



מרצה : פרופסור אבישי מנדלבאום

מתרגל: איתמר זייד

תאריך הבחינה: 10.02.2011

מס' סטודנט _____

שם _____

הנדסת מערכות שירות – 096324

מועד א' – סמסטר חורף תשע"א 2011

פתרון

- המבחן נמשך שלוש שעות. באחריותך להחזירו במועד.
- המבחן בחומר סגור.
- מותר להשתמש במחשבון.
- המבחן כולל 29 עמודים, כולל עמוד זה, ודף התפלגויות מהסתברות בעמוד האחרון.
- כאשר נדרשים חישובים, ניתן לעגלם באופן שאינו פוגע בהבנת התשובה.
- מלאו את הפרטים הנדרשים בראש העמוד, בכתב ברור בבקשה.
- על התשובות יש לענות במקומות המיועדים, שאמורים להספיק: מלל מיותר יגרע ניקוד.
- יש לתת הסברים או הוכחות רק אם התבקשתם במפורש לעשות זאת.

ניקוד:	שאלה 1	_____ מתוך 10 נקודות	תרגילי בית
	שאלה 2	_____ מתוך 10 נקודות	הרצאות/תרגולים/בחינות קודמות
	שאלה 3	_____ מתוך 20 נקודות	יישום
	שאלה 4	_____ מתוך 10 נקודות	תיאוריה

סה"כ _____ מתוך 50 נקודות

בהצלחה!

שאלה 1 תרגילי בית (10 נקודות)

חלק 1.1. מודל נוזלים

Tele-SHOP היא חברה העוסקת במכירות המתבצעות באמצעות מוקד שירות טלפוני. ניתן לתאר את הגעת השיחות למוקד השירות ע"י תהליך פואסון לא הומוגני בזמן עם פונקציית הקצב $\lambda(t)$. המוקד פועל בשעות 16:00-24:00. נסמן אינטרוול זמן זה ב- $[0, T]$.

ממוצע זמן שירות של שיחה הינו $\frac{1}{\mu}$ דקות וממוצע סבלנותו של לקוח (עד נטישה) הינו $\frac{1}{\theta}$.

הניחו כי עלות העסקת שרת הינה c ש"ל לדקה ועלות המתנה הינה h ש"ל לדקה. עוד הניחו כי הרווח מסיום שירות הינו r ש"ל והקנס על נטישת לקוח הינו s ש"ל.

נסמן ב- $Q(t)$ את מספר הלקוחות במערכת בזמן $0 \leq t \leq T$,

וב- $N(t)$ את מספר המוקדנים במערכת בזמן $0 \leq t \leq T$.

1.1.1 כתבו משוואה דיפרנציאלית עבור $Q(t)$, $0 \leq t \leq T$.

תשובה:

$$\frac{dQ(t)}{dt} = \lambda(t) - \mu(Q(t) \wedge N(t)) - \theta(Q(t) - N(t))^+, t \geq 0$$

1.1.2 תארו בקיצור איך ניתן, בעזרת גיליון אלקטרוני בלבד, לפתור בקירוב (מדויק ככל הנדרש) את המשוואה הדיפרנציאלית בסעיף 1.1.1.

תשובה:

נתחיל עם $Q(0)$ נתון. לאחר מכן, עבור $n = 1, 2, \dots$, נחשב,

$$Q(t_n) = Q(t_{n-1}) + \lambda(t_{n-1})\Delta - \mu(Q(t_{n-1}) \wedge N(t_{n-1}))\Delta - \theta(Q(t_{n-1}) - N(t_{n-1}))^+ \Delta$$

ונקבל את $Q(0), Q(\Delta), Q(2\Delta), \dots$. ככל ש- Δ קטן יותר, כך הקירוב מדויק יותר.

1.1.3 רשמו ביטוי לאחוז הנוטשים באינטרוול $[0, T]$.

תשובה:

$$P\{Ab\} = \frac{\int_0^T \theta(Q(t) - N(t))^+ dt}{\int_0^T \lambda(t) dt}$$

1.1.4 על מנת לקבוע את פונקצית האיוש $\{N(t), t \geq 0\}$ המשיגה רווח יומי מקסימלי של מוקד השירות, יש לפתור בעיית אופטימיזציה.

רשמו את פונקצית המטרה של בעיית האופטימיזציה ואת אילוצי הבעיה.

תשובה:

נסמן:

c – עלות מוקדן ליח' זמן, h – עלות המתנת לקוח ליח' זמן, r – רווח מסיום שירות, s – קנס על נטישת לקוח.

פונקצית המטרה:

$$C(N) = \int_0^T [r\mu(Q(t) \wedge N(t)) - (s\theta + h)(Q(t) - N(t))^+ - cN(t)] dt$$

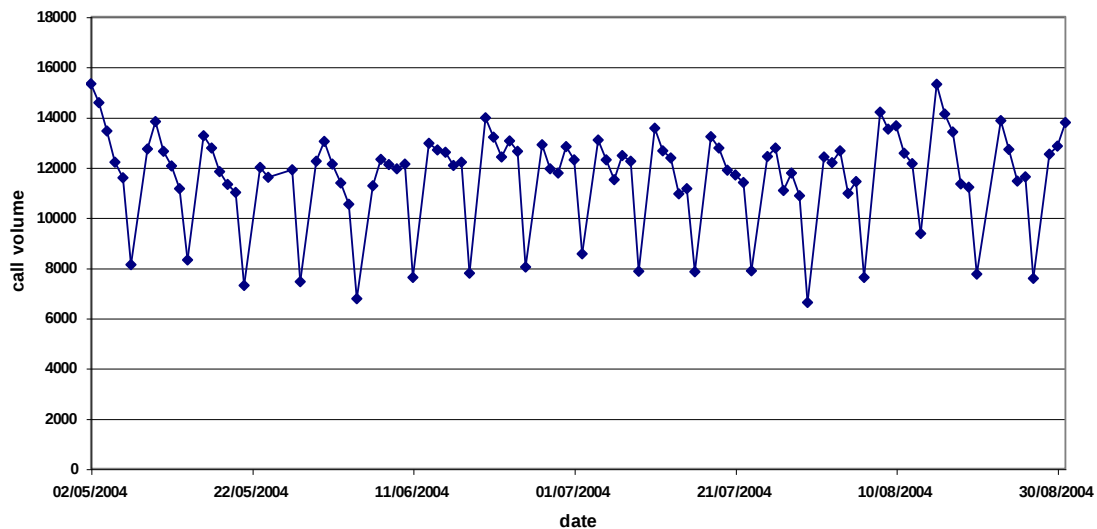
המטרה, להביא למקסימום את $C(N)$ (על $N(t), 0 \leq t \leq T$), תחת האילוצים:

- $\frac{dQ(t)}{dt} = \lambda(t) - \mu(Q(t) \wedge N(t)) - \theta(Q(t) - N(t))^+$
- $N(t) \geq 0, 0 \leq t \leq T$

חלק 1.2 מופעים ותחזיות.

נתונים סיכומיים של נתוני שיחות טלפון, במוקד של חברת תקשורת ישראלית, על פני שלושה חודשים. הנתונים כוללים, לכל אינטרוול זמן של חצי שעה, את מספר השיחות הנכנסות למוקד. מוגדרים 33 אינטרוולים במהלך יום עבודה. כמו בתרגיל הבית, נסמן את N_{jk} להיות מספר השיחות הנכנסות ביום j באינטרוול k . בעוד F_{jk} הוא תחזית מספר השיחות שיכנסו ביום j באינטרוול k .

1.2.1 הגרף הבא מציג את סך השיחות הנכנסות ליום בתקופה 02/05/2004-30/08/2004 (בימים א'-ו', ללא שבת). אילו מסקנות (בעלות משמעות תפעולית) ניתן להסיק מהגרף?



תשובה: ניתן לראות מבנה שבועי בו יש בימי שישי הרבה פחות שיחות וביום ראשון המספר הגבוה ביותר של שיחות. באופן כללי נראה שמספר השיחות יורד הדרגתית מראשון עד שישי. להבחנות אלו יש כמובן משמעות תפעולית בהקשר של החלטות איוש.

1.2.2 תארו במדויק את אחת משיטות התחזית (Forecasting) שהוצגו בכיתה או בתרגילי הבית. ציינו באיזו שיטה בחרתם.

(השתמשו בתשובתכם בסימונים N_{jk} ו- F_{jk})

תשובה: למשל ממוצע נע (4) - נגדיר את התחזית הבאה על סמך ממוצע של ארבעת האינטרוולים האחרונים מאותו סוג (למשל יום ג' בין 10:00 ל- 10:30):

$$F_{j,k} = \frac{N_{j-6,k} + N_{j-12,k} + N_{j-18,k} + N_{j-24,k}}{4}$$

חלק 1.3 אמידת סבלנות.

להלן ארבע שיטות לאמידת סבלנות של לקוחות:

1. מודל אקספוננציאלי סטנדרטי.

הזמן שלקוח מוכן להמתין הוא מהתפלגות אקספוננציאלית עם פרמטר θ .

2. מודל אקספוננציאלי עם נטישות מיידיות.

לקוח המתקשר ומסתבר לו שהוא נדרש להמתין בתור, נוטש באופן מיידי בהסתברות p (באופן מעשי זמן הנטישה ייקח מספר שניות). במידה והלקוח לא נוטש מיידית, סבלנותו תהיה אקספוננציאלית עם פרמטר θ .

3. סבלנות בת שתי פאזות.

סבלנות $X + Y^d$, ז"א סכום של X ו- Y , כאשר X הינו משתנה מקרי (בפרט קבוע) המתאים לזמן בו הלקוח "מרגיש בסדר" ולא חושב על נטישה, המ"מ Y מתאים לזמן בו הלקוח מתחיל להיות קצר רוח, והתפלגות Y הינה אקספוננציאלית עם פרמטר θ . לבסוף, X ו- Y בלתי תלויים.

4. התנהגות אדפטיבית.

סבלנות הלקוחות משתנה במשך שעות היום (לדוגמא בשעות אשר הלקוחות מצפים להמתין זמן רב תהיה להם סבלנות ארוכה).

