



מרצה : ד"ר גלית יום-טוב
מתרגלת: נועה זיכלינסקי

תאריך הבחינה: 20.2.17

מס' סטודנט _____

הנדסת מערכות שירות – 096324

מועד א' – סמסטר חורף 2017

פתרון

- המבחן נמשך שלוש שעות. באחריותך להחזירו במועד.
- המבחן בחומר סגור.
- מותר להשתמש במחשבון מדעי רגיל.
- המבחן כולל 28 עמודים, כולל עמוד זה.
- 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.

שאלה 1	_____ מתוך 25 נקודות	בחינות קודמות/הרצאות/תרגולים
שאלה 2	_____ מתוך 25 נקודות	תרגילי בית
שאלה 3	_____ מתוך 33 נקודות	יישום
שאלה 4	_____ מתוך 17 נקודות	תיאוריה

סה"כ _____ מתוך 100 נקודות

בהצלחה !

שאלה 1 – הרצאות\תרגולים\בחינות קודמות (25 נקודות).

חלק 1. חוק Little (13 נקודות)

2.1.1 (9 נקודות) מצורפים בזאת 3 עמודים שהוצגו בהרצאה, המתארים את קיומו או אי-קיומו של

חוק Little במערכות שונות (בכל עמוד מוצגים 3 גרפים). הסבירו, עבור כל עמוד, האם נראה שחוק

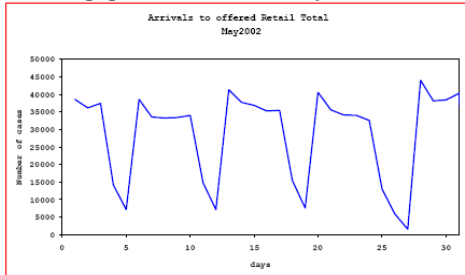
Little מתקיים, ומדוע הוא

צפוי/לא צפוי להתקיים במקרה

Little's Law for Retail calls, May 2002: US Bank

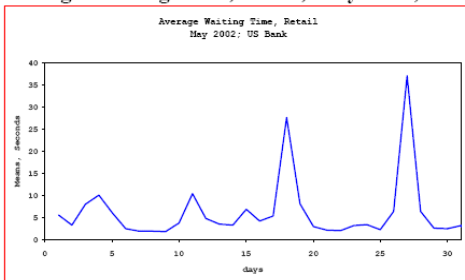
זה.

λ : Throughput Rate, Retail, May 2002; US Bank

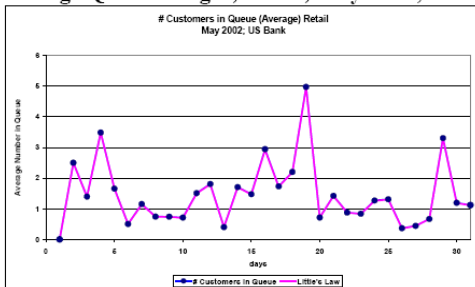


גרף 1: ימי חודש מאי 2002

W: Average Waiting Time, Retail, May 2002; US Bank (W_q)



L: Average Queue Length, Retail, May 2002; US Bank (L_q)



Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
λ	38476	36144	37414	14194	7107	38587	33572	33220	33349	34009	14807	7141	41293	37653	36872	35266
W	5.6	3.3	8.0	10.0	6.1	2.6	1.9	1.9	1.8	3.8	10.5	4.8	3.5	3.4	6.8	4.2
$\lambda \cdot W$	2.49	1.39	3.48	1.65	0.50	1.14	0.74	0.74	0.69	1.50	1.79	0.40	1.69	1.47	2.91	1.71
L	2.50	1.40	3.49	1.66	0.50	1.15	0.74	0.74	0.71	1.51	1.81	0.40	1.71	1.47	2.94	1.73

Date	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
λ	35338	15533	7530	40534	35493	34070	34005	32512	13100	5909	1558	43980	38163	38416	40284
W	5.4	27.6	8.1	3.0	2.1	2.1	3.2	3.5	2.3	6.4	37.0	6.5	2.7	2.5	3.2
$\lambda \cdot W$	2.19	4.96	0.71	1.41	0.87	0.83	1.25	1.30	0.35	0.44	0.67	3.29	1.18	1.11	1.47
L	2.20	4.97	0.71	1.42	0.88	0.84	1.27	1.31	0.37	0.44	0.67	3.30	1.20	1.12	1.48

תשובה: חוק Little מתקיים/לא מתקיים; צפוי/לא צפוי (סמנו בעיגול את התשובה הנכונה).

הסבר:

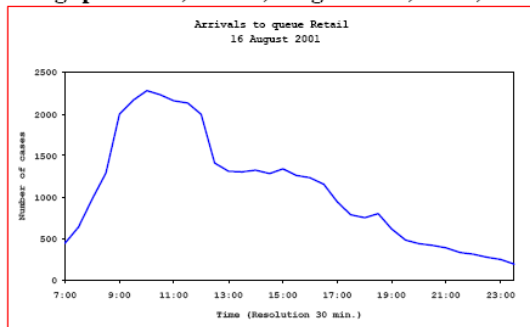
גרף 1 מאשש את קיומו של חוק Little ברזולוציה של ימים. בסוף כל יום מוקד השרות נסגר ולכן

ברזולוציה זו אנו מצפים שחוק Little יתקיים: גירסת אופק סופי המתחיל ומסתיים ריק.

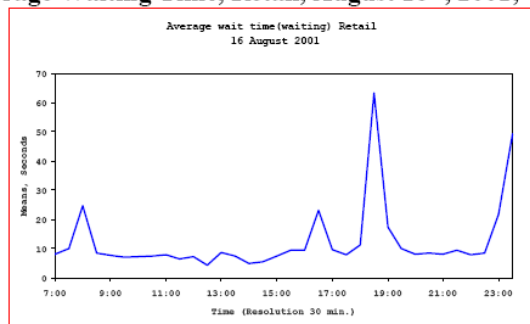
גרף 2: חצאי שעות על פני 2001/8/16

Little's Law for Retail calls, August 16th, 2001: US Bank

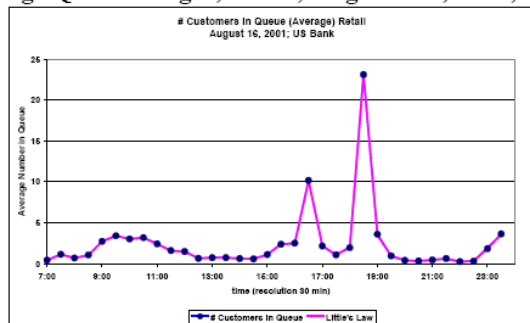
λ : Throughput Rate, Retail, August 16th, 2001; US Bank



W: Average Waiting Time, Retail, August 16th, 2001; US Bank



L: Average Queue Length, Retail, August 16th, 2001; US Bank



Time	7:00	7:30	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00
λ	443	639	987	1291	1998	2166	2278	2231	2158	2135	2000	1408	1311	1303	1323	1285	1340
W	1.7	3.2	1.2	1.5	2.4	2.8	2.4	2.6	2.0	1.3	1.3	0.8	1.0	1.0	0.8	0.8	1.5
$\lambda * W$	0.42	1.14	0.68	1.06	2.72	3.42	3.01	3.18	2.44	1.55	1.47	0.64	0.72	0.72	0.62	0.59	1.09
L	0.42	1.14	0.68	1.06	2.72	3.40	3.02	3.17	2.41	1.59	1.48	0.64	0.72	0.72	0.62	0.57	1.11

Time	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00	19:30	20:00	20:30	21:00	21:30	22:00	22:30	23:00	23:30
λ	1258	1235	1157	942	788	752	803	619	485	437	421	386	336	311	274	251	193
W	3.5	3.6	15.8	4.2	2.4	4.9	51.9	10.0	3.5	1.7	1.3	2.1	3.3	1.4	2.0	14.3	32.6
$\lambda * W$	2.422	2.45	10.2	2.173	1.06	2.05	23.16	3.43	0.95	0.41	0.314	0.44	0.62	0.24	0.30	2.00	3.50
L	2.37	2.49	10.17	2.16	1.07	1.94	23.11	3.59	0.95	0.40	0.31	0.45	0.62	0.24	0.30	1.83	3.63

תשובה: חוק Little מתקיים/לא מתקיים; צפוי/לא צפוי (סמנו בעיגול את התשובה הנכונה).

הסבר:

גרף 2 מאשש את קיומו של חוק Little גם ברזולוציה של חצאי שעות, מקרה זה הוא מפתיע, אך מתקיים כיוון שהמערכת גדולה ומגיעה מהר למצב יציב, וכן בגלל שזמני השרות וזמני ההמתנה קצרים (דקות) בהרבה מהרזולוציה בה אנו בוחנים את המערכת: גירסת מצב יציב.

