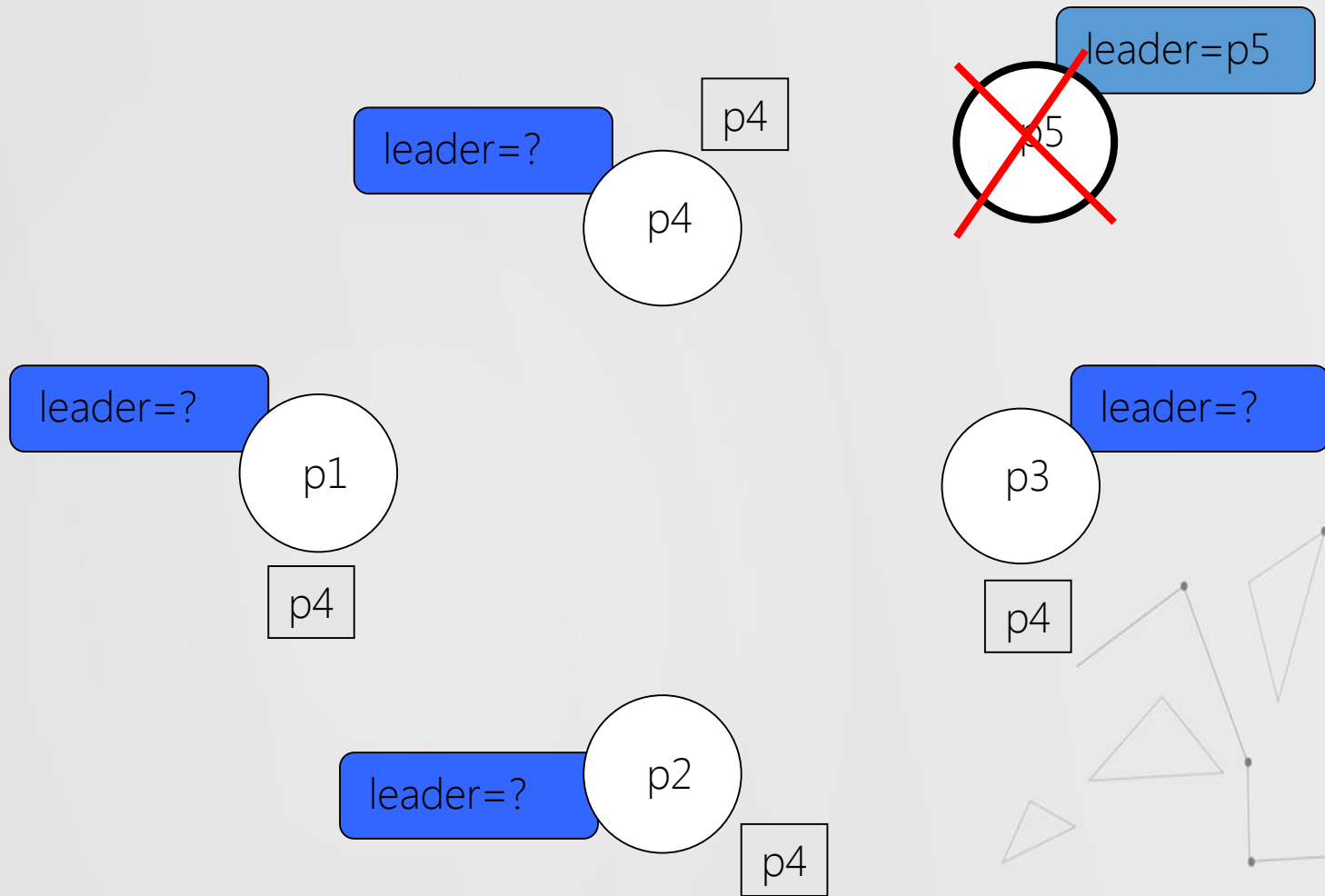
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



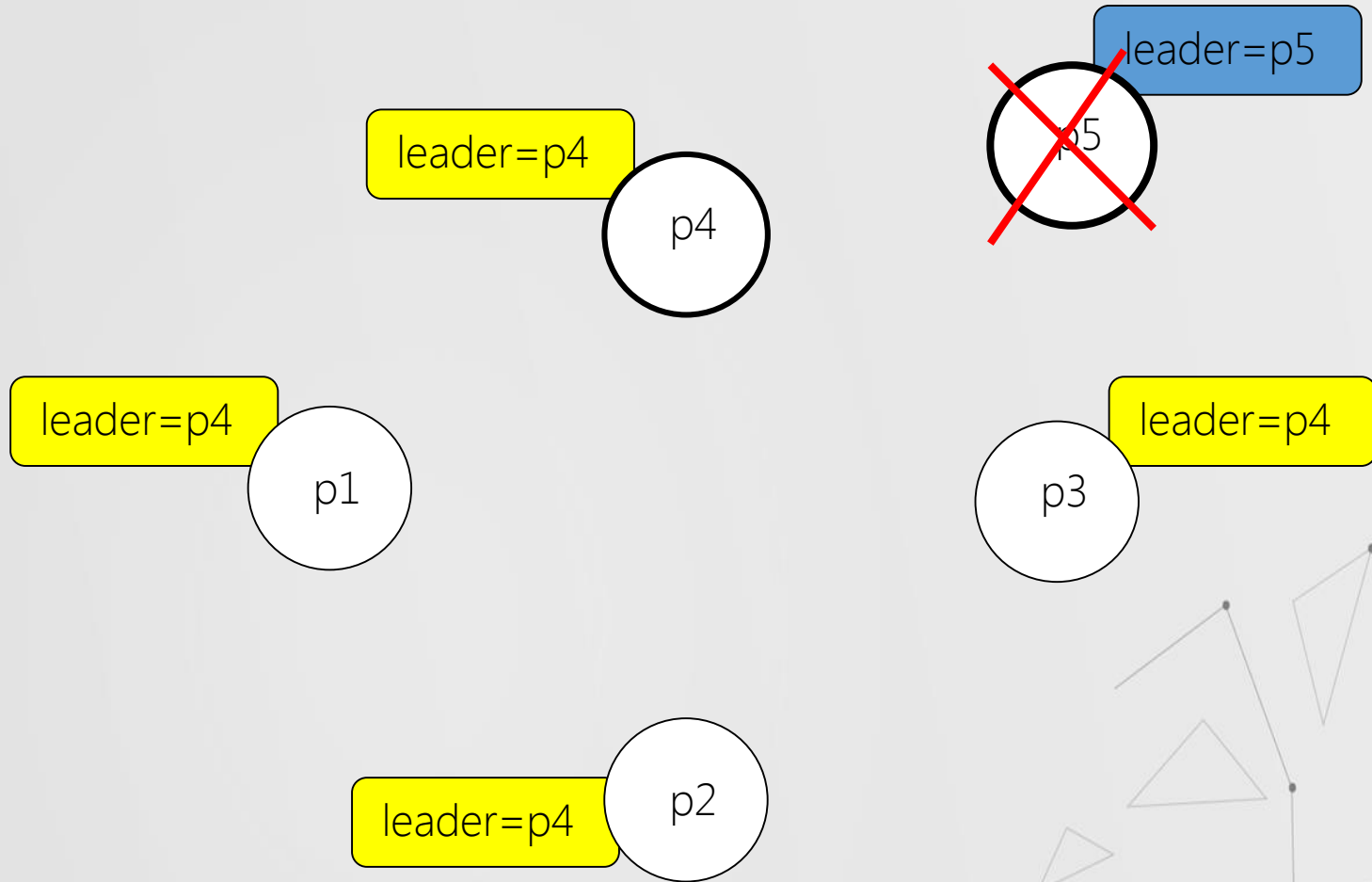
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