עליכם למדל את סבלנותם של לקוחות ממענה קולי טלפוני מסוים. כדי לבחור מודל מתאים לסוגים שונים של לקוחות, שורטטו ארבעת הגרפים הבאים:

- כל גרף מייצג סוג לקוחות מסוים (PS, IN, NE, NW).
- כל גרף מתאר פיזור נקודות, כאשר כל נקודה מחושבת כממוצע במשך רבע שעה.
- ערכי נקודה בגרף הם: ממוצע זמן ההמתנה של הלקוחות שהמתינו בתור, ברבע שעה מסוימת, מול אחוז הנוטשים של לקוחות אשר המתינו בתור באותה רבע שעה.
- בכל גרף משורטט קו ישר אשר נוצר על ידי רגרסיה ליניארית בין שני המשתנים.

Figure 1 NE customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

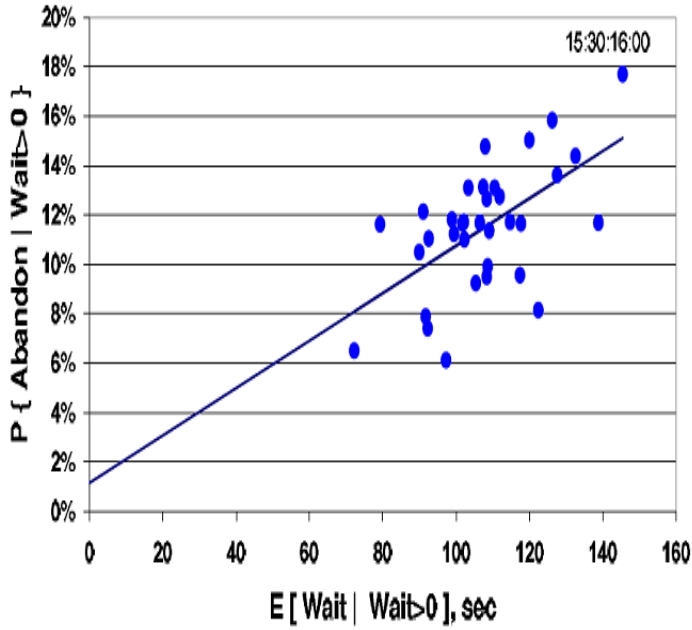


Figure 2 NW customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

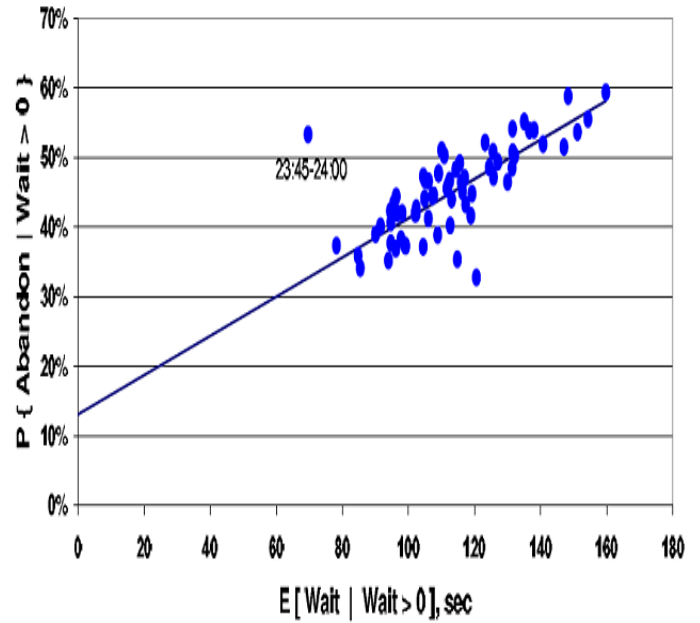


Figure 3 PS customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

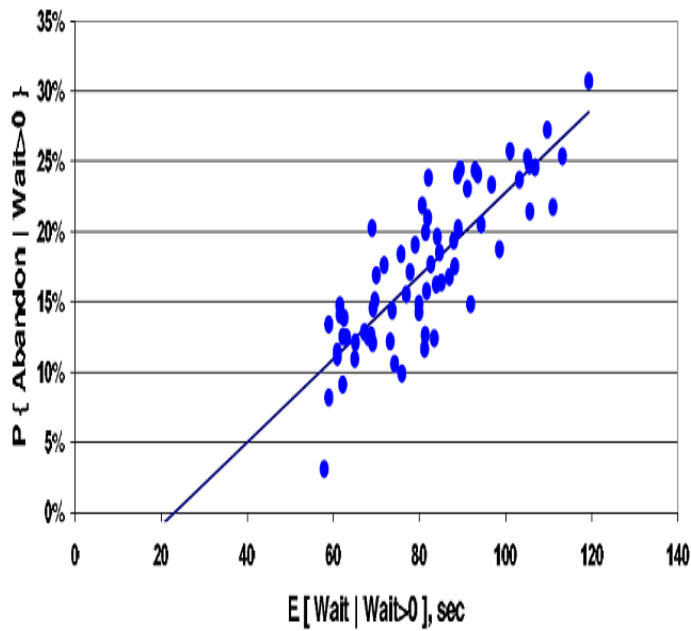
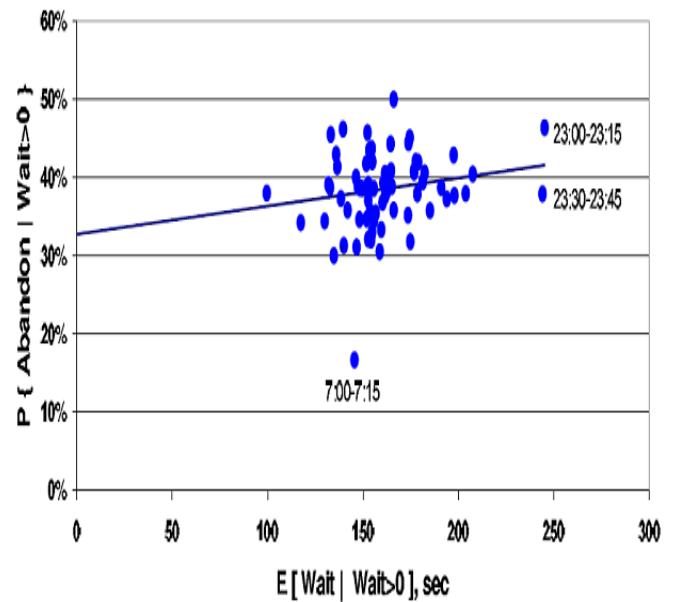


Figure 4 IN customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.



1.3.1 הסבירו בקצרה מדוע הנתונים שנבחרו עבור מידול הסבלנות (אותם נתונים אשר מופיעים בארבעת הגרפים), הם נתונים המדברים על לקוחות שהמתינו בתור בלבד, ולא על כלל הלקוחות. (במילים אחרות, מדוע לא נכללו הלקוחות ששורתו מיידית?)

תשובה: כאשר אנו אומדים סבלנות לקוחות אשר לא המתינו בתור, אלו לא תורמים מידע לגבי סבלנותם.

1.3.2 על סמך הגרפים הנ"ל, התאימו מודל סבלנות מתאים (מארבעת המודלים המוצגים בשאלה) לכל אחד מארבעת סוגי הלקוחות. הסבירו את תשובתכם.

לקוחות NE

תשובה: מאחר ובמודל אקספוננציאלי מתקיים הקשר $P(Abandon | Wait > 0) = \theta E(Wait | Wait > 0)$ עבור גרף בו קיים קשר ליניארי העובר בראשית הצירים נבחר את המודל אקספוננציאלי. (הגרף "כמעט" עובר דרך הראשית).

לקוחות NW

תשובה: מאחר וישנה נקודת חיתוך עם ציר Y בערך ב $P(Abandon | Wait > 0) \approx 0.12$, המודל המתאים יהיה מודל אקספוננציאלי עם קפיצה, בו ישנה הסתברות חיובית לנטוש ללא זמן המתנה משמעותי ולאחר מכן אנו מקבלים קו ישר, כלומר לקוחות אשר החליטו לא לנטוש ישר הם בעלי סבלנות אקספוננציאלית.

לקוחות PS

תשובה: מאחר וישנה נקודת חיתוך עם ציר ה X. בנקודה $E(Wait | Wait > 0) \approx 21\text{sec}$, אנו נבחר את מודל שתי הפאזות כאשר X הינו מ"מ דטרמיניסטי השווה ל 21 שניות לאחר מכן אנו עוברים לפאזה השנייה Y אשר יוצרת קו ישר ולכן מתאימה להתפלגות אקספוננציאלית.

לקוחות IN

תשובה: מאחר ואנו מקבלים קו ישר בעל שיפוע קטן מאד אנו למדים כי ההסתברות לנטישה של לקוחות אלו כמעט ולא תלויה בזמן ההמתנה שלהם ולכן המודל המתאים ביותר יהיה המודל האדפטיבי.

שאלה 2 הרצאות / תרגולים / בחינות קודמות (10 נקודות)

חלק 1.

חברת MBPF Finance Inc מבצעת הערכת בקשות להלוואות ולאחר מכן מאשרת או דוחה את הבקשות. החברה מעוניינת להעריך שינויים ברמת השירות הנובעים משינוי בתהליך הערכת הבקשות. מופע הבקשות להלוואות הינו בקצב של 1000 בקשות בחודש. עד לינואר 1998 (תחת "תהליך 1"), כל בקשה עוברת את אותו התהליך. נמצא כי בממוצע קיימות 500 בקשות בתהליך ההערכה, ובממוצע 20% מהבקשות מאושרות בתום התהליך. מכיוון שקיימים סוגי בקשות שונים וכי לא סביר להתייחס לכל הבקשות באותו אופן, הוחלט לבצע שינוי בתהליך העבודה. לפי "תהליך 2" כל הבקשות עוברות בדיקה ראשונית שבסופה תסווג כל בקשה לאחת מ-3 הקבוצות הבאות:

- A – בקשות שעבורן קיים סיכוי גבוה לאישור ועבורן נדרשת בדיקה קצרה נוספת.
- B – בקשות עבורן נדרשת בדיקה מפורטת.
- C – בקשות נדחות לאחר בדיקה ראשונית.