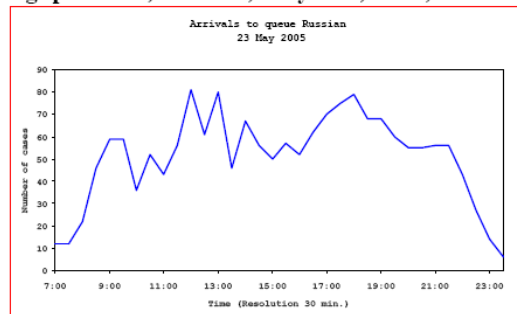
Little's Law for Russian calls, May 23rd, 2005: Israeli Telecom

גרף 3: חצאי

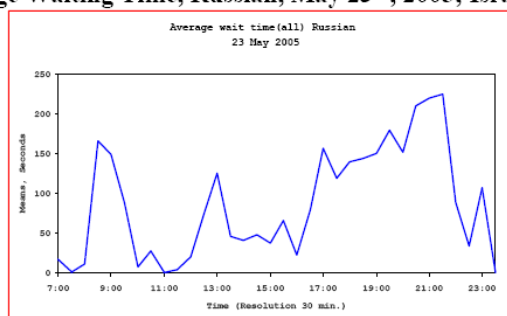
שעות על פני

2005/5/23

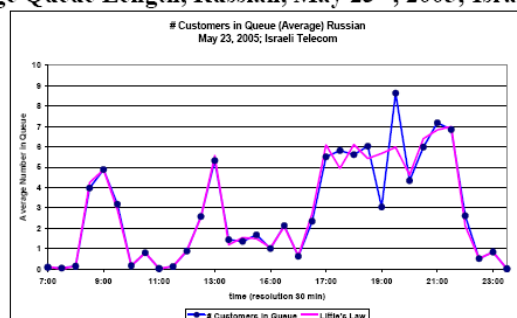
λ : Throughput Rate, Russian, May 23rd, 2005; Israeli Telecom



W: Average Waiting Time, Russian, May 23rd, 2005; Israeli Telecom



L: Average Queue Length, Russian, May 23rd, 2005; Israeli Telecom



Time	7:00	7:30	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00
λ	12	12	22	46	59	59	36	52	43	56	81	61	80	46	67	56	50
W	16.9	1.3	11.4	166.0	148.9	88.7	7.9	27.4	0.7	3.9	20.3	74.9	125.4	46.3	41.4	47.9	37.7
$\lambda * W$	0.11	0.01	0.14	4.24	4.88	2.91	0.16	0.79	0.02	0.12	0.91	2.54	5.57	1.18	1.54	1.49	1.05
L	0.08	0.04	0.14	3.97	4.88	3.18	0.16	0.79	0.02	0.12	0.88	2.57	5.32	1.44	1.36	1.67	1.00

Time	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00	19:30	20:00	20:30	21:00	21:30	22:00	22:30	23:00	23:30
λ	57	52	62	70	75	79	68	68	60	55	55	56	56	43	27	14	6
W	65.7	22.4	79.2	156.6	118.6	139.3	143.6	150.2	179.2	151.3	209.5	219.5	224.7	88.4	33.7	107.0	0.7
$\lambda * W$	2.08	0.65	2.73	6.09	4.94	6.11	5.42	5.68	5.97	4.62	6.40	6.83	6.99	2.11	0.51	0.83	0.00
L	2.13	0.62	2.34	5.51	5.82	5.61	6.03	3.04	8.63	4.34	5.99	7.18	6.85	2.61	0.51	0.83	0.00

תשובה: חוק Little מתקיים/לא מתקיים; צפוי/לא צפוי (סמנו בעיגול את התשובה הנכונה).

הסבר:

גרף 3 מראה לנו שאכן ברזולוציה של חצאי שעות, חוק Little לא תמיד מתקיים. הסיבה שהוא לא מתקיים כאן היא שמקרה זה מתאר מוקד קטן העובד בעומס גבוה. במוקד קטן זה יש לקוחות הממתינים יותר מחצי שעה, ולכן ברזולוציה שכזו יש השפעה הדדית של חצאי שעות עוקבים הגורמים לנוסחת Little לא להתקיים.

2.1.2. (4 נקודות) עבור המקרים בהם נוסחת Little לא מתקיימת, הציעו אלטרנטיבה שסביר שתעבוד, המקשרת בין קצב מופע, מספר השיחות במערכת וזמן שהיה במערכת.

תשובה:

ניתן להשתמש בנוסחה לחישוב ה-offered-load, שבעצם מהווה אנלוג לנוסחת Little במערכת תלוית-זמן:

$$L(t) = \int_{-\infty}^{\infty} \lambda(u) P(W > t - u) du$$

כאשר $L(t)$ – הוא ממוצע מספר האנשים במערכת בזמן t , W – משתנה מקרי המתאר את זמן שהיה במערכת, ו- $\lambda(u)$ – קצב המופע בזמן u .

חלק 2. תק"מ (12 נקודות)

קצב ההגעה לדלפק הקבלה בנמל תעופה הוא 282 נוסעים בשעה. קצב השירות בדלפק הוא 60 לקוחות בשעה. ישנם 5 פקידים (שרתים) העובדים בדלפק הקבלה. ניתן להניח כי זמני השירות מתפלגים מעריכית וכי תהליך הגעת הנוסעים הינו תהליך פואסון וכן כי זמני השירות בלתי תלויים בתהליך הגעת הנוסעים.

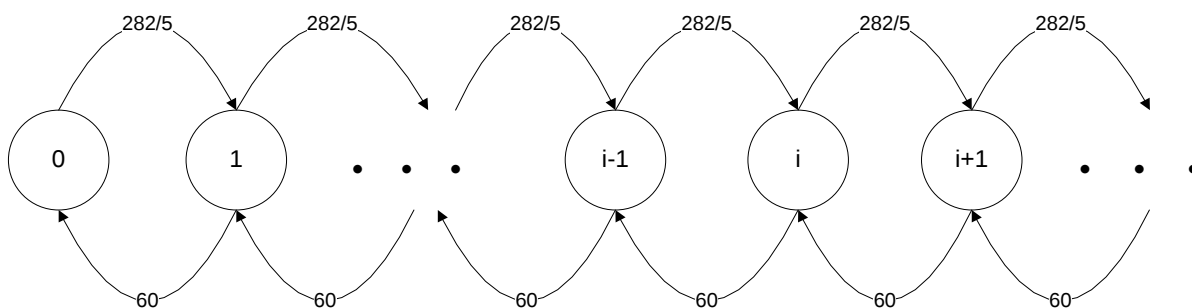
מנהל שדה התעופה בוחן שתי אפשרויות לניהול התורים:

1. חמישה תורי M/M/1 כאשר בכל אחד קצב הגעה של $282/5$ נוסעים לשעה, וקצב שירות של

60 נוסעים לשעה בכל אחד מהתורים.

2. תור M/M/5 יחיד בו קצב ההגעה הינו 282 נוסעים בשעה, ובו עובדים חמישה פקידים כל אחד בקצב של 60 לקוחות בשעה.

להלן דיאגרמת קצבי המעבר של אחד מתורי ה M/M/1 הנ"ל:



2.2.1 (5 נקודות) פתח נוסחה לחישוב $E(L_{sys}^1)$, כאשר $L_{sys}^1 =$ מספר האנשים במערכת M/M/1 כולה (שירות+תור) במצב יציב. הצב בנוסחה את ערכי השאלה ומצא את תוחלת מספר האנשים במערכת.

תשובה: לפי שיטת החתכים

$$\begin{cases} \frac{282}{5} \pi_i = 60 \pi_{i+1} \\ \sum_{i=0}^{\infty} \pi_i = 1 \end{cases}$$

ולכן

$$\pi_i = 0.94^i \pi_0$$

$$1 = \sum_{i=0}^{\infty} 0.94^i \pi_0 = \frac{1}{1-0.94} \pi_0 \Rightarrow \pi_0 = 1-0.94 = 0.06 \Rightarrow \pi_i = 0.94^i \cdot 0.06$$

