

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών (CPU/process management/scheduling)

ΛΕΙΤΟΥΡΓΙΑ

- Εφαρμογή πολιτικής καταχώρισης της ΚΜΕ (και έμμεσα των άλλων πόρων του συστήματος) στις διεργασίες
- Προετοιμασία της ουράς του διανομέα, δηλαδή της ουράς των κατάλληλων διεργασιών
Αυτή περιέχει τις διεργασίες που μπορούν να διανεμηθούν μέχρι την επόμενη ενεργοποίηση του διαχειριστή της ΚΜΕ
 - ανακατανομή (περιγραφητών) διεργασιών
 - επιλογή κατάλληλων να διανεμηθούν διεργασιών
 - εισαγωγή τους στην ουρά του διανομέα
- **προτεραιότητα** = θέση του περιγραφητή της διεργασίας στην ουρά του διανομέα

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΕΝΕΡΓΟΠΟΙΗΣΗ

- αιτήσεις δημιουργίας και διαγραφής διεργασιών
- σε τακτά χρονικά διαστήματα
 - αφύπνιση (διακοπές χρονομέτρου)
 - για την αντιμετώπιση αδιεξόδων

πιθανόν περισσότερες από μία ουρές διανομέα

- ανάλογα με προτεραιότητα / απαιτήσεις πόρων / χρόνο αναμονής των διεργασιών
- **Διεργασίες συστήματος**
 - έχουν τη μεγαλύτερη προτεραιότητα
 - κατανεμημένες σε επίπεδα προτεραιοτήτων
- **Διεργασίες χρηστών**
 - εκτελούνται μόνον όταν οι διεργασίες συστήματος δε χρησιμοποιούν την ΚΜΕ
 - μπορεί να είναι κατανεμημένες σε επίπεδα
 - αυτές θα μας απασχολήσουν παρακάτω

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΣΤΟΧΟΙ

1. αποδεκτός βαθμός χρήσης πόρων (*resource utilization*) (βασικότερος σκοπός)
 2. μικρός χρόνος απόκρισης (*response time*) στους χρήστες
 - υψηλή ρυθμαπόδοση (*throughput*) και
 - μικρός χρόνος αναμονής
 3. μικρή επιβάρυνση (*overhead*)
 - λίγοι οι πόροι (χρόνος και χώρος) που ξοδεύει το σύστημα
- Σύγκρουση (2) και (3)
 - ο χρόνος απόκρισης στους χρήστες ελαχιστοποιείται σε βάρος του χρόνου επιβάρυνσης
 - η επιβάρυνση ελαχιστοποιείται σε βάρος του χρόνου απόκρισης στους χρήστες
 - συμβιβασμοί (*compromises*)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Περιορισμοί (Constraints)

- πλήθος χρηστών στο σύστημα
 - συνάρτηση του φόρτου του συστήματος
 - μακροχρόνιος χρονοπρογραμματιστής
- βαθμός πολυπρογραμματισμού
 - μια διεργασία μπορεί να διανεμηθεί μόνο αν οι πόροι που απαιτεί υπάρχουν διαθέσιμοι από το σύστημα
 - μεσοχρόνιος (πολυ)προγραμματισμός με σκοπό την ικανοποιητική χρήση των πόρων του συστήματος

Φιλοσοφίες

- Στατικές προτεραιότητες διεργασιών
- Δυναμικές προτεραιότητες διεργασιών

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

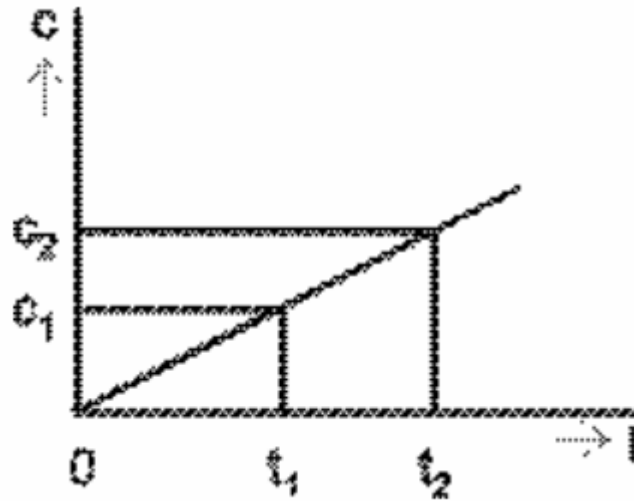
ΣΤΑΤΙΚΗ ΠΡΟΤΕΡΑΙΟΤΗΤΑ

- σταθερή σ' όλη τη διάρκεια ζωής μιας διεργασίας
- δίνεται όταν δημιουργείται η διεργασία
- κριτήρια
 - (προβλεπόμενος) χρόνος εκτέλεσης
 - (μέγιστος προβλεπόμενος) χώρος μνήμης που θα χρειαστεί
 - (προβλεπόμενη) δραστηριότητα ε/ε
 - “κόστος των πόρων” που καταχωρούνται στη διεργασία
- απλή φιλοσοφία
- μικρή επιβάρυνση
- εφαρμογή σε συστήματα ειδικού σκοπού (special purpose systems), π.χ.
 - στρατιωτικό σύστημα (προτεραιότητες ανάλογες βαθμού χρήστη)
 - εμπορικό σύστημα (προτεραιότητες ανάλογες κόστους ανά λεπτό)
 - σύστημα πραγματικού χρόνου (προτεραιότητες αντιστρόφως ανάλογες των διαστημάτων μεταξύ των απαιτήσεων εκτέλεσης των διεργασιών)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Υπολογισμός στατικών προτεραιοτήτων [Kleinrock]

- δύο διεργασίες p_1 και p_2 και αντίστοιχα κόστη
- κόστος: ανάλογο με το χρόνο από τη στιγμή δημιουργίας της μέχρι τη στιγμή της διαγραφής της



Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Υπολογισμός στατικών προτεραιοτήτων

- Έστω ότι οι p_1 και p_2 εκτελούνταν μόνες τους στο σύστημα

- χρησιμοποιούν t_1 και t_2 μονάδες χρόνου, αντίστοιχα

- συνολικός χρόνος $t = t_1 + t_2$

- η p_1 θα χρησιμοποιούσε το κλάσμα

$$\frac{t_1}{t_1 + t_2} = f_1$$

- του χρόνου, ενώ η p_2 θα χρησιμοποιούσε το κλάσμα

$$\frac{t_2}{t_1 + t_2} = f_2$$

- όπου $f_1 + f_2 = 1$

- στοιχίζουν c_1 και c_2 μονάδες κόστους, αντίστοιχα

- το συνολικό κόστος $c = c_1 + c_2$

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Υπολογισμός στατικών προτεραιοτήτων

- Έστω τώρα ότι οι p_1 και p_2 εκτελούνται μαζί στο σύστημα

(1) Αν η p_1 έχει μεγαλύτερη προτεραιότητα από την p_2

- $t_1' = t_1 \Rightarrow c_1' = c_1$

- $t_2' = t_1 + t_2 \Rightarrow t_2' = t_2 / (1 - f_1)$

- $\Rightarrow c_2' = c_2 / (1 - f_1)$

- $c' = c_1' + c_2' \Rightarrow c' = c_1 + c_2 / (1 - f_1)$

- $\Rightarrow c' = c_1 + c_2 + c_2 f_1 / (1 - f_1)$

(2) Αν η p_2 έχει μεγαλύτερη προτεραιότητα από την p_1

- $c'' = c_1 + c_2 + c_1 f_2 / (1 - f_2)$

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Υπολογισμός στατικών προτεραιοτήτων