נמצא כי בממוצע 25% מכלל הבקשות הן מסוג A, 25% מסוג B ו- 50% מסוג C. בממוצע, 70% מהבקשות מסוג A ו- 10% מהבקשות מסוג B מאושרות בתום הבדיקה הנוספת. מספר הבקשות הממוצע בתהליך הינו כדלקמן: 200 בקשות בבדיקה הראשונית, 25 בבדיקה לבקשות מסוג A ו- 150 בבדיקה לבקשות מסוג B.

2.1.1 חשבו את ממוצע זמן בדיקת בקשה לפי תהליך 1.

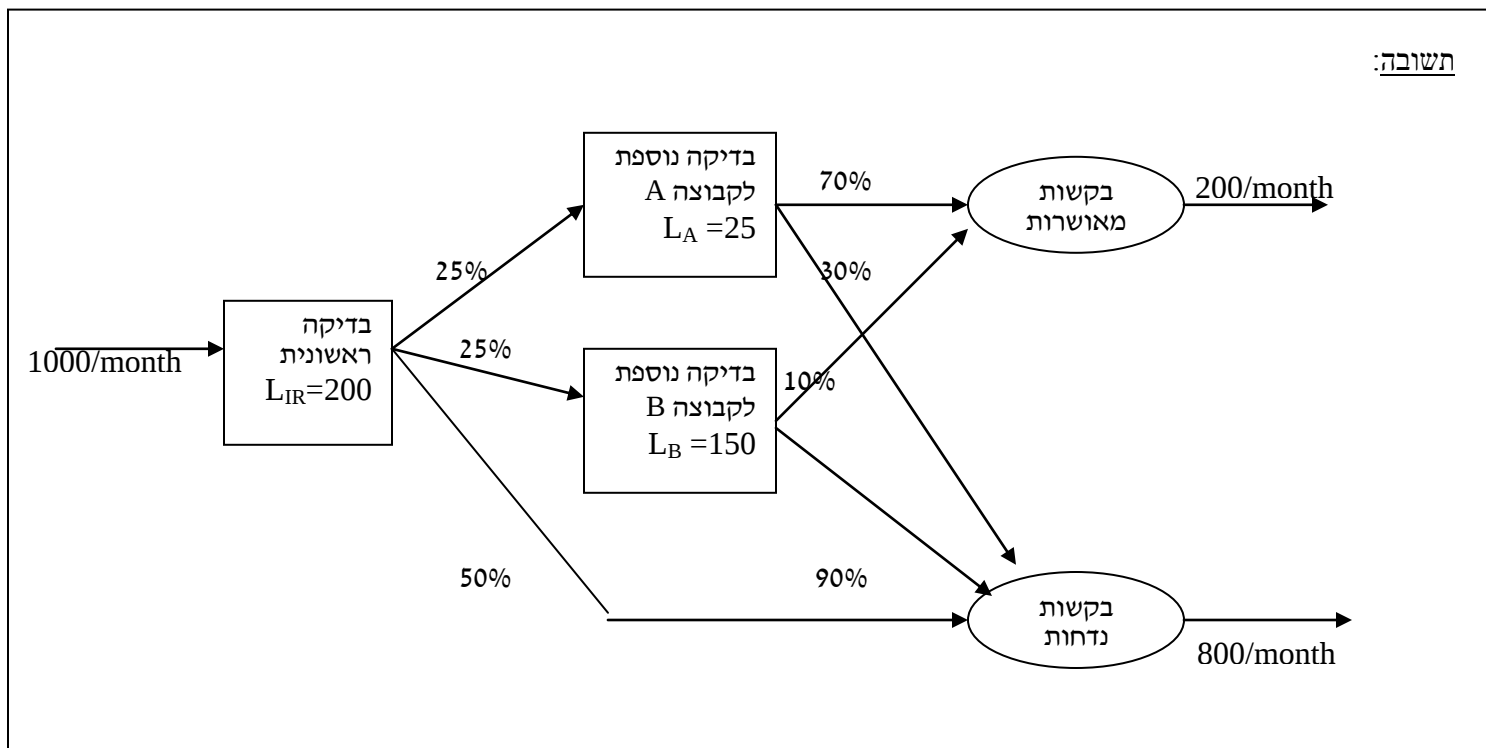
תשובה:

לפי חוק Little ,

$$W_1 = \frac{500}{1000} \text{ month} = 15 \text{ days}$$

2.1.2 השלימו את התרשים המתאר את תהליך 2 ע"י תוספת שמות הפעילויות, אחוזי מעבר, קצבי מעבר וממוצע מספר בקשות בתחנה.

תשובה:



2.1.3 חשבו את ממוצע זמן בדיקת בקשה לפי תהליך 2.

תשובה:

דרך 1:

$$W_2 = \frac{375}{1000} \text{ month} = 11.25 \text{ days}$$

דרך 2:

$$W_{IR} = \frac{200}{1000} \text{ month} = 6 \text{ days}; \quad W_A = \frac{25}{1000 \times 0.25} \text{ month} = 3 \text{ days};$$

$$W_B = \frac{150}{1000 \times 0.25} \text{ month} = 18 \text{ days};$$

$$\Rightarrow W_2 = 0.5 \cdot 6 + 0.25 \cdot (6 + 3) + 0.25 \cdot (6 + 18) = 11.25 \text{ days}$$

כלומר, שיפור של 4 ימים בממוצע זמן השירות.

2.1.4 האם זמן השירות הממוצע עבור בקשות שאושרו השתפר במעבר מתהליך 1 לתהליך 2? הציגו את חישוביכם.

תשובה:

דרך 1: $\frac{25\% \times 70\%}{20\%} = \frac{7}{8}$ מהבקשות שאושרו הן מסוג A, ו- $\frac{1}{8}$ מסוג B. לכן,

$$W_{Approved} = \frac{7}{8} \cdot 9 + \frac{1}{8} \cdot 24 = 10.875 \text{ days}$$

דרך 2:

$$L_{Approved} = L_{IR, App} + L_{A, App} + L_{B, App} = 200 \cdot 20\% + 25 \cdot 70\% + 150 \cdot 10\% = 72.5$$

$$\lambda_{Approved} = 1000 \cdot 20\% = 200 \text{ per month}$$

$$\Rightarrow W_{Approved} = \frac{72.5}{200} \text{ month} = 10.875 \text{ days}$$

← זמן העיבוד עבור הבקשות שאושרו השתפר.

2.1.5 הניחו כי 8 סוכנים עובדים בתחנת הבדיקה הנוספת לבקשות מסוג B וכי עומס העבודה מתחלק ביניהם באופן שווה. כל סוכן יכול לשרת 40 בקשות לחודש. בנוסף, בממוצע, 20% מהבקשות שמסווגות לאחר הבדיקה הראשונית לקבוצה A דורשות למעשה בדיקה מפורטת יותר, ונשלחות לבדיקה עבור קבוצה B (לאחר בדיקתן בקבוצה A). מהי כעת נצילות סוכן בתחנת הבדיקה מסוג B? הציגו את חישוביכם.

תשובה:

$$\rho_B = \frac{250 + 250 \cdot 20\%}{8 \cdot 4} = 94\%$$

2.1.6 תחת ההנחות של סעיף 2.1.5, מהו פרופיל התעסוקה של סוכן בתחנה B? כלומר, מהו אחוז הזמן בטווח הארוך בו עסוק הסוכן בהערכת בקשות שהגיעו ישירות מהבדיקה הראשונית, בטיפול בבקשות שהגיעו מתחנת בדיקה A ומהו אחוז זמן חוסר-התעסוקה?

תשובה:

בקשות מבדיקה ראשונית : 78%.

בקשות מ-A : $94\% - 78\% = 16\%$

בטל: $100\% - 94\% = 6\%$

להלן כתבה שפורסמה בחודש מאי 2010 באתר האינטרנט של העיתון גלובס:



זירת העסקים של ישראל

תיק איש

פורטל פיננסי

חדשות

ראשי

שוק ההון

וול סטריט ועולם

היי טק

נדל"ן

סביבה

נתח שוק וצרכנות

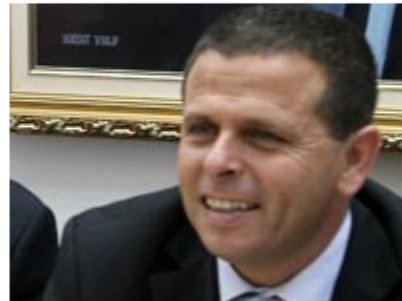
דין וחשבון

עמוד הבית < נתח שוק

אושר בטרומית: חברות יענו תוך 3 דקות במוקד שירות הלקוחות

הכנסת אישרה היום בקריאה טרומית את הצעת החוק של ח"כ איתן כבל, הקובעת כי חברות שלהן למעלה מ-10,000 לקוחות יחויבו לתת מענה אנושי במוקד השירות בתוך 3 דקות - או שיתחייבו לחזור ללקוח בתוך 3 שעות

14:27, 26/05/2010



נושאים: [איתן כבל](#) [הצעת חוק](#) [שירות לקוחות](#)

חברות הסלולר והאינטרנט יחויבו לענות ללקוחות בתוך 3 דקות: מליאת הכנסת אישרה היום (ד') בקריאה טרומית את הצעת החוק של ח"כ איתן כבל, הקובעת כי חברות שלהן למעלה מ-10,000 לקוחות יחויבו לתת מענה אנושי במוקד שירות הלקוחות בתוך 3 דקות, או שיתחייבו לחזור ללקוח שהשאיר הודעה במענה קולי בתוך 3 שעות.

חברת סלולר מסוימת, המפעילה מוקד שירות טלפוני, מעוניינת להתאים (לשפר) את רמת השירות שלה לחוק החדש. כיום, מספר המוקדנים העובדים במוקד השירות הוא 75, משך שיחה ממוצע הינו 6 דקות, בעוד הסבלנות הממוצעת של לקוח היא 12 דקות. קצב המופע משתנה על פני היום, וניתן להניח שהוא קבוע למקוטעין כפי שיוסבר בהמשך. אנליסט העובד בחברה הסלולארית הנוכחית ממדל את המוקד הטלפוני הנ"ל.