מאחר ו $P(L_{sys}^1 = i) = \pi_i = 0.94^i \cdot 0.06$ נקבל כי $L_{sys}^1 \sim Geo_0(0.06)$ ולכן

$$E(L_{sys}^1) = \frac{0.94}{0.06} = 15 \frac{2}{3}$$

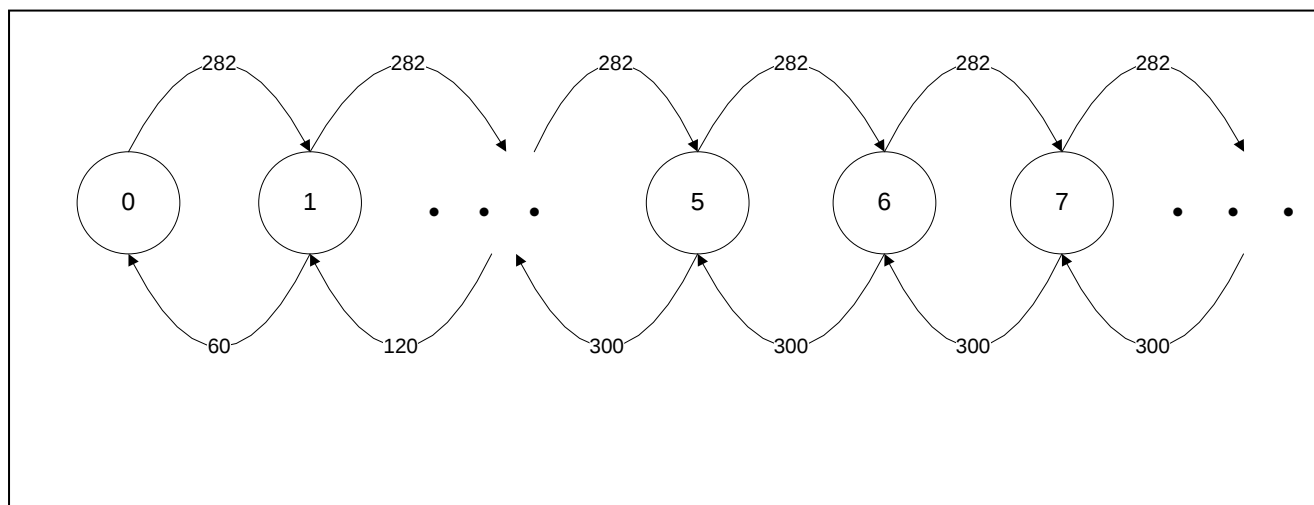
2.2.3 (2 נקודות) מצא את $E(W_q^1)$, כאשר W_q^1 - זמן ההמתנה בתור במצב יציב במודל M/M/1 הנ"ל.

תשובה:

לפי נוסחת Little, $E(W_{sys}^1) = \frac{E(L_{sys}^1)}{\lambda} = \frac{15 \frac{2}{3}}{\frac{282}{5}} = 0.2778 \text{ hour} = 16.7 \text{ min}$, כעת,

$$E(W_q^1) = E(W_{sys}^1) - E(\text{Service}) = 16.7 - 1 = 15.7 \text{ min}$$

2.2.4 (2 נקודות) השלם את דיאגרמת קצבי המעבר של התור M/M/5 הרלוונטי:



2.2.5 (3 נקודות) מצא את $E(W_q^5)$. כאשר W_q^5 – זמן ההמתנה בתור במערכת M/M/5 הנ"ל. עבור פתרון

סעיף זה ניתן להיעזר בדפי עזר מההרצאות אשר מופיעים בסוף שאלה 2.

תשובה:

$$E(W_q^5) = \frac{1}{\mu} E_{2,n} \left[\frac{1}{n} \frac{1}{1-\rho} \right] \text{ ולכן } \frac{W_q^5}{1/\mu} \sim \begin{cases} 0 & \text{w.p } 1 - E_{2,n} \\ \exp\left(\text{mean} = \frac{1}{n} \frac{1}{1-\rho}\right) & \text{w.p } E_{2,n} \end{cases}$$

נמצא את $E_{2,n}$:

$$R = \frac{\lambda}{\mu} = \frac{282}{60} = 4.7, \quad \rho = \frac{\lambda}{n\mu} = \frac{4.7}{5} = 0.94$$

$$\pi_0^{-1} = \frac{R^0}{0!} + \frac{R^1}{1!} + \frac{R^2}{2!} + \frac{R^3}{3!} + \frac{R^4}{4!} + \frac{R^5}{5!(1-\rho)} = 1 + 4.7 + \frac{4.7^2}{2} + \frac{4.7^3}{6} + \frac{4.7^4}{24} + \frac{4.7^5}{5! \cdot 0.06} = 372.9$$

$$\pi_0 = 0.00268$$

$$E_{2,n} = \frac{4.7^5}{5!} \frac{1}{1-0.94} \cdot 0.00268 = 0.854 \text{ ולכן}$$

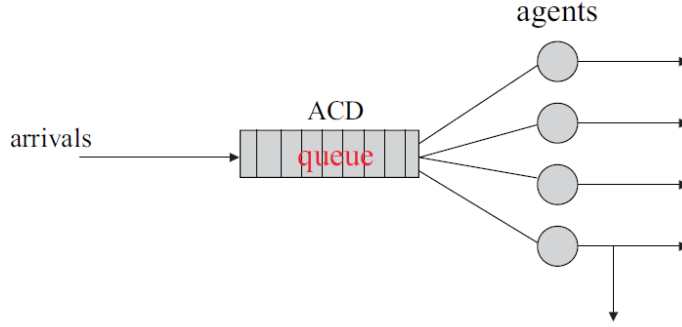
לסיכום,

$$E(W_q^5) = \frac{1}{60} \cdot 0.854 \cdot \frac{1}{5} \frac{1}{1-0.94} = 0.047 \text{ hours} = 2.85 \text{ min}$$

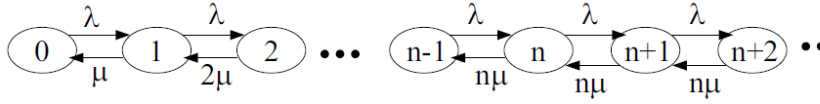
M/M/ n (Erlang-C) queue

- Poisson arrivals, rate λ ;
- n exponential servers, rate μ .

Widely used in call centers.



Transition-rate diagram



$$\lambda_j = \lambda, \quad j \geq 0,$$

$$\mu_j = (j \wedge n)\mu, \quad j \geq 1.$$

Agents' utilization

$$\rho = \frac{\lambda}{n\mu}.$$

Assume $\rho < 1$ ($R < n$) to ensure stability (as in M/M/1).

Steady-state distribution:

$$\begin{aligned} \pi_i &= \frac{R^i}{i!} \pi_0, & i \leq n, \\ &= \frac{n^n \rho^i}{n!} \pi_0, & i \geq n, \\ \pi_0 &= \left[\sum_{j=0}^{n-1} \frac{R^j}{j!} + \frac{R^n}{n!(1-\rho)} \right]^{-1}, \end{aligned}$$

where $R = \frac{\lambda}{\mu}$ is the **offered load**.

Erlang-C Formula (1917):

Delay probability:

$$P\{W_q > 0\} \triangleq E_{2,n} = \sum_{i \geq n} \pi_i = \frac{R^n}{n!} \frac{1}{1 - \rho} \cdot \pi_0.$$

Erlang-C computation: recursion, see Erlang-B below.

Number-in-queue:

$$P\{L_q = i\} = E_{2,n} \cdot (1 - \rho) \rho^i, \quad i > 0,$$

or

$$L_q = \begin{cases} 0 & \text{wp } 1 - E_{2,n} \\ \text{Geom}(1 - \rho) & \text{wp } E_{2,n} \end{cases}$$

Waiting time distribution:

$$\frac{W_q}{1/\mu} = \begin{cases} 0 & \text{wp } 1 - E_{2,n} \\ \exp\left(\text{mean} = \frac{1}{n} \cdot \frac{1}{1-\rho}\right) & \text{wp } E_{2,n} \end{cases}$$

Compare with M/M/1!

Departure process: Poisson(λ) in steady-state.

Proof via reversibility, as with M/M/1.

שאלה 2 – תרגילי בית (25 נקודות)

איוש בעזרת 4CC

ארבעה עשר מוקדנים עובדים במרכז השירות של GazolCo. רוב השיחות הן של לקוחות המתקשרים לשלם את חשבון הגז שלהם. מנתוני שיחות שניתחנו עולה כי זמן שירות טלפוני ממוצע הוא 4 דקות, קצב השיחות הנכנסות הוא 180 שיחות לשעה בממוצע, והסבלנות הממוצעת היא 3 דקות. אנו רוצים לעזור למנהל השירות לנתח את המערכת ולקבוע איוש מתאים בעזרת תוכנת 4CC.

2.1. (3 נקודות) מנהל המערכת רצה לבדוק מה יקרה אם קצב השיחות הנכנסות יגדל (בעוד מספר השרתים נשאר קבוע). להלן שתי טבלאות המתארות מה יקרה במצבים אלו. הטבלאות מבוססות על פלטים של התוכנה, אחד הופק בעזרת מודל M/M/N והשני בעזרת מודל M/M/N+M. איזה מהטבלאות מתאימה לאיזה מודל? הסבירו תשובתכם.