- Το χρονοπρόγραμμα (1) θα ήταν καλύτερο από (2) ανν:

$$\frac{c_2 f_1}{1 - f_1} < \frac{c_1 f_2}{1 - f_2} \Rightarrow \frac{f_1}{(1 - f_1) c_1} < \frac{f_2}{(1 - f_2) c_2}$$

- Προτεραιότητες σύμφωνα με την ποσότητα $\frac{f_i}{(1 - f_i) c_i}$
(υψηλή προτεραιότητα στις διεργασίες που η ποσότητα αυτή έχει μικρή τιμή), δηλαδή
 - ανάλογη με το κόστος εκτέλεσης της διεργασίας
 - αντιστρόφως ανάλογη με το ποσοστό του χρόνου που χρησιμοποιείται για την εκτέλεσή της

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΣΤΑΤΙΚΗ ΠΡΟΤΕΡΑΙΟΤΗΤΑ

ΠΡΟΒΛΗΜΑΤΑ

Δύσκολη ικανοποίηση των στόχων του χρονοπρογραμματισμού, γιατί:

- βασίζονται σε προβλέψεις των χαρακτηριστικών των διεργασιών
 - ανακρίβεια προβλέψεων
 - ορισμένες διεργασίες μπορεί μονοπωλούν άδικα την ΚΜΕ (υπερεκτίμηση του χρόνου εκτέλεσής τους)
- δε λαμβάνουν υπόψη πληροφορίες από το σύστημα
 - ο χρονοπρογραμματισμός δεν είναι ευέλικτος
 - γίνεται συνεχώς με τις ίδιες προτεραιότητες και όχι για μικρό διάστημα χρόνου
 - δε λαμβάνει υπόψη την μέχρι τώρα συμπεριφορά των διεργασιών
 - άγνωστος ο στιγμιαίος φόρτος (instantaneous load) των πόρων
- υποθέτουμε ότι οι διεργασίες χρησιμοποιούν ομοιόμορφα τους πόρους σ' όλη τη διάρκεια της ζωής τους
 - ακόμα και αν οι εκτιμήσεις είναι ακριβείς, αυτό δεν ισχύει
 - μειονεκτήματα αλγορίθμων μεγίστης απαίτησης
 - η ταξινόμηση των διεργασιών δε λαμβάνει υπόψη τις απαιτήσεις τους κάθε χρονική στιγμή

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΔΥΝΑΜΙΚΕΣ ΠΡΟΤΕΡΑΙΟΤΗΤΕΣ

- γνωστή ιστορία απαιτήσεων (demand history) της διεργασίας για έναν πόρο (δυναμική συμπεριφορά της διεργασίας στο σύστημα)
- *μελλοντικές απαιτήσεις της διεργασίας ~ ιστορία της διεργασίας*
- αναταξινόμηση των διεργασιών ανάλογα με τους πόρους που έχουν ήδη χρησιμοποιήσει
- η ιστορία είναι συνάρτηση του χρόνου
- οι προτεραιότητες πρέπει να υπολογίζονται συχνά κατά τη διάρκεια ζωής των διεργασιών
- καλύτερη επιλογή από τις στατικές προτεραιότητες
- μεγαλύτερη επιβάρυνση

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Κριτήρια για υπολογισμό δυναμικής προτεραιότητας

- ποσότητες των πόρων που χρησιμοποιήθηκαν στο παρελθόν
- ποσότητες των πόρων που χρησιμοποιούνται τη δεδομένη χρονική στιγμή
- αιτία που έχει πιθανώς καταστήσει τη διεργασία εμποδισμένη
- πλήθος των σφαλμάτων των τεμαχίων της διεργασίας
- χρόνος αναμονής (waiting time) της διεργασίας για κάποιον πόρο
- συνολικό χρόνος που έχει περάσει (elapsed time), από τη δημιουργία ή από την αρχή της αλληλεπίδρασης μιας διεργασίας

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Συχνότητα υπολογισμού δυναμικών προτεραιοτήτων (ΚΑΘΕ ΠΟΤΕ)

Κάθε φορά που συμβαίνει μια διακοπή ή μία παγίδα

από το βραχυχρόνιο χρονοπρογραμματιστή (διανομέα)

- μεγάλη ευελιξία
- μεγάλη επιβάρυνση
- μικρή συλλογή νέων πληροφοριών
- μικρή πιθανή αύξηση της επίδοσης συστήματος
- δυσανάλογη προς την επιβάρυνση
- ΟΧΙ

Σε αραιότερα διαστήματα

από το μεσοχρόνιο χρονοπρογραμματιστή (διαχειριστή της ΚΜΕ)

- προετοιμάζει ένα χρονοπρόγραμμα για το επόμενο χρονικό διάστημα (Osec)
- βασίζεται στην ιστορία των διεργασιών κατά τα προηγούμενα χρονικά διαστήματα
- *μειονέκτημα:* καταγραφή πληροφοριών επίδοσης συστήματος και χρήσης πόρων
- ανάγκη προσδιορισμού της συχνότητας υπολογισμού των προτεραιοτήτων

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Συχνότητα υπολογισμού δυναμικών προτεραιοτήτων

Ίδια συχνότητα για όλες τις διεργασίες

- σε κάθε ενεργοποίηση του διαχειριστή της ΚΜΕ
- συνολική ταξινόμηση της ουράς του διανομέα
- υποφερτή για μικρό πλήθος διεργασιών
- χρονοβόρα για μεγάλο πλήθος διεργασιών

καλύτερη η μερική ταξινόμηση

- προαγωγή ή υποβίβαση μερικών διεργασιών στην ουρά του διανομέα
- η ουρά του διανομέα πρέπει να διατηρείται ταξινομημένη

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Συχνότητα υπολογισμού δυναμικών προτεραιοτήτων

Διαφορετικές συχνότητες

- για μερικές διεργασίες (τις πιο κρίσιμες) σε κάθε ενεργοποίηση του διαχειριστή της ΚΜΕ
 - με σκοπό τον υποβιβασμό τους στην ουρά του διανομέα
- για τις υπόλοιπες διεργασίες αραιότερος υπολογισμός
- γενίκευση
 - κάθε διεργασία έχει το δικό της διαχειριστή
 - οι διεργασίες πραγματικού χρόνου υπολογίζουν τη δική τους προτεραιότητα και βρίσκονται συχνά στην κεφαλή της ουράς του διανομέα
 - διαφορετικός τρόπος για τις άλλες διεργασίες
- μια απλή λύση
 - όταν δημιουργείται ή αρχίζει η αλληλεπίδραση μιας διεργασίας
 - κάθε φορά που ξαναμπαίνει στη λίστα των ετοιμών (ready list)
- πραγματικότητα
 - ουρές πολλών επιπέδων
 - εκθετικός χρονοπρογραμματισμός

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΔΥΝΑΜΙΚΕΣ ΠΡΟΤΕΡΑΙΟΤΗΤΕΣ

Προσανατολισμένες στην εξυπηρέτηση των διεργασιών (service-oriented)

- ατομική αξιολόγηση (customer ή user oriented measure)
- πολιτικές: FCFS, RR, SPN, SRT
- λαμβάνουν υπόψη τους μόνο τη χρήση της ΚΜΕ κι όχι των άλλων πόρων του συστήματος

ΑΛΛΑ

βασικός σκοπός χρονικού προγραμματισμού είναι η διατήρηση του φόρτου σε όλους τους πόρους του συστήματος

- η υπερφόρτωση ενός πόρου καθυστερεί όλους τους άλλους
- ισορροπημένη φόρτωση πόρων (balanced loading of resources)
- οι διεργασίες πρέπει να επιλέγονται με τέτοιο έτσι ώστε να διατηρούνται απασχολημένες όλες οι μονάδες του υλισμικού
- αστάθμιστο (unbalanced) σύστημα: παράδειγμα
 - όλες οι διεργασίες χρησιμοποιούν τους δίσκους
 - καμία διεργασία δε χρησιμοποιεί την ΚΜΕ