חלק 3.1. מידול המוקד

3.1.1 איזה מודל סטוכסטי נראה לכם מתאים לניתוח מצב זה? מה הן ההנחות המתמטיות במודל זה?

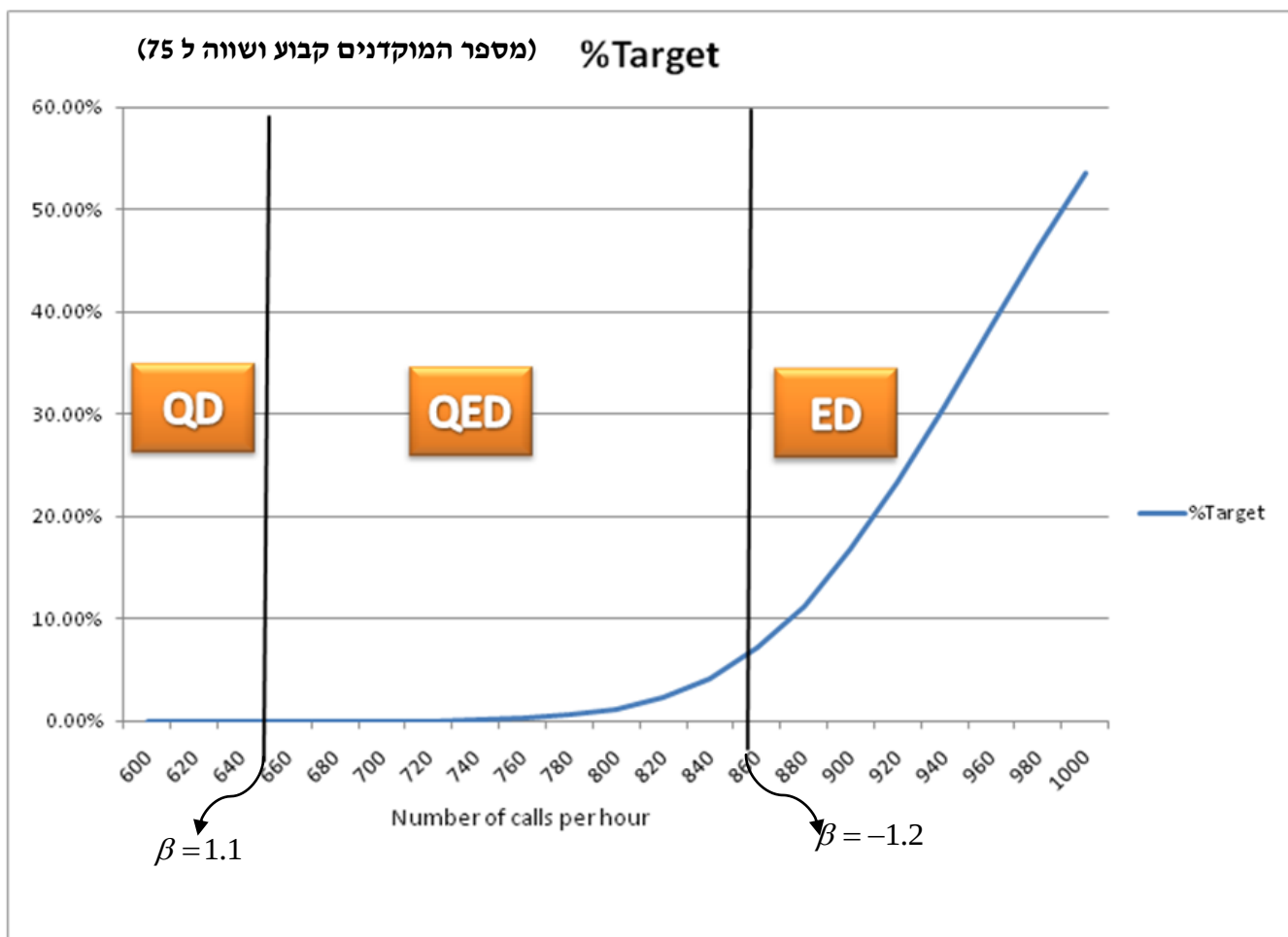
תשובה: מאחר וישנן נטישות המודל המתאים ביותר הינו מודל Erlang A. ההנחות הפורמאליות הדרושות לרלוונטיות מודל זה הן: התפלגות זמני השירות והסבלנות אקספוננציאלים, תהליך ההגעה הינו תהליך פואסון וישנה אי תלות בין זמני השירות לתהליך ההגעה. מניחים שקצב המופע קבוע על פני התקופות בהן משתמשים במודל.

הנהלת המוקד הטלפוני פונה אל האנליסט בנוגע להתאמת המוקד הטלפוני לחוק החדש. בשלב הראשון מעוניין האנליסט לבדוק מהו כיום אחוז הלקוחות אשר המתינו בתור יותר משלוש דקות. לשם כך נגדיר את המדד התפעולי הבא:

%Target = אחוז הלקוחות אשר המתינו בתור למעלה משלוש דקות.

שימו לב: בחישוב %Target אנו כוללים גם את הלקוחות שהמתינו יותר משלוש דקות וקיבלו שירות וגם את הלקוחות אשר המתינו יותר משלוש דקות ולבסוף נטשו את המוקד.

האנליסט יצר את הגרף הבא, אשר מציג את %Target כפונקציה של קצב המופע למוקד.



3.1.2 על הגרף מסומנים התחומים התפעוליים השונים (ED,QED,QD) עבור טווחי קצבי ההגעות השונים. (למשל, תחום ה QED תקף עבור $660 \leq \lambda \leq 860$). כיצד נקבעו התחומים הללו ?

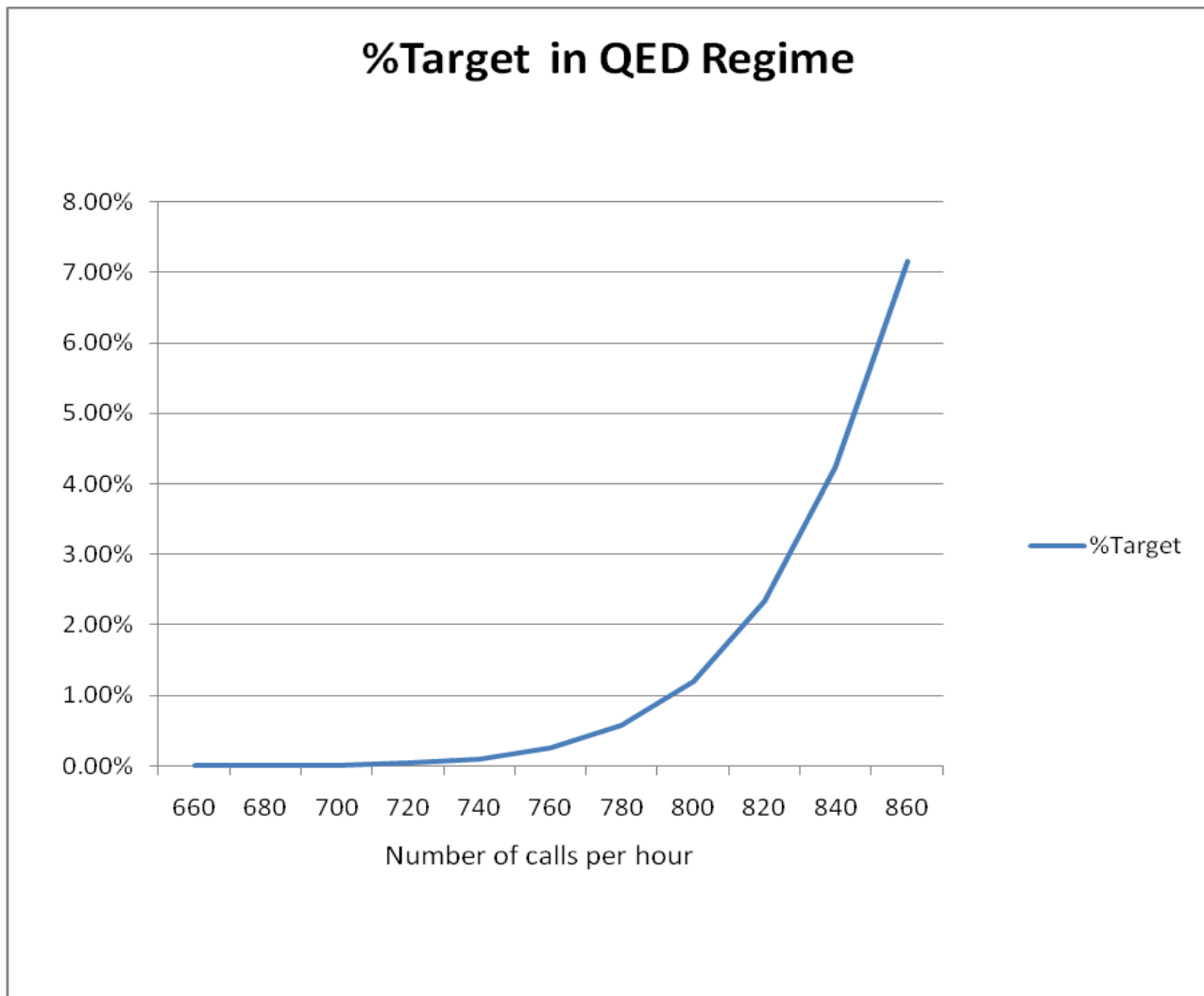
רמז: הניחו כי האיוש נקבע ע"י כלל השורש $n = R + \beta\sqrt{R}$, כאשר $R = \text{Offered Load}$ ו- n הינו מספר המוקדנים, התבוננו בטווחי β בתחומים הנ"ל.

תשובה:

תחום ה QD נקבע עבור $1.1 \leq \beta$, תחום ה QED נקבע עבור $-1.18 < \beta \leq 1.1$ ותחום ה ED נקבע עבור $\beta \leq -1.18$.

$$1.1 = \frac{75 - 660 \cdot \frac{1}{10}}{\sqrt{660 \cdot \frac{1}{10}}} \quad \text{למשל}$$

נתמקד כעת בתחום התפעולי QED, ונתבונן בגרף הבא המתאר אותו (Zoom של הגרף הקודם):



3.1.3 מהם טווחי %Target עבורם המוקד נמצא במשטר QED?