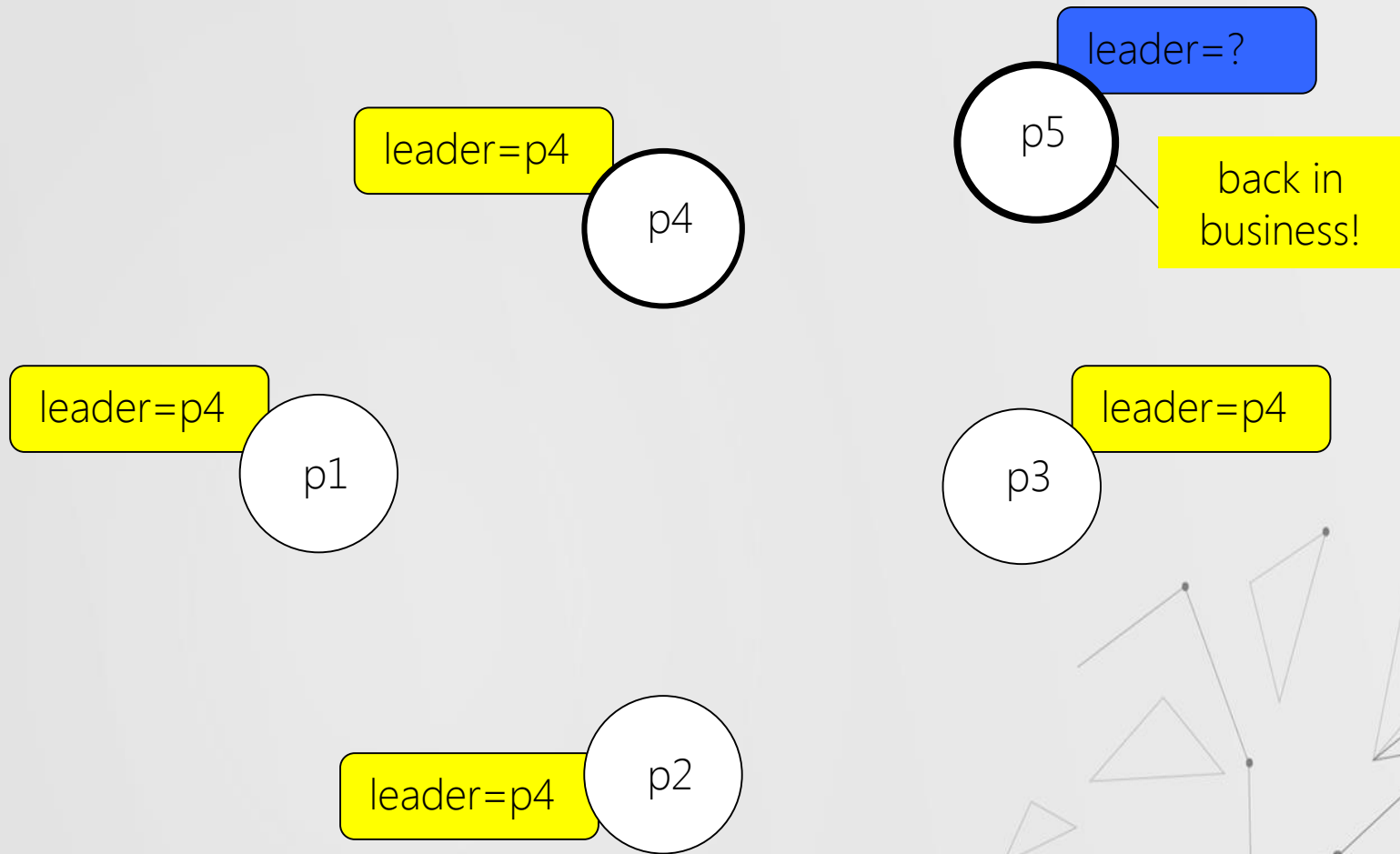
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



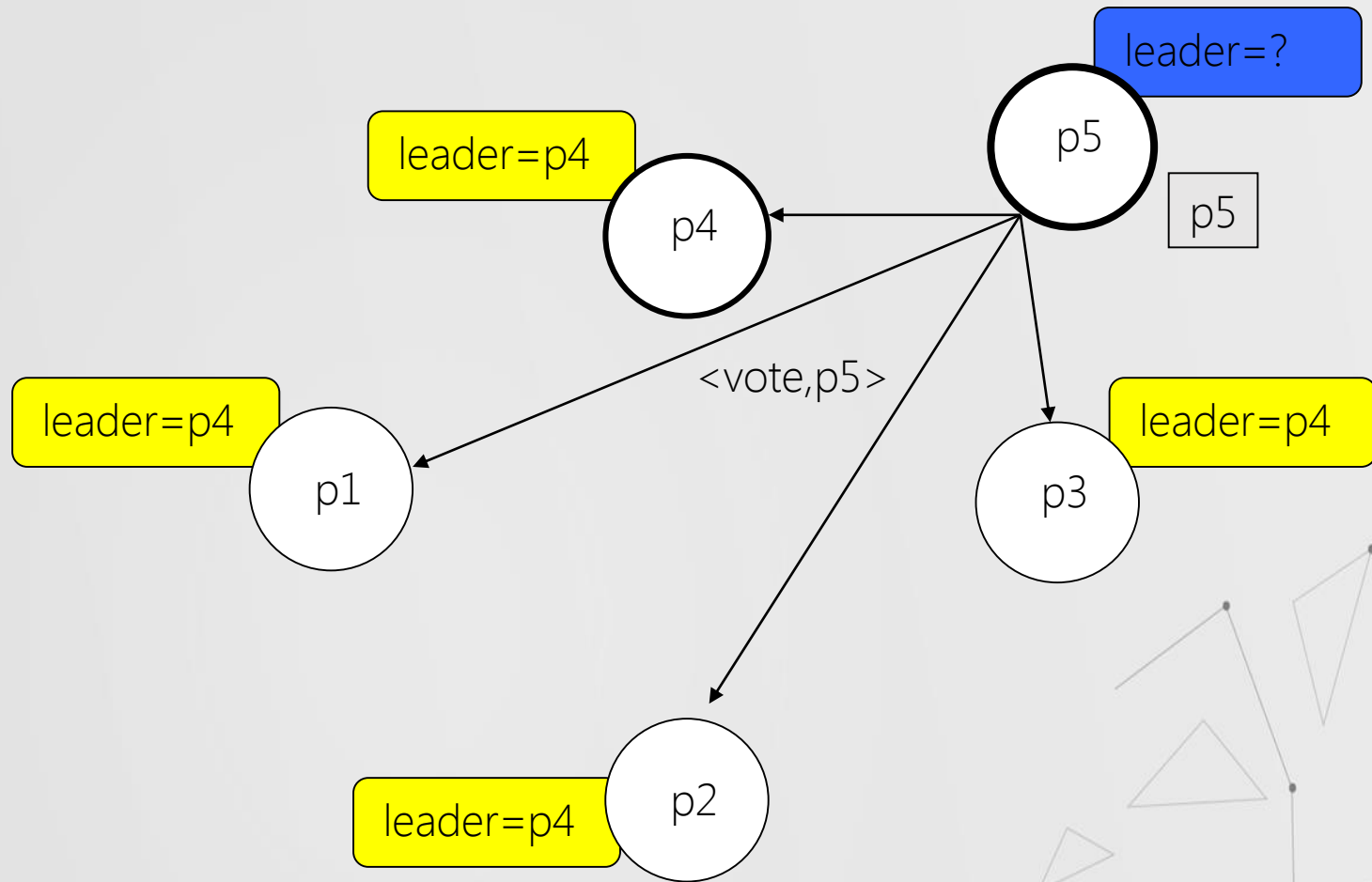
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