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Δυναμικές Προτεραιότητες

Προσανατολισμένες στη χρήση των πόρων (resource-oriented)

- συστημική αξιολόγηση (system oriented measure)
- πολιτικές: στάθμιση απαιτήσεων ΚΜΕ και ε/ε, ουρές πολλών επιπέδων, ...
- λαμβάνουν (προσπαθούν να λάβουν) υπόψη τους τη χρήση όλων των πόρων του συστήματος

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

FCFS (First Come First Served)

- οι διεργασίες μπαίνουν στην ουρά με τη σειρά άφιξής τους
- *μη-διακοπτόμενος*: οι διεργασίες εκτελούνται μέχρι να τελειώσουν ή να εμποδιστούν
- προτεραιότητα ανάλογη του χρόνου αναμονής
- τιμωρούνται οι σύντομες διεργασίες
 - πρέπει να περιμένουν μεγάλης διάρκειας διεργασίες που αφίχθηκαν νωρίτερα
- ευνοούνται οι περιορισμένες από ΚΜΕ διεργασίες γιατί
 - οι περιορισμένες από ε/ε διεργασίες εμποδίζονται γρήγορα
- δεν υπάρχει περίπτωση υποσιτισμού
- ελάχιστη επιβάρυνση
- πιθανόν μεγάλος χρόνος απόκρισης
- πολιτική προσανατολισμένη στην εξυπηρέτηση

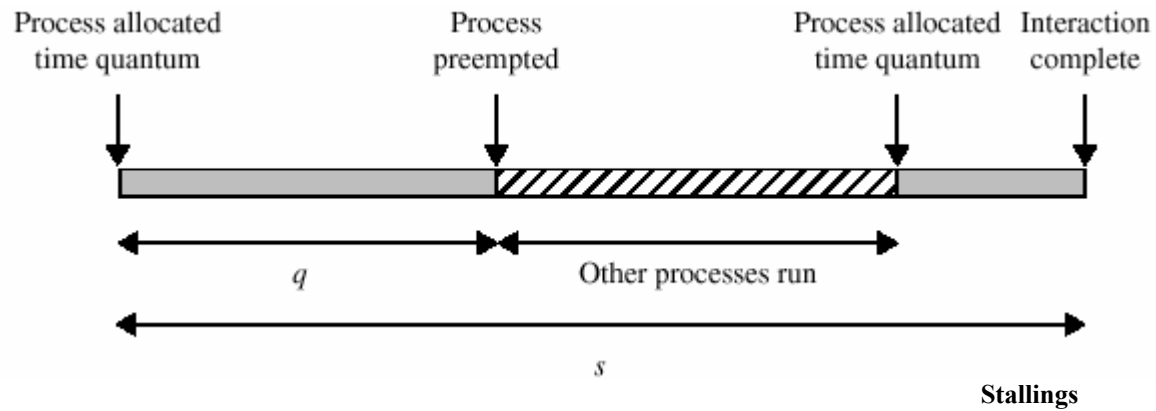
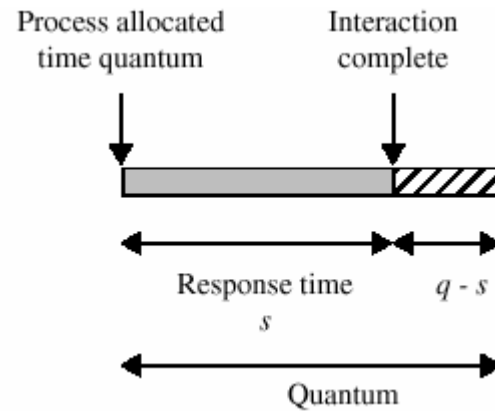
Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Γύρω-γύρω όλοι” (Round Robin, RR)

- *διακοπτόμενος FCFS*
- κάθε διεργασία μπορεί να εκτελεστεί για ένα συγκεκριμένο χρονικό διάστημα (quantum, time slice)
- όταν το διάστημα αυτό τελειώσει ο διανομέας βάζει την τρέχουσα διεργασία στο τέλος της ουράς του και διανέμει αυτήν που βρίσκεται στην κεφαλή της ουράς του
- υλοποίηση: χρονόμετρο διαστημάτων
- το μέγεθος του διαστήματος επηρεάζει την επίδοση
- μεγάλο διάστημα
 - οι σύντομες διεργασίες τιμωρούνται
 - μεγάλος χρόνος απόκρισης (→FCFS)
- μικρό διάστημα
 - οι σύντομες διεργασίες τελειώνουν σχετικά γρήγορα
 - μεγάλη επιβάρυνση (μεταγωγή συμφραζομένων)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ

RR (Round Robin)



Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Round Robin (RR)

- δίκαιος και για διεργασίες που είναι περιορισμένες από την ΚΜΕ και για αυτές που είναι περιορισμένες από ε/ε
- δεν υπάρχει περίπτωση υποσιτισμού
- επιβάρυνση εξαρτώμενη από το χρονικό διάστημα που δίνεται στις διεργασίες
- καλός χρόνος απόκρισης για τις σύντομες διεργασίες
- πολιτική προσανατολισμένη στην εξυπηρέτηση

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SPN (Shortest Process Next)

- οι διεργασίες μπαίνουν στην ουρά ανάλογα με τον (αναμενόμενο) χρόνο χρήσης της ΚΜΕ
- *μη-διακοπτόμενος*: οι διεργασίες εκτελούνται μέχρι να τελειώσουν ή να εμποδιστούν
- τιμωρούνται οι μεγάλης διάρκειας διεργασίες : οι σύντομες διεργασίες εκτελούνται πάντα πρώτες
- οι σύντομες διεργασίες μπορεί να περιμένουν πολύ ένα μια μεγάλη διεργασία έχει ήδη αρχίσει την εκτέλεσή της
- πιθανός υποσιτισμός (διεργασιών μεγάλης διάρκειας)
- μικρή επιβάρυνση
- καλός χρόνος απόκρισης για τις σύντομες διεργασίες
- πολιτική προσανατολισμένη στην εξυπηρέτηση
- απαιτεί προκαταβολική γνώση χρόνων χρήσης ΚΜΕ!

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SRT (Shortest Remaining Time first)

- *διακοπτόμενος SPN*
- η τρέχουσα διεργασία διακόπτεται εάν αφιχθεί μια διεργασία με μεγαλύτερη προτεραιότητα
- προτεραιότητα ανάλογη προς τον (αναμενόμενο) υπολειπόμενο χρόνο χρήσης της ΚΜΕ
- τιμωρούνται οι μεγάλης διάρκειας διεργασίες: οι σύντομες διεργασίες εκτελούνται πάντα πρώτες
- πιθανός υποσιτισμός (μεγάλης διάρκειας διεργασία)
- μεγαλύτερη επιβάρυνση από SPN
- καλός χρόνος απόκρισης για τις σύντομες διεργασίες
- πολιτική προσανατολισμένη στην εξυπηρέτηση
- απαιτεί προκαταβολική γνώση χρόνων χρήσης ΚΜΕ!

