

**אי ודאות בביקוש לשירות,
עם שימושים למוקדים טלפוניים וחדרי מיון**

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר
מגיסטר למדעים בסטטיסטיקה

שמרית ממן

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

יולי 2009

חיפה

תמוז תשס"ט

המחקר נעשה בהנחייתם של פרופסור אבישי מנדלבאום
וד"ר סרגיי זלטין בפקולטה לתעשייה וניהול.

אני מודה לפרופ' אבישי מנדלבאום ולד"ר סרגיי זלטין על הדרכתם
הצמודה והמקצועית במהלך העבודה על המחקר.

ברצוני להודות להוריי, אווה ורחמים, לאחותי ימית ולכל החברים על
התמיכה והעידוד.

לבסוף, אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

תקציר מורחב

מערכות שירות הינן חלק בלתי נפרד מחיי היום-יום שלנו, והן מהוות כיום חלק נרחב מכלכלתן של מדינות מערביות. במחקר זה אנו עוסקים במוקדי שירות טלפוניים ובחדרי מיון של בתי-חולים, אשר מהווים שער למערכת השירות הרלוונטית וחלון ראווה למה שקורה בתוכה.

ההנחה הסטנדרטית במידול מערכות שירות היא שתהליך מופע חיצוני של לקוחות הוא פואסוני, עם קצב מופע ידוע, ויישום הגישה הרווחת לאיוש מערכת שכזו נעשה לפי השלבים הבאים:

1. הסקת הפרמטר הפואסוני (קצב המופע) מנתונים היסטוריים, לפי אחת משיטות החיזוי (ממוצע נע MA או סדרות עיתיות מורכבות יותר).

2. הצבת הפרמטר במודל תורים שמתאים למערכת השירות (Erlang-A, Erlang-C), מודל לניתוב מבוסס מיומנויות וכדומה).

3. קביעת רמת האיוש לפי המודל התורי, כדי לעמוד ברמת שירות מוסכמת וקבועה מראש. נשאלת השאלה האם הנחת הפואסון הסטנדרטית נכונה ותקפה? באופן עקבי, מניתוח נתוני מוקדים טלפוניים, התשובה לשאלה זו שלילית, ונמצא כי השתנות תהליך המופע האמיתי גדולה יותר מזו שנצפה מתהליך פואסון סטנדרטי. אנו מייחסים את פיזור היתר לאי הודאות בקצב המופע. כלומר, ישנה השתנות הנוספת להשתנות הסטוכסטית של תהליך הפואסון הסטנדרטי.

בעבודה זו אנו חוקרים מערכות שירות עם אי ודאות בקצב המופע, הפועלות תחת משטר מכוון שירות ויעילות (מש"י) - Quality and Efficiency Driven (QED). משטר ה-QED משלב רמות שירות ויעילות גבוהות במערכת, דבר שהוא אפשרי עקב יתרון לגודל.

מודל התורים אותו אנו מנתחים הינו מודל $M^2 | M | n + G$, בו תהליך המופע הוא פואסוני, כאשר הסימון "?" מציין כי קצב המופע אינו ידוע. בנוסף, משכי השירות מפולגים מעריכית עם קצב שירות μ , במערכת n שרתים זהים סטטיסטית שעובדים במקביל ומשכי הסבלנות הם מהתפלגות כללית G .

את קצב המופע אנו ממדלים כמשתנה מקרי ומניחים שערכו משתנה מאינטרוול לאינטרוול (למשל, משעה לשעה) באופן בלתי תלוי ושווה התפלגות מהצורה:

$$(1) \quad \lambda + \lambda^c \cdot X, \quad c \leq 1,$$

כאשר X הוא מ"מ עם תוחלת 0 וסטטיית תקן סופית $\sigma(X)$, והוא מגלם את הסטוכסטיות בקצב המופע.

