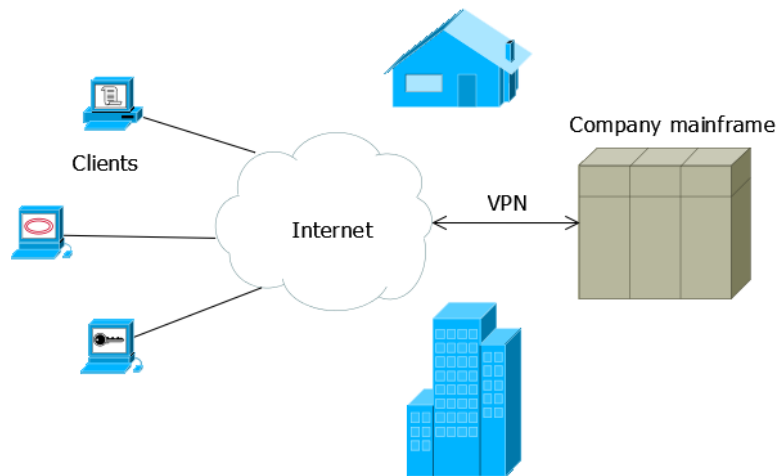


ΕΠΙΔΟΣΗ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΘΕΜΑ 3

Το τμήμα ενημέρωσης ενός οργανισμού παροχής συμβουλευτικών υπηρεσιών υποστηρίζεται από κεντρικό εξυπηρετητή (mainframe) στον οποίο οι πελάτες συνδέονται απομακρυσμένα μέσω εικονικού ιδιωτικού δικτύου (Virtual Private Network – VPN). Ο εξυπηρετητής περιλαμβάνει μια ΚΜΕ (CPU) και δύο δίσκους. Το φορτίο του συστήματος αποτελείται από διάφορες εφαρμογές (όπως HTTPS, SFTP, SMTP, SSH) που φτάνουν στο σύστημα (στη CPU) μέσω του VPN. Μετά την επεξεργασία της, κάθε αίτηση (εργασία) αποχωρεί από το σύστημα μέσω εξερχόμενης σύνδεσης. Οι εφαρμογές και τα δεδομένα που χρησιμοποιούν οι πελάτες δημιουργούν διάφορες απαιτήσεις σε υπολογιστικούς πόρους του συστήματος. Η επιβάρυνση λόγω του VPN θεωρείται αμελητέα.



Α. Χαρακτηρισμός φορτίου

Προκειμένου να μελετηθεί η επίδοση του συστήματος με τη βοήθεια μοντέλου αναμονής, θα πρέπει να προσδιοριστούν οι παράμετροι του φορτίου (workload characterization). Για τον σκοπό αυτό, πραγματοποιήθηκαν μετρήσεις με χρήση εργαλείων εποπτείας (monitors) του συστήματος. Τα δεδομένα των μετρήσεων περιγράφονται στη συνέχεια.

Σύνολο δεδομένων

Το αρχείο "mainframe.log" είναι ένα αρχείο καταγραφής (logfile), όπου έχουν αποτυπωθεί εργασίες (jobs), οι οποίες εκτέλεσαν συναλλαγές στο υπό μελέτη υπολογιστικό σύστημα, σε μια συνεχή περίοδο χρόνου 194.250 sec. Κάθε γραμμή αντιστοιχεί σε μια άφιξη – εργασία στο σύστημα και έχει 6 πεδία (χαρακτηριστικά):

1. Μέση απαίτηση (χρόνος) ανά επίσκεψη στη CPU
2. Μέση απαίτηση (χρόνος) ανά επίσκεψη στον δίσκο 1.

3. Αριθμός επισκέψεων στον δίσκο 1.
4. Μέση απαίτηση (χρόνος) ανά επίσκεψη στον δίσκο 2.
5. Αριθμός επισκέψεων στον δίσκο 2.
6. Μέση απαίτηση (χρόνος) ανά επίσκεψη στην εξερχόμενη σύνδεση.