תשובה:

מהגרף אנו מוצאים כי עבור $0 \leq \%Target \leq 7.2$ המערכת נמצאת תחת משטר QED

חלק 3.2. התאמת המוקד לחוק החדש: שלב א'

מעתה נדון בשעות הבוקר בהן קצב הגעת הלקוחות קבוע ושווה ל 1000 לשעה. מהגרף אנו למדים כי, כיום, בשעות אלו המערכת נמצאת תחת משטר ED ואחוז הלקוחות אשר ממתינים יותר מ 3 דקות גדול מ 50%. עקב החוק החדש, מספר זה אינו מקובל על מנהלי המוקד ולכן הם פונים אל האנליסט כדי שיקבע את מספר המוקדנים הדרוש בשעות אלו. ליתר דיוק, מנהל המוקד מטיל על האנליסט למצוא את מספר המוקדנים המינימאלי כך שאחוז הלקוחות הממתינים לא יעלה על 0.3%. (0.3% נבחר בשלב הראשון, כדי לקבל מושג על הנדרש לעמוד בחוק, וזאת כדי לתאר מצב שבו כמעט אין חזרה ללקוחות אחרי 3 דקות).

4CallCenters v2.23

File Table Settings Help

Performance Profiler | Staffing Query | Advanced Profiling | Advanced Queries | What-if Analysis

Performance Profiler Performance Profiler allows you to determine and optimize the Performance Level of your Call Center. Enter your call center's parameters below, then press 'Compute'.

Your Call Center's Parameters

- Number of Agents Answering Calls: 100
- Average Time to Handle One Call (mm:ss): 06:00
- Calls per 60 minute Interval: 1000
- Average Callers' Patience (mm:ss): 12:00

Settings

- Features: Abandons
- Basic Interval: 60 minutes
- Target Time: 03:00 (mm:ss)

Change Settings

Compute | Add to Table | Delete Rows | Clear All | Export | Import | Graph

	Basic Interval (minutes)	Target Time to Answer	Number of Agents	Average Handling Time	Calls per Interval	Average Patience	Agents' Occupancy	%Answer	%Abandon	Average Speed of Answer	%Answer within Target	%Abandon within Target	Average Queue Length
Results													
1	60.0	03:00.0	90.0	06:00.0	1,000.0	12:00.0	99.6%	89.7%	10.3%	01:16.8	87.3%	10.3%	20.7
2	60.0	03:00.0	91.0	06:00.0	1,000.0	12:00.0	99.5%	90.5%	9.5%	01:09.9	88.9%	9.4%	18.9
3	60.0	03:00.0	92.0	06:00.0	1,000.0	12:00.0	99.3%	91.4%	8.6%	01:03.2	90.3%	8.6%	17.2
4	60.0	03:00.0	93.0	06:00.0	1,000.0	12:00.0	99.2%	92.2%	7.8%	00:56.9	91.4%	7.8%	15.6
5	60.0	03:00.0	94.0	06:00.0	1,000.0	12:00.0	98.9%	93.0%	7.0%	00:51.0	92.5%	7.0%	14.0
6	60.0	03:00.0	95.0	06:00.0	1,000.0	12:00.0	98.7%	93.7%	6.3%	00:45.4	93.4%	6.3%	12.5
7	60.0	03:00.0	96.0	06:00.0	1,000.0	12:00.0	98.4%	94.4%	5.6%	00:40.2	94.2%	5.6%	11.2
8	60.0	03:00.0	97.0	06:00.0	1,000.0	12:00.0	98.0%	95.1%	4.9%	00:35.5	94.9%	4.9%	9.9
9	60.0	03:00.0	98.0	06:00.0	1,000.0	12:00.0	97.6%	95.7%	4.3%	00:31.1	95.6%	4.3%	8.7
10	60.0	03:00.0	99.0	06:00.0	1,000.0	12:00.0	97.2%	96.2%	3.8%	00:27.1	96.1%	3.8%	7.6
11	60.0	03:00.0	100.0	06:00.0	1,000.0	12:00.0	96.7%	96.7%	3.3%	00:23.5	96.7%	3.3%	6.6
12													

3.2.1 מהפלט של 4CC, מה הוא מספר המוקדנים המינימאלי שיבטיח כי לכל היותר 0.3% מהשיחות ימתינו יותר מ 3 דקות? נמקו במדויק.

תשובה:

נסמן את $\%Target$ להיות אחוז הלקוחות אשר המתינו יותר משלוש דקות (גם לקוחות אשר נטשו וגם לקוחות אשר קיבלו שירות). מתקיים כי

$$\%Target = (\%Abandon - \%Abandon\ Within\ Target) + (\%Answer - \%Answer\ Within\ Target)$$

או באופן פשוט יותר

$$100 - \%Target = \%Abandon\ Within\ Target + \%Answer\ Within\ Target \geq 99.7\%$$

ולכן מספר השרתים יהיה 95. (עבור 95 שרתים $100 - \%Target = 93.4\% + 6.3\% = 99.7\%$).

מעטה (ועד סיום חלק 3.2) הניחו כי מספר המוקדנים הנבחר הינו 95.

3.2.2 ע"ס פונקציית Garnett (מצורפת בסוף שאלה זו), מה תהיה ההסתברות להמתין בתור במוקד זה?

תשובה:

$$\beta = \frac{n-R}{\sqrt{R}} = \frac{n-\lambda/\mu}{\sqrt{\lambda/\mu}} = \frac{95-1000/10}{\sqrt{1000/10}} = -0.5$$

תחילה נמצא את β : $\beta = -0.5$

$$P(W_q > 0) \approx 0.8 \text{ ולכן } \beta = -0.5, GMR(0.5), \theta/\mu = 0.5$$

לכן נשתמש בעקומה $GMR(0.5)$, $\beta = -0.5$ ולכן $P(W_q > 0) \approx 0.8$

3.2.3 חשבו את $E(W_q | W_q > 0)$, פרטו את חישוביכם, והסבירו את משמעות מדד זה.

תשובה:

$$E(W_q) = E(W_q | W_q > 0)P(W_q > 0) + E(W_q | W_q = 0)P(W_q = 0) = E(W_q | W_q > 0)P(W_q > 0)$$

$$E(W_q | W_q > 0) = \frac{E(W_q)}{P(W_q > 0)}$$

ולכן $E(W_q) = \frac{E(L_q)}{\lambda} = \frac{12.5}{1000}$ לפי נוסחת Little. נמצא את $E(W_q)$. נקבל

$$E(W_q | W_q > 0) = \frac{12.5}{0.8 \cdot 1000} = 0.016 = 0.93 \text{ min}$$

כי $E(W_q | W_q > 0) = \frac{12.5}{0.8 \cdot 1000} = 0.016 = 0.93 \text{ min}$

מדד זה מציין את זמן ההמתנה הממוצע של לקוח שאכן התעכב בתור.

חלק 3.3. התאמת המוקד לחוק החדש: שלב ב'

חברת הסלולר החליטה להעסיק את אריק, מהנדס בוגר הפקולטה לתעשייה וניהול בטכניון, אשר למד את הקורס הנדסת מערכות שירות. אריק מעיר להנהלת החברה כי, לצורך החוק החדש, השיטה בה השתמשו עד עתה לחישוב מספר הלקוחות אשר המתניו יותר מ 3 דקות אינה מדויקת. אריק מציע שיטה אלטרנטיבית לבדיקה זאת. לפני שנפרט את השיטה המוצעת ע"י אריק, נציג את המאפיין Overflow אשר מופיע בתוכנת ה 4CC:

- בחירה ב "Overflows" היא עבור מערכת בה לקוחות הממתינים זמן מסוים (3 דקות במקרה זה) מפונים מהמערכת למערכת אחרת (מערכת אשר אחראית לחזור אל הלקוח לאחר שלוש שעות לכל היותר).
- קלט: Overflow Time Limit: כאן נקבע הזמן אשר לקוחות הממתינים יותר ממנו יופנו למערכת האחרת.
- פלט: %Overflow מציין את אחוז הלקוחות אשר עברו את הזמן המדובר (שלוש דקות) והופנו למערכת האחרת.

לשלמות, להלן פירוט ה- Feature "Overflow" בתוכנה 4CC (לקוח מ-Help).

אין צורך לקרוא את הפירוט הנ"ל.



Feature Overview : Overflows

What are "Overflows"?

At some call centers a technical limit has been set on the time a call can spend waiting in queue. Once this time limit is reached the call is "overflowed" and rerouted to some other destination, such as a different queue, an IVR or a voice box. Such an overflowing mechanism is meant to reduce congestion by offering callers a partial or temporary solution, hopefully before they become aggravated by a long wait.

When should I enable the "Overflows" feature?

If an overflowing mechanism is implemented at your call center and the fraction of calls "overflowed" is not negligible, then it is recommended that you enable the "Overflows" feature. However, since you cannot have both the "Overflows" and "Trunks" features enabled, you should decide which to enable according to the more dominant phenomena at your call center - overflowing or blocking.

What are the corresponding input parameters and performance indicators?

The input parameter corresponding to the "Overflows" feature is the "Overflow time", which is the time limit on calls waiting in queue.

There is a single performance indicator available only when the "Overflows" feature is enabled: %Overflow.