הפרמטר c נקרא מקדם אי הודאות, והוא מתאר את סדר גודל הרעש סביב הממוצע, λ . מקדם אי הודאות מקבל ערכים קטנים או שווים ל-1, כאשר אנחנו מתייחסים לחלוקה הבאה:

1. $c \leq 1/2$ - השתנות שאינה גדולה מההשתנות הפואסונית הטבעית. במקרה זה, על מנת להשיג ביצועי QED, יש לאייש את המערכת לפי כלל השורש, כפי שיתואר בהמשך.
הניתוח במקרים אלו מתאים לתוצאות בעבודותיהם של Halfin and Whitt (1981) ו- Garnett et al. (2002).
2. $1/2 < c < 1$ - השתנות מתונה. במקרה זה, אנו מציגים כלל איוש ומשטר חדש הנקרא QED-c, שיתוארו בהמשך.
3. $c = 1$ - השתנות קיצונית בסדר גודל של הפרמטר הממוצע. כאן אנו מראים כי לא ניתן להשיג ביצועי QED, במובן שיחד עם התכנסות ההסתברות להמתנה לקבוע $0 < \alpha < 1$, ההסתברות לנטוש גם כן מתכנסת לקבוע (להבדיל מהתכנסות ל-0 במשטר ה-QED). במקרה זה, אנו מציגים איוש המשיג רמת ביצועים המתאימה למשטר Efficiency Driven (ED) המאופיין ביעילות גבוהה יחד עם רמת שירות נמוכה יחסית לזו המושגת במשטר QED.

עבור $c > 1$ הממוצע הופך לזניח ביחס לרעש ולכן אנו לא דנים במקרים אלו.

בדיקת תקפות מודל קצב המופע נעשית על-ידי ניתוח נתוני הגעת שיחות למוקד שירות טלפוני של בנק ישראלי גדול. הניתוח כולל מודל רגרסיה ליניארית המספק אומדנים ל- c (התקבל כי $c \approx 0.8$) ו- $\sigma(X)$, וכן התאמת התפלגות Gamma Poisson Mixture לנתונים שלנו. בנוסף, נמצא כי עבור ערכי λ גדולים, המ"מ X מפולג בקירוב נורמלית, $Norm(0, \sigma(X))$. ניתוח נתוני הגעת חולים לחדר מיון בבית חולים ישראלי מראה התאמה למודל המופעים שלנו עם $c \approx 0.5$.

משטר QED-c ($1/2 < c < 1$)

פרק 4.2 מציג את משטר QED-c המאופיין ע"י כלל האיוש הבא:

$$(2) \quad n = R + \beta \cdot R^c, \quad -\infty < \beta < \infty, \quad 1/2 < c < 1.$$

הקבוע $R = \lambda / \mu$ נקרא העומס המוצע על המערכת, והוא מתאר את כמות העבודה הממוצעת, הנמדדת ביחידות זמן שירות, הנכנסת למערכת ביחידת זמן. הפרמטר β נקרא פרמטר איכות השירות.

ניתוח התנהגות גבולית ($\lambda \rightarrow \infty$) של תור $M^2 | M | n + G$ עם קצב מופע (1), כאשר $1/2 < c < 1$, ואיוש לפי (2) עם ערך c תואם, מראה כי ההסתברות להמתין מתכנסת לקבוע $0 < \alpha < 1$ וכי ההסתברות לנטוש וממוצע זמן ההמתנה המוצע (הזמן שימתין לקוח בעל סבלנות

אינסופית) מתכנסים ל-0 בקצב $1/n^{1-c}$. התכנסויות אלה, למעט קצבי ההתכנסות, תואמות להתכנסויות המתקבלות במשטר QED הקלאסי (קצב מופע דטרמיניסטי ואיוש לפי כלל השורש). בדיקת טיב הקירובים מראה התאמה טובה בין הקירוב שנמצא עבור ההסתברות להמתין לבין ערכה המדויק עבור ערכי c נמוכים (c קרוב ל-0.5), והתאמה טובה מאוד עבור ערכי c גבוהים (c מתקרב ל-1). בנוסף, ניתן לראות דיוק מצוין של הקירוב להסתברות לנטוש, עבור כל ערכי c .

משטר ED ($c = 1$)

נפתור בעיית Constraint Satisfaction למציאת רמת האיוש המינימלית, עבורה ההסתברות להתנתה אינה עולה על ערך קבוע מראש α . פתרון הבעיה עבור המקרה $c = 1$ מוצג עבור שני מקרים נפרדים:

א. X הוא משתנה מקרי בדיד - כלל האיוש המתאים הוא מהצורה של כלל השורש, תוך שינוי העומס המוצע בהתאם לתחום בו נמצאת ההסתברות להמתין אותה אנו שואפים להשיג.

ב. X משתנה מקרי רציף - כלל האיוש המתאים הוא מהצורה $n = R \cdot \delta$, כאשר δ הוא פרמטר איכות השירות, $\delta > 0$.

