

חקר ביצועים של רשתות PERT/CPM סטוכסטיות דינמיות

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

מגיסטר למדעים

בהנדסת תעשייה

יונית ברון

הוגש לסנט הטכניון-מכון טכנולוגי לישראל

ינואר 1996

חיפה

שבט תשנ"ז

המחקר נעשה בהנחיית פרופסור אבישי מנדלבאום בפקולטה להנדסת תעשייה וניהול.

תודתי לפרופסור אבישי מנדלבאום על עזרתו במהלך כל שלבי העבודה.

תודה מיוחדת להורי, בעלי, בתי ומשפחתי על תמיכתם במהלך כל תקופת לימודי.

תוכן העניינים

1	תקציר.
4	1 רשת PERT סטוכסטית דינמית.
4	1.1 סקר ספרות.
27	1.2 הצגת המודל.
27	1.2.1 תאור הפרויקט.
29	1.2.2 תאור המערכת.
31	1.2.3 דוגמאות.
33	1.3 מושגים להבהרה.
33	1.3.1 מסלול קריטי.
36	1.3.2 פעילות קריטית.
38	1.3.3 פרופיל שהיה.
43	1.4 דוגמת סימולציה.
43	1.4.1 תאור המודל.
44	1.4.2 מערכת הומוגנית.
44	1.4.2.1 זמני שרות קבועים.
44	1.4.2.2 זמני שרות לא קבועים.
47	שיטה א'.
52	שיטה ב'.
57	1.4.2.3 מופע לא קבוע.
60	1.4.3 מערכת לא הומוגנית.
63	2 מערכת טורים במקביל.
63	2.1 תאור המערכת.
63	2.1.1 המודל.
64	2.1.2 Lindley משוואות.

תוכן עניינים-המשך

65	2.1.3 הרחבת המשוואות.
66	2.2 תוצאות אנליטיות.
66	2.2.1 הגדרת משתנים מקריים קשורים.
67	2.2.1.1 תכונות מ"מ קשורים.
67	2.2.1.2 הוכחת על קשירות זמני שהיה.
69	2.2.1.3 תוצאות ומסקנות.
72	2.2.2 מופע דטרמיניסטי.
73	2.3 חסמים.
73	2.3.1 מופע פואסון.
75	מערכת n גדול וקבוע.
79	2.3.2 מופע דטרמיניסטי.
80	2.4 קרובים בעזרת שימוש בסימולציות ובחסמים.
80	2.4.1 מספר טורים ואורך טור וקטן.
85	2.4.2 טורים ארוכים.
85	2.4.2.1 השפעת הסנכרון ההתחלתי.
88	2.4.2.2 השפעת הסנכרון הסופי.
91	2.4.3 מספר טורים גדול.
93	2.4.4 מספר טורים ואורך טור גדולים.
94	2.4.5 מופע דטרמיניסטי.
97	2.5 סיכום, תוצאות ומסקנות עיקריות.
99	3 סיכום והרחבות.
101	4 רשימת המקורות.
105	5 נספחים.
105	טבלאות.

רשימת הטבלאות:

מספר	שם הטבלה	עמוד
1.3.1 א	קריטיות המסלולים עבור מערכת דינמית.	35
1.3.1 ב	קריטיות המסלולים עבור מערכת סטוכסטית.	35
3.2. א	קריטיות משימות.	37
1.3.2 ב	ממוצע אורכי תורים.	38
5.1.1-5.1.7	השוואת הקרובים לתוצאות הסימולציה.	104

רשימת שרטוטים:

מספר	שם	עמוד
1.1 א	רשת פרויקט הכוללת תורי סנכרון.	4
1.1 ב	מערכת סטוכסטית דינמית.	17
1.2.1 א	מערכת סטוכסטית.	27
1.2.2 א	מערכת סטוכסטית דינמית.	29
1.2.3 א	רשת PERT עבור תהליך המעצר.	31
1.2.3 ב	רשת PERT.	32
1.2.3 ג	רשת תהליכית עבור חברת חשמל.	32
1.3.3 א	רשת טורית.	39
1.3.3 ב	רשת מקבילית.	40
1.3.3 ג	רשת Gantt למודל הדינמי.	42
1.4.1 א	רשת PERT.	43
1.4.1 ב	המערכת הדינמית.	44
2.1.1 א	רשת ה-PERT.	63

רשימת גרפים:

מספר	שם	עמוד
1.1 א	מודל סטוכסטי סטטי, זמני משימות מפולגות מעריכי.	9
1.1 ב	מודל סטוכסטי דינמי.	17
1.4.2.1 א	זמן שהית פרויקט במערכת הומוגנית כפונקציה של הנצילות.	44
1.4.2.2 א	זמן שהית פרויקט במערכת כפונקציה של נצילות ומקדמי השתנות.	45
1.4.2.2 ב	זמן שהית פרויקט במערכת כפונקציה של נצילות ומקדמי השתנות.	45
1.4.2.2 ג	אחוז זמן ההמתנה כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה א'.	47
1.4.2.2 ד	אחוז זמן הסנכרון כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה א'.	48
1.4.2.2 ה	חלוקת הזמן במקדמי השתנות שונים כפונקציה של הנצילות, ע"פ שיטה א'.	50
1.4.2.2 ו	אחוז זמן ההמתנה כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה ב'.	52
1.4.2.2 ז	אחוז זמן הסנכרון כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה ב'.	53
1.4.2.2 ח	חלוקת הזמן במקדמי השתנות שונים כפונקציה של הנצילות, ע"פ שיטה ב'.	55
1.4.2.3 א	זמן שהיה ממוצע של פרויקט כפונקציה של מקדמי השתנות השרות והמופע.	57
1.4.2.3 ב	אחוז זמן ההמתנה כפונקציה של מקדמי ההשתנות, ע"פ שיטה ב'.	58
1.4.2.3 ג	אחוז זמן הסנכרון כפונקציה של מקדמי ההשתנות, ע"פ שיטה ב'.	59
1.4.3 א	זמן שהיה ממוצע במערכת לא הומוגנית.	62
2.4.1 א	חסמים תחתונים עבור n, m קטנים.	81
2.4.1 ב	תוחלת זמן שהיה במערכת m, n קטנים, החסם העליון והתחתון.	82
2.4.1 ג	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של אורך הטור.	82
2.4.1 ד	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של מספר הטורים.	82
2.4.1 ה	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של הנצילות.	83
2.4.1 ו	השוואת צפיפויות ברשת 2×2	84

מספר	שם	עמוד
2.4.1 ז	השוואת צפיפויות ברשת 2×8 .	84
2.4.2 א	זמן שהיה במערכת, החסם העליון והתחתון.	85
2.4.2 ב	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של מספר הטורים.	86
2.4.2 ג	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של אורך הטור.	86
2.4.2 ד	היחס $\bar{T}(n, m) / \bar{T}^*(n, m)$ כפונקציה של הנצילות.	86
2.4.2 ה	היחס $\bar{T}(n, m) / \bar{T}(1, m)$ כפונקציה של מספר הטורים.	88
2.4.2 ו	היחס $\bar{T}(n, m) / \bar{T}(1, m)$ כפונקציה של אורך הטור.	88
2.4.2 ז	היחס $\bar{T}(n, m) / \bar{T}(1, m)$ כפונקציה של הנצילות.	88
2.4.2 ח	תוחלת זמן שהיה כפי שמוערך ע"י הסימולציה.	90
2.4.3 א	זמן שהיה אופייני במערכת, והחסמים המתאימים.	91
2.4.3 ב	ההפרש המנורמל כפונקציה של הנצילות.	92
2.4.4 א	תוחלת זמן שהיה כפי שמוערך ע"י הסימולציה.	93
2.4.5 א	תוחלת זמן שהיה במערכות כפונקציה של מספר הטורים.	94
2.4.5 ב	תוחלת זמן שהיה במערכות כפונקציה של אורך הטור.	94

תקציר.

הגישות המקובלות לניהול, תכנון ובקרת פרויקטים מניחות, בד"כ, שפרויקטים עוברים במערכת באופן נפרד וב"ת זה בזה. ההתייחסות לפרויקטים בגישות אלה היא בד"כ דטרמיניסטית ותוצאת ניתוח של פרויקט בודד מהווה תחזית "ודאית" של מועד סיומו. במציאות, לעומת זאת, מטופלים בו זמנית מספר רב של פרויקטים המאופיינים באי-ודאות, וכך נוצרת מערכת דינמית. במערכת זו מתחרים פרויקטים על קבלת שרות מקבוצות משאבים מוגבלים, וכתוצאה מכך נוצרים תורים של פעילויות, השייכות לפרויקטים שונים. במחקר הנוכחי נחלק את התורים לשני סוגים:

- תורי המתנה למשאבים, עקב משאבים מוגבלים.
 - תורי סנכרון, הנוצרים עקב אילוצי קדימויות ואי-תלות בין הפעילויות.
- במחקר נבדוק השפעת פרמטרים שונים (כגון: נצילות, מקדמי השתנות וכד') על פרופיל השהיה של פרויקט במערכת.
- העבודה כוללת שני פרקים עיקריים:
- הפרק הראשון פותח בסקירת ספרות, בה אנו בונים באופן הדרגתי את המודל הדינמי. המודל הבסיסי הוא המודל הדטרמיניסטי. בתוספת התפלגויות לזמני השרות אנו מקבלים מודל סטוכסטי סטטי, ולבסוף ע"י הוספת מופע פרויקטים למערכת אנו מגיעים למודל הדינמי. לצורך המחקר הגדרנו מחדש מושגים מקובלים, כגון: מסלול קריטי, משימה קריטית וכד'. בנוסף, החלטנו כי ניתן לחלק את פרופיל השהיה של פרויקט במערכת לשלושה סוגי זמנים: זמן עיבוד, זמן המתנה וזמן סנכרון.
- הצענו שתי שיטות להגדרת זמנים אלו:
- שיטה א': שחרור המערכת מאילוצי משאבים וסנכרון תוך מדידת הרווח ברמת הביצועים.
- שיטה זו כוללת ארבעה שלבים הניתנים לביצוע בסדר שרירותי.
- שיטה ב': התייחסות לכל משימה בנפרד והגדרת מדד ביצועים השווה לסכום זמני שהיית המשימות בתחנות.
- בסיום הפרק הדגמנו את יישום השטות הנ"ל ובחינת פרמטרים נוספים במערכת קטנה, אך מעניינת מספיק, וכך ספקנו מסגרת לחקירה אמפירית המבוססת על סימולציות של רשתות תורים.

המסקנה העיקרית מהדוגמא היא, שזמן המתנה למשאבים מהווה חלק ארי מהזמן הכולל ע"פ שתי השיטות. לעומת זאת אחוז זמן הסנכרון נמוך יחסית.

שיטה א' מראה כי חלק מזמן הסנכרון נגרם ע"י ההמתנה, ולכן צמצום זמן ההמתנה יקטין באופן אוטומטי את זמן הסנכרון. על מנת לקצר את זמן ההשהיה במערכת נשקיע בצמצום זמן ההמתנה ע"י הורדת נצילות או הקטנת מקדמי ההשתנות (במופע ושרות). (אחת הדרכים להוריד את הנצילות היא הגדלת המשאבים).

מסקנה נוספת היא שהשתנות המופע אינה משפיעה, כמעט, על אחוז זמן הסנכרון בין המשימות. לעומת זאת על אחוז זמן ההמתנה משפיעים גם השתנות המופע וגם השרות.

ניתוח אנליטי של מערכת כללית קשה ובדרך כלל בלתי אפשרי, ולכן בפרק השני התמקדנו במערכת פשוטה וסמטרית. המערכת כוללת $m \times n$ משאבים, כל אחד מוקדש לפעילות מסוימת. הפרויקט מתפצל ל- n קבוצות של m משימות כ"א, כאשר כל אחת מ- n הקבוצות מתבצעת במקביל.

זמני המשימות מפולגים מעריכית, ומופע הפרויקטים למערכת מהווה תהליך חידוש.

בנינו שלושה חסמים לתוחלת זמן ההשהיה של פרויקט במערכת:

- חסם עליון המתקבל ע"י מערכת בעלת n מופעים ב"ת של פרויקטים.
- חסם תחתון המתקבל ע"י מערכת ללא מופע פרויקטים (מערכת סטוכסטית סטטית).
- חסם תחתון, המתקבל ע"י מערכת בעלת קבוצה אחת ($m=1$).

בעזרת שימוש בחסמים הנ"ל ובסימולציות קיבלנו קרובים אנליטיים לתוחלת זמן שהיה של

פרויקט במערכת, עבור מקרי גבול של הפרמטרים m, n :

• $n \rightarrow \infty, m$ קבוע.

• $m \rightarrow \infty, n$ קבוע.

• $m \rightarrow \infty, n \rightarrow \infty$.

בכל מקרה התאמנו נוסחאות אנליטיות מקורבות עבור מופע פואסוני ומופע דטרמיניסטי.

כמו כן, בדקנו את השפעת הפרמטרים השונים על התוחלת, כגון: נצילות המשאבים, גודל

הקבוצה (m), מספר הקבוצות (n) והשתנות המופע.

הבחנו בין שני סוגי סנכרון הקיימים במערכת:

- סנכרון התחלתי, הנוצר עקב מופע משותף של פרויקטים למערכת.
- סנכרון סופי, הנוצר עקב האילוץ שפרויקט מסיים את עיבודו רק בסיום עיבוד כל משימותיו.

כדוגמא למסקנות, אנו ממליצים להשתמש בקרוב הבא לתוחלת זמן שהיה של פרויקט במערכת מקבילית בעלת מופע פואסוני (עבור כל $m > 10$, ולכל n):

$$\bar{T}(n, m) = \frac{m}{\mu - \lambda} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right)$$

$$\frac{\bar{T}(n, m)}{1/\mu} = \frac{m}{1 - \rho} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right)$$

כאשר:

λ - קצב מופע הפרויקטים.

$1/\mu$ - ממוצע משך ביצוע משימה.

ρ - נצילות משאב.

קרוב זה נותן, על פי רוב, טעות של עד 5% מהתוחלת כפי שחושבה ע"י הסימולציות.

מהקרוב ניתן לראות, כי תוחלת זמן שהיה של פרויקט גדל בקצב ליניארי כפונקציה של m

ובקצב לוגריתמי כפונקציה של n . אם נשייך את מספר המשימות בקבוצה (m) לתורי המתנה

למשאבים, ואת מספר הקבוצות (n) לתורי סנכרון, הרי שמשאבים מוגבלים משפיע באופן ליניארי,

בעוד שפער הסנכרון משפיע באופן לוגריתמי. לכן, כדי לשפר ביצועי מערכת כדאי להשקיע יותר

במשאבים מוגבלים מאשר בהקטנת הסנכרון.

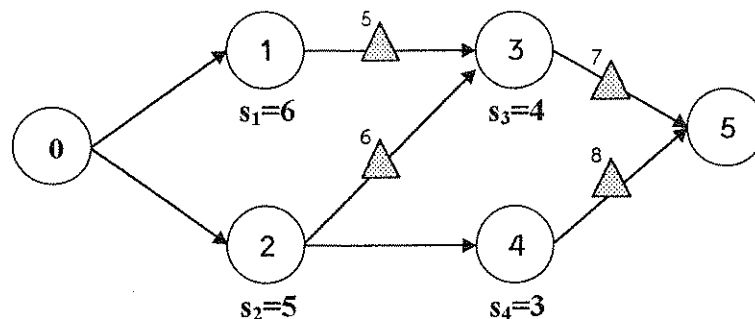
1.1 סקר ספרות.

ניהול פרויקטים בהנדסת תעשייה הוא נושא נרחב ועשיר. בד"כ, בכל הספרות העוסקת בניהול פרויקטים, הפרויקט מורכב ממשימות בלתי תלויות, ומקדימויות ביניהם אשר מתוארות כרשת בה הקשתות מייצגות קדימויות. כל משימה צורכת משאב (שרות) מסוים (לרוב משאבים שונים עבור משימות שונות). זמני ביצוע כל משימה יכול להיות קבוע וידוע (דטרמיניסטי) או מפולג ע"פ התפלגות כלשהי (סטוכסטי), ושוני זה משפיע רבות על מאפייני הפרויקט וניהולו. נתחיל להתמקד במערכת הדטרמיניסטית, בה זמני המשימות קבועים וידועים.

מערכת דטרמיניסטית:

דוגמא של מערכת ניתן לראות בשרטוט 1.1 א. המערכת מכילה 4 תחנות שרות, כאשר בכל תחנה מתבצעת משימה שונה. פרויקט המגיע למערכת, מורכב מ- 4 משימות שונות, משימה בכל תחנה. קיימות קדימויות בין ביצוע המשימות, ונתארן בעזרת רשת AON (Activity On Node). זמן ביצוע משימה i הוא s_i . לשם נוחות, הוספנו שתי תחנות דמה: תחנה 0- ממנה הפרויקט מתפצל למשימותיו, ותחנה 5- בה הפרויקט מסיים את עיבודו. מכיון שזמני המשימות אינם קבועים, נוצרים תורים הנקראים תורי סנכרון (הפרויקט מחכה עד שאילוצי הקדימויות יתמלאו).

▲ - תורי סנכרון



שרטוט 1.1 א: רשת פרויקט הכוללת תורי סנכרון.
Figure 1.1 a: Project Network with Synchronization Queues.

בדוגמא זו זמן ביצוע הפרויקט הוא 10 יח"ז, והמסלול הקריטי הוא 0-1-3-5.

בתור 6 מחכים 1 יח"ז, ובתור 8 - 2 יח"ז.

מאמרים רבים עוסקים ברשת דטרמיניסטית המתארת את שלבי ביצוע הפרויקט והקדימויות בין הפעילויות השונות.

בשנות ה-60 עסקו המאמרים בעיקר בבעיה של מינימום משך הפרויקט או מינימום עלות, ללא התחשבות בגורם הזמן.

בשנים האחרונות נכנס לפונקציה המטרה גם הגורם של עלות המימון תוך התחשבות בגורם הזמן. עבור פרויקטים הנמשכים זמן רב, ההשקעה מוחזרת בסיום הפרויקט, ולכן חשוב לקחת בחשבון גורם הזמן.

רוב המאמרים מתייחסים לשני סוגי תשלומים הנעשים במשך ביצוע הפרויקט (Cash Flows):

• זרימת תשלומים חיובית (Cash Inflows): תשלומים על ביצוע המשימות. התשלומים יכולים להינתן

בסיום ביצוע המשימות (Progress Payment) או בסוף ביצוע הפרויקט.

• זרימת תשלומים שלילית (Cash Outflows): עלות מלאי, הזמנת חומרים ועלות עונש על סיום מאוחר של

הפרויקט.

ע"מ ליצור ערך אחיד לכל התשלומים נהוג להוון אותם לפי נקודת זמן משותפת (זמן סיום הפרויקט או תחילתו).

כאמור, בתחילה התייחסה פונקציה המטרה התייחסה רק לקיצור זמן הפרויקט, אך עתה הוכנס גם הערך הכספי של הזרימות והיוונום לנקודת זמן משותפת, ולכן המטרה הפכה למקסימום הרווח, או מקסימום ערך הפרויקט.

לדוגמא, מאמרם של **[13] Patterson, Doersch** מתאר את הבעיה בעזרת תכנות בינארי, שפונקציה

המטרה שלה שואפת למקסימום ערך התשלומים (המהווה לזמן תחילת הפרויקט).

הפונקציה מכילה שני גורמים: תשלומי הפעילויות ועלות עונש עבור איחור זמן סיום הפרויקט,

והאילווצים כוללים אילווצי זמן סיום הפרויקט, אילווצי קדימויות ואילווצי הון בכל תקופה ותקופה.