הצעתו האלטרנטיבית של אריק:

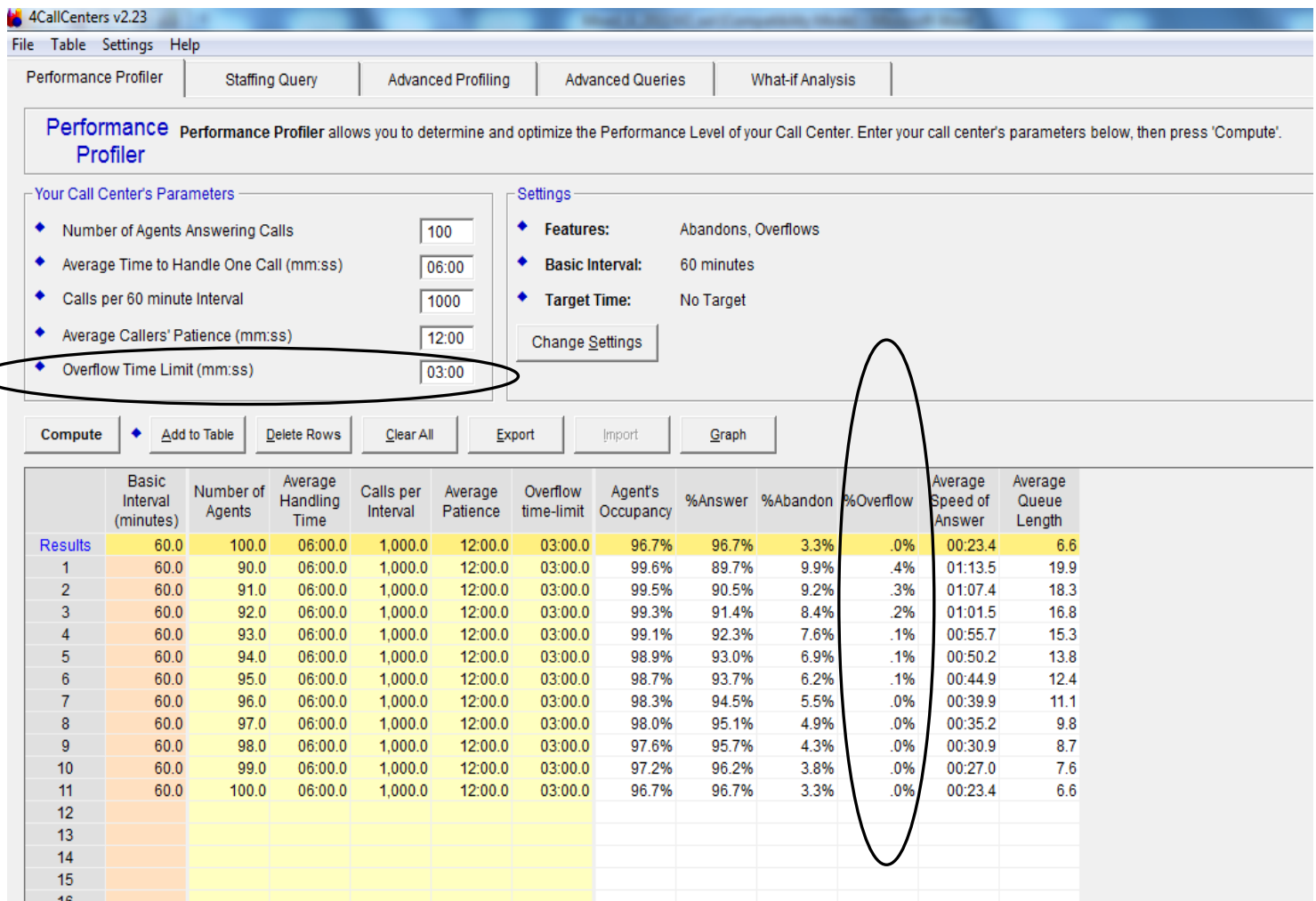
במקום לבדוק את אחוז הלקוחות אשר המתינו במערכת יותר מ 3 דקות בעזרת ה- Target-Time Feature בוחר אריק להשתמש במודל עם Overflows, וקובע את ה Overflow Time Limit להיות 3 דקות.

3.3.1 אריק טוען כי שיטתו תתאר טוב יותר את המתרחש במציאות. מדוע הוא צודק? פרטו!

תשובה:

אריק צודק מאחר ואותם לקוחות אשר הופנו למערכת אחרת (שם יחזרו אליהם לאחר לא יותר משלוש שעות) יוצאים מהמערכת ומפנים מקום ללקוחות שיופיעו אח"כ (דומה הדבר לנטישה אחרי 3 דקות). תופעה זו מתרחשת במודל עם ה Overflows ולא במודל המקורי עם ה Target Time (שם נשארים במערכת להמתין גם יותר מ 3 דקות).

להלן פלט ה 4CC של המודל הנ"ל, הכולל Overflows:



3.3.2 מהו כעת מספר המוקדנים המינימאלי הנדרש כדי לעמוד באילוף הבא: אחוז השיחות אשר ימתינו בתור יותר משלוש דקות לא יעלה על 0.3%? נמקו! השוו עם המקרה הקודם (שאלה 3.2.1).

תשובה:

נחפש את מספר המוקדנים המינימאלי עבורו %Overflow לא יהיה יותר מ 0.3%. מספר המוקדנים יהיה עתה 91. קיבלנו מספר מוקדנים קטן יותר ($91 < 95$) מאחר ולא שווים במערכת יותר משלוש דקות: התורים קצרים יותר $\leq$ נדרשים פחות מוקדנים.

3.3.3 במודל במודל של אריק, כפי שתואר לעיל, מניחים כי הלקוח עוזב את התור אחרי 3 דקות. ישום מתאים של הפרוטוקול מחייב כמובן הודעה קולית, לאחר שהלקוח המתין שלוש דקות: הודעה זו מודיעה ללקוח שמוקדן יחזור אליו תוך 3 שעות לכל היותר. זהו כמובן מצב לא רצוי (שירות לא טוב), ואריק מצביע עליו לפני הנהלת המוקד. מה ניתן לעשות אחרת? תנו תשובה אחת אותה ניתן לנתח בעזרת 4CC, ותשובה שנייה חופשית.

תשובה אותה ניתן לנתח בעזרת 4CC:

תשובה:

המוקד יכול לחזות בעת הגעת הלקוח (לפי מספר הלקוחות שבתור) מהו זמן ההמתנה הצפוי בתור (נניח שהאמד מדויק). במידה וזמן זה יעלה על 3 דקות, יקבל הלקוח הודעה מיידית (מיד בזמן הגעתו) על כך שיחזור אליו מוקדן תוך 3 שעות לכל היותר. במילים אחרות המודל של 4CC מכסה גם את המקרה שבו ההודעה נתנת בתחילת ההמתנה, ולא לאחר 3 דקות המתנה. בפרט, כל המדדים הרלוונטיים להמתנות ונטישות זהים בשני התרחישים (רק מדדים התלויים באורך התור משתנים, כי בתרחיש אחד מצטרפים לתור ועוזבים אחרי 3 דקות לכל היותר, ובתרחיש האלטרנטיבי לא מצטרפים לתור כאשר צפויה המתנה של יותר מ 3 דקות.

תשובה חופשית:

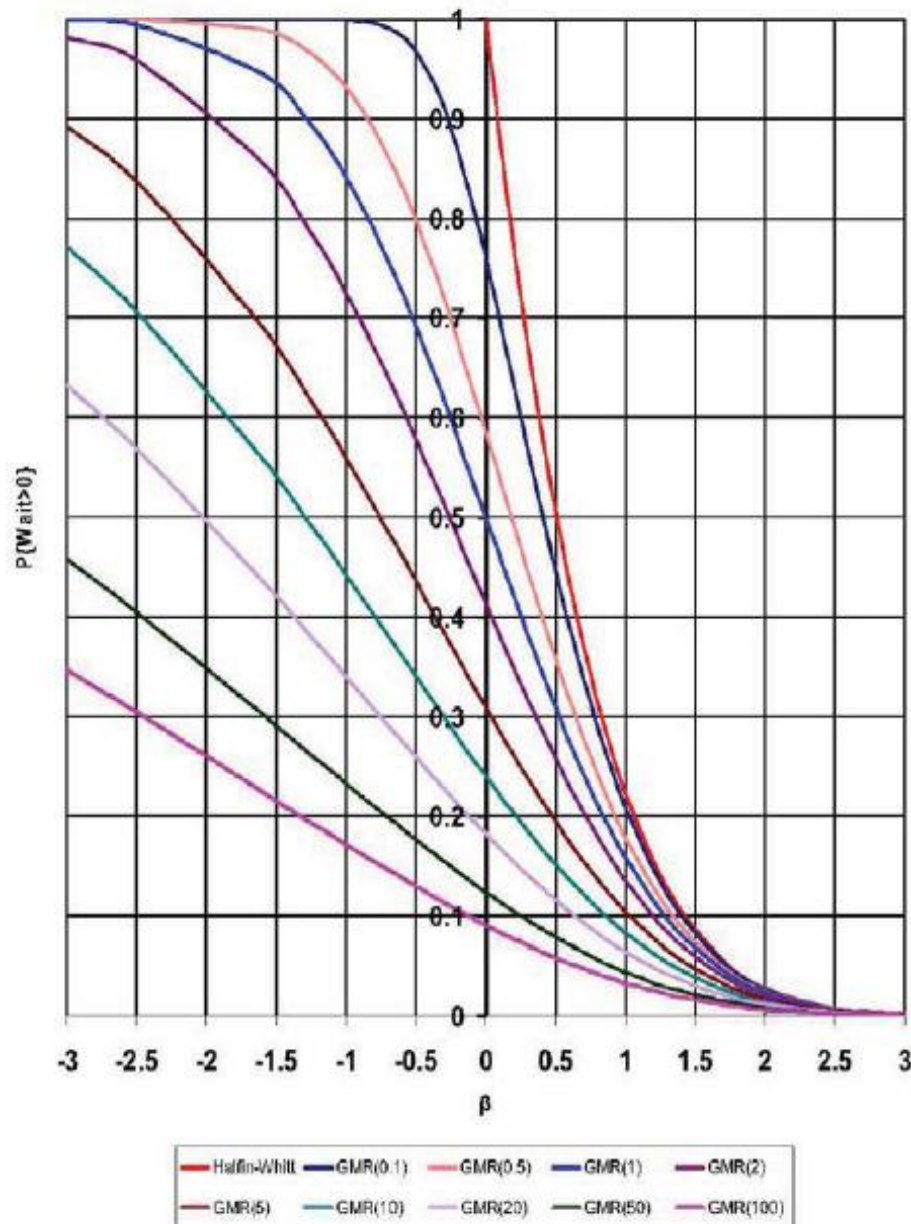
תשובה:

בעת קבלת ההודעה יוצגו בפני הלקוח שתי אפשרויות:

1. הלקוח יעזוב את המערכת ומוקדן יחזור אליו.
 2. הלקוח יכול לבחור להישאר בתור ולקבל שירות כאשר יגיע תורו.
- כאן נדרש לבנות מודל חדש, מעבר לתוכנת 4CC.