טבלה 1

Calls Per Hour	Average Speed of Answer (secs)	Agent's Occupancy (%)	P(Wait>0)
180	57.8	85.7%	$1-0.518=0.482$
185	79.6	88.1%	$1-0.447=0.553$
190	113.5	90.5%	$1-0.37=0.63$
195	171.3	92.9%	$1-0.286=0.714$
200	289.2	95.2%	$1-0.197=0.803$
205	647.0	97.6%	$1-0.101=0.899$
210	n/a	n/a	

טבלה 2

Calls Per Hour	Average Speed of Answer (secs)	Agent's Occupancy (%)	P(Wait>0)
180	9.2	80.7%	$1-0.699=0.300$
185	10.6	82.2%	$1-0.666=0.333$
190	12.0	83.7%	$1-0.632=0.367$
195	13.5	85.1%	$1-0.598=0.401$
200	15.1	86.3%	$1-0.564=0.435$
205	16.8	87.5%	$1-0.530=0.469$
210	18.6	88.7%	$1-0.497=0.502$

תשובה :

טבלה 1 הופקה בעזרת מודל M/M/N וטבלה 2 בעזרת מודל עם נטישות M/M/M+M. אנו מגלים זאת כי המודל עם הנטישות תמיד יציב, בעוד המודל ללא נטישות לא יציב אם קצב הגעת שיחות גדול מ-210 שיחות לשעה.

2.2. (3 נקודות) הסבירו מדוע בשורה האחרונה של טבלה 1 מופיע הערך n/a . מהו קצב ההגעה המקסימלי בו תתקבל תשובה מספרית.

תשובה :

מודל Erlang-C יציב כל עוד $\lambda > \mu$, ולכן $60/4 * 14 = 210$ עבור כל קצב הגעה קטן ממש מ-210 לקוחות לשעה נקבל שהמערכת יציבה והתוכנה תחשב מדדי שירות, ועבור כל ערך גדול או שווה ל-210 נקבל שהמערכת לא יציבה.

2.3. (4 נקודות) ציינו מה אתם מצפים שיקרה למדדים הבאים, במידה ופרמטר הסבלנות θ יעלה : אחוז הנוטשים, ממוצע זמן ההמתנה לשירות (ASA), ממוצע אורך התור, נצילות השרתים. סמנו את התשובה הנכונה והסבירו תשובתכם בקצרה.

תשובה : כאשר פרמטר הסבלנות עולה :

אחוז הנוטשים : יעלה / ירד

ממוצע זמן ההמתנה לשירות (ASA) : יעלה / ירד

ממוצע אורך התור : יעלה / ירד

נצילות השרתים : יעלה / תרד

הסבר : כאשר פרמטר הסבלנות עולה, הסבלנות הממוצעת של לקוח יורדת ולכן יותר לקוחות נוטשים. כתוצאה מכך פחות לקוחות יישארו לחכות לשירות מה שיגרום לתור לקטון ולעומס לרדת. לכן, גם זמן ההמתנה ונצילות השרתים ירדו.

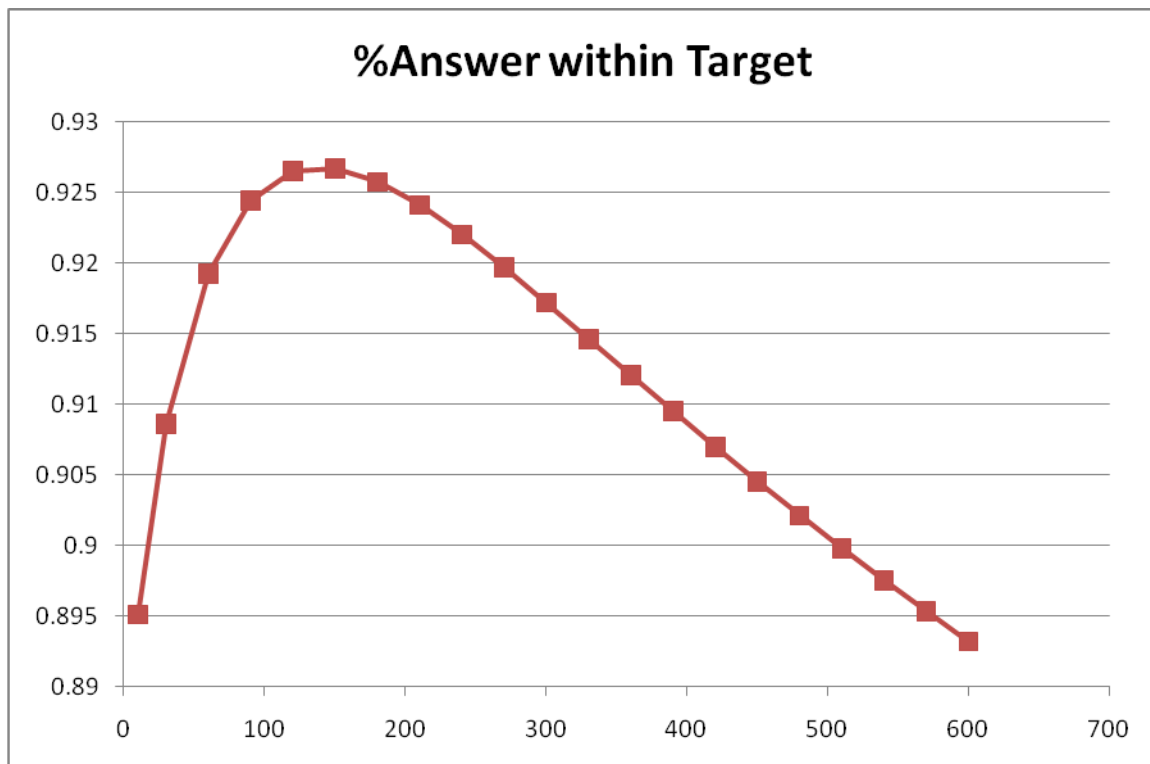
2.4. (4 נקודות) מה יקרה לאחוז הנוטשים אם סבלנות הלקוחות תעלה לאינסוף או תרד ל-0? לאיזה מודל המערכת תשאף במקרים אלו? הסבירו תשובתכם.

תשובה:

$\theta \rightarrow \infty$: כאשר פרמטר הסבלנות עולה לאינסוף, ממוצע הסבלנות שואף ל-0, ולכן כל לקוח שמגיע כאשר אין שרת פנוי עוזב מיד את המערכת. כלומר, % הנוטשים בגבול יהיה שווה לאחוז הלקוחות המגיעים לשרת עסוק - $P(W > 0) \rightarrow P(\text{Abandon})$. במצב זה המערכת שואפת למערכת M/M/N/N (Erlang-B).

$\theta \rightarrow 0$: כאשר פרמטר הסבלנות יורד ל-0, ממוצע הסבלנות שואף לאינסוף, אף לקוח לא נוטש (כלומר, $P(\text{Abandon}) \rightarrow 0$), ולכן המערכת שואפת למערכת M/M/N (Erlang-C).

2.5. (7 נקודות) כאשר בודקים את התנהגות של "אחוז הנענים תוך 1.5 דקות" כפונקציה של ממוצע הסבלנות מקבלים את הגרף הבא:

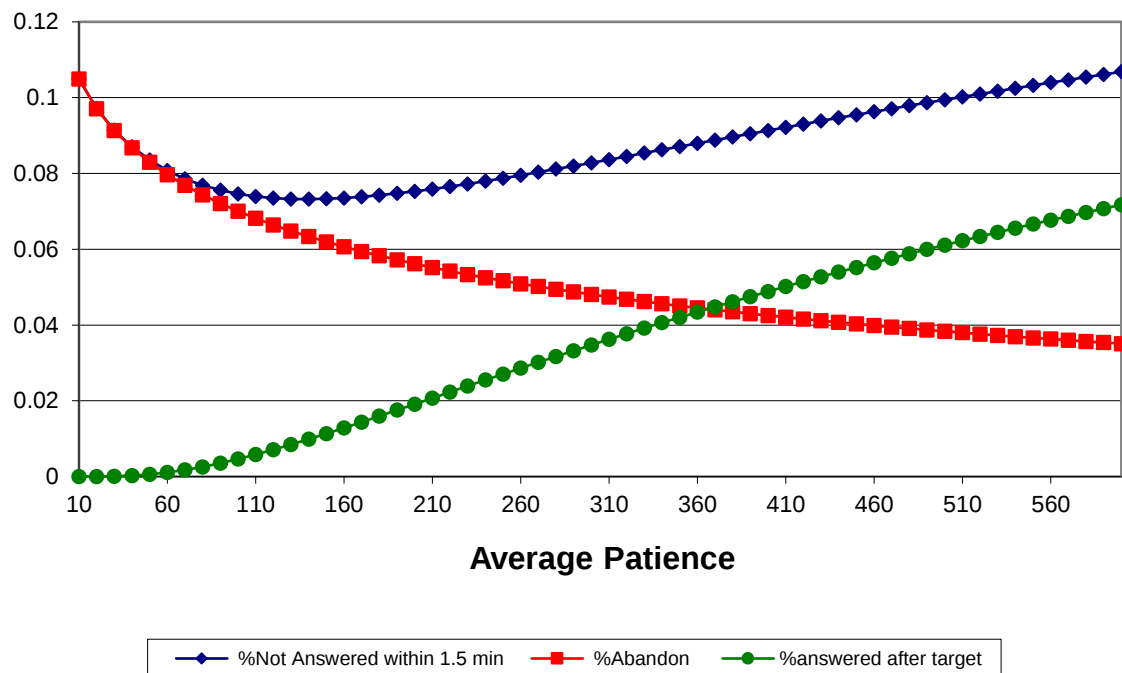


הסבירו מדוע לגרף צורה כזו.