ב- 1980, פורסם מאמרם של **[2] Aquilano, Smith**, אשר מתייחס לאילוצי משאבים.

המאמר מאחד את שתי השיטות המוכרות אז: שיטת המסלול הקריטי, CPM (Critical Path Method),

עם שיטת תכנון דרישות וחומרים, MRP (Material Requirements Planning).

שני המודלים ליניאריים ושניהם מתכננים לוח זמנים ע"ס יחסי קדימות:

• CPM - תכנון פעילויות, ציוד וכוח עבודה.

• MRP - תכנון דרישות לחומרים.

כהמשך למאמר הזה, מוסיפים **Smith-Daniels, Smith-Daniels [31]** את אילוצי ההון, עלויות המלאי

והזמנות הנעשות במשך ביצוע הפרויקט וגורם ההיוון.

הנחה חדשה נוספת, היא שכל התשלומים המתקבלים (החיוביים) מושקעים חזרה בפרויקט בתקופה

לאחר שהתקבלו. הנחה זו משמעותית ביותר, כיוון שעתה כמות משאב ההון תלוי בשיבוץ המשימות

ובסיומן.

אילוצי המודל מתייחסים לנקודות הבאות:

• אילוצי הון.

• אילוצי קדימות בין המשימות. (קיימים בכל המאמרים).

• אילוצי משאבים.

• עלות הזמנת חומרים והמלאי.

המאמר מראה, כי אי התחשבות בעלות הזמנת החומרים ועלויות המלאי יכול לעוות את התוצאה הנכונה.

הוספת העלויות הנ"ל יכולה לגרום לשינוי שיבוץ המשימות, כך ששתי משימות עבורן נדרש חומר יהיו

קשורות באמצעות הזמנה משותפת (כתוצאה מעלויות הזמנת החומרים), וכן משימות ידחו כתוצאה מהון

חסר (בגלל עלויות מלאי).

במאמרים שתוארו עד כה נקבעה ההנחה כי זמני המשימות קבועים, והבעיה, עקב כך, נהפכה לבעיית

שיבוץ (מתי להתחיל כל משימה בהתאם לאילוצים) ולא בכמה זמן לגמור כל משימה.

הרעיון של קיצור משימות (בתוספת תשלום, כמובן) הובא לראשונה במודל **CPM - Critical Path**

.Method

המודל מתאר רשת דטרמיניסטית, בה ניתן לקצר את זמן המשימה ע"י תוספת עלות עד שמגיעים לזמן המינימלי, הנקרא Crash Time (שבו העלות מקסימלית).

המודל מניח שני סוגי עלויות:

- עלות ישירה - עלות המשימות (העולה ככול שהמשימה מתקצרת).

- עלות עקיפה - עלות הקשורה לפרויקט (יורדת ככל שזמן הפרויקט מתקצר).

בגלל שתי העלויות הנ"ל נוצרת סתירה (Time - Cost Trade Off) וקיימת נקודת אופטימום, בה העלות מינימלית.

המודל מציג אלגוריתם למציאת אותה נקודה אופטימלית (כלומר משכי המשימות המביאים למינימום עלות) וכן מודל תכנות ליניארי המתאר את הבעיה.

ב- 1985, מביאים [29] Rosenblatt, Roll מאמר אשר גם הוא עוסק בנושא הערך העתידי של פרויקט.

המאמר קובע את זמן ביצוע הפרויקט, כאשר ניתן לצמצם את משך ביצועו.

חישוב הערך המהוון של העלויות בזמן סיום הפרויקט קובע את אורך זמן הפרויקט האופטימלי.

במאמרים קודמים למאמר זה (אותם לא ציינתי), בהם ניתנת האפשרות לקצר את זמן הפרויקט, הוצעו פרוצדורות אופטימליות והיוריסטיות לקיצור הפרויקט אבל ללא התחשבות בגורם הזמן.

מאמרים אשר התחשבו בזמן (כגון Patterson, Doersch, James) קובעים זמן אופטימלי לביצוע

הפעילויות, בהינתן זמני הפעילויות ועלותם מבלי לקחת בחשבון את אפשרויות קיצור הפעילויות.

כפי שניסיתי להראות, מאמרים רבים עוסקים ברשת דטרמיניסטית, ומתייחסים למגוון אילוצים: אילוצי

קדימות, אילוצי הון ואילוצי משאבים (חומרים, כוח-עבודה וציוד מוגבל).

פונקצית המטרה התחילה כמינימום משך הפרויקט, ובמשך הזמן עברה למקסימום הרווחים, בהתייחס

לעלויות ולתשלומים במשך הפרויקט ובסופו (עלויות הזמנת חומרים, עלויות מלאי, עלות עונש על סיום

מאוחר של הפרויקט) והיוון העלויות (בד"כ לתחילת ביצוע הפרויקט - NPV).

על האפשרות של קיצור פעילויות, מעטים המאמרים העוסקים בנושא.

בד"כ ארגונים לא מטפלים בפרויקט אחד בנפרד אלא קיים מצב בו מתבצעים מספר פרויקטים במקביל.

ולכן טבעי להרחיב את המודל הדטרמיניסטי ע"י הוספת פרויקטים במקביל המשתמשים באותם משאבים.

רשת זו נקראת מערכת מרובת-פרויקטים (Multi Project Scheduling).

מספר מאמרים עוסקים ברשתות מסוג זה, ואביא לדוגמא את מאמרם של Kurtulus, Narula מ-1985 [20].

המאמר מדבר על מערכת בעלת M פרויקטים בסה"כ, לאו דווקא זהים. כל פרויקט מכיל משימות ומערכת קדימויות. זמני המשימות, וכמות המשאבים הדרושה קבועה וידועה. כמו כן, קבועה גם כמות המשאבים הנתונה בתקופה, ואין הגבלה על מספר התקופות. המאמר עוסק במקרה בו עלויות אחור סיום פרויקטים שונות זו מזו ומציג מספר חוקי שיבוץ חדשים, בנוסף לחוקים הקיימים.

במציאות, בד"כ זמני משימות המרכיבות את הפרויקט אינם קבועים (דטרמיניסטים) וידועים מראש, אלא נוצרים ע"פ התפלגות כלשהי, ואז מושג של מסלול קריטי, והתפלגות משך הפרויקט מקבלים משמעות חדשה.

מערכת סטוכסטית סטטית:

אם ניקח את הרשת משרטוט 1.1 א, אך הפעם זמני המשימות יהיו מפולגות למשל מעריכי עם ממוצע s_i נקבל רשת סטוכסטית.

ישנן 3 אפשרויות למסלולים קריטיים: 0-1-3-5, 0-2-3-5, 0-2-4-5.

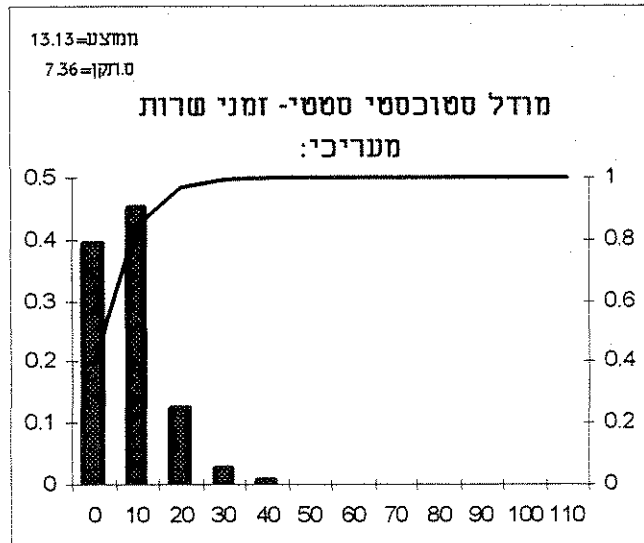
ע"פ ההגדרה המקובלת בספרות, המסלול הקריטי במערכת הוא המסלול בעל התוחלת הגבוהה ביותר של משך בצוע המשימות לאורכו (מסלול 0-1-3-5).

ערכנו סימולציות על המערכת הסטוכסטית: 20 חזרות, 10,000 פרויקטים בכל חזרה.

להלן התפלגות משך הפרויקט:

ממוצע משך הפרויקט: 13.13 ± 0.095 יח"ז (רווח סמך ברמת 95%).

ס.ת.קן: 7.36.



גרף 1.1 א : מודל סטוכסטי סטטי, זמני משימות מפולגות מעריכי.
Graph 1.1a: Static Stochastic Model, Task's Times Exponentially Distributed.

הספרות העוסקת במודל הסטוכסטי היא רחבה בהיקפה.

המודל הראשון שעסק בזמני פעילויות לא דטרמיניסטיות היה מודל ה-PERT.

המודל מבוסס על מספר הנחות סטטיסטיות:

- משך כל פעילות מפולג ע"פ התפלגות Beta, שצפיפותה נתונה ע"י :

$$f(t) = \frac{\Gamma(\alpha + \beta + 2)}{\Gamma(\alpha + 1)\Gamma(\beta + 1)} \frac{(t - t_o)^\alpha (t_p - t)^\beta}{(t_p - t_o)^{\alpha + \beta + 1}}$$

$$t_o < t < t_p \quad \alpha, \beta > -1$$

- לפעילות 3 הערכות למשך אופטימי (t_o), פסימי (t_p) וסביר (t_m) לקבלת קרוב "טוב" של הממוצע וסטית התקן של הפעילות.

$$t_e = (t_o + 4t_m + t_p) / 6 \quad \text{ממוצע הזמן הנצפה לפעילות הוא:}$$

$$\sigma_t = (t_p - t_o) / 6 \quad \text{וסטית התקן:}$$

- הפעילויות אינן תלויות.
- כל מסלול ברשת מפולג נורמלי והמסלולים ב"ת סטטיסטית.
- כדי לחשב את ממוצע אורך המסלול הקריטי, נעשה סכום אורכי הפעילויות שעל המסלולים השונים, והממוצע הארוך ביותר מוכרז כמסלול הקריטי, ולפיו יקבע משך הפרויקט (וכנ"ל לגבי שונות הפרויקט).

מאמרם של Golenko, Ginzburg ב-1989 [14] מראה כי שימוש בהנחות הנ"ל, שהן הנחות בלתי נכונות בד"כ, מעוות את התוצאות וגורם לשגיאות רבות בחישוב המסלול הקריטי. לכן מובאות מספר הצעות שיפור להנחות ה-PERT (חלקן הצעות ממאמרים קודמים) כאשר מטרת ההצעות היא שיפור הדיוק של ההערכות תוך שמירה על פשטות הניתוח.

דרך אנליטית מדויקת, אשר אינה מניחה התפלגות Beta למשימות מובאת במאמרם של Cooper, Thompson, Charnes ב-1964 [10] העוסק ברשת סטוכסטית סטטית. במאמרם הם משתמשים בתכנות עם אילוצים הסתברותיים (Chance Constrained Programming) כדי לבדוק את תכונות רשת ה-PERT, ובייחוד התפלגות משך הפרויקט. את רשת ה-PERT הם כותבים בשיטת AOA (Activity On Arrow), כאשר ישנם m צמתים, ו- n פעילויות. עבור רשת דטרמיניסטית, בעיית המסלול הקריטי יכולה להיכתב בעזרת התכנות הליניארי הבא:

$$\text{Max } \sum_{j=1}^n t_j x_j$$

$$\text{S.T. } \sum_{j=1}^n \varepsilon_{ij} x_j = a_i$$

$$x_j \geq 0, \quad (i=0, 1, \dots, m)$$

t_i - משך זמן משימה i .

$a_0 = -1$, $a_n = 1$ ו- $a_i = 0$, לכל האחרים.

ε_{ij} - מקבל שלושה ערכים: 1 אם צומת i הוא היעד של פעילות j .

-1 אם צומת i היא המקור של פעילות j .

0 - אחרת.

x_j - משתנה ההחלטה (0 או 1) המציין אם הפעילות j נמצאת על המסלול הקריטי.

כאשר מסתכלים על הבעיה הדואלית, מקבלים:

$$\text{Min } \sum_{i=0}^m u_i a_i$$

$$\text{S.T. } \sum_{i=0}^m u_i \varepsilon_{ij} \geq t_j$$

u_i - הזמן שהגענו לצומת i .

הערך $\sum_{i=0}^m u_i a_i$ נותן חסם עליון עבור משך הפרויקט.

נניח, עתה, כי במקום זמנים דטרמיניסטים של ביצוע הפעילויות (t_j) , מתוארים הזמנים ע"י מ"מ

מהתפלגות ידועה. ניתן לראות, ממאמרים קודמים, כי גם u_i שואפים למ"מ וניתן, בעיקרון, לחשב את

התפלגותם בעזרת התפלגות t_j .

במאמר מפתחים דרך למציאת u_i , $i=0, \dots, m$, כאשר נוספות הסתברויות למשך הפעילויות (β_j) . דרך זו

נקראת תכנות עם אילוצים הסתברותיים (Chance Constrained Programming):

$$\text{Min } \sum_{i=0}^m u_i a_i$$

$$\text{S.T. } P\left(\sum_{i=0}^m u_i \varepsilon_{ij} \geq t_j\right) \geq \beta_j$$

$\Downarrow$

$$\text{Min } \sum_{i=0}^m u_i a_i$$

$$\text{S.T. } \sum_{i=0}^m u_i \varepsilon_{ij} \geq F_j^{-1}(\beta_j)$$

מערכת זו שקולה למערכת הדטרמיניסטית.

התפלגות משך הפרויקט זהה להתפלגות המסלול הקריטי ולכן פתרון המערכת הנ"ל יכולה להיעשות

בדרך רקורסיבית, כאשר ישנם רק שני סוגי חישובים:

1. חישוב התפלגות $s+t$, סכום שני מ"מ (קונוולוציה של שתי ההתפלגויות).

2. חישוב התפלגות $\max(s,t)$, מקסימום של שני מ"מ (מכפלת שתי ההתפלגויות - בהנחה כי s, t ב"ת).

בדרך זו ניתן לחשב את התפלגות משך הפרויקט.

השיטה ב-[9] אנליטית מדויקת, אך קשה לחישוב עבור רשתות מורכבות.

דרך נוספת ניתן למצוא במאמרם של Signal, Pritsker ו- Solberg 1979 [30].

המאמר מתחשב בתלות בין המסלולים השונים (כתוצאה מפעילויות משותפות), בניגוד לשיטת ה- PERT

ומנסה להתגבר על תלות זו.

המאמר מציג את השימוש בגישת Monte- Carlo לחישוב התפלגות משך הפרויקט, (Conditional Monte Carlo) ומשפר את השיטה.

שיטת Conditional Monte Carlo:

עבור M מסלולים ברשת, התפלגות משך הפרויקט Y_c מוגדרת ע"י:

$$\Pr(Y_c \leq t) = \Pr(Y_1 \leq t, Y_2 \leq t, \dots, Y_M \leq t)$$

הבעיה היא שהמסלולים תלויים ולכן ע"פ הסתברות שלמה מתנים בכל הפעילויות ששייכות ליותר

$$F_{yc}(t) = \iint \Pr(Y_c \leq t \mid x_{i1}, x_{i2}, \dots, x_{in}) dF_{x_{i1}}(x_{i1}) \dots dF_{x_{in}}(x_{in}) \quad : (x_{i1} \dots x_{in})$$

וכך יוצרים מסלולים ב"ת.

את ההסתברות לפעילויות המותנות מקבלים מסימולציה.

המאמר מראה, ע"ס אותה שיטה, כי ניתן להתנות בקבוצת פעילויות אחרת, ובכך להקטין את מספר הסימולציות ולהקטין את השונות של התוצאה.

הם מגדירים Uniformly Directed Cutsets (UDCs) : קבוצת פעילויות כך שאין בקבוצה שתי

פעילויות ששייכות לאותו מסלול (אבל יתכן שפעילות בקבוצה תהיה שייכת לכמה מסלולים).

נבחר את קבוצה ה-UDC הגדולה ביותר, ואת ההתניה נעשה על קבוצת הפעילויות המשלימה שלה.

בצורה כזו מפחיתים את מספר הפעילויות, מספר הסימולציות, והשונות יורדת.

באופן דומה המאמר מציג אלגוריתמים למציאת אינדקס קריטיות של מסלול- ההסתברות שמסלול

מסוים ברשת יהיה הארוך ביותר, התפלגות הזמן המינימלי של הרשת ואינדקס אופטימלי - ההסתברות

שמסלול יהיה קצר משאר המסלולים.

עד עתה, אי-הודאות של זמני המשימות התבטאה בעזרת משתנים-מקריים, בשימוש בהנחות

סטטיסטיות חזקות ובגישות הסימולציה.

כאשר הפעילויות מפולגות מעריכי, ניתן לקבל בצורה מדויקת את התפלגות משך הפרויקט, כפי שמופיע

במאמרם של Kulkarni ו- Adlakha 1986 [19].

מאמרם עוסק ברשת PERT עם זמני פעילויות שהם משתנים מקריים בלתי תלויים המפולגים מעריכי.

המאמר מראה כיצד ניתן לתאר את רשת ה-PERT הסטוכסטית בעזרת שרשרת מרקוב רציפה בזמן

(Continuous Time Markov Chain - CTMC) בעלת מטריצת יוצר אינפיטיסימלי משולשת עליונה.

זמן סיום הפרויקט ברשת ה-PERT הוא הזמן עד הספיגה בשרשרת מרקוב.

מדדי הביצוע המרכזיים במאמר (כפי שכבר ראינו ממאמרים קודמים):

• זמן שהיה של פרויקט במערכת - התפלגות ומומנטים.

• אינדקס קריטיות של מסלולים - ההסתברות שמסלול מסוים ברשת יהיה הארוך ביותר.

• אינדקס קריטיות של פעילויות - ההסתברות שפעילות תשתייך למסלול הקריטי.

מבנה השרשרת מאפשר לפתח אלגוריתמים פשוטים, יחסית, לניתוח מדויק של המערכת.

עבור כל מדד מפותחים שני סוגי אלגוריתמים רקורסיביים: קדמי ואחורי.

לדוגמא, האלגוריתמים למציאת התפלגות זמן השהיה T של פרויקט במערכת, הם כדלהלן:

האלגוריתם האחורי (Backward Algorithm):

נגדיר: $P_i(t) = P\{X(t)=N / X(0)=i\}$; $0 \leq i \leq N$.

אזי $P(T \leq t) = F(t) = P_0(t)$ (N מצב סופג, 0 מצב התחלתי).

המשוואות הדיפרנציאליות עבור $P_i(t)$:

$$P_i'(t) = \sum_{j \geq i} q_{ij} p_j(t)$$

$$P_i(0) = \delta_{iN} \quad 0 \leq i \leq N$$

האלגוריתם הקדמי (Forward Algorithm):

נגדיר: $*P_i(t) = P\{X(t)=j / X(0)=0\}$; $0 \leq j \leq N$.

אזי $F(t) = *P_N(t)$.

המשוואות הדיפרנציאליות עבור $*P_i(t)$:

$$*P_j'(t) = \sum_{i \leq j} q_{ij} *P_i(t)$$

$$*P_j(0) = \delta_{j0} \quad 0 \leq j \leq N$$

במאמר מוסיפים דרכים לקרב רשת PERT שבה זמני הפעילויות אינם מעריכיים, אלא ב"ת מהתפלגות כלשהי, ע"י שרשרת סמי-מרקובית.

כמו כן, ישנה הרחבת השיטה עבור אילוצי משאבים.