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Πολιτικές στάθμισης ΚΜΕ και ε/ε

- *υψηλή προτεραιότητα σε διεργασίες περιορισμένες από ε/ε*
- επικάλυψη λειτουργιών των συσκευών ε/ε με τη χρήση άλλων πόρων
- αύξηση του βαθμού χρήσης των πόρων
- παράδειγμα:
 - αν μια περιορισμένη από ε/ε διεργασία δε διανέμεται γρήγορα καταλαμβάνει άσκοπα τη μνήμη για μεγάλο διάστημα
- ποιες διεργασίες είναι περιορισμένες από ε/ε και ποιες περιορισμένες από ΚΜΕ?
 - μέτρηση απαιτήσεων ΚΜΕ και/ή ε/ε (π.χ. στο διάστημα ανάμεσα σε διαδοχικές εκτελέσεις του διαχειριστή της ΚΜΕ)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

Πολιτικές στάθμισης ΚΜΕ και ε/ε

Προτεραιότητα διεργασιών

- ανάλογη των απαιτήσεων I/O
- αντιστρόφως ανάλογη των απαιτήσεων ΚΜΕ
 - πειραματικές ενδείξεις: οι εντολές ε/ε συγκεντρώνονται σε ορισμένα σημεία του προγράμματος μιας διεργασίας
- αντιστρόφως ανάλογη του χρόνου από τη στιγμή που η διεργασία εξέδωσε την τελευταία εντολή I/O
- αντιστρόφως ανάλογη του κλάσματος του κβάντου χρόνου που χρησιμοποίησε η διεργασία πριν εκδώσει την τελευταία εντολή ε/ε
- ανάλογη με το λόγο (μέτρο δραστηριότητας ε/ε):

$$\frac{\text{συνολικός χρόνος εκτέλεσης} + \text{χρόνος αναμονής}}{\text{χρόνος αναμονής}}$$

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SPN, SRT, ΣΤΑΘΜΙΣΗ ΚΜΕ και Ε/Ε

Πολιτικές που χρειάζονται να γνωρίζουν προκαταβολικά (να προβλέπουν) τα χαρακτηριστικά των διεργασιών

ΥΠΟΘΕΣΗ

Η συμπεριφορά μιας διεργασίας στην επόμενη διανομή της θα είναι ή ίδια με τη συμπεριφορά της κατά την προηγούμενη διανομή της

ΑΛΛΑ

- εκρήξεις (bursts) εκτέλεσης μιας περιορισμένης από ε/ε διεργασίας
 - ελάττωση προτεραιότητας
 - ανάλογη με τη διάρκεια της έκρηξης χρήσης της ΚΜΕ
- εκρήξεις ε/ε μιας περιορισμένης από την ΚΜΕ διεργασίας
 - αύξηση προτεραιότητας
 - πάλι ανάλογη με τη διάρκεια της

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SPN, SRT, στάθμιση ΚΜΕ και Ε/Ε

ΜΙΑ ΛΥΣΗ

- *Ιστορικός μετρητής εξομάλυνσης (damped historic count)*
- λαμβάνεται υπόψη η δραστηριότητα της διεργασίας όχι μόνο κατά το τελευταίο χρονικό διάστημα, αλλά και κατά τα προηγούμενα χρονικά διαστήματα
- *Πρόβλημα:* φτωχή εξυπηρέτηση σε μια σύντομη περιορισμένη από ε/ε διεργασία αν αναμειχθεί με πολλές περιορισμένες από ΚΜΕ διεργασίες

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SPN, SRT, στάθμιση ΚΜΕ και E/E

- Προτεραιότητα ανάλογη της μελλοντικής συμπεριφοράς της διεργασίας (π.χ. χρονικό διάστημα που η διεργασία θα χρησιμοποιήσει την ΚΜΕ μέχρι να εμποδιστεί)
- Διακοπτόμενη ή μη διακοπτόμενη πολιτική
- Προσανατολισμένη στην εξυπηρέτηση ή στη χρήση των πόρων
- Άγνωστη μελλοντική συμπεριφορά της διεργασίας
- Πρόβλεψη βάσει της ιστορίας της διεργασίας
- Επόμενο διάστημα ~ εκθετικός μέσος όρος προηγούμενων χρονικών διαστημάτων
- τ_v : εκτίμηση χρόνου χρήσης ΚΜΕ για την v -οστή διανομή
(αρχικά $\tau_1 = \tau_0$)
- T_v : πραγματικός χρόνος χρήσης ΚΜΕ στην v -οστή διανομή
$$\tau_{v+1} = aT_v + (1-a)\tau_v, 0 < a < 1$$
- a = ιστορικός μετρητής εξομάλυνσης (*damped historic count*) (σχετικό βάρος τρέχουσας και παρελθούσης συμπεριφοράς)
 - $a=0 \Rightarrow \tau_{v+1} = \tau_v$
ο πραγματικός χρόνος δεν παίζει κανένα ρόλο
 - $a=1 \Rightarrow \tau_{v+1} = T_v$
μόνο ο τελευταίος πραγματικός χρόνος παίζει ρόλο

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

SPN, SRT, στάθμιση ΚΜΕ και E/E

- Αναπτύσσοντας τον παραπάνω τύπο:

$$\tau_{v+1} = aT_v + (1-a)aT_{v-1} + \dots + (1-a)^i aT_{v-i} + \dots + (1-a)^v a\tau_0$$

- Συνήθως $a=1/2 \Rightarrow \tau_{v+1} = (T_v + \tau_v)/2$
 - εκτιμήσεις: $\tau_1 = \tau_0$, $\tau_2 = T_1/2 + \tau_0/2$, $\tau_3 = T_2/2 + T_1/4 + \tau_0/4$,
 $\tau_4 = T_3/2 + T_2/4 + T_1/8 + \tau_0/8$, ...
- προτεραιότητα = αντιστρόφως ανάλογη της διάρκειας του (εκθετικού μέσου όρου του) χρόνου που χρησιμοποιεί η διεργασία προτού εμποδιστεί

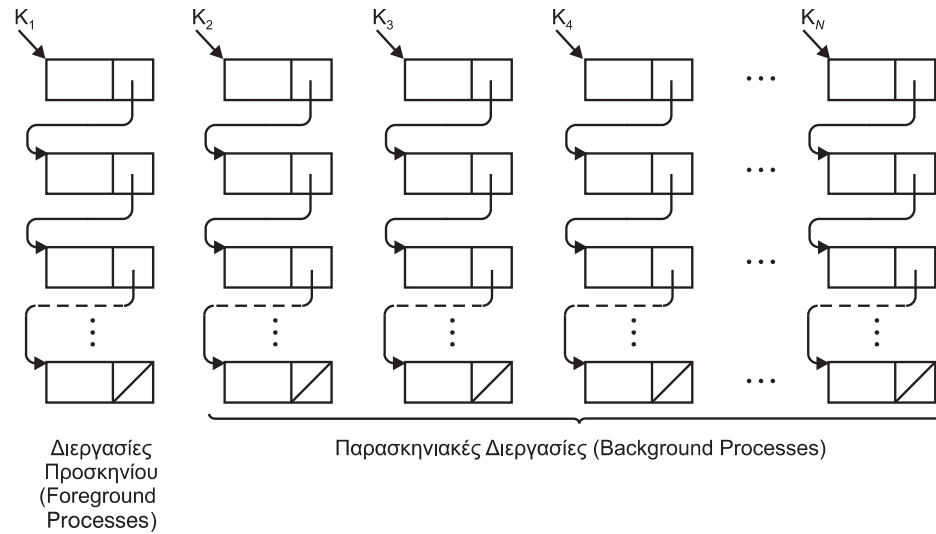
Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ (MULTI-LEVEL QUEUES)

- **N ουρές διανομής**
- Διανέμεται η πρώτη εργασία της ουράς του υψηλότερου επιπέδου
- Κάθε ουρά εξυπηρετείται με μια ορισμένη πολιτική (συνήθως RR ή FCFS)
- Οι περιορισμένες από ε/ε διεργασίες έχουν μεγαλύτερη προτεραιότητα
- Διεργασίες περιορισμένες από την ΚΜΕ ή την ε/ε~ ικανότητα να χρησιμοποιήσουν το χρόνο της ΚΜΕ
 - Οι περιορισμένες από ε/ε διεργασίες ~ αυτές που δε χρησιμοποιούν ολόκληρο το τεμάχιο χρόνου τους (εμποδίζονται λόγω ε/ε)
 - Περιορισμένες από ΚΜΕ διεργασίες ~ αυτές που εξαντλούν ολόκληρο το τεμάχιο χρόνου τους (κάνουν μεγάλη χρήση του χρόνου της ΚΜΕ μεταξύ των απαιτήσεών ε/ε τους)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ



Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

- **Οι διεργασίες ΔΕΝ μετακινούνται από μια ουρά σε μια άλλη**
 - παράδειγμα ($N=2$),
 - 1^η ουρά : διεργασίες αλληλεπίδρασης
 - 2^η ουρά: άλλες διεργασίες
- **Οι διεργασίες μετακινούνται από μια ουρά σε μια άλλη (ουρές ανάδρασης (feedback queues))**
 - Περιοδικά από τον διαχειριστή της ΚΜΕ
 - Παράδειγμα (XDS 940): τέσσερις ουρές
 - 1^η ουρά: διεργασίες αλληλεπίδρασης με τερματικό
 - 2^η ουρά: διεργασίες άλλης ε/ε
 - 3^η ουρά: διεργασίες που δεν εξαντλούν το κβάντο χρόνου
 - 4^η ουρά: διεργασίες που εξαντλούν το κβάντο χρόνου
 - Οι διεργασίες που ελευθερώνονται προσθέτονται αρχικά στις ουρές 1 και 2 (ανάλογα με την ε/ε που τις εμπόδιζε)
 - Την πρώτη φορά που εξαντλούν το κβάντο τους προσθέτονται στο τέλος της 3^{ης} ουράς
 - Αν συνεχίσουν να εξαντλούν το κβάντο τους προσθέτονται στο τέλος της 4^{ης} ουράς
 - Αν εμποδιστούν επανέρχονται στην 1^η ή 2^η ουρά

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

- **Βασικός κανόνας:** διανέμεται πάντα η διεργασία που βρίσκεται στην υψηλότερη ουρά που δεν είναι άδεια
- **Πρόβλημα:** οι διεργασίες που βρίσκονται στις χαμηλότερες ουρές μπορεί να διανέμονται σπάνια ή καθόλου
 - υποσιτισμός= μερικό αδιέξοδο (partial deadlock)

ΛΥΣΕΙΣ (εμπειρικές, ad hoc) για τις διεργασίες στις χαμηλές ουρές

- Καταχώριση μεγαλύτερου κβάντου χρόνου
 - π.χ. διπλάσιο κβάντο όσο κατεβαίνουμε επίπεδο
= εκθετικός χρονοπρογραμματισμός (exponential scheduling)
 - αν όμως δεν διανεμηθούν ποτέ?
- Μόνιμη παραμονή στην μνήμη (συνήθως 1 ή 2 διεργασίες)
 - ελάτωση διαθέσιμης μνήμης
 - μικρότερη πιθανότητα διανομής διεργασιών υψηλότερης προτεραιότητας
- Κατανομή του χρόνου της ΚΜΕ στις ουρές
 - μεγαλύτερα ποσοστά στις υψηλότερες ουρές
 - περιοδική διανομή εργασιών κάθε ουράς
- Αύξηση προτεραιοτήτων

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

ΛΥΣΕΙΣ (εμπειρικές - ad hoc) για τις διεργασίες στις χαμηλές ουρές

- **Αύξηση προτεραιοτήτων**

- Προαγωγή

- Μιας διεργασίας στην κεφαλή της ουράς ενός επιπέδου στο τέλος της ουράς του επόμενου επιπέδου
 - Περιοδική προαγωγή

- *Γενική Αμνηστία (General Amnesty)*

- Όλες οι διεργασίες που βρίσκονται στην κατώτερη ουρά για κάποιο χρονικό διάστημα
 - προάγονται σε κάποια ανώτερη ουρά
 - όσες εξακολουθήσουν να είναι κακές σύντομα θα ξαναυποβιβαστούν στις κατώτερες ουρές

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

▪ Αύξηση προτεραιοτήτων

▪ *Περιορισμένη Αμνηστία (Limited Amnesty) ή υπό όρους αναστολή (parole)*

- μια διεργασία που βρίσκεται στην κατώτερη ουρά περισσότερο για κάποιο χρονικό διάστημα προάγεται στην προηγούμενή της ουρά
- πιθανόν να μη φθάσει στις ανώτερες ουρές και να μην εξυπηρετηθεί
- υλοποίηση
 - ένας αριθμός προτεραιότητας σε κάθε διεργασία
 - περιοδική αύξηση για τις διεργασίες που παραμένουν στο ίδιο επίπεδο για ένα χρονικό διάστημα

Παρατηρήσεις

- Περιορισμένες από ΚΜΕ και ε/ε διεργασίες: "προσεγγίζονται" ανάλογα με την ποσότητα κβάντου που μπορεί να χρησιμοποιεί μια διεργασία
- Δεν λαμβάνεται υπόψη η χρήση άλλων πόρων (π.χ. μνήμης)
- Υποθέτουμε ότι οι περιορισμένες από ΚΜΕ διεργασίες χρησιμοποιούν σε μεγάλο βαθμό και την μνήμη!

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΟΥΡΕΣ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

Παραλλαγές

- πιο ευαίσθητες (sensitive) στο φόρτο των πόρων
 - Αρχική ταξινόμηση με βάση στατικά χαρακτηριστικά
 - Μετακίνηση σε ουρές με βάση (και) απαιτήσεις μνήμης
 - Μικρές απαιτήσεις: υψηλότερα επίπεδα
 - Μεγάλες απαιτήσεις: χαμηλότερα επίπεδα

Παράμετροι

- το πλήθος των ουρών
- η πολιτική της κάθε ουράς
- η μέθοδος υποβιβασμού σε κατώτερες ουρές
- η μέθοδος προβιβασμού σε ανώτερες ουρές, και
- η ουρά στην οποία εισάγεται μια διεργασία
- προσομοίωση οποιασδήποτε άλλης πολιτικής
- στην πράξη χρησιμοποιούνται παραλλαγές των ουρών πολλών επιπέδων

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΜΕΤΡΗΣΗ ΕΞΥΠΗΡΕΤΗΣΗΣ ΔΙΕΡΓΑΣΙΩΝ

Εξυπηρέτηση διεργασιών

- χρήση πόρων υλισμικού
 - (ενεργητικοί) πόροι χρόνου (ΚΜΕ, ελεγκτές ε/ε)
 - (παθητικοί) πόροι χώρου (μνήμη, δίσκος, ...)
- χρήση πόρων λογισμικού (εξυπηρέτηση παγίδων/διακοπών)

Χρήσεις/συχνότητα μέτρησης χρήσης πόρων

- καθορισμός προτεραιοτήτων (ουρές πολλών επιπέδων)
 - σε κάθε διανομή
- λογιστική χρέωση και
- προσδιορισμός φόρτου συστήματος
 - ενεργητικός ή φόρτος χρόνου (active ή time load)
 - παθητικός ή φόρτος χώρου (passive ή space load)
- στα μεσοδιαστήματα μεταξύ ενεργοποιήσεων του διαχειριστή της ΚΜΕ

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΜΕΤΡΗΣΗ ΕΞΥΠΗΡΕΤΗΣΗΣ ΔΙΕΡΓΑΣΙΩΝ

Συνολικό ποσό πόρων (resource count)