רמז : ניתן לחלק את הלקוחות שלא נענים תוך 1.5 דקות לשתי קבוצות. הגדירו מהן, וכיצד הן עוזרות להסברת התופעה הנ"ל.

תשובה :

ניתן לחלק את הלקוחות שלא נענים תוך 1.5 דקות, ללקוחות שנטשו (לפני או אחרי 1.5 דקות) וללקוחות שלא נטשו וקיבלו שרות לאחר יותר מ- 1.5 דקות. אחוז הלקוחות הנוטשים יורד כפונקציה של ממוצע הסבלנות. אחוז הלקוחות שלא נטשו ומקבלים שרות לאחר יותר מ- 1.5 דקות עולה ככל שממוצע הסבלנות עולה (כיוון שאם ללקוחות יש יותר סבלנות יהיו יותר לקוחות בתור $\leq$ זמני ההמתנה יעלו $\leq$ אחוז אלו שלא נוטשים ומקבלים שרות לאחר יותר מ- 1.5 דקות עולה). לכן, חיבור שתי הפונקציות המונוטוניות הנ"ל גורם לכך שאחוז הלקוחות שלא קיבלו שירות תוך 1.5 דקות כפונקציה של ממוצע הסבלנות יראו לפי הגרף למטה. מכאן נובע שאחוז הלקוחות שכן קיבלו שרות תוך 1.5 דקות היא פונקציה כפי שרואים בשאלה.



2.6. (4 נקודות) מנהל המוקד גילה שבימים מסוימים (לאחר שליחת חשבונות ללקוחות), קצב השיחות למוקד גדול פי כמה מימים רגילים. ליתר דיוק, ביום לאחר שליחת החשבונות קצב ההגעה הוא 540 לקוחות לשעה (פי 3 מיום רגיל) וביום למחרת 360 לקוחות לשעה (פי 2 מיום רגיל); לאחר מכן הקצב חוזר לרמתו הרגילה.

מנהל המוקד רוצה לחשב את כמות המוקדנים הדרושה כך שאחוז הנוטשים ישאר ברמתו של יום רגיל (קרי כ- 5%). לשם כך הוא יצר בתוכנת 4CC את הפלט הבא:

Staffing Query tool helps you decide how many Agents your Call Center requires. Enter your call center's **Performance Goal** (inactive Performance Indicators appear in parenthesis) and other parameters, then press 'Compute'.

Your Call Center's Parameters

- Performance Goal: Agent's Occupancy 50%
- Average Time to Handle One Call (mm:ss): 04:00
- Calls per 60 minute Interval: 540
- Average Callers' Patience (mm:ss): 03:00

Settings

- Features: Abandons
- Basic Interval: 60 minutes
- Target Time: 00:10 (mm:ss)

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	Agent's Occupancy	%Answer	%Abandon	Average Speed of Answer
Lower										
Upper										
1	60.0	00:10.0	14.0	04:00.0	180.0	03:00.0	80.7%	94.2%	5.8%	00:09.2
2	60.0	00:10.0	15.0	04:00.0	180.0	03:00.0	77.0%	96.2%	3.8%	00:06.0
3	60.0	00:10.0	26.0	04:00.0	360.0	03:00.0	87.5%	94.8%	5.2%	00:08.6
4	60.0	00:10.0	27.0	04:00.0	360.0	03:00.0	85.4%	96.1%	3.9%	00:06.4
5	60.0	00:10.0	37.0	04:00.0	540.0	03:00.0	91.6%	94.2%	5.8%	00:09.8
6	60.0	00:10.0	38.0	04:00.0	540.0	03:00.0	90.3%	95.3%	4.7%	00:07.9
7										

Ready 2/1/2010 1:06 PM

הוא טוען שניתן לראות שבמערכת שלנו קיים יתרון לגודל באיוש. האם הוא צודק? הסבירו.

תשובה:

המנהל צודק. כפי שניתן לראות מהנתונים, כאשר קצב ההגעות עולה פי 2 מספר השרתים עולה רק פי 1.8; כאשר קצב ההגעות עולה פי 3, מספר השרתים הדרוש על מנת לשמור על אותה רמת שירות עולה רק פי 2.5. זוהי בדיוק התופעה של "יתרון לגודל".

שאלה 3 – יישום (33 נקודות)

להלן קטע מתוך אתר המועצה הישראלית לצרכנות (אין צורך לקרוא את כל הטקסט):

חוק שלוש הדקות - הסוף להתייבשות בהמתנה למוקדי הלקוחות

החל מ-17.6.2015, החברות מחויבות לספק מענה אנושי תוך 3 דקות, או אפשרות להשאיר הודעה תוך 2 דקות. לאחר מכן החברות מחויבות לחזור לצרכן תוך 3 שעות. על מי חל החוק ומה לעשות במקרה שאין מענה?



שלום לזמני המתנה ארוכים במוקדי שירות לקוחות? בשנת 2009 ניסחה המועצה לצרכנות הצעת חוק הקובעת כי יכלול בחוק "3 הדקות" גם שירותים נוספים, כגון: טעויות בחשבון, בירור חשבון ועוד, כך שצרכן המנסה לברר את חשבונו יקבל אף הוא שירות מהיר ויעיל. עד להרחבה שהציעה המועצה, תחולת החוק היתה על תקלות בלבד. הצעת החוק לפיה כל פניה, לא משנה באיזה נושא, תיענה תוך 3 דקות מקסימום.

במקרה שצרכן פנה לאחד העוסקים שהחוק חל עליהם וזמן ההמתנה הצפוי לקבלת מענה אנושי עולה על שלוש דקות, על העוסק להודיע לצרכן בהודעה מוקלטת, לא יאוחר משתי דקות מתחילת השיחה, כי זמן ההמתנה הצפוי עולה על שלוש דקות וכי יש באפשרותו לעבור לשירות השארת הודעה, או להמתין לקבלת מענה אנושי.

במקרה שהצרכן בחר לעבור לשירות השארת הודעה – הצרכן ישאיר בהודעה מוקלטת את שמו ומספר טלפון להתקשרות עמו והעוסק יחזור לצרכן במענה אנושי לא יאוחר משלוש שעות ממועד השארת ההודעה.

במקרה שהצרכן בחר להמתין למענה אנושי – העוסק יודיע לצרכן בהודעה מוקלטת את מקומו בתור ואת זמן ההמתנה המשוער ויודיע לצרכן כי באפשרותו לעבור לשירות השארת הודעה בכל רגע. אם הצרכן בחר להשאיר הודעה, יחזור אליו העוסק במענה אנושי לא יאוחר משלוש שעות ממועד השארת ההודעה.

במקרה שהצרכן השאיר הודעה לעוסק פחות משעתיים לפני תום יום העבודה, רשאי העוסק לחזור לצרכן במענה אנושי ביום העבודה הבא של העוסק ולא יאוחר משלוש שעות ממועד תחילת העבודה.

אם הצרכן לא ענה – במקרה שהצרכן לא ענה, יודיע לו העוסק בהודעה קולית או במסרון SMS את טווח השעות שבמהלכן יחזור לצרכן ובלבד שלא יהיה מאוחר משלוש שעות מהמועד שבו מסר את ההודעה כאמור.

אם הצרכן לא ענה בפעם השנייה – במקרה שהעוסק חזר אל הצרכן בפעם השנייה והצרכן לא ענה, העוסק יודיע לו בהודעה קולית או במסרון שעליו לפנות לעוסק מחדש.

חברת סלולר ידועה, 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 הני"ל.

חלק 3.1. מידול המוקד (6 נקודות)

3.1.1 (3 נקודות) מתוך המודלים הבסיסיים שנלמדו בכיתה, איזה מודל סטוכסטי נראה לכם מתאים לניתוח מצב זה? מה הן ההנחות המתמטיות במודל זה?