הייחוד העיקרי בשיטה שהוצגה הוא שהאלגוריתמים פשוטים, וניתנים לתכנות בקלות.

ברשתות לא-מעריכיות עוסקים Kang, Serfozo [17]. מאמרם עוסק במערכת מקבילית (משימות

במקביל) ומוצא את התפלגות זמן סיום הפרויקט.

באופן כללי, התפלגות זמן פרויקט במערכת הומוגנית מקבילית, עם משימות ב"ת מאופיינת באופן הבא:

$$M_n = \max \{ X_1, \dots, X_n \}$$

המאמר מנתח את ההתפלגות הגבולית של המקסימום תוך שימוש בעובדה שעבור n קבוע וגדול, ניתן

$$P\{M_n \leq x\} \cong G(a_n x + b_n) \quad \text{להשתמש בקרוב:}$$

בהסתמך על מאמרים קודמים ידוע שההתפלגות הגבולית של המקסימום $G(x)$ יכולה להיות אחת

משלושת האפשרויות הבאות:

$$G(x) = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (\text{Gumbel distribution})$$

$$G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \text{ for some } \alpha > 0 \quad (\text{Type II distribution}) \end{cases}$$

$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x \leq 0, \text{ for some } \alpha > 0 \quad (\text{Type III distribution}) \\ 1 & x > 0 \end{cases}$$

בהסתמך על ההתפלגויות הנ"ל המאמר [16] מראה כי התפלגות זמן סיום הפרויקט היא Gumbel

(האפשרות הראשונה) עבור המקרים המעניינים הבאים ומוצא את הקבועים a_n, b_n :

- זמני המשימות ב"ת מפולגים ארלנג.
- זמני המשימות ב"ת כאשר כל משימה היא סכום של תת-משימות מפולגות ארלנג.
- זמני המשימות ב"ת כאשר כל משימה היא סכום של תת-משימות מפולגות נורמלי.
- זמני המשימות ב"ת מפולגים Phase-Type - התפלגות זמן ספיגה של שרשרת מרקוב, בזמן רציף.
- רשת כלשהיא בעלת משימות מפולגות מעריכי (ע"ס מאמרם של Kulkarni ו- Adlakha מ-1986).

תאור רשת PERT סטוכסטית בעזרת שרשרת מרקוב רציפה בזמן נעשה גם במאמרם של

Buss, Rosenblatt מ-1995 [9]. במאמרם הם מוסיפים את אלמנט השהיית הפעילויות, ע"מ לקבל

מקסימום ערך נוכחי של פרויקט, אך בניגוד למאמרים הראשונים הפעם מדובר על רשת סטוכסטית.

המאמר מנתח את המקרים הבאים:

1. רשת בעלת שתי משימות מעריכיות במקביל-

- תוחלת ערך נוכחי (Expected Present Value - EPV) של הפרויקט במקרה של השהיית משימה

בזמן דטרמיניסטי.

- ההשהיה האופטימלית.

2. רשת כללית (עדיין זמני המשימות מעריכים) -

- תוחלת הערך הנוכחי של הפרויקט בלי השהיות

- עם משימה אחת מושהית בזמן דטרמיניסטי (כך שזמן זמן המשימה כבר לא מעריכי).

- איזה משימה כדאי להשהות ובכמה (שיטה מקורבת).

המאמר מראה כי המשימות המועמדות להשהיה יכולות להימצא על המסלול הקריטי, בניגוד מוחלט

למערכת דטרמיניסטית.

ניתן לסכם ולומר כי ישנם מספר גדול של מאמרים העוסקים ברשת בעלת זמני משימות סטוכסטיות. רוב

המאמרים מנסים לקבל את התפלגות זמן השהיה של פרויקט במערכת ולהגדיר אינדקס קריטיות של

מסלולים ומשימות (המושג "מסלול קריטי" שמוגדר במדויק ברשת דטרמיניסטית, לא ברור ברשת

סטוכסטית וניתן להגדירו במספר אופנים).

בד"כ מציאת התפלגות זו ואינדקס הקריטיות, באופן כללי, קשה ואומנם ישנם אלגוריתמים המטפלים

ברשתות כלליות אבל הם קשים ליישום (למשל Charnes, Cooper, Thompson).

לכן מוסיפים הנחות המפשטות את הניתוח כגון:

- התחשבות במסלולים ב"ת (מודל PERT ו-Solberg, Signal Prisker).

- זמני משימות מפולגות מעריכי וכך מעבר לרשת מרקוב (Adlakha, Kulkarni ו-Buss, Rosenblatt).

- מציאת התפלגות גבולית של רשתות מקביליות (Kang, Serfozo).

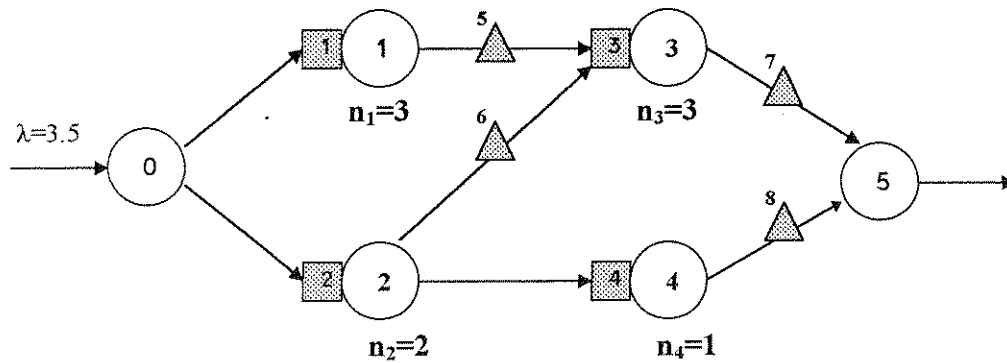
במאמרים הנ"ל לא מצאנו מושגים כמו "קיצור פעילויות", ו"אילוצי משאבים" שאותם מצאנו בשפע במערכות דטרמיניסטיות.

עד עתה עסקנו במערכת סטוכסטית בפרויקט יחיד, המגיע למערכת וכל המשאבים פנויים עבורו, כלומר אין תורי המתנה למשאבים הנוצרים כתוצאה מפרויקטים קודמים שעוד לא סיימו. לכן, במקביל למערכת מרובת פרויקטים עם זמנים דטרמיניסטים, נסיף עתה מופע פרויקטים (הזמן בינומפעי בד"כ נלקח מהתפלגות נתונה).

מערכת כזאת נקראת **מערכת סטוכסטית דינמית או רשת תורים** (Queueing Network). פה קשה הרבה יותר לנתח את ביצועי המערכת, כיוון שבנוסף להמתנות הרגילות, נוצרים תורי המתנה כתוצאה ממופע אקראי של פרויקטים למערכת. גם מושגים כמו מסלול קריטי ואינדקס קריטיות מקבלים משמעות בלתי ברורה כיוון שכל פרויקט, בהגיעו, רואה מערכת עם מספר שונה של פרויקטים. בשרטוט 1.2 מובאת המערכת שהוצגה בשרטוט 1.1, אך הפעם (בנוסף לזמני המשימות המפולגות מעריכי) נוסף גם מופע פרויקטים. הזמן הבינומפעי מפולג מעריכי עם ממוצע 3.5 יח"ז. בנוסף, לשם הדוגמא, הוספנו את מספר השרתים (יחידות המשאב) הנמצאים בכל תחנת שרות - n_i . נוצרים שני סוגי תורים: תורי משאבים ותורי סנכרון (עליהם נפרט בהמשך). להלן המערכת הדינמית:

■ - תור המתנה למשאבים.

▲ - תור סנכרון.



שרטוט 1.1 ב: מערכת סטוכסטית דינמית.
Figure 1.1b: Dynamic Stochastic Model.

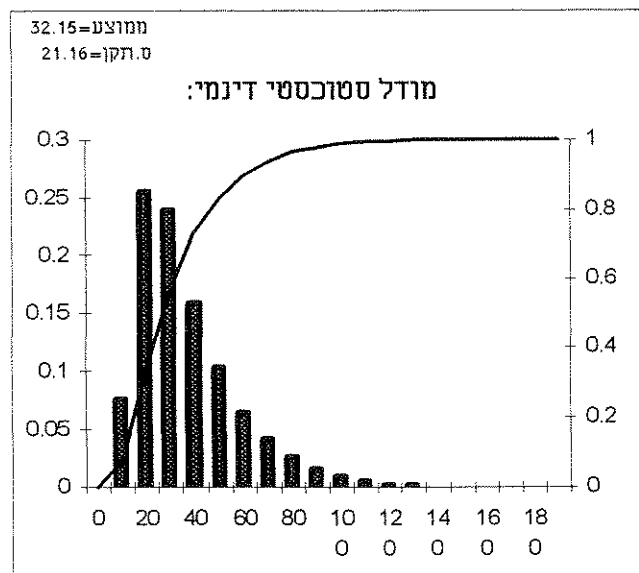
נערכו סימולציות על המערכת הדינמית: 20 חזרות, 20000 יח"ז כל חזרה. סטטיסטיקות הושמטו מ-

50000 יח"ז הראשונות.

להלן התפלגות משך הפרויקט:

ממוצע 32.15 ± 1.5 יח"ז (רווח בר סמך של 95%).

ס.תקן = 21.16 .



גרף 1.1 ב: מודל סטוכסטית דינמי.
Graph 1.1b: Dynamic Stochastic Model.

ברור כי דינמיות המערכת מעלה בהרבה את ממוצע זמן השהיה (תוחלת זמן השהיה במערכת ניתנת כמובן לשינוי בהתאם למופע הפרויקטים).

קיימת ספרות ענפה העוסקת ברשתות תורים.

בספרות זו הוקדש רוב המחקר לרשתות פשוטות של משימות במקביל, כאשר הטיפול ברשת הוא נדיר ומתרכז בהיבטים איכותיים, כגון יציבות (Massey, Baccelli ו-Towsley [7]).

רשת פשוטה ובסיסית מתקבלת כאשר פרויקט מתפצל, עם הופעתו, למספר משימות ב"ת המעובדות במקביל ע"י שרתים הומוגניים. הפרויקט מסיים את עיבודו כאשר כל משימותיו סיימו את עיבודן.

רשת זו, הנקראת בספרות "Split And Match" או "Fork-Join", נדונה למשל ע"י

Mandelbaum ו-Avi-Itzhak [24], Baccelli, Makowski ו-Shwartz [6], Tantawi, Nelson ו-[25], ו-

Tantawi ו-Towsley, Nelson [26].

Mandelbaum ו-Avi-Itzhak [24] מניחים פיצול הפרויקט למספר משימות קבוע k , כאשר כל משימה

מעובדת במערכת תור-שרת. עבור מופע פרויקטים כללי וזמני שרות כלליים ($GI/G/1^k$ -split), התפלגות

זמן השהיה של פרויקט היא התפלגות מקסימום זמני שהיית המשימות $T_n^{(k)} = \max(w_{1,n}, w_{2,n}, \dots, w_{k,n})$,

כאשר $w_{i,n} = \eta_{i,n} + v_{i,n}$ זמן שהיה של משימה i בפרויקט ה- n .

המאמר מתבסס על נוסחאות Lindley: $\eta_{i,n+1} = [\eta_{i,n} + v_{i,n} - t_n]^+$

כאשר:

$\eta_{i,n} =$ זמן שהיה בתור של משימה i השייכת לפרויקט ה- n .

$v_{i,n} =$ זמן שרות למשימה i השייכת לפרויקט ה- n .

$t_n =$ זמן בין הופעת פרויקט ה- n לבין הופעת פרויקט ה- $n+1$.

המאמר מסיק, כי בד"כ זמני שהיית המשימות ($w_{i,n}$) תלויים עקב הזמן הבין מופעי המשותף. במקרה

פרטי יותר, במערכת $D/M/1^k$ -split, קיימת אי תלות בין זמני שהיית המשימות. במצב יציב,

כל $1 \leq i \leq K$ $w_{i,n}$ מעריכים (נובע מ- $D/M/1$). ולכן $T_n^{(k)}$, במצב יציב הוא מקסימום בין מעריכים.

ע"י ניתוח אנליטי המאמר מסיק כי ממוצע זמן השהיה של פרויקט הוא מכפלת שני איברים : ממוצע זמן השהיה במערכת $D/M/1$, וגורם תזמון- הטור ההרמוני הגדל כמו $\ln K$ $(1/b * \sum_{i=1}^k 1/i)$.

במאמרם של Nelson ו-Tantawi [25] מוצע שימוש בשיטת אומדן חדשה, הנקראת "Scaling Approximation", שבעזרתה מנותחת מערכת התורים ההומוגנית המכילה $K > 2$ שרתים, בעלת מופע פואסוני וזמני שרות מעריכים.

ע"י הזנחת השפעת התורים, מתקבלת מערכת סטטית בה זמני השהיה של פרויקט הוא מקסימום בין K משתנים מקריים, ב"ת, ש"ה מעריכים וכך מתקבל החסם התחתון $T_k \geq H_k * (1/\mu)$.

T_k - ממוצע זמן שהית הפרויקט המערכת המקורית.

H_k - הטור ההרמוני $(\sum_{i=1}^k 1/i)$.

μ - קצב השרות בכל תחנה.

חסם העליון מתקבל ע"י הנחת אי תלות בין K המערכות : $T_k \leq H_k * T_1$.

הוכחת החסם נעשית ע"י שימוש בתכונות "משתנים מקריים קשורים" (associated random variables),

ובייחוד באי שוויון : $P[\max_{1 \leq i \leq n} X_i > x] \leq 1 - \prod_{i=1}^n P(X_i \leq x)$.

מהות שיטת האומדן במאמר נובעת מהעובדה המעניינת שהחסם העליון והתחתון לממוצע זמן השהיה

של פרויקט במערכת, גדלים בקצב $O(\ln K)$, ולכן גם T_k חייב לגדול באותו קצב.

כך ע"י שימוש בתוצאה מדויקת עבור $K=2$, כתנאי התחלה, ושימוש בתוצאות אמפיריות, מתקבל ביטוי

מקורב לזמן השהיה הממוצע, במערכת עם $K > 2$:

$$T_k \cong \left[\frac{H_k}{H_2} + \frac{4}{11} \left(1 - \frac{H_k}{H_2} \right) \rho \right] T_2, \quad K \geq 2,$$

כאשר $\rho = \lambda/\mu$.

השפעת גורם הסנכרון כבדה, במיוחד כאשר הסנכרון נעשה לפני העיבוד. (במאמרים שנסקרו עד כה,

הסנכרון בין המשימות נעשה אחרי העיבוד).

מאמרו של **Harrison [16]** מתאר מערכת הרכבה, בעלת שרת אחד המרכיב פריט מ- k סוגי רכיבים שונים. k הסוגים באים ב- k מופעי קלט, המהווים תהליכי חידוש ב"ת. השרת מחכה עד שכל k הרכיבים יהיו זמינים ואז מרכיב את הפריט. זמני השרות (ההרכבה) הם מ"מ ב"ת ש"ה, וב"ת במופעי הקלט. השרת בבטלה כאשר לפחות בסוג אחד אין רכיבים. הסנכרון, כפי שתואר, נעשה במודל זה לפני העיבוד.

Harrison מגדיר W_n^k - זמן ההמתנה בתור של רכיב k -י בפריט ה- n , ומראה כי $W_n = [W_n^1, W_n^2, \dots, W_n^k]$ וקטור ההמתנה בתור אינו מתכנס בהתפלגות לגבול מסוים (כיוון שהמופעים ב"ת). כלומר, מחיר הסנכרון כבד במערכת זו וגורם לאי יציבותה, כיוון שיש k מופעי קלט ב"ת והשרת נאלץ לחכות עד שכל k הרכיבים יהיו זמינים, ועד אז תהליך הרכבת המוצר לא יכול להתחיל.

Courcoubetis, Simon, Reiman [3] ו-**Courcoubetis, Reiman, Baccelli [12]** עוסקים אף הם בסנכרון הנעשה לפני העיבוד.

שני המאמרים מנתחים רשת בעלת N שרתים ותורים, כאשר לרשת מגיעים שני סוגי משימות: משימות רגילות ("Ordinary" Customer) המעובדות ע"י כל שרת בנפרד, ומשימות סינכרון ("N-Tuple" Customer) המצריכות עיבוד בו-זמנית ע"י כל השרתים. קיימים N מופעי פואסון של משימות רגילות (לכל שרת בנפרד, וללא תלות בין המופעים), ומופע נוסף של משימות סנכרון, המהווה תהליך חידוש. ברגע שמשימת הסנכרון מגיעה לאחד השרתים, השרת בטל עד ששאר $N-1$ השרתים יקבלו את המשימה, ואז השרות נעשה במקביל.

המערכת מוגדרת יציבה (Stability) אם המאורע ($W^p=0$) הוא נשנה חיובי, כאשר W^p הוא זמן ההמתנה של משימת הסנכרון ה- p . במילים אחרות, תוחלת תקופת התעסוקה של המערכת סופית. לאחר ניתוח המערכת, מסקנת המאמר כי תנאי ליציבות הוא מהצורה $\gamma + \beta < \alpha$, כאשר α - תוחלת הזמן הבין-מופעי של משימות הסנכרון.

β - תוחלת המקסימום בין N השרתים של מספר המשימות הרגילות המגיעות בין שתי משימות סינכרון עוקבות.

γ - תוחלת זמן שרות של משימת סינכרון.

מציאת תנאי היציבות מתבססת על חסמים עליונים ותחתונים של מערכות G/G/1

במאמר נוסף בנושא, מנתחים Courcoubetis ו-Reiman [11] מערכת דומה, אך הפעם הם מניחים כי ישנן אינסוף משימות רגילות, הממתינות בתור משותף.

כמו כן מופע משימות הסנכרון פואסוני, וזמני השרות מעריכים.

הם מציעים שיפור המערכת: תור מיוחד למשימות הסנכרון, וכאשר מגיעים ל-K* משימות, שולחים את כולן לעיבוד.

המאמר מוסיף עלות המתנה ליח"ז של משימת הסנכרון, ורווח עבור כל משימה רגילה שעובדה מסתיים. ומוצא את ה-K* האופטימלי, עבורו מושג מקסימום רווח.

Baccelli, Makowski, Shwartz [6] ו-Baccelli, Makowski [4] ו-Baccelli, Makowski [5]

מוצאים חסמים עליונים ותחתונים פשוטים לסטטיסטיקות שונות במערכת Fork-Join הכוללים גם חסמים לתנאי יציבות המערכת.

במאמרים משנת 1987 משתמשים Baccelli, Makowski [4] ו-Baccelli, Massey, Towsley [7]

ב"משתנים מקריים קשורים" (Associated random variables) ובתכונותיהם. הם מוכיחים כי ניתן לחסום את זמן ההשהיה הממוצע של פרויקט במערכת בעלת k תחנות GI/G/1 במקביל, במערכת בעלת k מופעים ב"ת.

במאמרים אלו גם מופיעים תנאים הכרחיים ומספיקים ליציבות המערכת ומראים כי מופע דטרמיניסטי או זמני שרות דטרמיניסטיים מקטינים את זמן ההשהיה במערכת.

במאמר המשך, בשנת 1988, ממשיכים Baccelli, Makowski [5] לתת חסמים תחתונים ועליונים

למערכות Fork-Join נוספות. המאמר מפתח נוסחאות רקורסיביות למציאת זמנים שונים במערכת עבור כל רשת כללית, בינהן:

$$y_{n+1}^j = \max\{\max_{i \in P(j)} \{y_{n+1}^i + \sigma_{n+1}^i\}, y_n^j + \sigma_n^j - \tau_{n+1}\}$$

$$y_0^j = \max_{i \in P(j)} \{y_0^i + \sigma_0^i\}$$

$$r_n = \max_{i \in P(B+1)} \{y_n^i + \sigma_n^i\}$$

כאשר:

y_n^i - זמן ההמתנה בין מופע הפרויקט ה- n לבין תחילת שירותו בתחנה i .