(όλοι οι χρόνοι είναι σε msec)

Ομαδοποίηση (Clustering)

Τα δεδομένα των μετρήσεων εμφανίζουν υψηλή μεταβλητότητα ως προς ορισμένα χαρακτηριστικά, οπότε ένα μοντέλο βασισμένο στη συνολική μέση συμπεριφορά του φορτίου δεν θα ήταν αντιπροσωπευτικό. Συνεπώς, θα γίνει ομαδοποίηση των μετρήσεων, ώστε να προκύψουν ομάδες (clusters) με χαμηλή μεταβλητότητα στο εσωτερικό τους, οι οποίες θα αντιστοιχούν σε διαφορετικές κατηγορίες πελατών.

Για την ομαδοποίηση και την εξαγωγή των κατηγοριών, θα χρησιμοποιηθεί το εργαλείο "JWAT" του πακέτου JMT και, συγκεκριμένα, η εφαρμογή "Workload Analysis". Για την εισαγωγή των δεδομένων, στο tab "Input" θα θέσετε το αρχείο "mainframe.log" ως αρχείο εισόδου (input file), και το αρχείο "mainframe.jwatformat" ως αρχείο μορφοποίησης (fileformat), το οποίο επιτρέπει την ανάγνωση του logfile. Ως μέθοδο δειγματοληψίας "Workload sampling method" θα διαλέξουμε τυχαία δειγματοληψία ("Random sampling") με τιμή 50.000 παρατηρήσεις.

Η ομαδοποίηση θα πραγματοποιηθεί με χρήση του αλγορίθμου k-Means. Οι παράμετροι της ομαδοποίησης ορίζονται στο tab "Clustering". Στο πεδίο "variables" επιλέξτε όλες τις μεταβλητές ώστε η ομαδοποίηση να γίνει με βάση όλα τα διαθέσιμα χαρακτηριστικά του αρχείου καταγραφής.

Στις παραμέτρους του k-Means (clustering options) επιλέξτε:

- Number of clusters: 4
- Αριθμός επαναλήψεων (iterations): 20
- Μετασχηματισμός (transformation): $(\text{value} - \text{min}) / (\text{max} - \text{min})$

Αποτελέσματα

Στη συνέχεια, πατήστε "επίλυση" (solve). Στο επόμενο tab "Clustering information", θα δείτε τα αποτελέσματα της ομαδοποίησης για 2, 3 και 4 ομάδες (Num. of clusters). Στον ίδιο πίνακα, περιλαμβάνονται δύο κριτήρια της ποιότητας του clustering: ο δείκτης Ratio και ο δείκτης Overall Mean Square Ratio (OMSR). Ο δείκτης Ratio εκφράζει τον λόγο της διακύμανσης των μεταβλητών μεταξύ των ομάδων ως προς τη διακύμανσή τους εντός των ομάδων (higher better).

Από τα αποτελέσματα, θα χρησιμοποιήσουμε αποκλειστικά την ομαδοποίηση σε k=3 ομάδες (το επιλέγετε στο Num. of Clusters), που αντιστοιχούν στις τρεις κατηγορίες εργασιών που θέλουμε να διακρίνουμε στο υπό μελέτη σύστημα. Θα εκτελέσετε 5 επαναλήψεις της ομαδοποίησης, με διαφορετική αρχικοποίηση κάθε φορά, και θα επιλέξετε την ομαδοποίηση εκείνη, που πετυχαίνει το υψηλότερο Ratio για k=3 ομάδες.

Για να ξαναγίνει αρχικοποίηση και εκτέλεση του αλγορίθμου πρέπει να πάτε πίσω στο tab "Input", να δεχθείτε το reset και να επαναλάβετε τη διαδικασία. Οι διαδοχικές εκτελέσεις του αλγορίθμου αποθηκεύονται στο tab "Clustering information" στον πίνακα "Clusterings" με τη σειρά εκτέλεσής τους (η πιο πρόσφατη εκτέλεση πιο κάτω). Τυχόν άκυρες εκτελέσεις (μηδενικά αποτελέσματα) μπορούν να αγνοηθούν και να διαγραφούν.