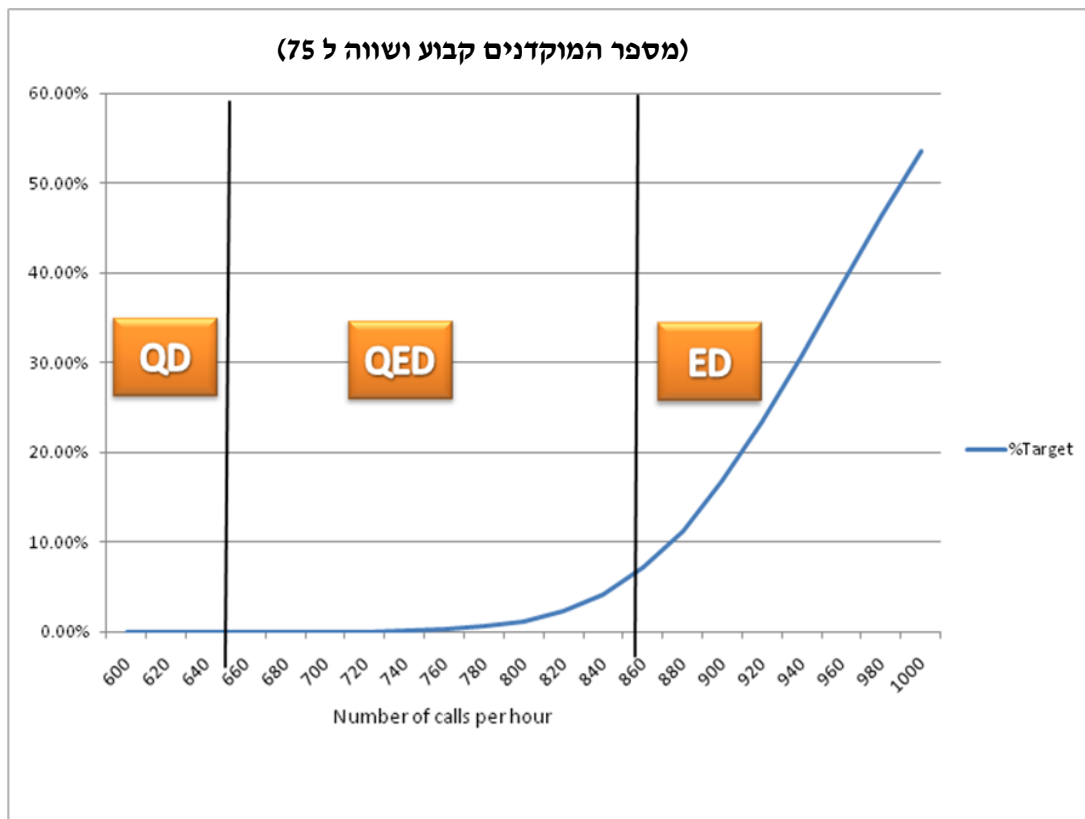
תשובה:

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 (3 נקודות) 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 נקבע כל אחד מהתחומים הללו ?

רמז: הניחו כי האיוש נקבע ע"י כלל השורש $n = R + \beta\sqrt{R}$, התבוננו בטווחי β בתחומים הנ"ל.

תשובה:

תחום ה QD נקבע עבור $\beta \leq -1.1$,

תחום ה QED נקבע עבור $-1.18 < \beta < 1.1$

ותחום ה ED נקבע עבור $\beta \geq 1.18$. למשל $1.1 = \frac{75 - 660 \cdot \frac{1}{10}}{\sqrt{660 \cdot \frac{1}{10}}}$

חלק 3.2. התאמת המוקד לחוק החדש: שלב א' (14 נקודות)

מעתה נדון בשעות הבוקר בהן קצב הגעת הלקוחות קבוע ושווה ל 1000 לקוחות בשעה. 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. בשלב ראשון אין לחברה יכולת טכנית לקבל הודעות של לקוחות במענה הקולי. על כן, מנהל המוקד מטיל על האנליסט למצוא את מספר המוקדנים המינימאלי כך שאחוז הלקוחות הממתינים יותר מ- 3 דקות לא יעלה על 0.3%. (0.3% נבחר כך שהחברה לא תעבור על החוק. כלומר, כדי ליצור מצב שבו אין צורך להציע ללקוחות את האופציה להשארת הודעה).

להלן פלט של תוכנת 4CC

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	Agent's 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 (4 נקודות) 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

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

$$עבור 95 שרתים \quad 100 - \%Target = 93.4\% + 6.3\% = 99.7\%$$

ולכן, מספר השרתים יהיה 95.

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

3.2.2 (3 נקודות) על סמך פונקציית Garnett המצורפת 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 :

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

3.2.3 (4 נקודות) חשבו את $E(W_q | W_q > 0)$, פרטו את חישוביכם, והסבירו את משמעות מדד זה.

תשובה:

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

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

$$E(W_q) = \frac{E(L_q)}{\lambda} = \frac{12.5}{1000} \text{ לפי נוסחת Little. ונקבל כי}$$

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

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

3.2.4 (3 נקודות) מהי תוחלת זמן ההמתנה בתור של לקוחות שנטשו את המערכת?

תשובה:

$$E(W_q) = E(W_q | Abandon)P(Abandon) + E(W_q | Service)P(Service)$$

$$\Rightarrow E(W_q | Abandon) = \frac{E(W_q) - E(W_q | Service)P(Service)}{P(Abandon)} =$$

$$\frac{E(W_q) - (Average\ Speed\ of\ Answer) \cdot P(Service)}{P(Abandon)}$$

$$E(W_q | Abandon) = \frac{45 - 45.4 \cdot 0.937}{0.063} = 39 \text{ sec}$$

חלק 3.3. התאמת המוקד לחוק החדש: שלב ב' (13 נקודות)

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 הפעלה מלאה של החוק החדש, השיטה בה השתמשו עד עתה לחישוב מספר הלקוחות אשר המתנינו יותר מ 3 דקות נותנת **הערכת יתר** לזמני המתנה בתור.

3.3.1 (3 נקודות) הסבר מדוע אריק צודק.

תשובה:

בשיטה בה השתמשו עד עכשיו, הלקוחות שבוחרים באפשרות שיחזרו אליהם נלקחים בחשבון גם מבחינת זמני ההמתנה הארוכים שלהם (יותר מ- 2 הדקות שהמתנינו בפועל) וגם מבחינת העומס שהם יוצרים על המערכת, הגורם לזמני ההמתנה של לקוחות אחרים להיות ארוכים יותר.

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 הממתנינים 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 : קובע את הזמן אשר לקוחות הממתנינים יותר ממנו יופנו למערכת האחרת.
- פלט: %Overflow מציין את אחוז הלקוחות אשר עברו את הזמן המדובר (שלוש דקות) והופנו למערכת האחרת.

הצעתו האלטרנטיבית של אריק:

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

3.3.2 (3 נקודות) האם השיטה שמציע אריק מדויקת? הסבר.

תשובה:

השיטה שמציע אריק אינה מדויקת, משום שהוא מזניח את העובדה שלא כל הלקוחות בוחרים באפשרות שיחזרו אליהם. חלק מהלקוחות בוחרים להמשיך ולהמתין ובכך מגדילים את העומס על המערכת. שיטתו של אריק היתה מדויקת אם כל הלקוחות שהיו מקבלים הודעה, היו בוחרים באפשרות שיחזרו אליהם (במקום להמתין עוד). כלומר, הוא מניח שללקוחות אשר קיבלו הודעה 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:

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
- Overflow Time Limit (mm:ss): 03:00

Settings

- Features: Abandons, Overflows
- Basic Interval: 60 minutes
- Target Time: No Target

Change Settings

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

	Basic Interval (minutes)	Number of Agents	Average Handling Time	Calls per Interval	Average Patience	Overflow time-limit	Agent's Occupancy	%Answer	%Abandon	%Overflow	Average Speed of Answer	Average Queue Length
Results	60.0	100.0	06:00.0	1,000.0	12:00.0	03:00.0	96.7%	96.7%	3.3%	.0%	00:23.4	6.6
1	60.0	90.0	06:00.0	1,000.0	12:00.0	03:00.0	99.6%	89.7%	9.9%	.4%	01:13.5	19.9
2	60.0	91.0	06:00.0	1,000.0	12:00.0	03:00.0	99.5%	90.5%	9.2%	.3%	01:07.4	18.3
3	60.0	92.0	06:00.0	1,000.0	12:00.0	03:00.0	99.3%	91.4%	8.4%	.2%	01:01.5	16.8
4	60.0	93.0	06:00.0	1,000.0	12:00.0	03:00.0	99.1%	92.3%	7.6%	.1%	00:55.7	15.3
5	60.0	94.0	06:00.0	1,000.0	12:00.0	03:00.0	98.9%	93.0%	6.9%	.1%	00:50.2	13.8
6	60.0	95.0	06:00.0	1,000.0	12:00.0	03:00.0	98.7%	93.7%	6.2%	.1%	00:44.9	12.4
7	60.0	96.0	06:00.0	1,000.0	12:00.0	03:00.0	98.3%	94.5%	5.5%	.0%	00:39.9	11.1
8	60.0	97.0	06:00.0	1,000.0	12:00.0	03:00.0	98.0%	95.1%	4.9%	.0%	00:35.2	9.8
9	60.0	98.0	06:00.0	1,000.0	12:00.0	03:00.0	97.6%	95.7%	4.3%	.0%	00:30.9	8.7
10	60.0	99.0	06:00.0	1,000.0	12:00.0	03:00.0	97.2%	96.2%	3.8%	.0%	00:27.0	7.6
11	60.0	100.0	06:00.0	1,000.0	12:00.0	03:00.0	96.7%	96.7%	3.3%	.0%	00:23.4	6.6
12												
13												
14												
15												
16												

3.3.3 (2 נקודות) מהו כעת מספר המוקדנים המינימאלי הנדרש על פי אריק כדי לעמוד באילוץ 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 שוהים במערכת יותר משלוש דקות : התורים קצרים יותר $\Rightarrow$ נדרשים פחות מוקדנים.