σ_n^i - זמן שרות של פרויקט n בתחנה i .

τ_n - זמן בינומפעי בין פרויקט n לפרויקט $n+1$.

r_n - זמן שהיה של פרויקט n במערכת.

$P(j)$ - קבוצת המקדימות המיידיות של תחנה j .

בתחילת המאמר מסווגים המחברים את סוגי הסנכרון הקיימים במערכות ל- 5 קבוצות:

1. Arrival-arrival synchronization - זמני המופע לשתי תחנות מקיימים אילוצים מסוימים. לדוגמא כאשר זמני המופע לשתי תחנות זהה.
2. Departure-departure synchronization - זמני העזיבה משתי תחנות מקיימים אילוצים מסוימים. לדוגמא כאשר זמן העזיבה מתחנה k לא יכולה להיות לפני זמן העזיבה מתחנה j עבור כל פרויקט.
3. Service-service synchronization - זמני תחילת השרות בשתי התחנות מקיימים אילוצים מסוימים.
4. Departure-arrival synchronization - זמני עזיבה מתחנה וזמני כניסה לתחנה מקיימים אילוצים מסוימים. לדוגמא, זמני העזיבה וזמני הכניסה לתחנה העוקבת מיידית זהים.
5. Service-arrival synchronization - זמני תחילת השרות בתחנה וזמני המופע לתחנה מקיימים אילוצים מסוימים. דוגמא נפוצה היא כאשר תחילת השרות של תחנה j צריכה להתחיל לא לפני הכניסה לתחנה k , עבור כל פרויקט.

ב- 1989 ממשיכים Baccelli, Makowski, Shwartz [6] לתת חסמים עליונים ותחתונים למערכות,

ומפתחים תנאי יציבות של המערכת.

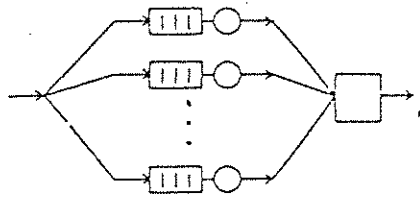
גם פה מסיקים כי עבור מערכת Fork-Join בעלת k תחנות במקביל, ממוצע זמן השהיה במערכת עולה

בקצב לוגריתמי ביחס למספר התחנות, k .

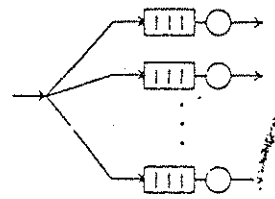
השוואה חשובה ומעניינת נעשית במאמר במאמרם של Tatawi ו-Towsley, Nelson [26] בין 4 מבנה

מערכות שונות: C/NS, C/S, D/NS, D/S כאשר C- תור מרכזי,

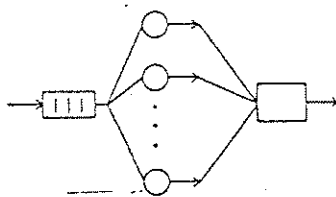
D - תור נפרד לכל שרת. S- פיצול המשימות לעיבוד במקביל, NS- עיבוד המשימות בטור במעבד מסוים.



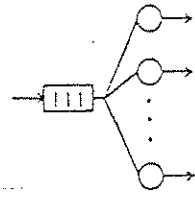
1. Distributed / Splitting



2. Distributed / No Splitting



3. Centralized / Splitting



4. Centralized / No Splitting

במערכת מס' 1 (D/S) הפרויקט מתפצל מיד עם הופעתו למשימות, וכל משימה נשלחת למערכת תור-שרת

משל עצמה. הפרויקט מסיים את עיבודו לאחר סיום עיבוד כל משימותיו.

במערכת מס' 2 (D/NS) הפרויקט אינו מתפצל למשימות, אלא נשלח כולו לאחת ממערכות תור-שרת.

במערכת מס' 3 (C/S) הפרויקט מתפצל מיד עם הופעתו למשימות, אבל כל המשימות עומדות בתור מרכזי

אחד המשותף לכל השרתים. (הפרויקט מסיים את עיבודו לאחר סיום עיבוד כל משימותיו).

במערכת מס' 4 (C/NS) הפרויקט אינו מתפצל למשימות אלא משורת ע"י שרת אחד ועומד בתור מרכזי

המשותף לכל השרתים.

מחקירה אמפירית של סוגי המערכות ומתוצאות אנליטית ממאמרים קודמים מסיקים כי תור מרכזי

יחיד תמיד עדיף, ובהעדר תקורה על פיצול פרויקט למשימות, עדיף לפצל משימות לעיבוד במקביל.

במקרה של תקורה, מקבלים פונקצית ממוצע זמן שהיה של פרויקט שהיא פונקציה קמורה של מספר

המשימות, בעלת מינימום יחיד.

ברשתות יותר מורכבות עוסק Harrison [15]. ברשימות לימוד (Teaching Notes) שלו הוא מראה,

אומנם על רשת ספציפית, כי ככל שמקדם השונות של מופע הפרויקטים וזמני השרות גדלים, גדל גם זמן

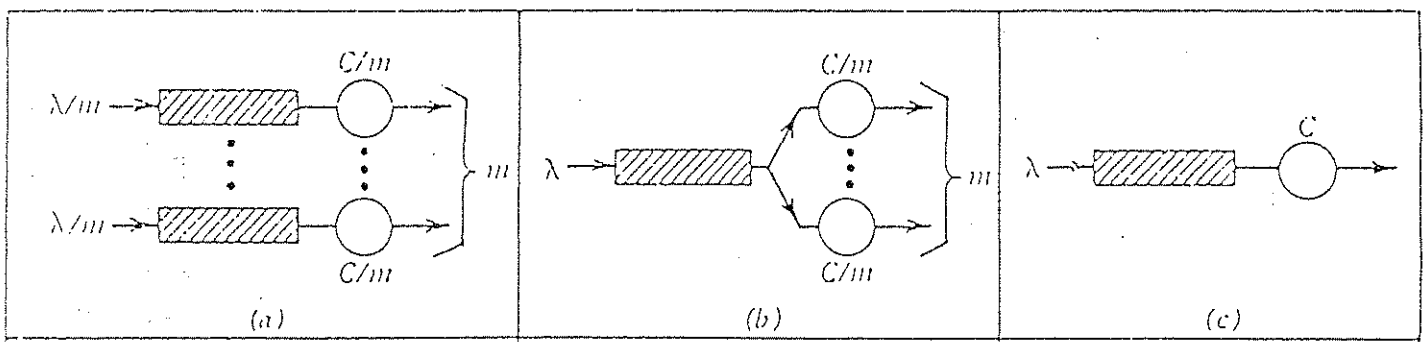
השהות הממוצע של פרויקט במערכת.

ברשתות בהן קיימים יותר מסוג אחד של פרויקטים, ישנן סימולציות ברשימות המראות כי כדאי לתת עדיפות לפרויקטים בעלי שארית זמן שרות קצר יותר.

גם שם מוזכרת המקביליות, דהיינו הגדלת מספר השרתים העובדים במקביל (אפילו אם משאירים את נצילות המערכת כמקודם, כלומר מגדילים את קצב המופע באותו יחס).

זווית ראייה משלימה מוצגת במאמרו של Kleinrock [18]. המאמר משווה בין שלושה מבנים של

מערכות:



במערכת (a) ישנן m מערכות G/G/1 נפרדות ובלתי תלויות.

מערכת טובה יותר היא מערכת (b), בה המופעים התאחדו למופע אחד, בעל תור משותף (היתרון הברור

הוא, שבמערכת (b) לא יתכן מצב שתור אחד מלא, ושאר התורים ריקים, ושרתיהם בטלים).

מערכת (c) היא העדיפה ביותר, בה m השרתים מאוחדים לשרת אחד עם קצב עבודה הגדול פי m .

(היתרון במערכת זו, שלא יתכן מצב בו שרת אחד עובד ושרת אחר בטל).

מסקנה חשובה של Kleinrock היא שאפשר להקטין את זמן השהיה הממוצע של פרויקט במערכת ע"י

הקטנת קצב המופע, או הגדלת קצב השרות, אך השינוי יגרום לירידת היעילות של המערכת.

לעומת זאת ע"י הגדלת קצב השרות וקצב המופעים בו זמנית, זמן השהיה יקטן באותו פרופורציה

ויעילות המערכת תישאר כמקודם.

ברשתות התורים הכלליות ביותר מחסוג בו אנו דנים עוסקת Nguyen [27]. מאמרה מתייחס לרשת PERT סטוכסטית דינמית בעלת שרתים הומוגניים ואילוצי קדימויות קבועים.

במאמר מביאה Nguyen הגדרה אינדוקטיבית לגבי משך הזמן הכולל (זמן ההמתנה + זמן השרות) שפרויקט רואה, כשמגיע למערכת:

$$\phi_j(t) = \phi(t) + \max_{i \in \varphi(j)} T_i(t)$$

$$T_j(t) = \max_{i \in \varphi(j)} T_i(t) + W_j(\phi_j(t))$$

$$T(t) = \max_{j \in \varphi(j+1)} \{T_j(t)\}$$

כאשר: $\phi_j(t)$ - זמן הגעת הפרויקט הבא, אחרי זמן t לתחנה j .

$\phi(t)$ - זמן הגעת הפרויקט הבא אחרי זמן t לתחנה 1.

$T_i(t)$ - הזמן בין הופעת הפרויקט הבא אחרי זמן t לסיום עיבוד בתחנה i .

$\varphi(j)$ - קבוצת הממקדימות המידיות של תחנה j (ראה הגדרה בסעיף 1.2.1)

$W_j(t)$ - סה"כ זמן השרות של הפרויקטים הממכים בתור (והפרויקט

שבעיבוד) בתחנה i .

$T(t)$ - זמן המתנה ושרות של הפרויקט המגיע לאחר זמן t .

Nguyen מראה ששרשרת כזו הפועלת בעומס גבוה (Heavy Traffic) ניתנת לקירוב בעזרת תנועת בראון

משוקפת רב-מימדית (Reflected Brownian Motion).

כדי לחשב את זמן השהיה של פרויקט, משתמש המאמר בעקרון ה-"Snapshot": "צילום" המערכת בעת הופעת פרויקט.

עבור מערכת בעומס גבוהה, התנודות בכמות העבודה במערכת זניחות יחסית לזמן שהפרויקט שוהה במערכת, ולכן ברגע הופעתו, ניתן ל"הקפא" את המערכת, ולחשב את זמן השהיה ע"י שימוש בחישוב מסלולים קריטיים במערכת PERT סטוכסטית רגילה, רק שבמקום זמני השרות בצמתים, יילקחו

כמויות העבודה בכל צומת $(W_i(t))$.

לסיכום סקר הספרות, מחקרים בנושא רשתות תורים מתמקדים בד"כ ברשתות בסיסיות ופשוטות, כגון Fork-Join. חלק מהמאמרים עוסקים ברשתות סטטיות (רשתות PERT), וכאשר מכלילים את הרשתות לדינמיות, הניתוח האנליטי נעשה בתחומים צרים בלבד (יציבות או גבולות ב- Heavy Traffic).

הקשיים התאורטיים והדוגמאות המעשיות ממחישים היטב את חשיבות המודל של רשת PERT סטוכסטית דינמית, מהיבטים מחקרניים וישומיים גם יחד.

עבודה זו תנסה לתרום לקידום שני היבטים אלו.

1.2 הצגת המודל.

נתונה מערכת המורכבת מפרויקטים ומשאבים.

1.2.1 תאור פרויקט.

כל פרויקט מכיל J משימות, ממוספרות ע"י $j=1...J$.

סדר ביצוע המשימות מתואר ע"י רשת קדימויות, רשת AON בה הקדימויות נתונות ע"י

הקשתות/חיצים, והמשימות ע"י הצמתים. זמני המשימות ב"ת מפולגות מהתפלגות כלשהיא. סיום

ביצוע הפרויקט הוא סיום ביצוע כל משימותיו.

נוסיף לרשת שני צמתי דמה:

צומת 0 - צומת התחלה, המציין את תחילת הפרויקט (אליו מגיע הפרויקט וממנו מתפצל למשימותיו).

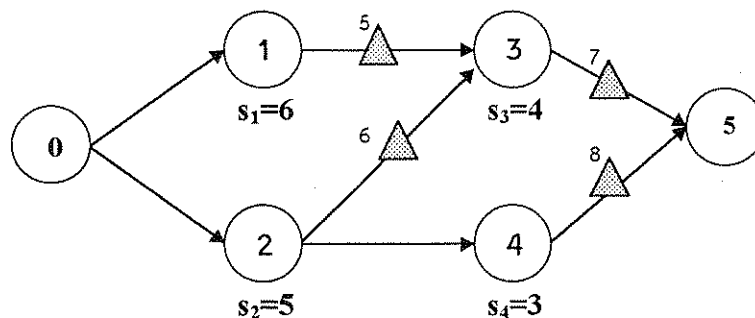
צומת $J+1$ - צומת סיום, המציין את סיום הפרויקט.

נוסיף לרשת את תורי הסנכרון הנוצרים במערכת עקב הקדימויות (עליהם נפרט בהמשך), ועתה נוצרת

לנו מערכת סטוכסטית סטטית (ראה סקר ספרות).

דוגמא:

▲ - תורי סנכרון.



שרטוט 1.2.1 א: מערכת סטוכסטית.

Figure 1.2.1a: Stochastic Network.

הפרויקט בדוגמא מכיל 4 משימות. בתחילה מבוצעות משימות 1,2. לאחר ביצוע משימה 2, יכולה

להתבצע משימה 4. כדי לבצע את משימה 3, חייבים לסיים את משימות 1 ו-2. הפרויקט מסתיים כאשר

מסתיימות משימות 3 ו-4, וזאת כאשר 1 ו-2 הסתיימו לפני כן.

התפלגות זמן ביצוע משימה j כללית, בעלת ממוצע s_j ומקדם שונות c_j . (אין הנחות נוספות על ההתפלגות). קיימת אי-תלות בין זמני ביצוע המשימות.

משימה i נקראת "מקדימה מיידית" (m "מ") של משימה j , אם בסיום משימה i הפרויקט יכול לעבור לביצוע משימה j , ואז משימה j נקראת "עוקבת מיידית" (e "ע") של משימה i .

נסמן:

$P(j)$ - קבוצת המ"מ של משימה j .

$PP(j)$ - מספר המ"מ של משימה j .

$S(j)$ - קבוצת הע"מ של משימה j .

$PS(j)$ - מספר הע"מ של משימה j .

בדוגמא: $PS(2)=2$, $S(2)=\{3,4\}$, $PP(3)=2$, $P(3)=\{1,2\}$

כאשר למשימה j מספר משימות ע"מ $(PS(j)>1)$, אזי נקרא לה משימת פיצול (Fork).

כאשר למשימה j מספר משימות מ"מ $(PP(j)>1)$, אזי נקרא לה משימת איחוד (Join).

בדוגמא: משימה 2 היא משימת פיצול, ומשימה 3 היא משימת איחוד.

במקרה של משימת איחוד, שלה יותר ממ"מ אחת, נוצרים תורי סנכרון (סנכרון) בו המשימות מחכות עד ששאר המשימות המ"מ יגמרו את עיבודן.

בדוגמא, נוצרים 4 תורי סנכרון, ממוספרים 5-8.

1.2.2 תאור המערכת.

כל משימה j בפרויקט מבוצעת בתחנת שרות, שגם אותה נסמן ב- j . תחנת השרות יכולה להכיל n_i יחידות משאב.

עתה, נוסף מופע פרויקטים למערכת המתואר ע"י תהליך חידוש עם קצב λ . המופע בלתי תלוי בזמני ביצוע המשימות.

נוצרת רשת PERT סטוכסטית דינמית המכילה תורים נוספים של משימות.

עתה, במערכת ישנם שני סוגי תורים:

תור המתנה לשרת - כאשר שרת עסוק עם משימה של פרויקט מסוים ופרויקטים נוספים מגיעים. תורים אלו נוצרו עקב דינמיות המערכת.

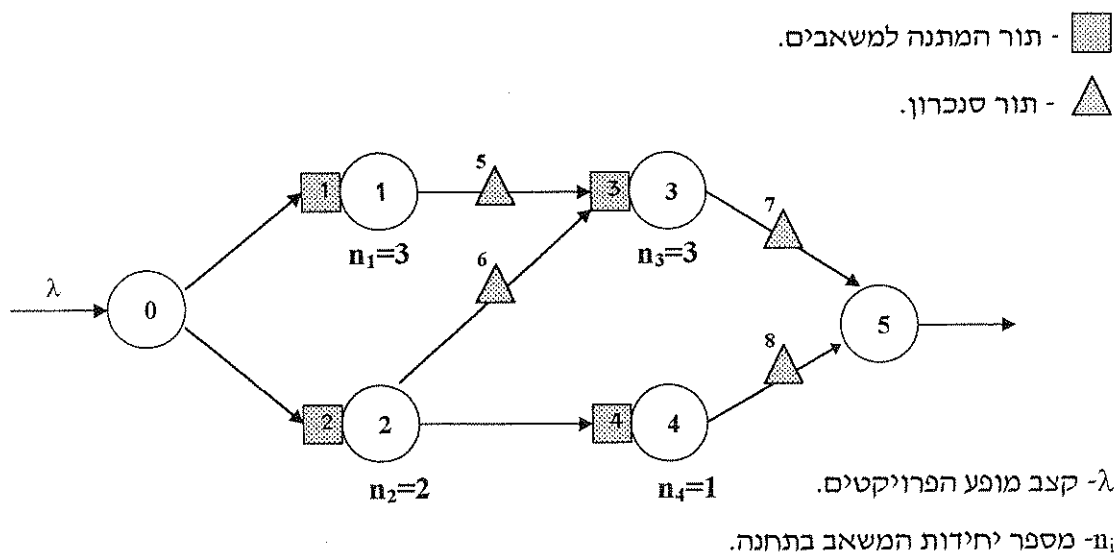
תור סנכרון - משימות מ"מ המחכות למשימות אחרות. תורים אלו נוצרו עקב מבנה הפרויקט.

כמקובל בתורת התורים, נתייחס בהמשך ליחידות המשאבים כאל "שרתים".

לפני כל תחנה j ($j=1 \dots J$) נוסף תור למשימות המחכות לשרתים בתחנה j .

עבור כל תחנה j , המהווה תחנת איחוד, נוסף עוד $PP(j)$ תורים, המשמשים לסנכרון כל המשימות המ"מ $i \in P(j)$, i .

דוגמא: רשת PERT סטוכסטית דינמית.



שרטוט 1.2.2 א : מערכת סטוכסטית דינמית.
 Figure 1.2.2a: Dynamic Stochastic Network.

נסמן את המספר הכולל של התורים ב- K , ונתייחס לתור k , $k=1,2,\dots,K$.

כל תור מכיל מספר מקומות בלתי מוגבל.

מדיניות השרתים היא FCFS (First Come First Served). מדיניות זו חד-משמעית בצמתים שאינם צמתי איחוד. בצומת איחוד זמן הגעת המשימה נחשב כזמן הגעת אחרון המשימות המ"מ, אשר צריכות להיות מתוזמנות לצורך ביצוע משימת האיחוד.

הערה: תחת המדיניות הנ"ל, הסדר בין הפרויקטים נשמר בכל תחנה.