- που έχει λάβει η διεργασία R, μέχρι τη χρονική στιγμή t

$$R(t) = \sum_{i=1}^n w_i R_i(t)$$

- n: πλήθος των τύπων των πόρων
- $R_i(t)$: πλήθος των μονάδων του πόρου i που έχει χρησιμοποιήσει η διεργασία μέχρι τη χρονική στιγμή t
- w_i : (σχετικό) κόστος (βάρος) ανά μονάδα του πόρου i
 - π.χ. - **ενεργητικοί πόροι**: απόλυτος χρόνος
 - **παθητικοί πόροι**: το ελάχιστο χρονικό διάστημα, μέσα στο οποίο είναι δυνατό να χορηγηθεί μια μονάδα πόρου (μνήμη: bits/sec, εκτυπωτής: cps ή lpm, κ.λπ.)
 - **R(t)**= (σταθμισμένο) άθροισμα των χρόνων που χρησιμοποίησε η διεργασία σ' όλους τους πόρους του συστήματος

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΒΑΘΜΟΣ ΠΟΛΥΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ

- υπάρχουν συνδυασμοί διεργασιών που είναι αδύνατο να εκτελεστούν ταυτόχρονα κι έτσι
- αποκλείουν την ικανοποιητική χρήση πόρων
- Διαχειριστής ΚΜΕ
 - *Επιλογή ενός συνόλου διεργασιών που*
 - Μπορούν να εκτελεστούν ταυτόχρονα
 - Επιτυγχάνουν την καλύτερη δυνατή χρήση των πόρων του συστήματος
 - Εφαρμογή πολιτικής χρονοπρογραμματισμού
 - $R(i, j)$ =φόρτος διεργασίας i για τον πόρο j
 - $R(j)$ = άθροισμα φόρτου ενός συνόλου διεργασιών για τον πόρο j
 - $W(j)$ =συνολικός φόρτος που μπορεί να εξυπηρετήσει ο πόρος j
 - $R(j)/W(j)$ = επίπεδο φόρτου του πόρου j (≤ 1)
 - + προτεραιότητες διεργασιών

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΒΑΘΜΟΣ ΠΟΛΥΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ

Περιοριστικός Πόρος (Limiting Resource)

- πόρος με το μεγαλύτερο φόρτο
 - ο βαρύτερα χρησιμοποιούμενος πόρος
 - αδύνατη η πλήρης χρήση των πόρων
- *συνήθως η μνήμη*
- καθορίζει το *βαθμό πολυπρογραμματισμού* = αριθμό διεργασιών στο σύνολο των καταλλήλων
- μεγάλος βαθμός πολυπρογραμματισμού
 - μεγαλύτερη χρήση των άλλων πόρων (π.χ. χρόνος της ΚΜΕ)
 - μεγάλος ανταγωνισμός = καθυστέρηση

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΒΑΘΜΟΣ ΠΟΛΥΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ

Σ' ένα σύστημα γενικού-σκοπού συνήθως η μνήμη αποτελεί τον κρίσιμο (περιοριστικό) πόρο

- βαθμός πολυπρογραμματισμού ~ μέγεθος μνήμης
- χρόνος εκτέλεσης διεργασιών ~ μέγεθος μνήμης
- αλώνισμα ~ υπερφόρτωση (*overload*) μνήμης
 - ~ μεγάλος βαθμός πολυπρογραμματισμού ~ μεγάλη λίστα κατάλληλων
- ο διαχειριστής της μνήμης ΔΕΝ εξασφαλίζει ότι ο χώρος εργασίας μιας διεργασίας είναι αρκετά μεγάλος για να χωρέσει το σύνολο εργασίας της
- ο διαχειριστής της ΚΜΕ αποφασίζει ποιες διεργασίες πρόκειται να εκτελεστούν στο μέλλον (λίστα κατάλληλων)
 - χρησιμοποιεί πληροφορίες από χαμηλότερα επίπεδα του συστήματος, δηλαδή
 - τον πυρήνα
 - τους διαχειριστές των περιφερειακών συσκευών
 - το διαχειριστή της μνήμης
 - ετοιμάζει τη λίστα του διανομέα
 - ώστε να είναι δυνατή η αποτελεσματική η διεκπεραίωσή της

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Υπερφόρτωση (overload) συστήματος

- όριο που αφορά στο βαθμό (κατα)μερισμού των πόρων με μεγάλη ζήτηση (π.χ. μνήμη)
- πέρα από αυτό το όριο η επίδοση του συστήματος υποβιβάζεται (π.χ. αλώνισμα)
- εξαρτάται από το βαθμό πολυπρογραμματισμού
- αντιμετώπιση παρόμοια με τα αδιέξοδα

Έγκριση - Ανακάλυψη - Επανόρθωση υπερφόρτωσης

- Επιτρέπεται ο μεγαλύτερος δυνατός βαθμός πολυπρογραμματισμού
- Αν ανακαλυφθεί υπερφόρτωση
 - αφαίρεση μίας ή περισσότερων διεργασιών από αυτές που έχουν πολυπρογραμματιστεί
- Αν ανίχνευθεί υποφορτισμός (underload)
 - ενεργοποίηση μιας ανασταλμένης ή νέας διεργασίας

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Αποφυγή υπερφόρτωσης

- Συμβιβασμός μεγάλου βαθμού πολυπρογραμματισμού και προβλημάτων συναγωνισμού των διεργασιών για πόρους(π.χ. αλώνισμα)
- Δημιουργία (από το διαχειριστή της ΚΜΕ) της λίστας κατάλληλων η οποία αποτελείται από ένα υποσύνολο των ετοιμών διεργασιών
- Διαθέσιμη στο διανομέα για το επόμενο χρονικό διάστημα
- Σταθερές συνθήκες φόρτου για το διάστημα αυτό
- Μειονέκτημα: πιθανόν να μένουν αδρανείς πόροι

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Έγκριση (Επέμβαση χειριστή)

- παλαιότερα συστήματα (ATLAS, T.H.E.)
- ο χειριστής παρατηρούσε "μεγάλη δραστηριότητα" στο τύμπανο
- αδρανοποιούσε μία ή περισσότερες διεργασίες
- κατάλληλη μέθοδος μόνο όταν υπερφόρτωση συμβαίνει πολύ σπάνια

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Ανακάλυψη - Επανόρθωση υπερφόρτωσης

- Λόγος εικονικού προς φυσικό χώρο
 - εικονικός χώρος όλων των διεργασιών που πολυπρογραμματίζονται
χωρητικότητα της φυσικής μνήμης
 - αν ο λόγος αυτός είναι μεγαλύτερος από μια ορισμένη τιμή, τότε το σύστημα είναι υπερφορτωμένο
 - αν ο λόγος αυτός είναι μικρότερος από μια ορισμένη τιμή, τότε το σύστημα είναι υποφορτωμένο
 - Υπόθεση:
αν το πλήθος των τεμαχίων μιας διεργασίας που βρίσκονται στην κύρια μνήμη παραμένει σταθερό, τότε η συχνότητα των σφαλμάτων των τεμαχίων της διεργασίας είναι σταθερή
 - η υπόθεση δεν ισχύει πάντα

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Ανακάλυψη - Επανόρθωση υπερφόρτωσης

- Χρόνος κύκλου
 - αλγόριθμοι επόμενου ταιριάσματος και δεύτερης ευκαιρίας
 - χρόνος που χρειάζεται για να συμπληρωθεί ένας κύκλος διάσχισης της κυκλικής λίστας των ελεύθερων ή των κατειλημμένων περιοχών, αντίστοιχα
 - π.χ. επόμενο ταίριασμα
 - αν ο χρόνος αυτός είναι μικρότερος από μια ορισμένη τιμή, τότε αδρανοποιείται μια διεργασία (γιατί?)
 - αν ο χρόνος αυτός είναι μεγαλύτερος από μια ορισμένη τιμή, τότε ενεργοποιείται μια διεργασία (γιατί?)