3.3.4 (3 נקודות) בהמשך לסעיף הקודם, נניח כי אותם מוקדנים אשר עונים לשיחות, הם אלו אשר חוזרים ללקוחות אשר המתינו יותר משלוש דקות. מהי כעת נצילות המוקדנים? (הניחו שמשך שיחה ללקוחות שחזרו אליהם זהה למשך שיחה של לקוחות שנשארו בתור).

תשובה :

עבור איוש של 91 מוקדנים :

לפי השורה השלישית בפלט, פרופורציית הזמן שהמוקדנים עסוקים בשיחות נכנסות היא $\rho_1 = 0.995$. בנוסף, ל- 0.3% מהשיחות הנכנסות צריכים לחזור (אחרי שהמתינו יותר משלוש דקות), כלומר

$$\rho_2 = \frac{0.003 \cdot 1000}{91 \cdot 10} = 0.0033$$

ולכן, $\rho = \rho_1 + \rho_2 = 0.9983$

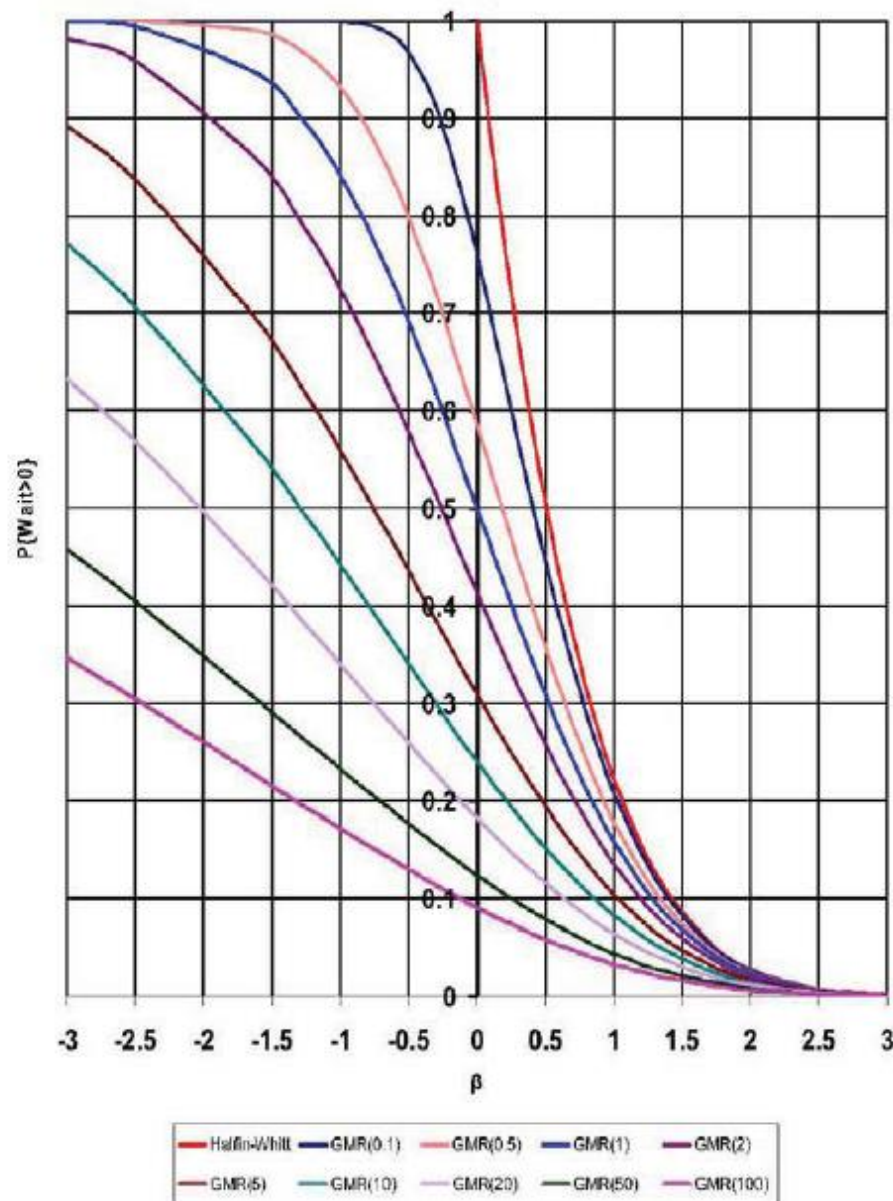
3.3.5 (2 נקודות) עבור יישום הפרוטוקול החדש נדרש לבצע הערכה של זמן ההמתנה הצפוי ללקוח ברגע הגעתו למערכת. הצע שתי דרכים לביצוע הערכה זו.

תשובה :

1. חישוב ממוצע ההמתנה של מספר לקוחות שנכנסו לשירות בזמן האחרון (למשל, בחצי שעה האחרונה).
2. על פי אורך התור ברגע ההגעה של הלקוח, ניתן לחשב את תוחלת משך ההמתנה הכולל לפי התפלגות Gamma (בהנחה שזמני השירות אקספוננציאליים).

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 $\theta/\mu = x$. 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 (17 נקודות)

חלק 1. אמידה מוטה

נניח שזמנים בין-מופעים של אוטובוסים לתחנה הם בלתי-תלויים, שווי-התפלגות, ומתוארים על ידי משתנה מקרי (מ"מ) S . נניח שלמ"מ S יש תוחלת סופית $E(S)$ ושונות סופית $\sigma^2(S)$. בכיתה הראינו שעבור נוסע המופיע באקראי לתחנה, הזמן S_e מרגע מופע הנוסע עד רגע מופע האוטובוס הבא הוא בעל תוחלת:

$$E(S_e) = \frac{1}{2} E(S) \cdot [1 + C^2(S)]$$

כאשר $C(S)$ הוא מקדם ההשתנות של המ"מ S . (בהקשרים של שירות, S_e נקרא גם "זמן שארית השירות").

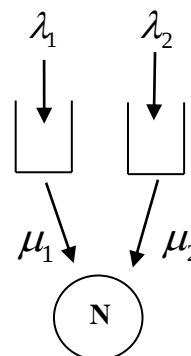
4.1 (3 נקודות) הסבירו בקיצור את התופעה המתוארת על ידי הנוסחה ל- $E(S_e)$ (אותה כינינו "אמידה מוטה"), תוך שימוש במקרה פרטי (התפלגות מיוחדת) הממחיש אותה.

תשובה:

התופעה: דגימה של אינטרוול זמן ע"י בחירה של נקודת זמן בו, גורמת להטיה לטובת אינטרוולי זמן ארוכים.
לדוגמא: כאשר האינטרוולים הם זמנים בין מופעים של אוטובוסים, המתפלגים מעריכית (exp), נניח עם ממוצע $\frac{1}{\mu}$, ממוצע S_e גם הוא $\frac{1}{\mu}$ (מיידית, על סמך חוסר זיכרון); שיקול נאיבי היה צופה כי הממוצע של S_e הוא רק $\frac{1}{2} \cdot \frac{1}{\mu}$.

חלק 2. עדיפויות

נתונה מערכת תורים בארכיטקטורת V, המתוארת בציור הבא:



הפרמטרים של מודל המערכת נתונים ע"י :

- ♦ λ_1 (λ_2) - קצבי מופע לקוחות מסוג 1 (2),
- ♦ μ_1 (μ_2) - קצבי שירות לקוחות מסוג 1 (2),
- ♦ θ_1 (θ_2) - קצבי נטישה של לקוחות מסוג 1 (2),
- ♦ N - מספר השרתים בתחנה

המערכת פועלת לפי מדיניות שהיא *preemptive*, ז"א ניתן בכל רגע להפסיק שירות של לקוח ולהחזירו להמתנה בתור מהסוג שלו, אבל אסור "לזרוק" לקוחות מהמערכת. בנוסף, המדיניות הינה *work conserving*, דהיינו, השרתים חייבים לשרת כאשר יש לקוחות בתור (ז"א לא ייתכן מצב שבו יש שרתים פנויים ובעת ובעונה אחת יש גם לקוחות הממתנים בתור).

לצורך בנית **מודל נוזלים** למערכת הנ"ל, נגדיר את המשתנים הבאים :

$X_1(t)$ = סה"כ מספר הלקוחות מסוג 1 במערכת (בשירות או בתור), בזמן $0 \leq t$.

$X_2(t)$ = סה"כ מספר הלקוחות מסוג 2 במערכת (בשירות או בתור), בזמן $0 \leq t$.

4.2.1 (4 נקודות) כתבו משוואה דיפרנציאלית עבור $X_1(t)$ ו- $X_2(t)$, $0 \leq t$, תחת ההנחה שהתחנה

נותנת עדיפות ללקוחות מסוג 1.