נסמן $X_k(t)$ - גודל התור ה- k (כלומר, מספר המשימות הממתונות, בטרם שירות) בזמן t , $k=1,\dots,K$. מצב המערכת בזמן t מוגדר ע"י הוקטור המקרי $X(t)=(X_1(t),\dots,X_K(t))$, והתפתחות המערכת מתוארת ע"י התהליך הסטוכסטי $X=\{X(t), t \geq 0\}$.

תנאי הכרחי ובד"כ מספיק ליציבות המערכת הוא $\forall i, \lambda \times s_i < 1$, לכן נניח בהמשך $\lambda < \min_i (1/s_i)$, (ראה Baccelli, Massey ו-Towsley [7]). למשל, כאשר המופע פואסוני וזמני השירות מעריכים, X הוא תהליך קפיצה מרקובי והתנאי הכרחי ומספיק לארגודיות.

1.2.3 דוגמאות.

רשתות דינמיות שימושיות ביותר בתעשייה ובניהול פרויקטים.

להלן 3 דוגמאות למחקרים מעשיים בהם יש שימוש במודל רשתות דינמיות.

מאמרו של Larson [21] מתאר מחקר שנעשה בעיר ניו-יורק ב-1988, בנושא שיפור תהליך מעצר ושיפוט,

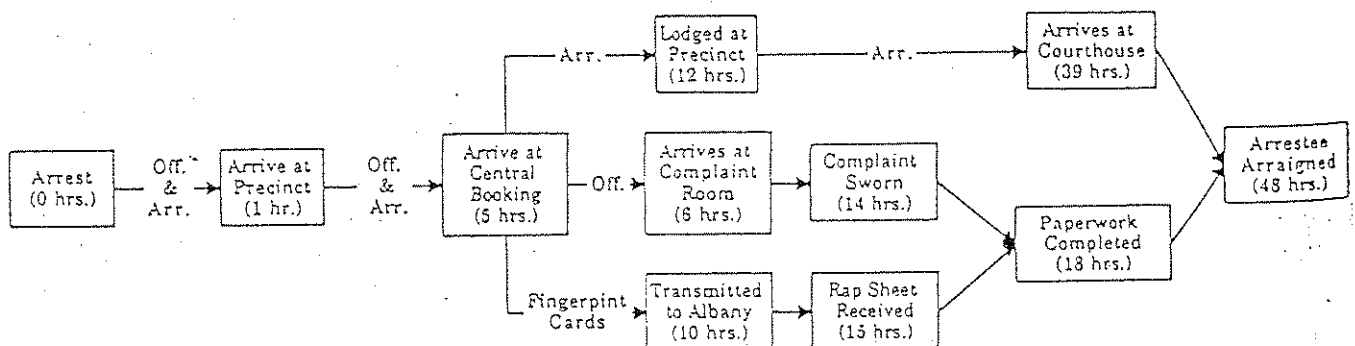
בגלל האיחורים הרבים בשיפוט ותנאי המעצר הגרועים.

במחקר תארו את תהליך המעצר והשיפוט ע"י שימוש ברשת PERT סטוכסטית דינמית, ובסיומו הגיעו

להמלצות לטווח הקרוב והרחוק, כגון: הגדלת המשאבים (כו"א וחדרי מעצר), הקטנת מספר צרכנים,

הקמת ניהול מרכזי ובסיסי מידע משותפים.

להלן רשת ה-PERT:



שרטוט 1.2.3 א: רשת PERT עבור תהליך המעצר.

Figure 1.2.3a: PERT Network.

מאמרים של Nguyen, Mandelbaum, Adler ו-Schwerer [1] עוסק בתאור מערכת פיתוח מוצרים

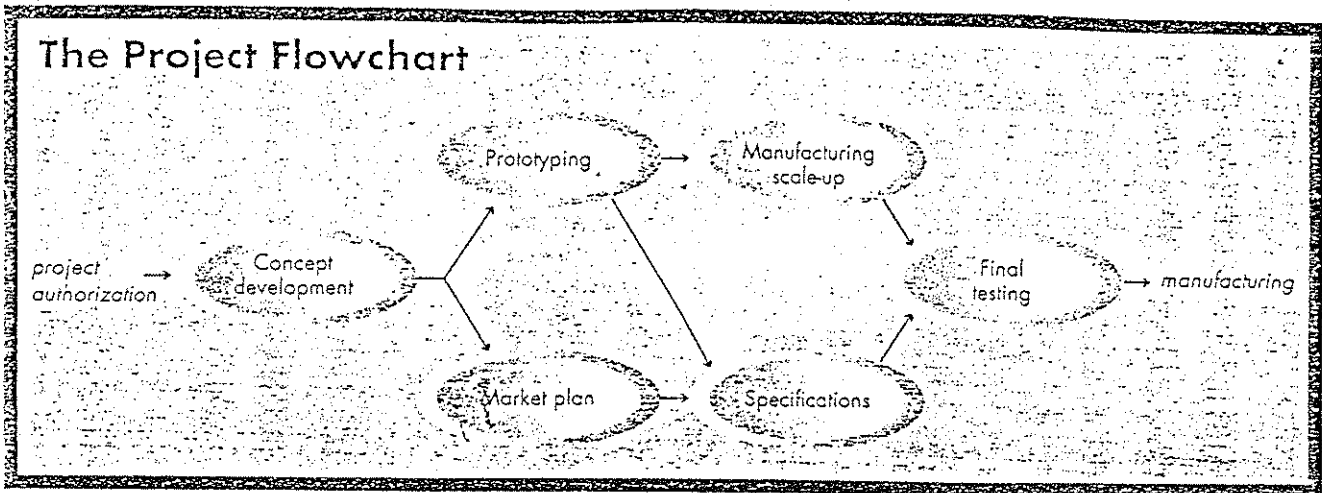
במפעל אנונימי, המכונה התאגדות הכימיקלים (מחלקת פלסטיקה) ע"י רשת PERT סטוכסטית דינמית.

החידוש במאמר שהוא עוסק במערכת עם מספר פרויקטים שונים המשתמשים במשאבים משותפים, לכן

נקודת המבט הפרויקטלית המקובלת משתנה לראיה תהליכית. הראיה התהליכית מציבה את המשאבים

בצמתים, במקום הפעילויות.

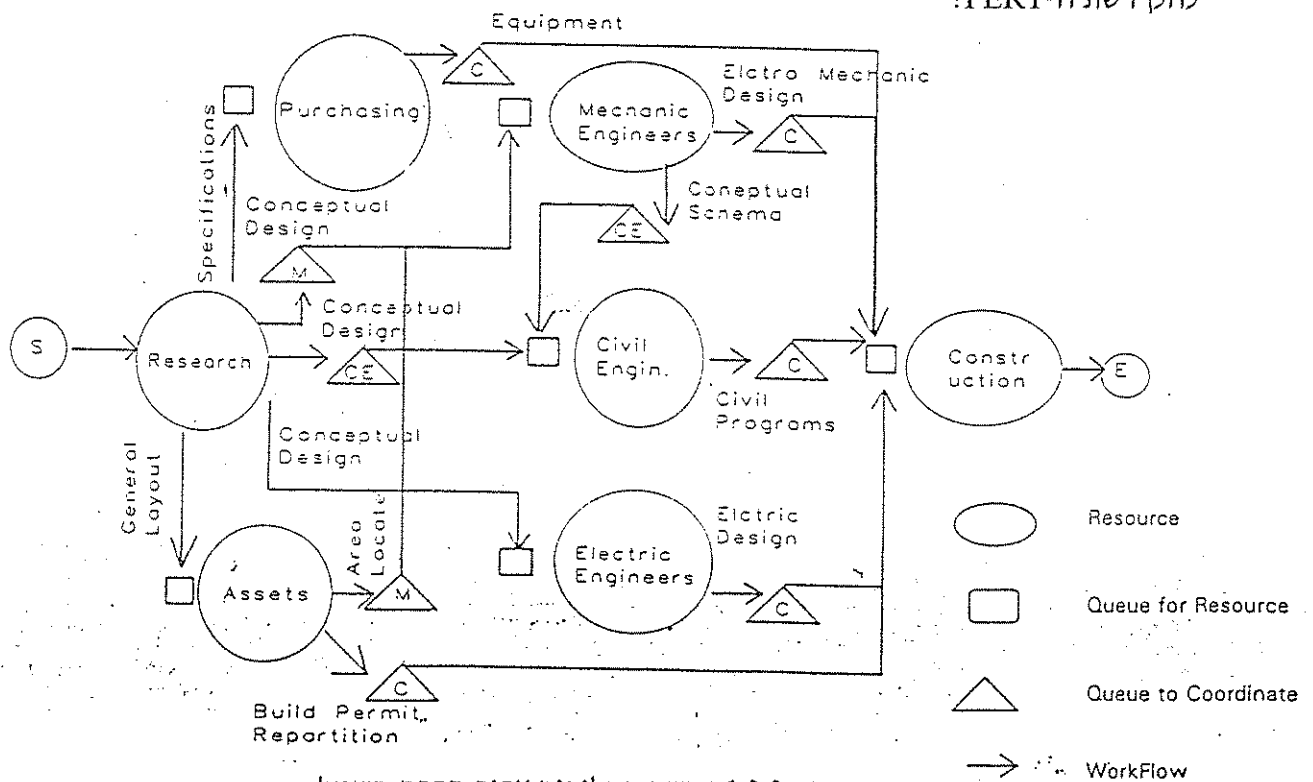
הפרויקטים השונים מגיעים לצמתים ומתחרים על קבלת השרות.



שרטוט 1.2.3 ב: רשת PERT.
Figure 1.2.3b: PERT Network.

עבודת התזה של למברג [35] עוסקת בתאור תהליך תומך תחנות המשנה בחברת החשמל לישראל. במערכת מורכבת, כדוגמת חברת החשמל, קיימת בעיה של תכנון לוחות זמנים וזימון משאבים בהיבט רב-פרויקטלי, שכלי הניהול הקיימים בתחנה אינם מאפשרים להתמודד עמה. לכן בעבודת התזה נעשה ניסיון לתאר את המערכת ולבדוק שינויים במערכת ע"י פיתוח מודל סטוכסטי דינמי, וסימולציות.

להלן רשת ה-PERT:



שרטוט 1.2.3 ג רשת תהליכית עבור חברת חשמל.
Figure 1.2.3c: Processing Network Representation.

1.3 מושגים להבהרה.

בסעיף זה ננסה להבהיר מושגים המקובלים ברשתות PERT (דטרמיניסטיות וסטוכסטיות), ונדגים את השימוש במושגים ואת תרומתם עבור המערכת הדינמית מסעיף 1.2 (שרטוט 1.2.2 א).

1.3.1 מסלול קריטי

במערכת דטרמיניסטית, המושג "מסלול קריטי" ברור ומדויק: המסלול ברשת שסכום זמני המשימות עליו בעל ערך מקסימלי. אפשרות אחרת להגדרה: רצף של משימות, שכל דחייה של ביצוע אחת מהן תגרום לדחייה זהה של מועד סיום הפרויקט.

במערכת סטוכסטית ידוע לכל משימה ממוצע זמן ביצועה וכן שונותה. ההנחה במערכת סטוכסטית בעלת מספר מספיק גדול של פעילויות, שכל מסלול מפולג נורמלי עם שני פרמטרים: תוחלת המתקבלת מסכום תוחלות המשימות על המסלול, ושונות המתקבלת מסכום השונות. עתה המושג "מסלול קריטי" מוגדר באופן הבא: מסלול בעל התוחלת הגבוהה ביותר של משך הפעולות לאורכו.

להגדרה זו הרבה בעיות: יתכנו מצבים בהם נידחה משימות הנמצאות על מסלולים לא קריטיים, ולמרות זאת נגרום לדחיית משך זמן הפרויקט (בניגוד למקרה הדטרמיניסטי), ההגדרה אינה מתחשבת במסלולים תלויים ברשת, וכד' (ראה סקר ספרות).

הסבר אנליטי לבעיות ההגדרה:

נסמן π_i $i=1,2,\dots$ - מסלול i ברשת.

T - זמן שהיה של הפרויקט.

$$\begin{aligned} T &= \max_i \{\pi_i\} \\ E(T) &= E(\max_i \{\pi_i\}) \geq \max_i \{E(\pi_i)\} \end{aligned} \quad (1)$$

אי השוויון מתקיים מהסיבה הבאה:

$$\begin{aligned}
 & \max_i \{\pi_i\} \geq \pi_i \quad \forall i \\
 & E \max_i \{\pi_i\} \geq E \pi_i \quad \forall i \\
 & \quad \downarrow \\
 & E \max_i \{\pi_i\} \geq \max_i \{E \pi_i\}
 \end{aligned}
 \tag{2}$$

ולכן החלפת התפלגויות בממוצעים מעוותת את התוצאות.

במערכת דינמית המושג "מסלול קריטי" עוד יותר מעורפל כיוון שכל פרויקט המגיע למערכת פוגש מערכת שונה, המכילה מספר שונה של פרויקטים ממתינים. חישוב המתבסס על ממוצעים בלבד (בדומה למערכת סטוכסטית) אינו אפשרי, בד"כ, כיוון שלא ניתן לחשב אנליטית את המספר הממוצע של הפרויקטים הממתינים לפני כל תחנה. אי לכך הגדרה אנליטית, בדומה להגדרה הדטרמיניסטית והסטוכסטית, קשה למצוא, אך ניתן למצוא הגדרה דומה. לכן נגדיר:

"מסלול קריטי" - המסלול בעל הסיכוי הגבוה ביותר להיות הארוך ביותר.

"אינדקס קריטיות" של מסלול - הסיכוי של מסלול להיות קריטי.

אלגוריתם למציאת מסלול קריטי:

א. בצע N סימולציות שונות (N מספר שרירותי, גדול מספיק).

עבור כל סימולציה:

1. א. הרץ את המערכת עד שתגיע למצב יציב.

2. א. הרץ עבור עוד B פרויקטים (B מספר שרירותי, גדול מספיק).

עבור כל פרויקט, מצא את המסלול בו הפרויקט שהה זמן מרבי

וסמן מסלול זה.

3. א. חשב עבור כל מסלול את הסיכוי להיות הקריטי, אינדקס קריטיות (P_i כלומר, עבור כל

מסלול i , פרופורצית מספר הפעמים שהמסלול היה בפועל מסלול קריטי P_i).

ב. חשב ממוצעי אינדקס קריטיות עבור כל מסלול (ממוצע של N התוצאות, עבור כל מסלול).

ג. המסלול הקריטי יהיה זה בעל ממוצע האינדקס הגבוה ביותר.

נערכו סימולציות עבור המערכת משרטוט 1.2.2 א'.

הסימולציה כללה 50 חזרות, כל חזרה 20,000 יח"ז. המידע מ-10,000 יח"ז הראשונות הושמט

והסטטיסטיקות חושבו עבור שאר יח"ז.

עבור כל פרויקט חושב זמן השהיה שלו במערכת, זמני השהיה בתורים והמסלול הקריטי. בתום כל חזרה

חושבו ממוצעי הזמנים והתפלגות המסלול הקריטי.

בסיום הסימולציה חושבו סטית התקן ורווחי הסמך (תמיד ברמת בטחון 95%).

להלן תוצאות האלגוריתם:

מסלול	ממוצע אינדקס קריטיות	חצי רווח סמך (רמת סמך 95%)	% מהממוצע
0-1-3-5	0.146	0.0067	4.58
0-2-3-5	0.104	0.0052	5
0-2-4-5	0.75	0.011	1.466
	1.00		

טבלה 1.3.1 א: קריטיות המסלולים עבור מערכת דינמית.

Table 1.3.1a: The Path's Criticality for the Dynamic Model.

כפי שרואים מהתוצאות, המסלול הקריטי במערכת הנ"ל הוא 0-2-4-5, אותו סימנו 75% מהפרויקטים

שהורצו.

הערה: האלגוריתם הנ"ל ניתן ליישום גם עבור רשת סטוכסטית או דטרמיניסטית.

ע"מ להדגים זאת, ניקח את הרשת הסטוכסטית משרטוט 1.2.1 א'.

להלן אינדקס הקריטיות עבור שלושת המסלולים:

מסלול	ממוצע אינדקס קריטיות	חצי רווח סמך (רמת סמך 95%)
0-1-3-5	0.47	0.0074
0-2-3-5	0.26	0.006
0-2-4-5	0.27	0.0058
	1.00	

טבלה 1.3.1 ב: קריטיות המסלולים עבור מערכת סטוכסטית.

Table 1.3.1b: The Path's Criticality for the Stochastic Model.

התוצאה שהתקבלה, שמסלול 0-1-3-5 הוא הקריטי, אינה סותרת את החישוב האנליטי (אכן מסלול זה

בעל התוחלת הגבוהה ביותר).

הערה: יתכנו מקרים בהם האלגוריתם המוצע והחישוב האנליטי המקובל יסתרו זה את זה. לדוגמא, נתבונן בפרויקט המכיל שני מסלולים במקביל. כל מסלול מכיל משימה אחת בלבד. זמן ביצוע משימה א' הוא $\text{Gamma}(10.5, 1)$ ואילו משך ביצוע משימה ב' הוא $\text{Gamma}(4, 2.5)$. תוחלת ביצוע משימה א' היא: $10.5 = 10.5 \times 1$, ואילו תוחלת משימה ב' היא: $10 = 4 \times 2.5$. ע"פ החישוב האנליטי המסלול הקריטי הוא המסלול הראשון, המכיל את משימה א'. לעומת זאת, סימולציות הראו כי המסלול השני (המכיל את משימה ב') סומן עבור 55.43% מהפרויקטים, כלומר כתוצאה מהפעלת האלגוריתם נסיק שהמסלול השני הוא הקריטי, בניגוד לחישוב האנליטי. ההסבר לכך הוא מקדם ההשתנות הגבוה של משימה א' ($CV=1$) לעומת מקדם ההשתנות של משימה ב' ($CV=0.632$). למרות שממוצע הזמן של משימה א' נמוך יותר, הפיזור הגבוה גורם לכך שמשימה א' ארוכה יותר אצל רוב הפרויקטים. כאשר אנו רוצים לשפר ביצועי מערכת ולהוריד את זמן השהיה בה, כדאי לשפר משימות הנמצאות על המסלול הקריטי. הבעיה עתה, באיזה משימה כדאי להתחיל?

1.3.2 משימה קריטית

חשוב לדעת איזה משימה היא הקריטית ביותר, כדי להשקיע בה את מירב המשאבים. לכן נגדיר:

"משימה קריטית" - משימת צוואר בקבוק. שיפור בה יגרום לשיפור מרבי במערכת (ניתן לקחת מדדים שונים לשיפור מצב המערכת. אנו משתמשים בתוחלת זמן השהיה של פרויקט).

"אינדקס קריטיות" של משימה - הסיכוי שמשימה קריטית.

סביר להניח שמשימה קריטית היא המשימה הנמצאת מירב הפעמים על המסלול הקריטי, כיוון שאז שיפור בה ישפר את המסלול הקריטי הרבה פעמים (ובעקבות כך ישפר את המערכת). נבדוק עתה הנחה זו.

האלגוריתם למציאת משימה קריטית זהה לאלגוריתם מציאת המסלול הקריטי. גם פה אפשר להריץ סימולציות רבות ועבור כל פרויקט לסמן את המשימות על המסלול הקריטי. בסוף התהליך, לחשב איזה משימה סומנה מירב הפעמים.

חישוב תחליפי לאלגוריתם, הנותן תוצאות זהות הוא החישוב הבא:

חישוב אינדקס קריטיות של משימה:

עבור כל משימה, סכם את אינדקס הקריטיות של המסלולים עליהם היא נמצאת.