Erlang-A: The Garnett Delay-Functions

$P\{W_q > 0\}$ vs. the QOS parameter β , for varying patience θ/μ .



GMR(x) describes the asymptotic probability of delay as a function of β when $\frac{\theta}{\mu} = x$. Here, θ and μ are the abandonment and service rate, respectively.

Note: **Erlang-C** = limit of **Erlang-A**, as patience $\uparrow$ indefinitely.

שאלה 4 תיאוריה (10 נקודות)

מודל פשוט למערכת שירות, על פני אינטרוול זמן $[0, T]$, מאופיין על ידי אבני הבניין הבאות:

$$L(t) = \text{מספר הלקוחות במערכת בזמן } t, 0 \leq t \leq T;$$

$$A(t) = \text{מספר הלקוחות שהגיעו למערכת עד זמן } t, 0 \leq t \leq T;$$

$$W_1, W_2, \dots, W_{A(T)} = \text{משכי השהייה של לקוחות במערכת.}$$

4.1 חלק

4.1.1 מהו L = מספר (ממוצע) של לקוחות על פני אינטרוול הזמן $[0, T]$?

$$L = \frac{1}{T} \int_0^T L(t) dt$$

4.1.2 מהו λ = קצב המופע (הממוצע) למערכת, על פני אינטרוול הזמן $[0, T]$?

$$\lambda = \frac{A(T)}{T}$$

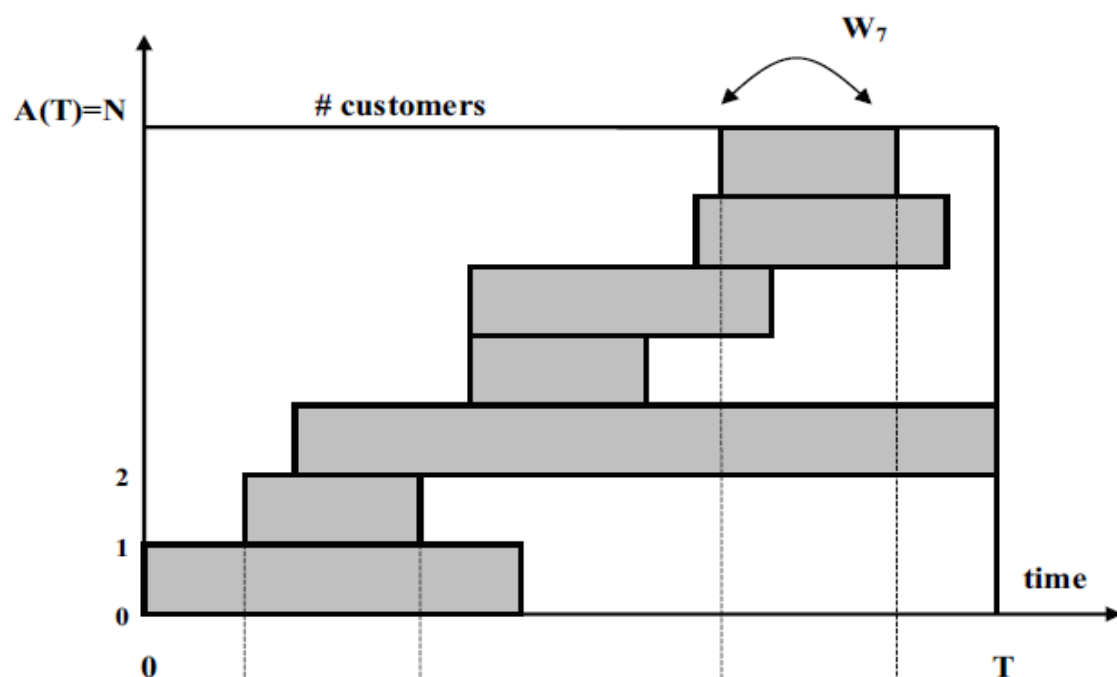
4.1.3 מהו W = ממוצע משך השהייה של לקוח במערכת?

$$W = \frac{1}{A(T)} (W_1 + \dots + W_{A(T)})$$

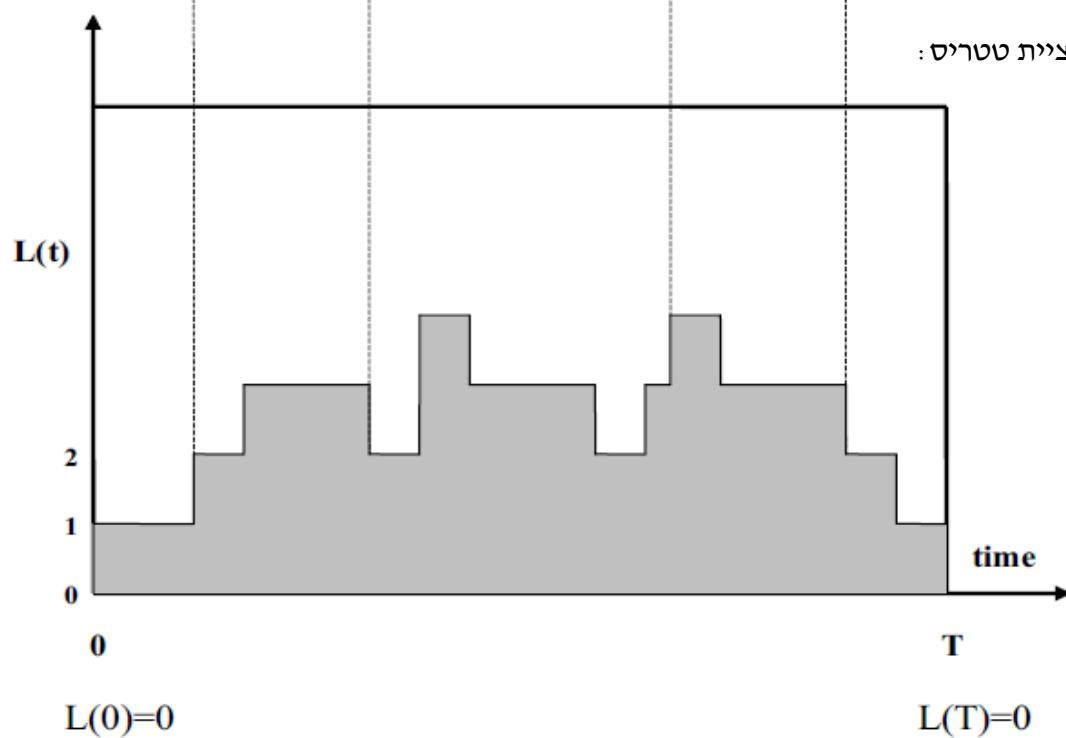
4.2 חלק

נסחו והוכיחו את נוסחת Little למערכת ה"ל". השתמשו בדיאגרמת גאנט להוכחתכם, ונסחו במדויק תנאים המספיקים לנוסחת Little להתקיים. (רמז: השתמשו בטרנספורמציות טורים).

פתרון מדפי ההרצאה:



טרנספורמציות טריס:



Note: Vertical cut = number of customers in the system.

S = Shaded area (units: customer $\times$ hours), measures total waiting.

$W = \frac{S}{N}$, divides waiting among customers
(the **customer's** view).

$L = \frac{S}{T}$, divides waiting over time
(the **manager's** view).

$\lambda = \frac{N}{T}$, implies $L = \lambda \cdot W$
(which adds the **server's** view).

More formally:

$$\text{Area} = \int_0^T L(t)dt = \sum_{k=1}^{A(T)} W_k \quad (\text{whenever } L(0) = L(T) = 0) .$$

Define

$$\left. \begin{aligned} L &= \frac{1}{T} \int_0^T L(t)dt \\ W &= \frac{1}{A(T)} \sum_{k=1}^{A(T)} W_k \\ \lambda &= \frac{A(T)}{T} \end{aligned} \right\} \implies L = \lambda W$$

4.3 חלק

מושג ה Offered-Load, אותו סימנו ב- $R = \{R(t), t \geq 0\}$, משחק תפקיד מרכזי בניתוח מערכת שירות.

$$4.3.1 \quad R(t) = \int_{-\infty}^t \lambda(u) P(S > t-u) du, \quad t \geq 0$$

אחת מ-4 הצגותיו של R נתונה על ידי:

כאשר $\lambda(t), t \geq 0$ הוא פונקצית קצב מופע.

S הוא משתנה מקרי המייצג משך שירות הנדרש על ידי מופע.

מדוע ניתן להתייחס להצגה הנ"ל כאל הכללה של נוסחת Little?

רמז: הציבו $\lambda(t) = \lambda$ (קצב מופע קבוע).

תשובה:

$$R(t) = \lambda \int_{-\infty}^t P(S > t-u) du = \lambda \int_{t-u=x}^{\infty} P(S > x) dx = \lambda E(S)$$

הסבר: במערכת $M/G/\infty$, או במודל קופסא-שחורה של מערכת שירות, $R = \lambda E(S)$ מתאר את ממוצע

מספר הלקוחות במערכת (במצב יציב). זהו מקרה פרטי של $R(t), t \geq 0$, המתאר את ממוצע מספר הלקוחות

במערכת, כפונקציה של זמן $t \geq 0$, במערכת $M_t/G/\infty$.

4.3.2 כיצד משתמשים ב $\{R(t), t \geq 0\}$ כבסיס להמלצת איוש, ומה מצפים ממדדי תפעול המערכת תחת המלצה זו?

תשובה:

$$N(t) = \left\lceil R(t) + \beta \sqrt{R(t)} \right\rceil$$

כאשר $N(t)$ = מספר השרתים המומלץ לזמן $t \geq 0$.

β נקבע לפי מודל סטציונארי מתאים.

מצפים למדדי תפעול שיהיו פחות או יותר קבועים בזמן, ומזדהים עם המודל הסטציונארי המתאים.