תשובה :

$$\frac{dX_1(t)}{dt} = \lambda_1 - \mu_1(X_1 \wedge N) - \theta_1(X_1 - N)^+$$

$$\frac{dX_2(t)}{dt} = \lambda_2 - \mu_2(X_2 \wedge (N - X_1)^+) - \theta_2[X_2 - (N - X_1)^+]^+$$

בכיתה הוכחנו את נוסחת חניצין-פולטשיק לחישוב זמני ההמתנה הממוצעים, ובתרגול הוכחתם הרחבה של נוסחה זו לרשת עם עדיפויות, כמו שהוצגה לעיל, כאשר יש שרת יחיד. (דפי ההרצאה והתרגול מצורפים).

4.2.2 (4 נקודות) בדף ההוכחה של נוסחת חניצין-פולטשיק מסומנים שני ריבועים ב-1 ו-2. הסבר על מה מבוסס כל אחד מהמעברים.

תשובה :

1. מבוסס על תכונת Poisson Arrivals See Time Averages (PASTA)

2. מבוסס על חוק ליטל

4.2.3 (4 נקודות) נגדיר :

W_1^p - ממוצע זמני המתנה של לקוחות מסוג 1 (עדיפות גבוהה) במדיניות preemptive

W_2^p - ממוצע זמני המתנה של לקוחות מסוג 2 (עדיפות נמוכה) במדיניות preemptive

W_1^n - ממוצע זמני המתנה של לקוחות מסוג 1 (עדיפות גבוהה) במדיניות non-preemptive

W_2^n - ממוצע זמני המתנה של לקוחות מסוג 2 (עדיפות נמוכה) במדיניות non-preemptive

הניחו שהעומס הכולל זהה. השלימו את הסימן המתאים ($=$, $>$, $<$) והסבירו בקצרה.

תשובה :

$$W_1^p \quad < \quad W_1^n$$

$$W_2^p \quad > \quad W_2^n$$

$$W_1^p - W_2^p \quad < \quad W_1^n - W_2^n$$

הסבר :

במדיניות preemptive הקבוצה עם העדיפות הגבוהה כלל לא מושפעת מקבוצת העדיפות הנמוכה. במדיניות non-preemptive הקבוצה עם העדיפות הנמוכה מעכבת את הקבוצה עם העדיפות הגבוהה. דבר זה בא לידי ביטוי בשארית זמן השירות (מחכים לסיומי שירות גם של לקוחות עם עדיפות נמוכה). לכן, זמן ההמתנה של קבוצת העדיפות הגבוהה גדול יותר במדיניות non-preemptive, ובהתאם כיוון שעומס העבודה זהה, זמן ההמתנה של קבוצת העדיפות הנמוכה קטן יותר. כלומר, הפערים בין זמני ההמתנה של שתי קבוצות העדיפות יהיו קטנים יותר במדיניות preemptive.

4.3.4 (2 נקודות) בהמשך לסעיף הקודם. כיצד ישתנה היחס השלישי במערכת מרובת שרתים (n שואף לאינסוף)?

תשובה : אם יש הרבה שרתים, ההסתברות למצוא שרת פנוי גדלה באופן משמעותי ואז הסיכוי שנצטרך לחכות למישהו וודאי למישהו בעדיפות נמוכה יותר קטנה, ולכן אנו צופים שהפער בין המדיניות הללו יהיה קטן יותר ויותר ככל ש- n גדל.

Pollaczek-Khintchine formula

M/G/1 Queue: Poisson arrivals,
generally distributed (iid) service durations.

Theorem. (Pollaczek-Khintchine)

$$E(W_q) = E(S) \cdot \frac{\rho}{1 - \rho} \cdot \frac{1 + C^2(S)}{2}.$$

Proff:

Derivation of Pollaczek-Khintchine

For customer $n = 1, 2, \dots$, denote

$W_q(n)$ = waiting-time of n -th customer.

$R(n)$ = residual service time, at time of the n -th arrival;
(= 0, for arrivals without waiting).

$L_q(n)$ = # of customers in queue, at time of n -th arrival.

$\{S_n\}$ = sequence of service-times.

$$W_q(n) = R(n) + \sum_{k=n-L_q(n)}^{n-1} S_k, \quad n \geq 1.$$

$$EW_q(n) = ER(n) + E(S_1) \cdot EL_q(n), \quad \text{by Wald,}$$

$$E(W_q) = E(R) + E(S_1)E(L_q), \quad n \uparrow \infty, \text{ assuming}$$

$$\exists \text{ limit } + \boxed{1}$$

$$= E(R) + \lambda E(S_1)E(W_q), \quad \text{by } \boxed{2}$$

$$E(W_q) = E(R) + \rho E(W_q), \quad \rho < 1 \Leftrightarrow \exists \text{ steady-state,}$$

$$E(W_q) = E(R)/(1 - \rho).$$

Left to calculate $E[R]$?

Via **Biased Sampling** (see next page):

- ρ = Prob. of arriving to a busy server. (**PASTA+Little**)

$$- E(R) = (1 - \rho) \cdot 0 + \rho \cdot E(S) \cdot \frac{1 + C^2(S)}{2}. \quad \text{q.e.d.}$$

Table 1: Discrete and Continuous Distributions

Distribution	Notation	Probability Function	Parameters	Support	Moment Gen. Function	Expectation	Variance	Comments
Binomial	$N \sim \text{Bin}(n, p)$	$p(k) = \binom{n}{k} p^k q^{n-k}$	$0 \leq p = 1 - q \leq 1; n \geq 1$	$k = 0, 1, \dots, n$	$[p \cdot e^t + q]^n$	np	npq	
Multi-nomial	$X \sim \text{Mult}(n, p)$	$p(x) = \frac{n!}{x_1! \dots x_k!} \prod_{i=1}^k p_i^{x_i}$ $x = (x_1, \dots, x_k)$	$0 \leq p_i; \sum_{i=1}^k p_i = 1; n \geq 1$ $p = (p_1, \dots, p_k)$	$0 \leq x_i; \sum_{i=1}^k x_i = n$		$E(X_i) = np_i$ $i = 1, \dots, k$	$\text{cov}(X_i, X_j) = -np_i p_j$ $(i \neq j)$	$X_i \sim \text{Bin}(n, p_i)$
Negative Binomial	$X \sim NB(m, p)$	$p(k) = \binom{k-1}{m-1} p^m q^{k-m}$	$0 \leq p = 1 - q \leq 1; m \geq 1$	$k = m, m+1, \dots$	$\left[\frac{p \cdot e^t}{1 - q \cdot e^t} \right]^m$	$\frac{m}{p}$	$\frac{mq}{p^2}$	Number of trials till m -th success
Geometric	$X \sim G(p)$	$p(k) = q^{k-1} \cdot p$	$0 \leq p \leq 1$	$k = 1, 2, \dots$	$\frac{p \cdot e^t}{1 - q \cdot e^t}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$X \sim NB(1, p)$
Poisson	$X \sim \mathcal{P}(\lambda)$	$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda \geq 0$	$k = 0, 1, \dots$	$\exp\{-\lambda(1 - e^t)\}$	λ	λ	
Hyper-geometric	$X \sim HG(N, M, n)$	$p(k) = \frac{\binom{N}{k} \binom{M-k}{n-k}}{\binom{N+M}{n}}$	$N, M > 0;$ $1 \leq n \leq N + M$	$\max(0, n - M) \leq k$ $k \leq \min(n, N)$	$\sum_k e^{tk} p(k)$	$\frac{n \cdot N}{N+M}$	$n \cdot \frac{N}{N+M} \cdot \frac{N}{N+M} \cdot \frac{M}{N+M} \cdot \frac{M}{N+M} - \left(\frac{n \cdot N}{N+M}\right)^2$	N - special elements M - ordinary ones n - sample size k - special in sample
		Density Function						
Beta	$X \sim \text{Beta}(\alpha, \beta)$	$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1} (1-t)^{\beta-1}$	$\alpha, \beta > 0$	$0 \leq t \leq 1$	Complicated	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha \cdot \beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\alpha = \beta = 1 \rightarrow \text{Uniform}$
Gamma	$X \sim \text{Gamma}(\alpha, \lambda)$	$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}$	$\alpha, \lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-\alpha}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ $= (x-1)\Gamma(x-1)$
Erlang	$X \sim \text{Erlang}(n, \lambda)$	$f(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$	$n = 1, 2, \dots; \lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-n}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$X \sim \text{Gamma}(n, \lambda)$
Exponential	$X \sim \text{Exp}(\lambda)$	$f(t) = \lambda e^{-\lambda t}$	$\lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-1}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$X \sim \text{Gamma}(1, \lambda)$
Normal	$X \sim N(\mu, \sigma^2)$	$f(t) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}$	$-\infty < \mu < \infty; \sigma > 0$	$-\infty < t < \infty$	$\exp\left\{t\mu + \frac{1}{2}t^2\sigma^2\right\}$	μ	σ^2	