נדגים זאת על המערכת הדינמית משרטוט 1.2.2 א'.

המשימה	אינדקס קריטיות
1	0.146
2	$0.104 + 0.75 = 0.854$
3	$0.146 + 0.104 = 0.25$
4	0.75

טבלה 1.3.2 א: קריטיות משימות.

Table 1.3.2a: Tasks' Criticality.

כפי שרואים, משימה 2 היא בעלת אינדקס קריטיות הגבוה ביותר. אך האם משימה 2 היא הקריטית?

טבלה 1.3.2 ב מביאה את ממוצעי אורכי תורי השהיה (תורי ההמתנה + המשימות בעיבוד) במערכת.

תור	ממוצע	חצי דיוח סגור (95%)
1	3.13	0.063
2	6.57	0.318
3	1.37	0.09
4	18.25	1.49

טבלה 1.3.2 ב: ממוצעי אורכי התורים

Table 1.3.2b: Queue-Length Average.

אם נסתכל על ממוצע אורכי התורים הנוצרים במערכת, תור ההמתנה לתחנה 2 גדול, אך תור ההמתנה

לפני תחנה 4 גדול הרבה יותר.

סימולציות שנעשו על המערכת הראו כי אם מקטינים את ממוצע השרות בתחנה 2 ביח"ז, מקבלים שיפור

של זמן השהיה במערכת של 26.2%, לעומת זאת ע"י הקטנת ממוצע השרות בתחנה 4 מקבלים שיפור של

41.2%, כלומר תחנה 4 מהווה, ללא ספק, תחנת צוואר בקבוק.

המסקנה המתבקשת מדוגמא זו, שאינדקס קריטיות של משימות אינו מהווה מדד אמין עבור קריטיות.

מדד אחר, כפי שהוזכר כבר, הוא ממוצע אורך תור המתנה, או זמן שהיה בתור. ככל שאורך תור ההמתנה

לפני התחנה גדול יותר, הרי תחנה זו מהווה צוואר בקבוק וכדאי להשקיע בה.

כלומר, ע"פ הצעתנו, כדי למצוא את המשימה הקריטית או תחנת צוואר בקבוק במערכת דינמית נבצע את השלבים הבאים:

1. נבחר מסלול קריטי - ע"פ האלגוריתם מסעיף 1.3.1.
2. נבחר משימה קריטית - משימה הנמצאת על המסלול הקריטי, שלה אורך תור הגדול ביותר. קריטריון זה יהיה זמן שהיה בתחנה מקסימלי, אך מאחר וקצב מופע הפרויקטים זהה לכל התחנות, הרי זמן שהיה ואורך תור מדדים שקולים. עבור מנהל ייצור או מפעיל תחנה בָּרוּר כי קל יותר לבדוק את אורך התור בתחנה מאשר זמן שהיה.

1.3.3 פרופיל שהיה (Time Profile).

בסעיפים קודמים ראינו שישנם שני סוגי תורים בהם הפרויקט שוהה: המתנה למשאבים וסנכרון. בנוסף, ברור שישנו זמן בו הפרויקט מעובד בתחנות השונות. מטרתנו בסעיף זה לנסות ולהגדיר חלוקת זמן של הפרויקט, שתנסה ללמד אותנו כמה אחוזים משקיע פרויקט בממוצע בהמתנה, עיבוד וסנכרון. מעבר לכך, המושג סנכרון דורש הסבר. יתכנו כמה סוגי סנכרון וחלקנו את זמן הסנכרון הבסיסי לשני סוגים:

סנכרון תכני (Design) - סנכרון הקיים ברשת בעלת אילוצי קדימויות. האילוצים גורמים שמשימה i אינה יכולה להתחיל עד סיום משימה j ולכן נוצר זמן מבוטל של משימה i שאינה יכולה להתחיל מיד בעיבוד. לדוגמא, ברשת המופיעה בשרטוט 1.2.2, משימה 4 אינה יכולה להתחיל בעיבוד מיד עם הופעת הפרויקט. היא צריכה להמתין עד סיום עבוד משימה 2, וזהו זמן מבוטל. הסנכרון קיים עבור כל משימה שלה משימה מקדימה מיידית אחת לפחות (שימוש בסימנים מתחילת הסעיף, קבוצת המשימות להן יש סנכרון תכני: $\{i \mid PP(i) \geq 1\}$).

ניתן לקרוא לסנכרון זה גם **סנכרון התחלה**, כיוון שהוא נוצר בגלל שלא ניתן להתחיל את כל המשימות בו זמנית.

סנכרון תפעולי (Operative) - סנכרון הנוצר ברשת בגלל אילוצי הקדימות ואי תלות בין זמני המשימות השונות.

כאשר משימות i, j מ"מ למשימה k וזמני השרות של i, j אינם תלויים (אינם שווים בהסתברות 1) נוצר מצב שאחת מהמשימות צריכה לחכות לחברתה בתורי הסנכרון (סומנו בשרטוטים במשולש). לדוגמא, ברשת המופיעה בשרטוט 1.2.2, משימות 1, 2 מ"מ למשימה 3. לכן עבור כל פרויקט, הקצרה מבינהן תצטרך לחכות לחברתה.

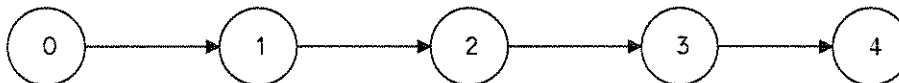
הסנכרון קיים עבור כל משימה המהווה מקדימה מיידית למשימה אחרת, שלה מספר משימות מ"מ גדול מאחת (קבוצת המשימות בעלות סנכרון תפעולי: $\{i \mid \exists j, i \in P(j) \text{ PP}(j) > 1\}$). ניתן לקרוא לסנכרון זה גם **סנכרון סוף**, כיוון שנוצר בגלל שלא ניתן לסיים את כל המשימות בו זמנית.

בנוסף, ברור שבמערכת בעלת אופי סטוכסטי ודינמי הזמן המושקע בסנכרון גדל. בהמשך נמשיך לדון ולהסביר את סוגי הסנכרון.

ישנן רשתות שבהן המושגים זמן המתנה, עיבוד וסנכרון מקבלים משמעות אינטואיטיבית וניתן אפילו לנתחן אנליטית, תחת הנחות מסוימות. למשל:

רשת טורית - בה המשימות מבוצעות אחת אחר השניה. (שרטוט 1.3.3 א),

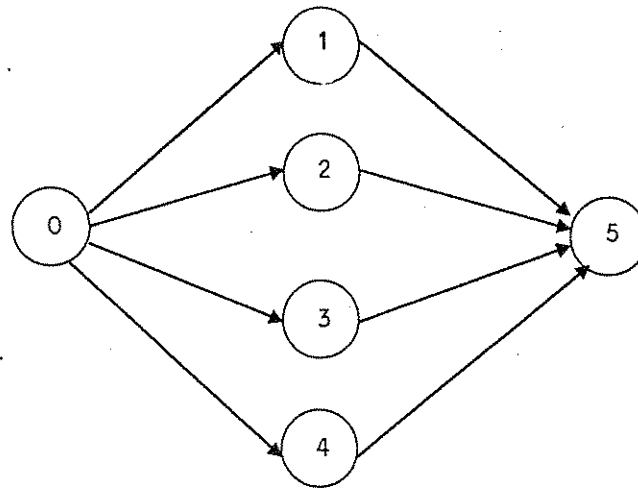
ברשת זו הסנכרון התכני קיים (כיוון שלא כל המשימות יכולות להתחיל מיד עם הופעת הפרויקט), אך לא קיים סנכרון תפעולי (אין תורי סנכרון).



שרטוט 1.3.3 א: רשת טורית.
Figure 1.3.3a: Serial Network.

רשת מקבילית - בה הפרויקט מתפצל, מיד עם הופעתו, למספר משימות המעובדות במקביל. הפרויקט מסיים את עיבודו לאחר סיום עיבוד כל משימותיו (ראה שרטוט 1.3.3 ב).

ברשת זו הסנכרון התפעולי קיים (כיוון שיש יותר ממספר מ"מ למשימה, ואם זמני המשימות אינם שווים בהסתברות 1, ייווצר מצב שמשימה תגמור לפני רעותה). אך סנכרון תכני לא קיים (כיוון שכל המשימות מתחילות ביחד).



שרטוט 1.3.3 ב: רשת מקבילית.
Figure 1.3.3b: Parallel Network.

עבור רשת כללית, הגדרת הזמנים וחקירת פרופיל הזמן איננה מיידית.

אנו מציעים שתי שיטות .

שיטה א'.

שחרור המערכת מאילוצי משאבים וסנכרון תוך מדידת הרווח ברמת הביצועים.

שחרור המערכת יתבצע בארבעה שלבים , כאשר הסדר בינם שרירותי (כל סדר ביצוע הגיוני, אך נותן משקל שונה לאילוצים והמתנות. לכן התוצאות המתקבלות שונות בסדרים שונים).

• **שלב א:** שחרור המערכת מאילוצי משאבים מוגבלים, ע"י הוספת מספר בלתי מוגבל של שרתים בכל תחנה. תוספת כזו מחסלת את תורי ההמתנה למשאבים ויוצרת למעשה תחנה בודדת של $G/G/\infty$. מופע הפרויקטים בתחנה זו זהה למופע הקודם וזמן השרות נתון ע"י זמן ביצוע של פרויקט ברשת PERT סטוכסטית סטטית עם דיאגרמת קדימויות וזמני ביצוע כמקודם, אבל ללא תורי המתנה בין ביצוע המשימות.

כדאי לשים לב כי ישנן משימות הממתינות בתורי משאבים שונים, אך צריכות לאחר מכן להסתכרן. בגלל תורי ההמתנה השונים נוצר מצב שההבדל בין זמני הסיום של המשימות גדל ובעקבותיו גם זמני סנכרון. לכן שחרור המערכת מאילוצי משאבים יתרום גם להקטנת זמן סנכרון זה (שהוא בעצם זמן סנכרון "מדומה" הנוצר עקב זמני ההמתנה).

• **שלב ב:** שחרור המערכת מאילוצי סנכרון הנגרמים כתוצאה ממערכת בעלת אופי סטוכסטי ודינמי ע"י שינוי ההתפלגויות (ההתפלגות זמני ביצוע המשימות) לדטרמיניסטיות.

- **שלב ג:** שחרור המערכת מאילוצי סנכרון הנוצרים מאילוצי הקדימויות (סנכרון תכני או סנכרון התחלה) ע"י מעבר לרשת מקבילית.
- **שלב ד:** שחרור המערכת מאילוצי סנכרון הנגרמים מאי תלות בין זמני המשימות (סנכרון תפעולי או סנכרון סוף) ע"י דרישה שכל זמני השרות בתחנות יהיו זהים בהסתברות 1.

בסיום השלבים הנ"ל נקבל מערכת, בה כל המשימות מתחילות ומסתיימות ביחד. כלומר, כל הזמן הפרויקט בעיבוד.

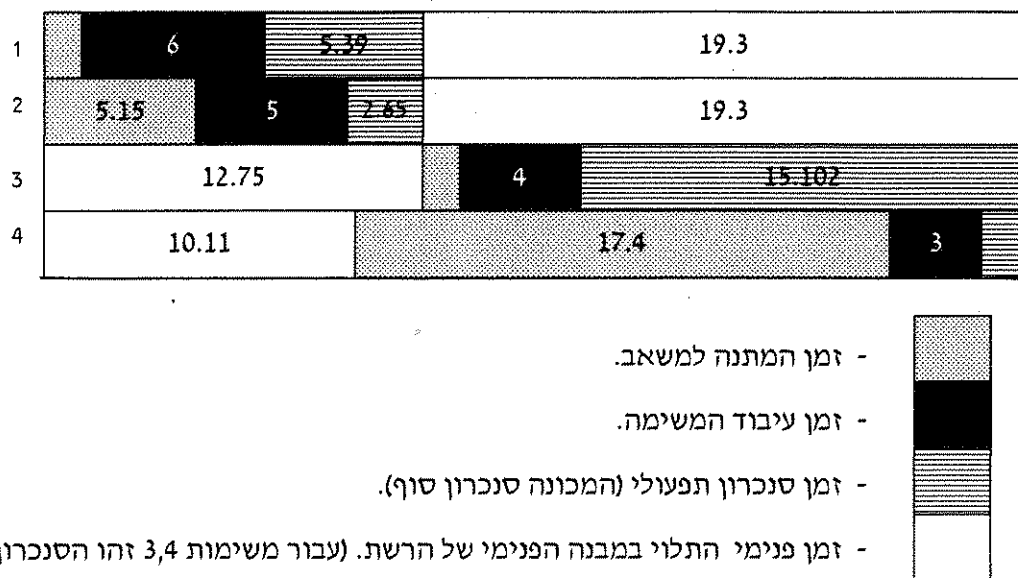
הערה: ניתן, כמובן, לבצע רק חלק מהשלבים.

שיטה ב'.

אפשרות אחרת להגדרת הזמנים הוא מנקודת מבט תפעולית. כל יחידת זמן של כל משימה מלווה בעלות. נגדיר מדד ביצועים השווה לסכום זמני שהיית המשימות בתחנות. זמן משימה יכול להתחלק עתה באופן חד-משמעי לזמן עיבוד, לזמן שהמשימה מחכה בתור למשאב ולזמן שהמשימה מחכה בתור לסנכרון עם משימות אחרות.

ניתן לתאר את המהלך הממוצע של הפרויקט בעזרת רשת Gantt ע"י שימוש בחלוקה הנ"ל.

לדוגמא, נתאר את רשת ה-Gantt עבור המערכת בשרטוט 1.2.2 א' :



שרטוט 1.3.3 ג: רשת Gantt למודל הדינמי
 Figure 1.3.3c : Gantt Network for the Dynamic Model.

כל שורה מתייחסת למשימה שונה והמספרים המופיעים בתוך המלבנים הם ממוצע הזמן שהפרויקט שוהה.

כדאי לשים לב לשני מאפיינים חשובים ברשת:

1. ברשת מסומן גם למשימה 1 וגם ל-2 זמן סנכרון. עבור כל פרויקט, ברור שרק אחת מהמשימות 1 ו-2 מחכות, אך לעולם לא שתיהן. רשת ה-Gantt מציגה ממוצעים. ולכן, בממוצע, שתי המשימות ימתינו (וכנ"ל לגבי משימות 3,4).
2. כפי שרואים מהרשת, משימות 1,2 מסיימות תמיד את ביצועהן ביחד (כנ"ל לגבי משימות 3,4), ותמיד משימה 3 מתחילה לאחר משימות 1,2 ומשימה 4 לאחר משימה 2. האילוצים הללו מתקיימים עבור כל פרויקט, ולכן מתקיימים גם עבור הממוצע.

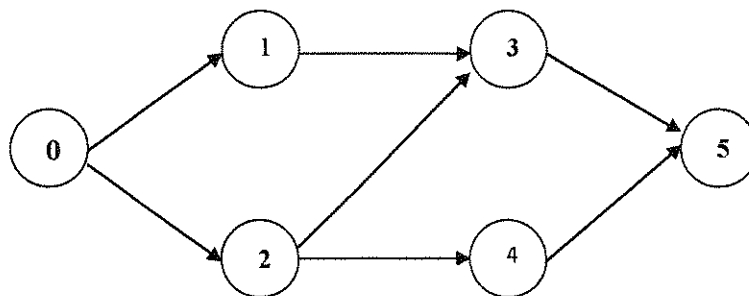
בסעיף הבא נביא דוגמת סימולציה. בדוגמא ניישם את שתי השיטות לחקירת פרופיל השהיה של הפרויקט במערכת ונשווה בינם. כמו כן נחקור את השפעת הפרמטרים השונים על תוחלת זמן השהיה, כגון: נצילות השרתים, ומקדמי השתנות. הדוגמא יכולה להוות מסגרת יעילה לחקירת מערכת תורים כלשהיא ולהצעות שיפור.

1.4 דוגמת סימולציה.

1.4.1 תאור המודל.

המערכת מכילה 4 תחנות שרות, כאשר בכל תחנה נמצא שרת יחיד ומתבצעת משימה שונה. פרויקט המגיע למערכת, מורכב מ- 4 משימות שונות, משימה בכל תחנה. את הקדימויות בין ביצוע המשימות נתאר בעזרת רשת PERT שלהלן.

לשם נוחות, הוספנו שתי תחנות שרות: תחנה 0 - ממנה הפרויקט מתפצל למשימותיו, ותחנה 5 - בה הפרויקט מסיים את עיבודו.



שרטוט 1.4.1 א: רשת PERT.
Figure 1.4.1 a: PERT Network.

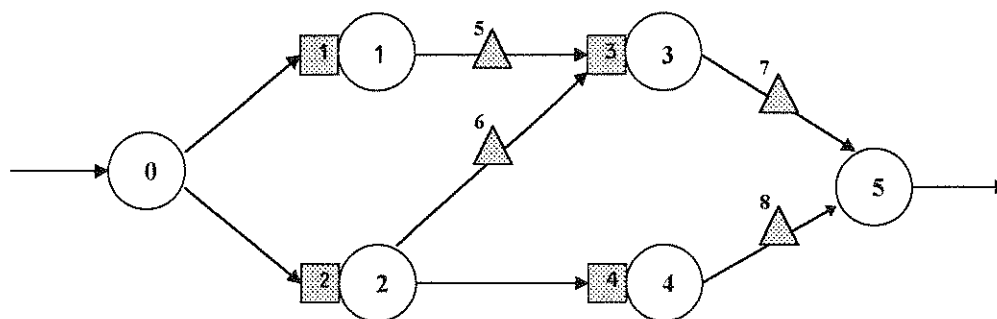
מופע הפרויקטים למערכת הוא פואסוני עם קצב $\lambda = 0.1$ וזמני השרות מפולגים Gamma (בהמשך, הפרמטרים יותאמו לניסויים).

כפי שהוסבר בפרק הקודם, נוצרים שני סוגי תורים: תורי סנכרון ותורי המתנה למשאבים.

להלן המערכת הדינמית:

■ - תור המתנה למשאבים.

▲ - תור סנכרון.



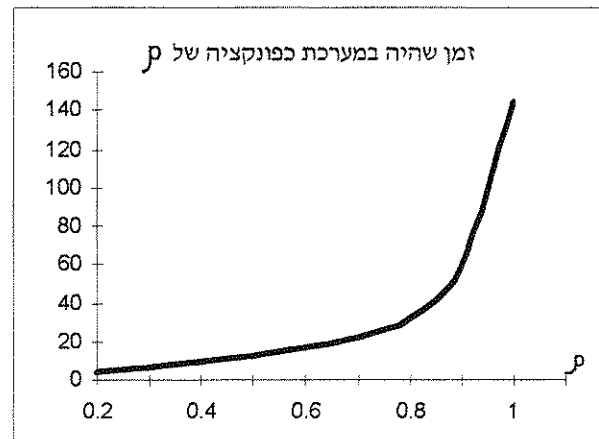
שרטוט 1.4.1 ב: המערכת הדינמית
Figure 1.4.1 b: Dynamic Model.

1.4.2 מערכת הומוגנית.

1.4.2.1 זמני שרות קבועים.

נעשו סימולציות (כפי שהוסבר בפרק 1.3.1) עבור מערכת בעלת 4 שרתים הומוגנים, כאשר זמני שרות בתחנות קבועים וידועים.

להלן זמן שהיה הממוצע של פרויקט במערכת כפונקציה של נצילות השרתים.