תחת כללי איוש אלה, ההסתברות לנטישה מתכנסת לקבוע וממוצע משך ההמתנה הוא בסדר גודל של משך שירות. תכונות אלה עקביות עם המקובל עבור ממשטר ה-ED. ניסויים נומריים מראים התאמה מצוינת של הקירובים לערכיהם המדויקים של מדדי הביצוע.

משטר QED ($c \leq 1/2$)

עבור אי ודאות מסדר $c \leq 1/2$ נמצא כי כלל השורש פותר את בעיית ה-Constraint Satisfaction שהוצגה קודם לכן. כלל השורש הוא מהצורה

$$(3) \quad n = R + \beta \cdot \sqrt{R}, \quad -\infty < \beta < \infty.$$

הקשר בין ההסתברות להמתין α ופרמטר איכות השירות β מתואר על-ידי פונקצית גרנט (Garnett), נסמנה ב- $\alpha(\beta)$.

נמצא כי עבור $c < 1/2$, ההתנהגות הגבולית של תור מסוג $M \mid n + G$ שקולה לזו של תור $M \mid M \mid n + G$ (ללא רכיב האקראיות בקצב המופע). אם כך, כלל השורש, המהווה פיתרון לבעיה המוצגת עבור $M \mid M \mid n + G$ (Zeltyn and Mandelbaum (2005)), תקף גם כאשר קיימת אי ודאות בקצב המופע מהצורה (1) עם $c < 1/2$, ומשיג רמת ביצועים של משטר QED. עבור $c = 1/2$, שוב פותר כלל השורש את לבעיית האיוש, אך כעת הקשר בין ההסתברות להמתין ופרמטר איכות השירות הינו

$$P\{W_q > 0\} = E[\alpha(\beta - X)].$$

מערכות שירות עם קצב מופע תלוי-זמן ואקראי

הניתוח האנליטי בעבודת המחקר נעשה עבור מערכות תורים בהן תהליך המופע הוא סטציונארי, כלומר אינו תלוי בזמן. בפועל, קצבי תהליכי המופע הם תלויי זמן ונשאלת השאלה האם ניתן להיעזר בכללי האיוש המתאימים לתורים סטציונאריים גם עבור תורים תלויי-זמן. מפאת מורכבות מערכות תלויות זמן, אנו משתמשים בסימולציות המהוות גישה אלטרנטיבית לניתוח אנליטי. במאמרם של Feldman et al. (2007) מתואר האלגוריתם Iterative Staffing Algorithm (ISA), שמטרתו היא לקבוע פונקצית איוש (רמת איוש תלויה זמן) המשיגה יציבות בזמן של ההסתברות להמתין על ערך שנקבע מראש. לקוד הסימולציה המיישם את האלגוריתם הוסף רכיב האקראיות בקצב המופע, והורצו סימולציות, עבור ערכי c שונים, על מנת לבדוק האם כללי האיוש המתאימים למקרים הסטציונאריים המתוארים בפרק 4 מתאימים גם לתורים תלויי זמן. סיכום תוצאות הסימולציות מראה שאכן ניתן להתאים את כללי האיוש הסטציונאריים למציאות המשתנה בזמן. ביתר פירוט:

1. ניתן לייצב את ההסתברות להמתנה על ערכי מטרה הנקבעים מראש, תוך שימוש באלגוריתם ה-ISA.
2. מושגת רמה סבירה של יציבות בזמן של מדדי ביצוע נוספים של המערכת: ההסתברות לנטישה, ממוצע אורך תור ונצילות השרתים.
3. מוצעים כללי איוש תלויי-זמן האנלוגיים לכללי האיוש הסטציונאריים. עבור כל מקרה שנבחן נמצא כי, יחד עם ייצוב ההסתברות להמתנה α , פרמטר איכות השירות β גם הוא יציב בזמן והיחס בין α ו- β ניתן לתיאור כבמקרה הסטציונארי המתאים. למשל, עבור $1/2 < c < 1$, האנלוג לכלל האיוש הסטציונארי שתואר קודם לכן הוא למעשה

$$n_t = R_t + \beta \cdot R_t^c, \quad t \geq 0,$$

כאשר R_t הוא האנלוג תלוי הזמן של העומס המוצע $R = \lambda / \mu$, דהיינו המספר הממוצע של שרתים עסוקים במערכת $G + M | \infty + M_t^?$ (אינסוף שרתים) מתאימה.