4.3.3 נחזור לסעיף 4.1. נניח שהמערכת מתארת יום בחדר מיון בבית חולים. נניח כי $\{W_i\}$ הנ"ל הם בעצם סכום של שני

משתנים, דהיינו:

$$W_i = Q_i + S_i, i=1, \dots, A(T)$$

כאשר

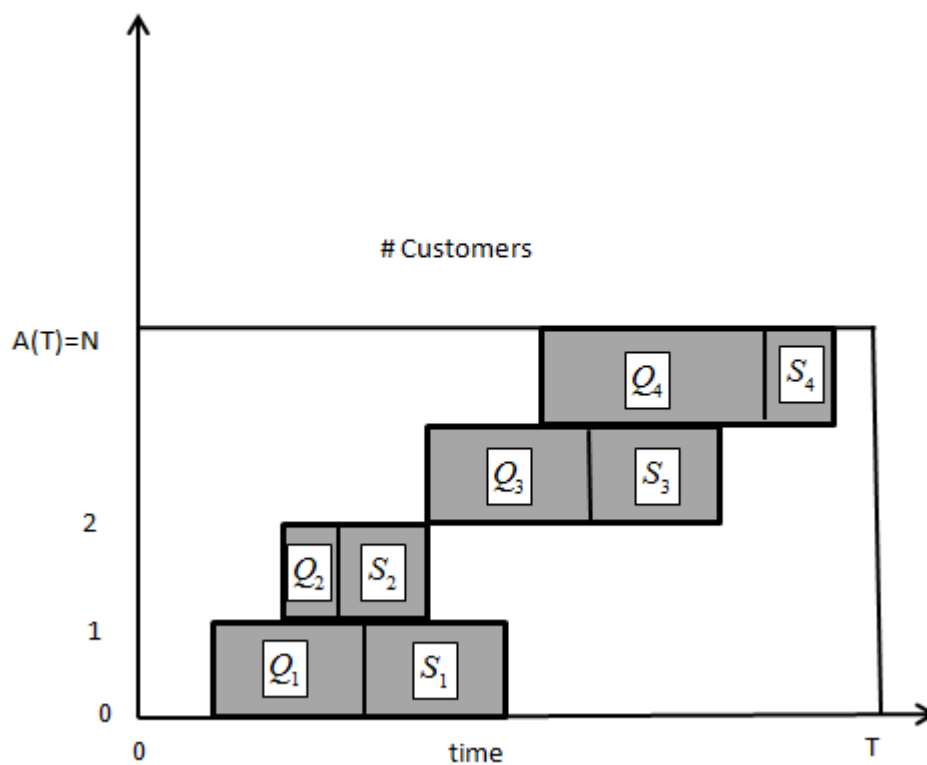
Q_i = זמן המתנה למיטה במיון של לקוח i (מרגע מופע הלקוח למיון).

S_i = משך האשפוז של לקוח i במיון.

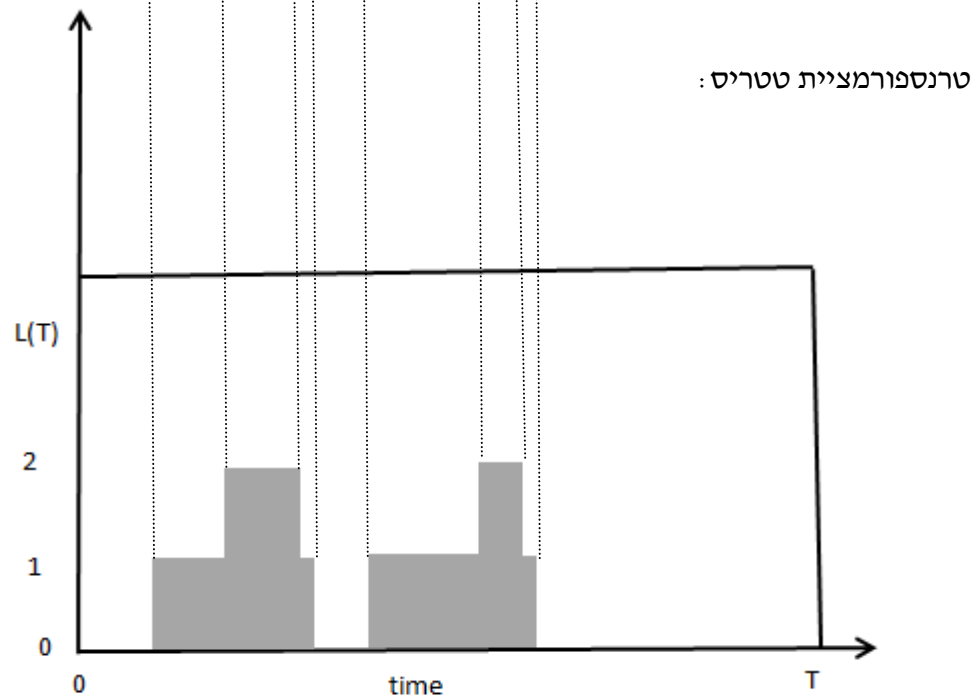
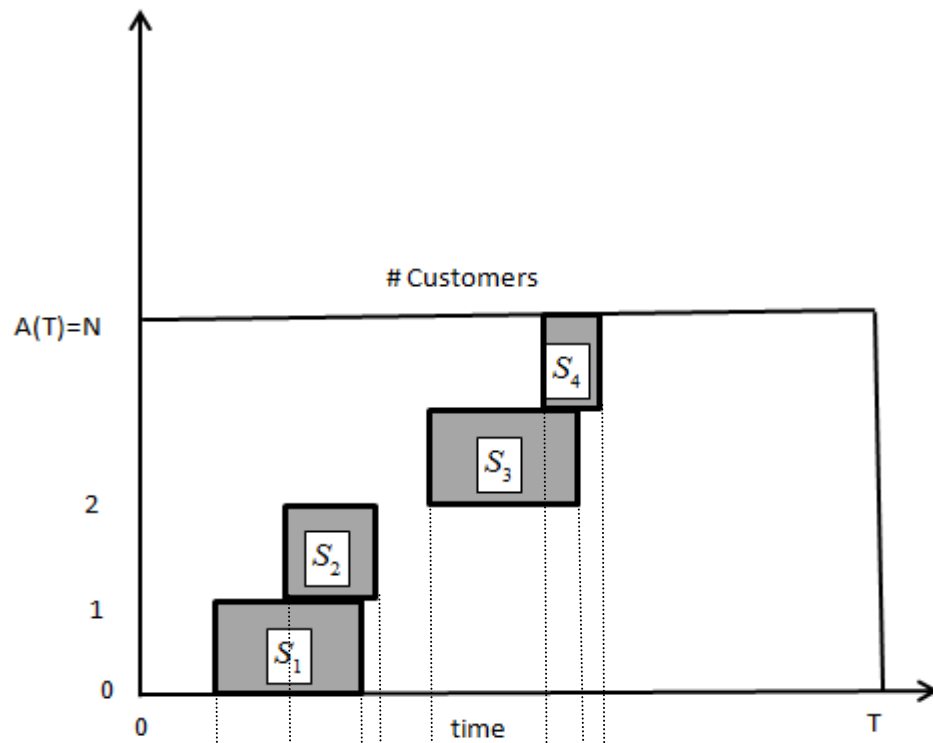
בעזרת הגאנט שציירתם בסעיף 4.2, הדגימו מה הוא המספר המינימאלי של מיטות שנדרש כדי שאף לקוח לא ימתין. מה הקשר של מספר זה למושג ה- Offered-Load?

תשובה:

נדגים על הגאנט הבא:



תחילה, נתבונן במערכת בה ישנן אינסוף מיטות (כלומר אין תורים במערכת) ולמערכת זו נבצע טרנספורמציה טריס. כך נקבל את מספר המיטות המינימאלי הנדרש בכל רגע כך שלא תהיינה המתנות.



לכן המספר המינימאלי של מיטות שנדרש כדי שאף לקוח לא ימתין הוא 2.
הקשר למושג ה- Offered-Load הוא : מספר המיטות המינימאלי שווה ל $\max_{0 \leq t \leq T} \{R(t)\}$ כאשר $R(t)$ הוא
ה Offered-Load בזמן t.

הסבר : אחת מהאינטרפטציות של ה- Offered-Load היא כדלהלן : מספר הלקוחות במערכת שקולה עם
אינסוף שרתים, אנו "דאגנו" למערכת עם אינסוף שרתים ברגע שהורדנו את זמני ההמתנה לאפס.

4.3.4

השתמשו בתובנות של סעיף 4.3.3 כדי לתאר, עקרונית ובקצרה, אלגוריתם אותו ניתן להוסיף לתוכנת

SEESTAT, כדי לחשב את ה- Offered-Load במוקד שירות טלפוני.

איזה קושי עלול להיווצר בחישוב זה, איך ניתן להתגבר עליו, ובאילו תנאים? (רמז: נטיות)

תשובה:

תיאור האלגוריתם : הרעיון הכללי של האלגוריתם יהיה "הורדת זמני ההמתנה לאפס". כלומר, נשמור
את הנתונים הבאים : זמן הגעת כל לקוח ומשך השירות שלו. לצורך החישוב נניח כי הלקוח מקבל שירות
ברגע שהוא נכנס אל המערכת (כמו במערכת עם אינסוף שרתים). במערכת זו מספר הלקוחות הנמצאים
במערכת הוא ראליזציה של ה- Offered-Load. מיצוע של ראליזציות, בעזרת האופציות הסטנדרטיות
של SEESTAT, יתן את פונקציה ה- Offered-Load. כפי שהגדרנו.
הקושי שעלול להיווצר בחישוב זה הוא : אין בידנו נתונים לגבי זמן השירות של הלקוחות אשר נטשו את
המערכת. לקוחות אלו לא היו נוטשים במערכת עם אינסוף שרתים ולכן נצטרך "להשלים" את זמני
השירות שלהם. תחת התנאי שזמן השירות בלתי תלוי בסבלנות הלקוחות ובזמן ההגעה, ניתן להשלים
את זמני השירות של הלקוחות הנוטשים כממוצע (או הגרלת מ"מ) של זמני השירות של הלקוחות אשר
לא נטשו את המערכת.