גרף 1.4.2.1 א: זמן שהית פרויקט במערכת הומוגנית כפונקציה של הנצילות.
Graph 1.4.2.1 a: Throughput Time in a Homogeneous System, as Function of Utilization.

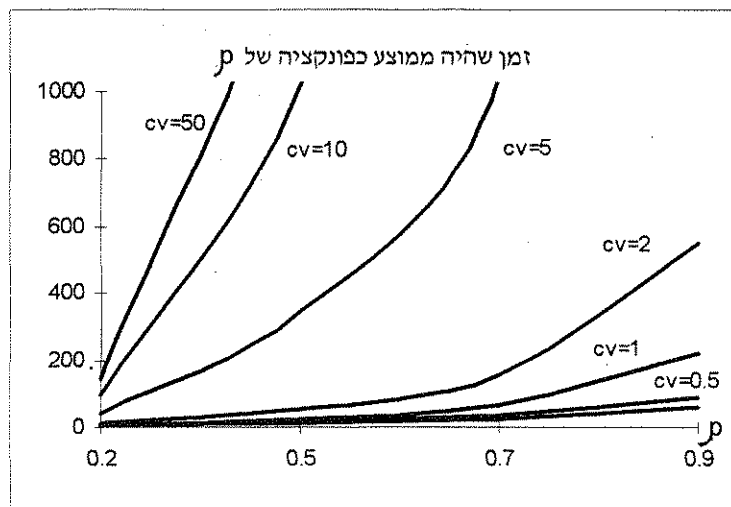
כפי שניתן לצפות מראש, ככל שנצילות השרתים עולה, עולה גם זמן שהיה הממוצע של פרויקט, (וכמובן גם תורי ההמתנה המככים לשרתים).

1.4.2.2 זמני שרות לא קבועים.

עבור מערכת לא דטרמיניסטית, נעשו סימולציות עבור הנצילות 0.2, 0.5, 0.7, 0.9, ועבור מקדמי שונות $cv = 0, 0.5, 1, 2, 10, 50$. בחרנו בהתפלגות Gamma עבור זמני השרות (בהסתמך על מאמרם של C.A. Leonhard, J.S. Davis [21]).

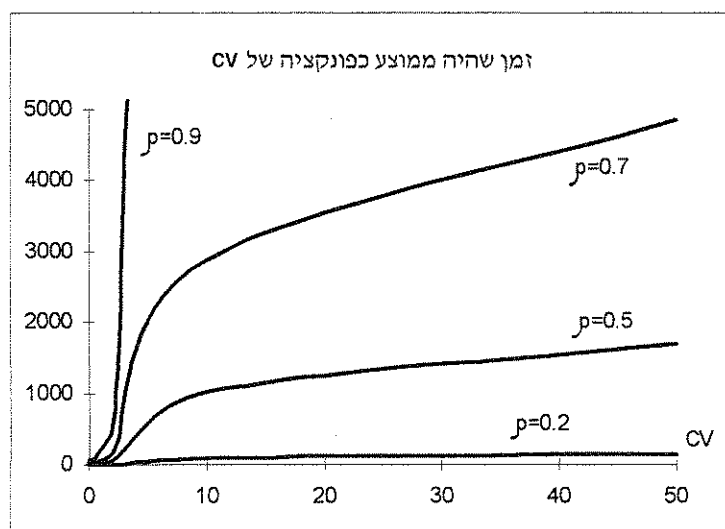
המאמר מוצא, בהסתמך על חקירת מפעלים, כי ההתפלגות הטובה ביותר לתאור זמני עיבוד משימה היא התפלגות Gamma.

להלן זמן שהיה הממוצע של פרויקט במערכת עבור כל מקדם השתנות, כפונקציה של הנצילות.



גרף 1.4.2.2 א: זמן שהיה פרויקט במערכת כפונקציה של נצילות ומקדם השתנות.
Graph 1.4.2.2 a: Throughput Time, as Function of Utilization, CV Fixed.

ניתן לראות כי עבור כל מקדם השתנות, זמן ההשהיה של הפרויקט עולה.



גרף 1.4.2.2 ב: זמן שהיה פרויקט במערכת כפונקציה של נצילות ומקדם השתנות.
Graph 1.4.2.2 b: Throughput Time, as Function of CV, Utilization Fixed.

מסקנות:

- עבור נצילות קבועה, ומקדם השתנות גדל - זמן השהיה בד"כ עולה.
- עבור מקדם השתנות קבוע, ונצילות גדלה - זמן השהיה עולה.
- ככל שמקדם ההשתנות גדל, והנצילות גדלה - קצב עליית זמן השהיה עולה.

מטרתנו היא לשפר את המערכת, כלומר: הורדת תוחלת זמן השהיה של פרויקט.

ע"ס הסימולציות והגרפים, ניתן לחקור את השגת המטרה בשני אופנים:

1. הורדת נצילות השרתים.

2. הקטנת מקדם השתנות השרות (ומאוחר יותר נראה כי גם הקטנת שונות המופע מורידה את זמן השהיה).

כפי שהוסבר בפרק 1.3, ניתן לחלק את זמן השהיה של פרויקט במערכת לסכום של שלושה

זמנים: זמן המתנה למשאב, זמן סנכרון של משימות וזמן עיבוד.

אנו נדגים את שתי השיטות שהוצעו על המערכת הנוכחית.

תכנון הסימולציה:

כל סימולציה כללה 50 חזרות. אורך כל חזרה נקבע בהתאם לפרמטרים של הסימולציה, כך

שנקבל אורך רווח סמך קטן מספיק.

גם הזמן עד שהמערכת מגיעה למצב יציב נקבע בכל סימולציה בהתאם לפרמטרים (ע"י ניסוי

ותהייה).

עבור כל פרויקט חושב אחוז הזמן שהשקיע בעיבוד, סינכרון והמתנה. בסיום כל חזרה חושבו

ממוצעי האחוזים ובסיום הסימולציה חושבו הממוצעים ורווחי הסמך (ברמת בטחון של 95%).

שיטה א'.

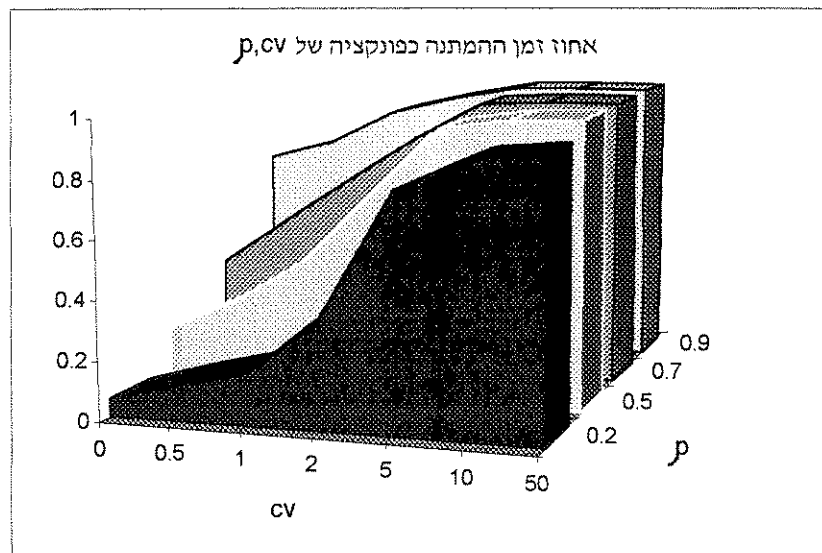
בחרנו לבצע את שני השלבים הראשונים אשר נאמרו בסעיף 1.3.3 :

1. הפיכת המערכת הדינמית למערכת סטוכסטית, וכך שחרור המערכת מתורי ההמתנה.
2. הפיכת המערכת הסטוכסטית לדטרמיניסטית, וכך שחרור המערכת מתורי הסנכרון שנוצרו עקב האופי הדינמי והסטוכסטי של המערכת.

הערה: עקב הומוגניות המערכת, ביטלנו באופן זה גם את תורי הסנכרון התפעולי.

בחרנו לא לבטל את הסנכרון התכני (כלומר לא לשנות את רשת הקדימויות לרשת מקבילית) כדי שיהיה ניתן להשוות עם שיטה ב', בה לא מתחשבים בסנכרון התכני.

נערכו סימולציות כדי לבדוק את השפעת מקדם ההשתנות של השרות והנצילות עבור מערכת הומוגנית על אחוז הזמן שהפרויקט משקיע בהמתנה, עיבוד וסנכרון. להלן תוצאות הסימולציות עבור אחוז זמן ההמתנה:

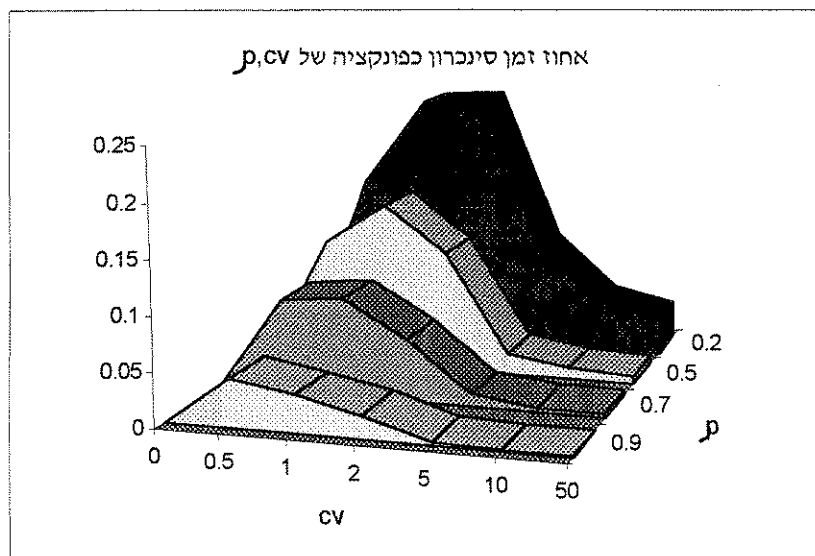


גרף 1.4.2.2 ג: אחוז זמן ההמתנה כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה א.
Graph 1.4.2.2 c: Percent of Waiting Time, as Function of Utilization & CV, according to the First Method.

מסקנות:

- עבור נצילות קבועה ו- cv גדל - אחוז זמן ההמתנה למשאבים עולה (העליה המשמעותית מתרחשת בערכי cv בינוניים).
- ככל שהנצילות עולה, קצב עליית אחוז ההמתנה קטן כפונקציה של מקדם ההשתנות (כלומר בנצילויות נמוכות עליה תלולה, ובנצילויות גבוהות עליה קלה).

להלן תוצאות הסימולציות עבור אחוז זמן הסנכרון:

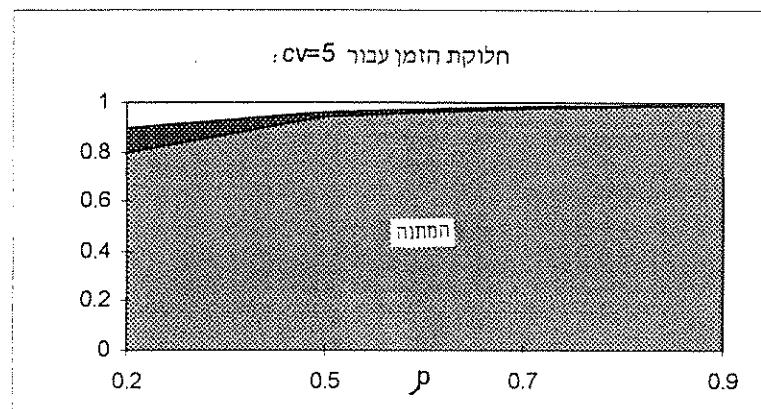
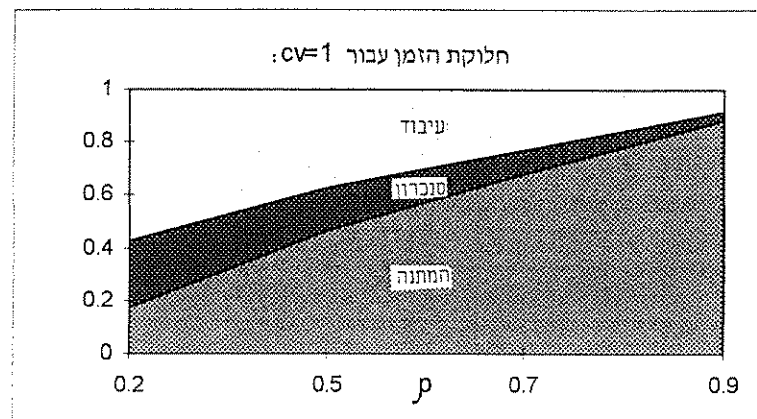
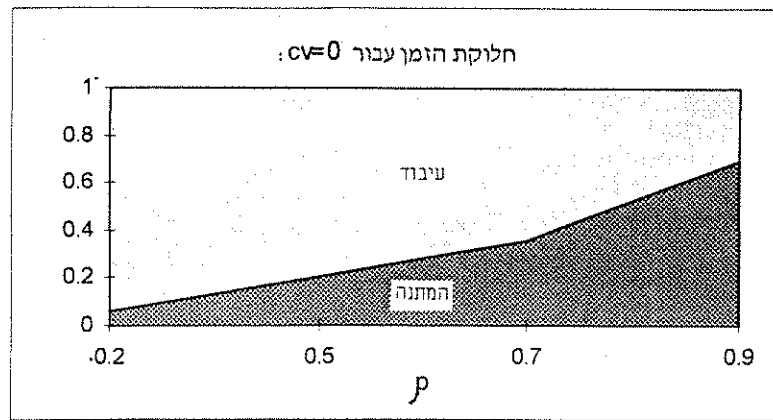


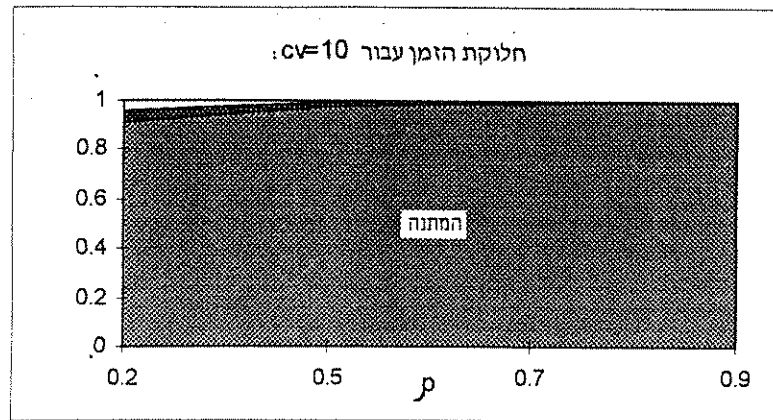
גרף 1.4.2.2 ד: אחוז זמן הסנכרון כפונקציה של נצילות ומקדם השתנות השרות, ע"פ שיטה א.
Graph 1.4.2.2 d: Percent of Synchronization Time, as Function of Utilization & CV, according to the First Method.

מסקנות:

- עבור נצילות קבועה, ו- cv עולה - אחוז זמן הסנכרון יורד.
- עבור מקדם השתנות קבוע, ונצילות עולה, אחוז זמן הסנכרון יורד.
- ככל שהנצילות עולה, קצב ירידת אחוז הסנכרון יורד.
- עבור מערכת דטרמיניסטית (שוונות 0), איו סנכרון (בגלל הומוגניות המערכת).
- עבור cv גבוהים, אחוז זמן הסנכרון אפסי.

על מנת להשוות את חלוקת הזמן עבור מקדם השתנות קבוע כפונקציה של הנצילות, נביא את הגרפים שלהלן:





גרף 1.4.2.2 ה: חלוקת הזמן במקדמי שונות שונים כפונקציה של הנצילות, ע"פ שיטה א.
Graph 1.4.2.2 e: Profile at Different CV, as Function of Utilization, according to the First Method.

מסקנות:

- במערכת בעלת $cv=0$, כלומר מערכת בעלת זמני שרות קבועים אין זמן סנכרון (בגלל הומוגניות השרתים). הזמן הדומיננטי הוא עיבוד והמתנה, המתרחשת כתוצאה ממופע פואסוני (ולא מופע דטרמיניסטי).
- ככל שמקדם ההשתנות עולה, אחוז זמן הסנכרון והעיבוד יורדים וזמן ההמתנה עולה.

סיכום שיטה א':

שיטה זו מחלקת את זמן השהיה של פרויקט ל-3 זמנים ע"י שחרור המערכת מאילוצי משאבים וסנכרון תוך מדידת הרווח ברמת הביצועים.

מאחר ואילוצי משאבים נגרמים עקב מחסור במשאבים, שחרור המערכת מאילוצים אלו התבצע ע"י הוספת אינסוף משאבים בכל תחנה.

הבעיה הייתה, מהם הגורמים לסנכרון. חלק מהסנכרון בין משימות, נגרם עקב דינמיות המערכת (השרות והמופע) וחלקו היה נשאר גם עם המערכת הייתה דטרמיניסטית לחלוטין.

בסעיף הקודם הזכרנו שסנכרון נוסף נגרם עקב אילוצי משאבים. בגלל שמשימות שונות, הצריכות להיות מסנכרות בעתיד, מחכות בתורים שונים (לדוגמא, משימה 1 ו-2), זמן סיום עיבוד המשימות שונה, ואז זמן הסנכרון עולה. אם היו מספיק משאבים, סנכרון זה לא היה נגרם, ולכן אולי נכון יותר לקרוא לו זמן המתנה (או זמן סנכרון "מדומה").

סנכרון נוסף שלא התייחסנו אליו בדוגמא זו, היה הסנכרון התכני הנגרם עקב אילוצי קדימות.

בשלב ראשון, שחררנו את המערכת מאילוצי המשאבים, כך שלא נותרו תורי המתנה (וכך, באופן אוטומטי, הורדנו את ה"סנכרון המדומה").

בשלב שני, הורדנו את אילוצי הסנכרון, ע"י הפיכת זמני השרות במערכת לזמנים קבועים (שווים). בגלל שהמערכת הומוגנית, הורדנו באופן אוטומטי את הסנכרון התפעולי (אך הסנכרון התכני נשאר).

המסקנה, כפי שראינו בגרפים, שזמן ההמתנה למשאבים מהווה חלק ארי מהזמן הכולל. לעומת זאת, אחוז זמן הסנכרון נמוך.

מטרתנו היא שיפור המערכת, כלומר: הורדת תוחלת זמן השהיה ו/או הגדלת אחוז זמן העיבוד. ע"פ שיטה זו, ע"מ לשפר את המערכת כדאי להשקיע את מירב המאמצים בהורדת זמן ההמתנה, לדוגמא ע"י הגדלת משאבים. הגדלה זו תוריד את זמן השהיה של הפרויקט, ותבטל תורי המתנה. יישאר זמן סנכרון המהווה אחוז קטן מהזמן הכולל.