Για την ομαδοποίηση που θα επιλεγεί ως βέλτιστη, στο tab "Clustering info" στον πίνακα "Cluster details", θα βρείτε το ποσοστό των εργασιών που αντιστοιχεί σε κάθε κατηγορία. Στο δεύτερο tab, "Cluster info", θα βρείτε τις συνολικές μέσες τιμές - κέντρα ("Center") της κάθε ομάδας για όλες τις μεταβλητές εισόδου. Με βάση τις τιμές αυτές, μπορούν να προσδιοριστούν οι παράμετροι του φορτίου για κάθε κατηγορία εργασιών.

B. Μελέτη ανοικτού δικτύου

1. Με βάση τα υπό μελέτη δεδομένα και τα χαρακτηριστικά των τριών κατηγοριών, όπως προέκυψαν από την ομαδοποίηση, υπολογίστε:

- α) τον συνολικό ρυθμό αφίξεων και τον ρυθμό αφίξεων κάθε κατηγορίας
- β) τη συνολική απαίτηση εξυπηρέτησης των εργασιών, για κάθε σταθμό του δικτύου και για κάθε κατηγορία

2. Επιπρόσθετα υπολογίστε:

- α) για κάθε σταθμό του δικτύου και για κάθε κατηγορία:

- τον ρυθμό απόδοσης
- τον βαθμό χρησιμοποίησης
- τον χρόνο παραμονής
- τον μέσο αριθμό εργασιών

- β) συνολικά για το δίκτυο:

- τον μέσο χρόνο απόκρισης ανά κατηγορία
- τον μέσο αριθμό εργασιών ανά κατηγορία

Γ. Μελέτη κλειστού δικτύου

Ο οργανισμός μελετά την επέκταση των πληροφορικών υποδομών του με την ανάπτυξη ενός εταιρικού διαλογικού δικτύου intranet, ώστε μελλοντικά να μπορεί να ανταποκριθεί σε αύξηση του φορτίου. Η μελέτη βασίζεται στον σχεδιασμό κλειστού δικτύου, με χρήση των δεδομένων του ανοικτού δικτύου. Επειδή δεν ικανοποιούνται πλήρως οι προϋποθέσεις του μοντέλου BCMP, συνιστάται η χρήση του εργαλείου JSIM του JMT. Το κλειστό δίκτυο θα οριστεί ως εξής:

Στην ΚΜΕ και στους δίσκους η απαίτηση εξυπηρέτησης θα ακολουθεί εκθετική κατανομή με μέση τιμή ίση με αυτή που υπολογίστηκε στο ανοικτό δίκτυο για την κάθε κατηγορία. Η CPU ακολουθεί κανονισμό PS και οι δίσκοι FCFS.

Η εξερχόμενη σύνδεση αντικαθίσταται από τερματικά με χρόνο σκέψης που θα ακολουθεί εκθετική κατανομή με μέση τιμή 100 φορές τον χρόνο εξυπηρέτησης στον εξερχόμενο σύνδεσμο για την κάθε κατηγορία του ανοικτού δικτύου.

Ως προς τον αριθμό εργασιών ανά κατηγορία, προβλέπεται να είναι πενταπλάσιος από τον μέσο αριθμό εργασιών της κάθε κατηγορίας στο ανοικτό δίκτυο στρογγυλοποιημένος προς τον επάνω ακέραιο (ceiling).

Για όλες τις κατηγορίες θεωρήστε ως σταθμό αναφοράς τα τερματικά.

Κατά την παράδοση του συστήματος θα υπογραφεί Συμφωνία Επιπέδου Υπηρεσίας (Service Level Agreement – SLA). Ένας από τους όρους του SLA ορίζει ότι ο μέσος χρόνος απόκρισης των αιτημάτων για κάθε κατηγορία δεν επιτρέπεται να υπερβαίνει ένα δεδομένο ποσοστό του μέσου χρόνου σκέψης στα τερματικά.

Με βάση τα αποτελέσματα της προσομοίωσης ποια πρέπει να είναι η ελάχιστη τιμή αυτού του ποσοστού ώστε το υπό εξέταση σύστημα να τηρεί το SLA;