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Ανακάλυψη - Επανόρθωση υπερφόρτωσης

- **Συχνότητα σφαλμάτων**
 - αλγόριθμοι που δεν υποφέρουν από την ανωμαλία του Belady
 - η συχνότητα σφαλμάτων ελαττώνεται όσο αυξάνονται οι διαθέσιμες περιοχές (πλαίσια) της μνήμης
 - Υπολογισμός της συχνότητας των σφαλμάτων
 - ο διαχειριστής της ΚΜΕ διατηρεί μετρητές εξομάλυνσης
 - χρόνος ΚΜΕ που ξοδεύεται για την εξυπηρέτηση των σφαλμάτων τεμαχίων και χρόνος ελεγκτή/διαύλου για τις μεταφορές των τεμαχίων
 - αν ο χρόνος αυτός είναι μεγαλύτερος από μια ορισμένη τιμή, τότε αδρανοποιείται μια διεργασία
 - ο χρόνος αυτός είναι μικρότερος από μια ορισμένη τιμή, τότε ενεργοποιείται μια διεργασία
- Άμεσος έλεγχος φόρτου μνήμης
- Εύκολη υλοποίησή
- Αποτελεσματικότητα
 - Λ.Σ. MCP των Burroughs 5500 [Wulf]
 - Πλήρης εξάλειψη του αλωνίσματος

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Αποφυγή υπερφόρτωσης με χρήση προτεραιοτήτων των διεργασιών

- κρίσιμες (critical) διεργασίες
- μη κρίσιμες (non-critical) διεργασίες
- υλοποίηση ελέγχου φόρτου = απαγόρευση αντικατάστασης των τεμαχίων μιας κρίσιμης διεργασίας από τεμάχια μιας μη κρίσιμης διεργασίας
- *Κανόνες αντικατάστασης*
 - τα τεμάχια μιας κρίσιμης διεργασίας δεν αντικαθίστανται
 - τα τεμάχια μιας κρίσιμης διεργασίας αντικαθιστούν τα τεμάχια μιας μη κρίσιμης διεργασίας
 - τα τεμάχια μιας μη κρίσιμης διεργασίας προσκομίζονται, μόνο αν
 - υπάρχει χώρος στη μνήμη ή
 - μπορούν να αντικαταστήσουν τεμάχια διεργασιών που περιμένουν την αλληλεπίδραση του χρήστη από το τερματικό
- οι κανόνες μπορούν να εφαρμόζονται με κάποια (καθ)υστέρηση (hysteresis) για
 - μεγαλύτερη ελάττωση άσκοπη ανταλλαγή τεμαχίων και
 - αύξηση χρήσης των πόρων του συστήματος

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Αποφυγή υπερφόρτωσης με χρήση συνόλου εργασίας

- ο βαθμός του πολυπρογραμματισμού αλλάζει με την πάροδο του χρόνου
- το σύστημα κάνει μια διεργασία αναστελλόμενη για σύνολο εργασίας όταν ο χώρος της διαθέσιμης μνήμης είναι μικρότερος από το σύνολο εργασίας της διεργασίας αυτής, μ' άλλα λόγια
- όταν το (προβλεπόμενο) άθροισμα των συνόλων εργασίας των διεργασιών της λίστας των καταλλήλων είναι μικρότερο από το μέγεθος της διαθέσιμης μνήμης
 - η εφαρμογή του παραπάνω κανόνα βοηθά στην πρόληψη του αλωνίσματος παρόλο που:
 - τα σύνολα εργασίας των διεργασιών δεν ορίζονται επακριβώς και
 - η αρχή της τοπικότητας δεν ισχύει πάντα

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Αποφυγή υπερφόρτωσης με χρήση συνόλου εργασίας

MULTICS

- Μια διεργασία φορτωνόταν πριν να διανεμηθεί:
 - φόρτωση του τμήματος περιγραφητή της διεργασίας
 - προκαταβολική φόρτωση όλων των σελίδων, τις οποίες είχε χρησιμοποιήσει η διεργασία στην τελευταία διανομή της
- Καμία διεργασία δεν μπορούσε να εκτελεστεί, αν το (προηγούμενο) σύνολο εργασίας της δεν μπορούσε να φορτωθεί στη μνήμη

MCP (Burroughs 5000)

- η διαθέσιμη μνήμη έπρεπε να χωράει τα σύνολα εργασίας όλων των διεργασιών που επρόκειτο να διανεμηθούν
- W_i = σύνολο εργασίας της διεργασίας δ_i (όπως το είχε προβλέψει ο μεταγλωττιστής)
- S = σύνολο διεργασιών στη λίστα κατάλληλων
- Η δ_i έμπαινε στη λίστα κατάλληλων αν $W_i \leq w \cdot M - \sum_{j \in S} W_j$
- w = συντελεστής βάρους (weighting factor)
 - μέτρο ασφάλειας της υπερφόρτωσης της μνήμης
 - $1.0 \leq w \leq 1.5$ καλύτερη τιμή $w \sim 1.3$

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Αποφυγή υπερφόρτωσης με χρήση συνόλου εργασίας

- Διάφορες τεχνικές καθορισμού των συνόλων εργασίας
- Στενή σχέση διαχειριστή της ΚΜΕ και διαχειριστή μνήμης
- Πρόβλημα: επιλογή παραθύρου χρόνου Δt [Denning]
 - $\Delta t \sim$ χρόνου εκτέλεσης της διεργασίας για να μην
 - αδικούνται οι διεργασίες που έχουν πάρει μικρό χρόνο εξυπηρέτησης από την ΚΜΕ (γιατί?)
- Διαχειριστής μνήμης: χρήση των συνόλων εργασίας σε μια μορφή πολιτικών LRU ή LFU
 - καταχωρητές ηλικίωσης ή μετρητές χρήσης
 - σύνολο εργασίας = όλα τα τεμάχια που έχουν τουλάχιστον ένα 1 στα n πσδ τους
 - αν δεν έχει γίνει αναφορά σ' ένα τεμάχιο για n διαδοχικές διακοπές του χρονομέτρου, το τεμάχιο παύει ν' ανήκει στο σύνολο εργασίας
 - Η παράμετρος n προσδιορίζεται πειραματικά για το κάθε σύστημα
- αποτελεσματική πρόληψη του αλωνίσματος
 - όταν τα σύνολα εργασίας
 - διατηρούνται σχετικώς σταθερά
 - έχουν προβλεφθεί σωστά

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΈΛΕΓΧΟΣ ΦΟΡΤΟΥ

Σύγκριση Έγκρισης - Αποφυγής υπερφόρτωσης

- *Η έγκριση-επανόρθωση υστερεί της αποφυγής [Wilkes]*
 - σύστημα (ελέγχου) οδηγούμενο από σφάλματα (error-driven control systems)
 - Τα συστήματα αυτά είναι συνεχή και δύσκολα εφαρμόζονται οι τεχνικές τους σε διακριτά συστήματα
 - Τα Λ.Σ. είναι διακριτά συστήματα (ειδικά όταν το πλήθος των διεργασιών στη λίστα του διανομέα είναι μικρό)
 - ένα σύστημα εικονικής μνήμης πρέπει να λειτουργεί μακριά από το σημείο αλωνίσματος
 - στην έγκριση-επανόρθωση το σύστημα μπαίνει και βγαίνει στην κατάσταση του αλωνίσματος κι έτσι
 - ξοδεύει άδικα πολλούς πόρους
- ΟΜΩΣ η έγκριση-επανόρθωση χρησιμοποιείται συνήθως στην πράξη με χρήση τεχνικών μέτρησης της συχνότητας των σφαλμάτων

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΠΡΟΤΕΡΑΙΟΤΗΤΕΣ ΝΕΩΝ ΔΙΕΡΓΑΣΙΩΝ

Ποια θα είναι η αρχική προτεραιότητα μιας νέας διεργασίας?