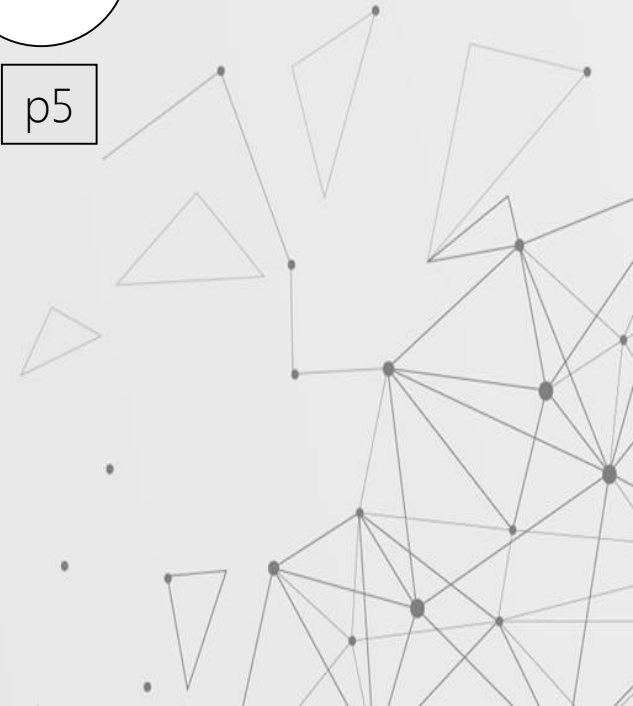
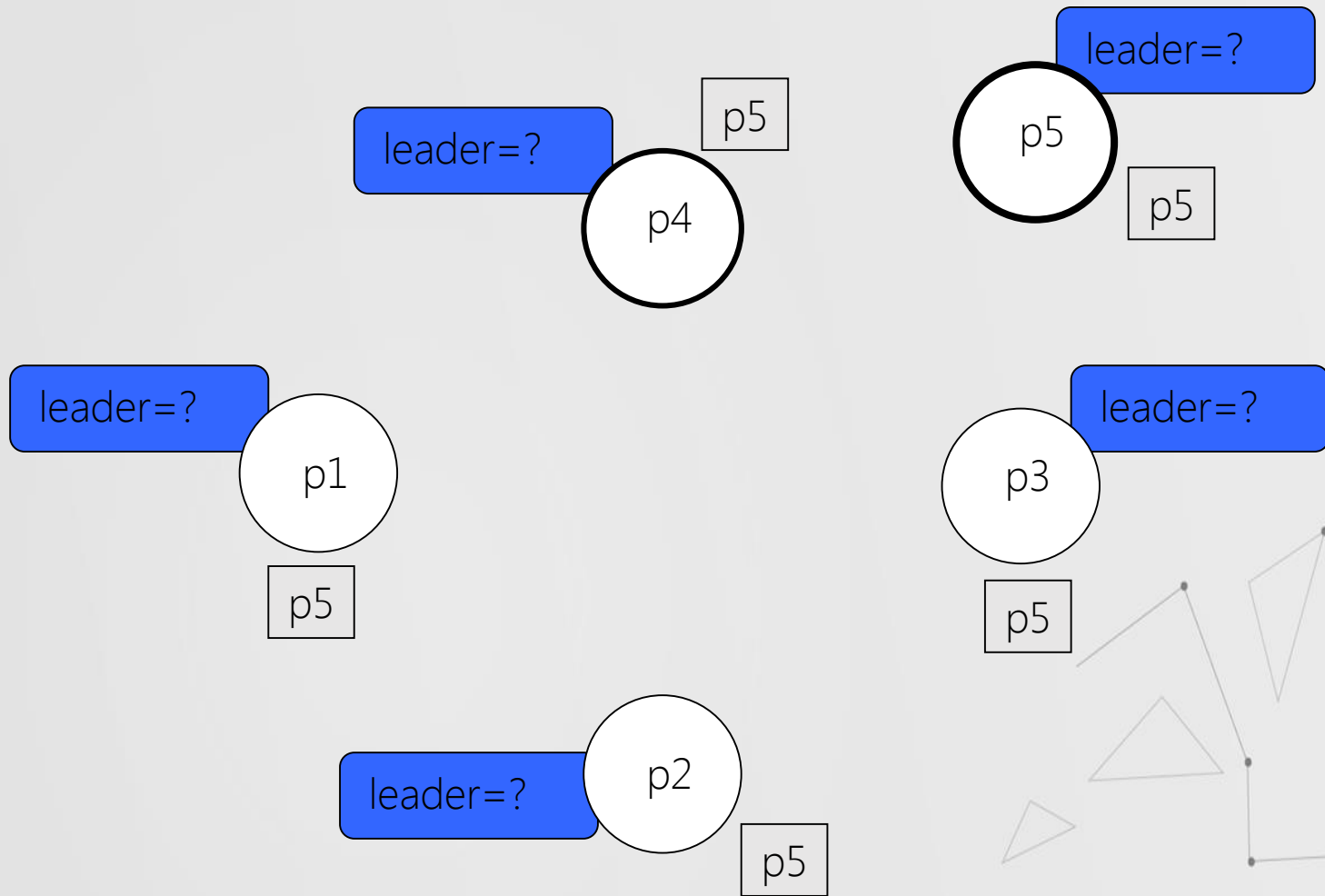
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



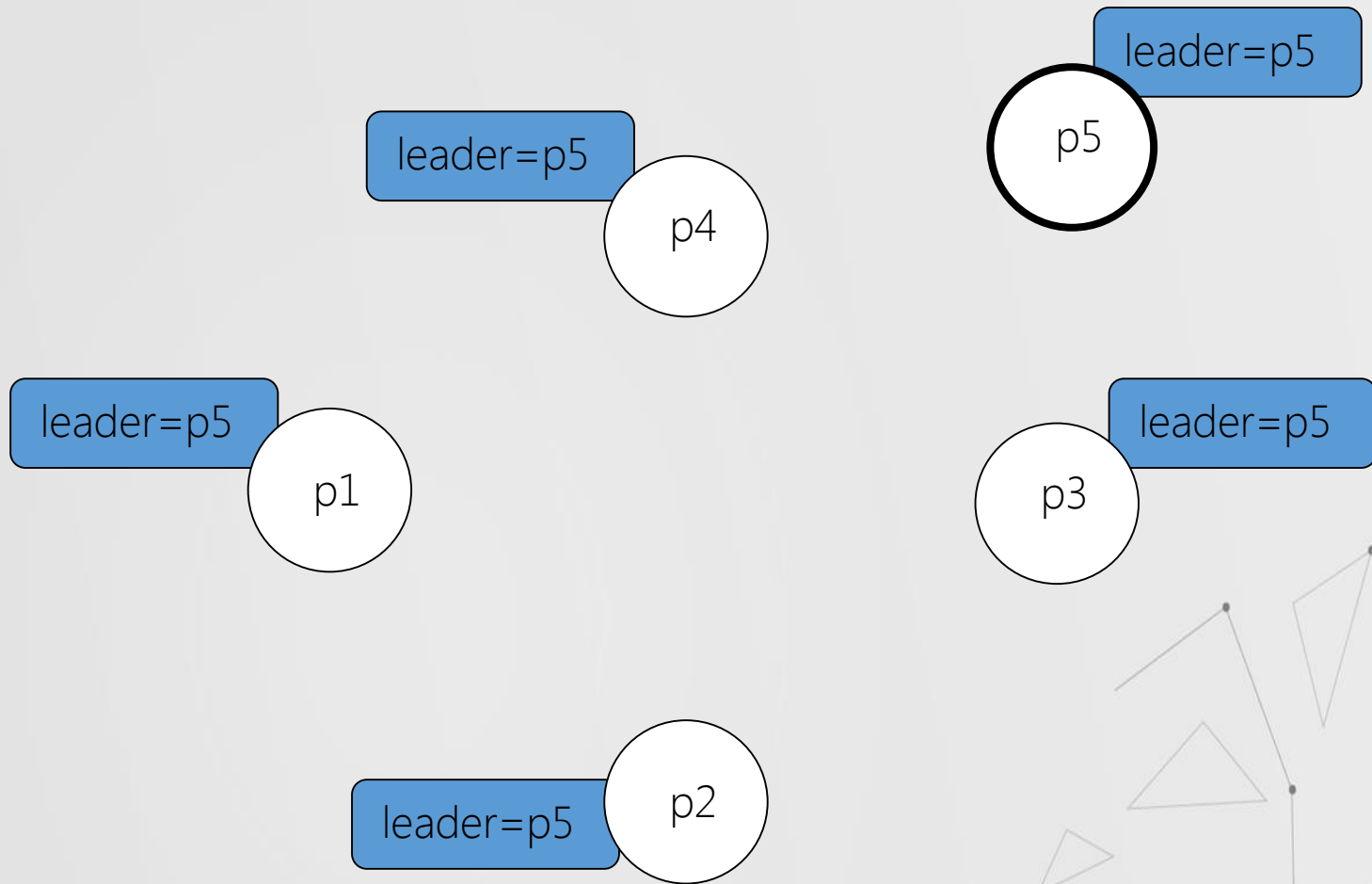
Αλγόριθμος ισχυροτέρου (Garcia-Molina)



Αλγόριθμος ισχυροτέρου (Garcia-Molina)



Αλγόριθμος ισχυροτέρου (Garcia-Molina)



Υποστήριξη για δυναμικά συστήματα

- Με φυσική πολυεκπομπή, μπορεί να υποστηριχθεί η δυναμική προσθήκη νέων διεργασιών χωρίς να απαιτείται προηγούμενη γνώση των αναγνωριστικών.
- Κάθε διεργασία που προστίθεται αρχίζει τη διαδικασία όπως στην περίπτωση επανεκκίνησης μετά από βλάβη.
- Δεν υπάρχει κανένα πρόβλημα με την αποχώρηση (ή εθελοντική αποχή) διεργασιών.
- Απλά δεν πρόκειται να εκλεγούν ως αρχηγοί για όσο διάστημα δεν συμμετέχουν στις εκλογές.