- Ορίζεται από τον χρήστη
 - Κατάλληλη για συστήματα ενός χρήστη ή ειδικού σκοπού
- Ανάλογη με τους πόρους που θα χρειαστεί η διεργασία
 - Προβλήματα ανάλογα με αυτά της στατικής προτεραιότητας
- Καθορίζεται από τη διεργασία-γονέα
 - Περιορισμοί για την αποφυγή εκχώρησης πολύ υψηλής προτεραιότητας
 - Προτεραιότητα σε σχέση με αυτήν του γονέα
 - Ελάττωση περιοχής επιτρεπτών προτεραιοτήτων όσο χαμηλώνει η ιεραρχία
 - Παράδειγμα: 1^ο επίπεδο 0 ως 99, 2^ο επίπεδο 0 ως 9 κ.ο.κ.
 - διεργασία-γονέας = 20,
 - διεργασία-παιδί = 20 ± 9
 - Μια διεργασία-παιδί κληρονομεί την συμπεριφορά του γονέα της

Διαχείριση/χρονοπρογραμματισμός ΚΜΕ/διεργασιών

ΠΡΟΤΕΡΑΙΟΤΗΤΕΣ ΝΕΩΝ ΔΙΕΡΓΑΣΙΩΝ

Ποια θα είναι η αρχική προτεραιότητα μιας νέας διεργασίας?

- Η μέγιστη δυνατή προτεραιότητα
 - Αν είναι μικρή περιορισμένη από ε/ε διεργασία θα τελειώσει γρήγορα
 - Αν είναι μεγάλη περιορισμένη από ΚΜΕ θα υποβιβαστεί
 - Αλλά θα έχει καθυστερήσει άλλες περιορισμένες από ε/ε διεργασίες

Ιδανική λύση: προτεραιότητα νέων διεργασιών

- Μεγαλύτερη από αυτή των περιορισμένη από ΚΜΕ διεργασιών
- Μικρότερη από αυτή των περιορισμένων από ε/ε διεργασιών

Χρονοπρογραμματισμός ΚΜΕ στο σύστημα UNIX

- ουρές πολλών επιπέδων
- κάθε ουρά περιέχει διεργασίες με την ίδια προτεραιότητα
- περιέχουν διεργασίες έτοιμες και φορτωμένες στη μνήμη
 - ο διαχειριστής της μνήμης αποφασίζει ποιες διεργασίες θα φορτωθούν στη μνήμη
- RR πολιτική για όλες τις ουρές
- κβάντο = 100msec (4.2 BSD)
- αποφυγή υποσιτισμού: λαμβάνονται υπόψη οι ηλικίες των διεργασιών
- προτεραιότητα = ακέραιος αριθμός
- μεγαλύτερος ακέραιος εκφράζει μικρότερη προτεραιότητα
- προτεραιότητα διεργασιών χρήστη > 0
- προτεραιότητα διεργασιών συστήματος < 0
 - `swapper`
 - έλεγχος ε/ε συσκευών ομάδων
 - διαχείριση αρχείων
 - έλεγχος ε/ε συσκευών χαρακτήρων
 - αλλάζουν μόνο αν αλλάξουν ορισμένες σταθερές
 - επαναμεταγλώττιση του Λ.Σ.

Χρονοπρογραμματισμός ΚΜΕ στο σύστημα UNIX

- Χαμηλότερη προτεραιότητα στις διεργασίες που χρησιμοποιούν την ΚΜΕ περισσότερο.
 - σε κάθε διακοπή του χρονομέτρου αυξάνεται κατά 1 ο μετρητής χρήσης της ΚΜΕ της τρέχουσας διεργασίας
 - υπολογισμός προτεραιοτήτων κάθε 1sec
 - **προτεραιότητα = προτεραιότητα_βάσης + μετρητής**
χρήσης της ΚΜΕ/2
 - δεν τιμωρείται πολύ μια διεργασία για την πρόσφατη συμπεριφορά της
 - προτεραιότητα βάσης == 0
 - κλήση συστήματος `nice` για > 0
 - υπερχρήστης < 0
 - ευνοούνται οι περιορισμένες από ε/ε διεργασίες
- Εντολές χειρισμού διεργασιών
(`ps`, `^Z`, `kill`, `&`, `bg`, `fg`, `nice`)

Χρονοπρογραμματισμός ΚΜΕ στο σύστημα UNIX

ΕΛΕΓΧΟΣ ΦΟΡΤΟΥ

- Γίνεται από τον χρονοπρογραμματιστή της ΚΜΕ
- Αν αποφασιστεί ότι το σύστημα είναι υπερφορτωμένο, τότε μεταφέρονται (από τον swapper) στο δίσκο ολόκληρες διεργασίες μέχρι να επανέλθει ο φόρτος σε κανονικά επίπεδα
- **Υπερφόρτωση όταν**
 - υπάρχει μεγάλος μέσος όρος φόρτου
 - πλήθος ελεύθερων πλαισίων μικρότερο από μία ορισμένη ελάχιστη τιμή `minfree`
 - μέσος όρος ελευθέρων πλαισίων (για ένα προηγούμενο χρονικό διάστημα) μικρότερος από μία επιθυμητή τιμή `desfree`
 - `desfree, minfree`: κλάσματα της διαθέσιμης μνήμης
 - `lotsfree > desfree > minfree`

Χρονοπρογραμματισμός ΚΜΕ στο σύστημα UNIX

ΕΛΕΓΧΟΣ ΦΟΡΤΟΥ

Ανταλλαγή τεμαχίων διεργασιών μόνο όταν:

- υπάρχει ανταγωνισμός μεταξύ πολλών διεργασιών
 - το μέγεθος της διαθέσιμης μνήμης είναι μικρό
 - υπάρχει χρόνια έλλειψη πλαισίων
- ή
- για άλλους διαφορετικούς λόγους
 - μια διεργασία παραμένει αδρανής για' αρκετό χρονικό διάστημα (≥ 20 sec)

Χρονοπρογραμματισμός ΚΜΕ στο σύστημα UNIX

ΕΛΕΓΧΟΣ ΦΟΡΤΟΥ

- Οι εκτελέσιμες διεργασίες σπάνια ανταλλάσσονται ολοκληρωτικά
 - μια διεργασία χρειάζεται μόνο το σύνολο εργασίας της στη μνήμη
 - ο `pagedemon` αντικαθιστά τις αχρησιμοποίητες σελίδες μιας διεργασίας
- Λίστα κατάλληλων
 - κάθε μερικά sec ο `swapper` εξετάζει αν υπάρχουν έτοιμες διεργασίες στο δίσκο
 - επιλέγεται η διεργασία που έχει παραμείνει για περισσότερο διάστημα στο δίσκο (εκτός αν είναι πολύ μεγάλη)
 - η διεργασία προσκομίζεται μόνο όταν υπάρχουν αρκετά διαθέσιμα πλαίσια
 - φορτώνονται πρώτα ο πίνακας των σελίδων και η δομή χρήστη (περιβάλλοντος) της
 - μπαίνει στη λίστα των κατάλληλων
 - ανάλογα του χρόνου που έχει παραμείνει στο δίσκο απασχολεί χώρο μνήμης μέχρι 1/2 του εικονικού της χώρου
- Δύο επίπεδα χρονοπρογραμματισμού (αλληλεπίδραση)
 - χρονοπρογραμματιστής μνήμης (μεσοχρόνιος)
 - χρονοπρογραμματιστής ΚΜΕ (βραχυχρόνιος)
 - μικρή προτεραιότητα διεργασίας
 - μεγάλη πιθανότητα να αφαιρεθούν οι σελίδες της
 - μεγάλη πιθανότητα να αντικατασταθεί εντελώς η διεργασία