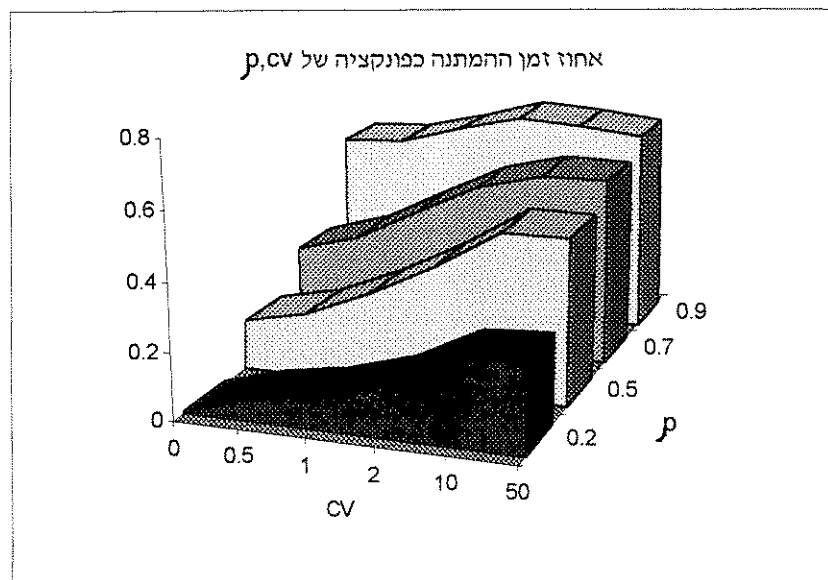
הערות:

- בכל השיטות, כפי שהוזכר, הקטנת זמן השהיה יושג ע"י הקטנת הנצילות ו/או הקטנת השונות.
- ניתן להמשיך ולשפר את המערכת ע"ס השלבים השונים עד קבלת מערכת בה זמן עיבוד מהווה 100%.

שיטה ב'.

נחשב עבור כל משימה את שלושת הזמנים (הם מוגדרים במדויק עבור כל משימה) ואז נסכם עבור כל המשימות.

ע"פ שיטה זו, סכום כל זמני שהית המשימות אינו זמן שהית הפרויקט במערכת אלא עלות שהית הפרויקט (לכל משימה משלמים יחידת כסף לכל יחידת זמן שהיה).
מטרתנו, עתה, לראות כיצד מתחלק זמן שהית הפרויקט לשלושת הזמנים הנ"ל.
להלן אחוז זמן ההמתנה של הפרויקט, כפונקציה של הנצילות ומקדם השתנות:

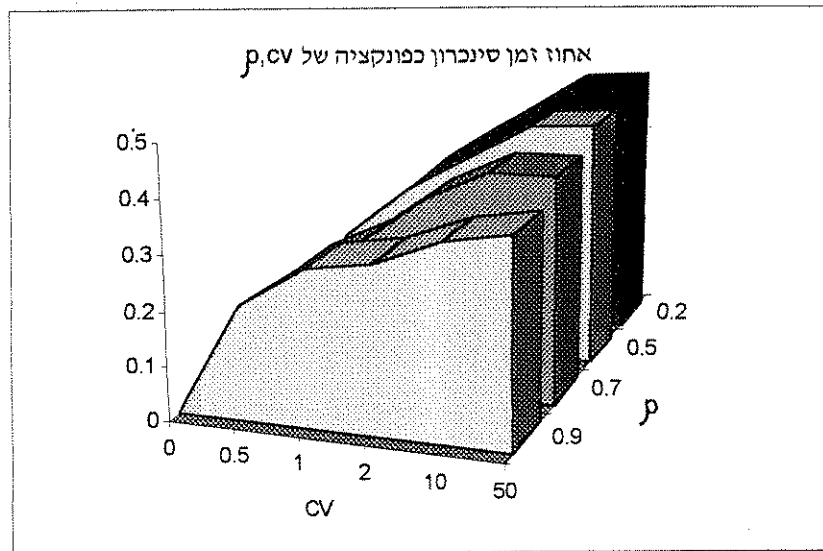


גרף 1.4.2.2 ו: אחוז זמן ההמתנה כפונקציה של הנצילות ומקדם השתנות השרות, ע"פ שיטה ב.
Graph 1.4.2.2 f: Percent of Waiting Time, as Function of Utilization & CV, according to the Second Method.

מסקנות:

- עבור נצילות קבועה, ו- cv גדל - אחוז זמן ההמתנה למשאבים עולה, עד למקדמי השתנות גבוהים, בהם אחוז ההמתנה מתייצב (ואפילו יורד מעט).
- קצב עלית אחוז זמן ההמתנה נמוך ואינה שונה בהרבה בין הנצילויות השונות.
- עבור CV קבוע, ונצילות עולה, אחוז זמן ההמתנה עולה. עבור מערכת בעלת זמני שרות קבועים ($cv = 0$) אחוז זמן ההמתנה גדל ככל שהנצילות עולה.

ניתן לשער שגם הנצילות וגם מקדם ההשתנות משפיעים על אחוז זמן ההמתנה כאשר הנצילות משפיע יותר מאשר מקדם ההשתנות.
להלן אחוז זמן הסנכרון של הפרויקט, כפונקציה של הנצילות ומקדם ההשתנות:



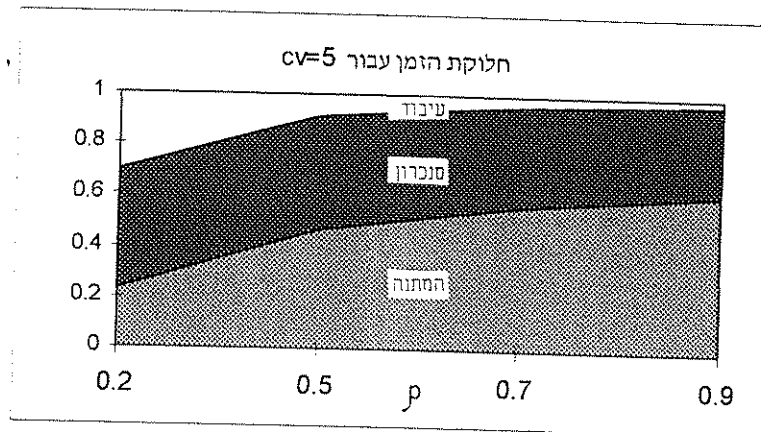
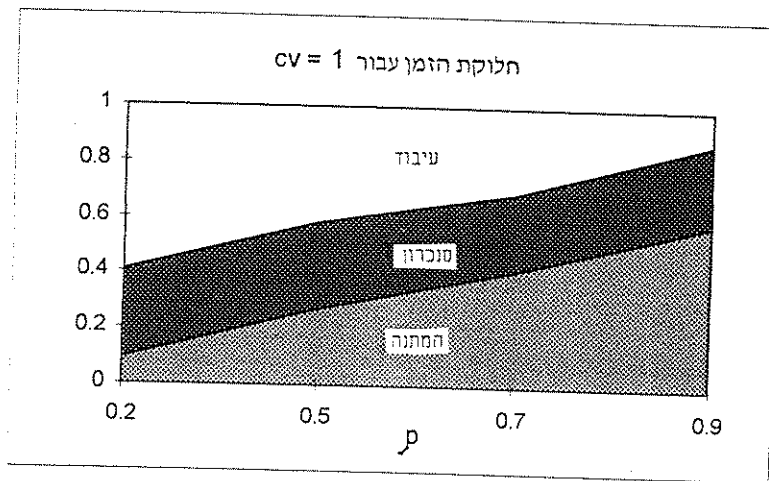
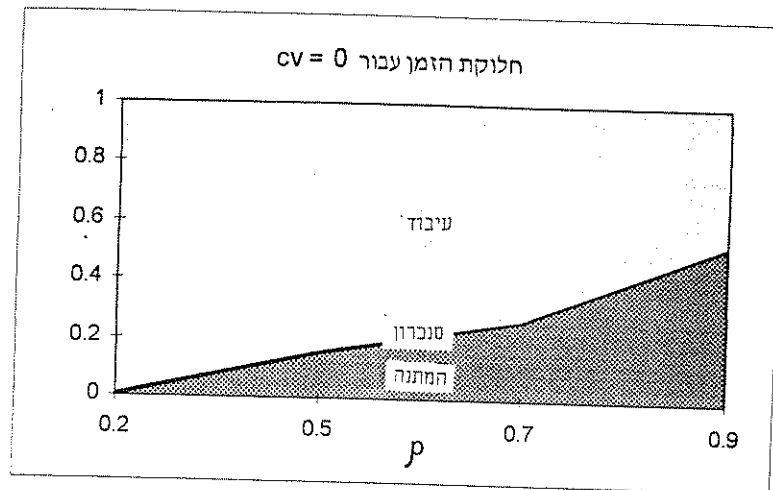
גרף 1.4.2.2 ז: אחוז זמן הסנכרון כפונקציה של הנצילות ומקדם השתנות השרות, ע"פ שיטה ב.
Graph 1.4.2.2 g: Percent of Synchronization Time, as Function of Utilization & CV, according to the Second Method.

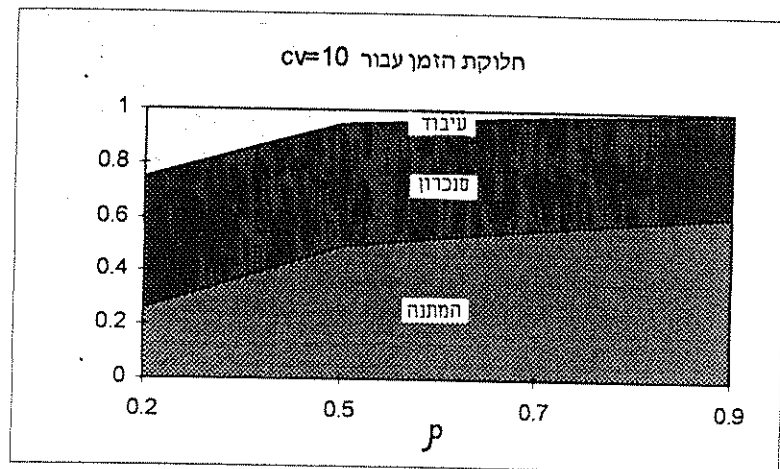
מסקנות:

- עבור נצילות קבועה, ו-cv עולה - אחוז זמן הסנכרון גדל.
- קצב עליית אחוז זמן הסנכרון גבוה (בייחוד במעבר בין מערכת עם $cv=0$ ומערכת עם cv נמוך) ואינה שונה בהרבה בין הנצילות השונות.
- עבור מקדם השתנות קבוע, ונצילות עולה, אחוז זמן הסנכרון יורד מעט.
- עבור מערכת $cv=0$, אחוז זמן הסנכרון אפסי, בכל הנצילות.
- עבור cv גבוהים, אחוז זמן הסנכרון מתייצב.

ההשערה היא שאחוז זמן הסנכרון מושפע ממקדם ההשתנות, ולא מהנצילות. עבור נצילות משתנה, אחוז זמן הסנכרון אינו משתנה באופן משמעותי. לעומת זאת עבור מקדם השתנות משתנה, האחוז מושפע.

כדי לבסס את ההשערות הנ"ל הבאנו את ארבעת הגרפים הבאים, אשר משווים עבור מקדם השתנות קבוע, את חלוקת הזמן (כפונקציה של הנצילות).





גרף 1.4.2.2 h: חלוקת הזמן במקדמי השתנות שונים כפונקציה של הנצילות, ע"פ שיטה ב.
Graph 1.4.2.2 h: Time Profile at Different CV, as Function of Utilization, according to the Second Method.

מסקנות:

- ככל שמקדם ההשתנות גדל, והנצילות גדלה - אחוז זמן העיבוד קטן.
- עבור מערכת בעלת מקדם השתנות נמוך, זמן ההמתנה והעיבוד דומיננטיים. דומיננטיות זו יורדת ככל שמקדם ההשתנות עולה, ולבסוף מגיעים לזמן המתנה וסנכרון דומיננטיים, וזמן עיבוד אפסי.
- עבור כל המערכות (cv קבוע), אחוז זמן הסנכרון נשאר קבוע אך אחוז ההמתנה עולה משמעותית על חשבון אחוז העיבוד. לכן ניתן להסיק כי הנצילות אינה משפיעה על אחוז הסנכרון אך לנצילות יש השפעה רבה על אחוז ההמתנה.

סיכום שיטה ב'.

בשיטה זו הגדרת הזמנים היא מנקודת מבט תפעולית. אנו הגדרנו מדד ביצועים השווה לסכום שהיית המשימות בתחנות. ניתן לחלק את זמן השהיה בכל משימה באופן חד-משמעי לזמן עיבוד, המתנה וסנכרון. עתה, נסכם את שלושת הזמנים עבור כל המשימות. שיטה זו מתאימה למודל בו ישנם 3 סוגי עלויות: עלות עיבוד משימה, עלות המתנה של משימה למשאב ועלות המתנה של משימה מעובדת (סנכרון). הערה: בשיטה זו הסנכרון התכני אינו נלקח בחשבון.

ע"ס תוצאות הסימולציות והגרפים שהובאו מסיקים כי אחוז זמן הסנכרון נשאר קבוע, כפונקציה של הנצילות. לכן הקטנת הנצילות לא תשפיע על אחוז זמן הסנכרון אלא על זמן ההמתנה (שירד ככל שהנצילות קטנה).

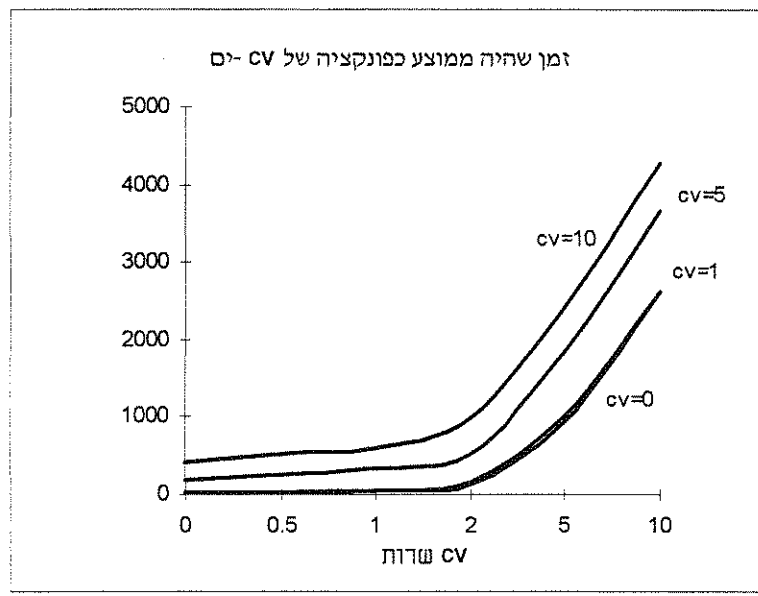
ע"מ להקטין את אחוז הסנכרון, כדאי להוריד את מקדמי ההשתנות של השרות. ע"מ לשפר את אחוז ההמתנה למשאבים, כדאי להשקיע במקדם ההשתנות אם המערכת בעלת נצילות נמוכה. עבור מערכת בעלת נצילות גבוהה, הקטנת מקדם ההשתנות לא תתרום להורדת אחוז ההמתנה, ולכן צריך להוריד נצילות.

אם נשווה את תוצאות שתי השיטות, ניתן לראות כי בשיטה ב' אחוז ההמתנה למשאבים רב, אך גם אחוז זמן הסינכרון גבוה.

בניגוד לכך, שיטה א' מדגישה את המשקל הרב של ההמתנה למשאבים. חלק מהסיבה להבדל בין שתי השיטות הוא אחוז הסינכרון "המדומה" (אותו סינכרון הקיים בגלל המתנת המשימות למשאבים). שיטה א' משייכת אחוז זה להמתנה ושיטה ב' - לסינכרון. עדיין, המסקנה העיקרית משתי השיטות היא שכדי לשפר את ביצועי המערכת רצוי להשקיע יותר בהוספת משאבים.

1.4.2.3 מופע לא קבוע.

עד עתה, סקרנו את ההשפעות שנבעו מנצילות שונות ומשרות בעל מקדם השתנות משתנה. על מנת להשלים את התמונה, צריכים גם לבדוק מהי השפעת המופע על תוחלת זמן ההשהיה. נערכו סימולציות בהן ממוצע הזמן הבין מופעי נשאר קבוע, או מקדם ההשתנות שלו השתנה. בכל הסימולציות הנ"ל לקחנו מערכת הומגנית עם נצילות שרתים 0.7. להלן זמן שהיה ממוצע של פרויקט במערכת כפונקציה של מקדם השתנות השרות והמופע:

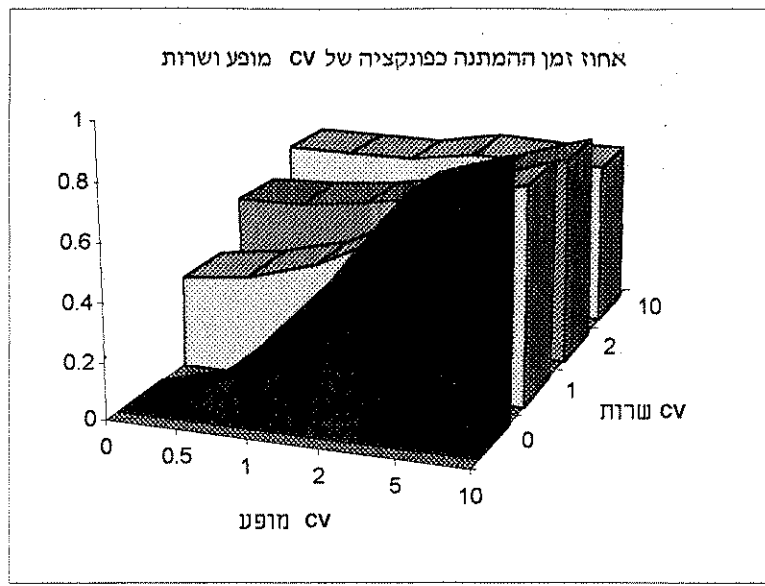


גרף 1.4.2.3 א: זמן שהיה ממוצע של פרויקט כפונקציה של מקדמי השתנות השרות והמופע.
Graph 1.4.2.3 a: Average Throughput Time, as Function of CV's.

מסקנות:

- עבור שרות קבוע, ו-CV של המופע עולה - זמן ההשהיה הממוצע של פרויקט במערכת עולה, אך לא בהרבה (בניגוד להשפעת CV של השרות).
- עבור מופע קבוע, ושרות משתנה - זמן ההשהיה עולה, אך קצב העלייה בכל המופעים דומה.

נבדוק עתה את ההשפעה של המופע והשרות על אחוז זמן ההמתנה והסנכרון, לפי השיטה השניה. להלן אחוז זמן ההמתנה:

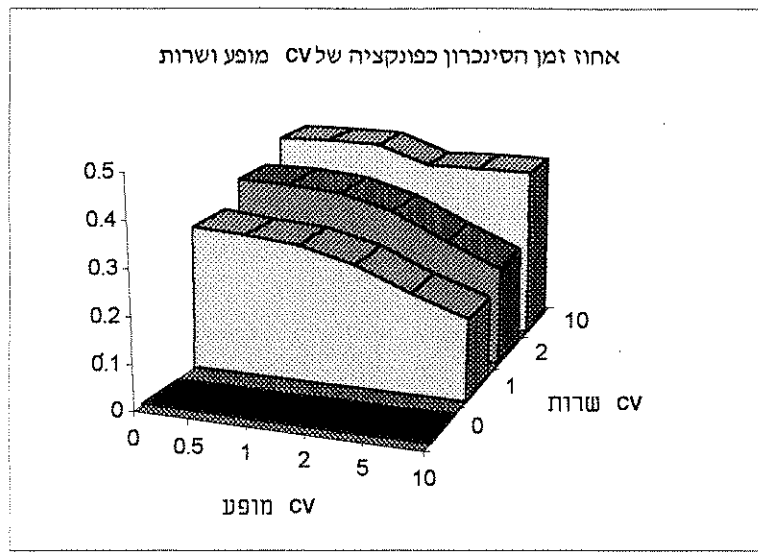


גרף 1.4.2.3 ב: אחוז זמן ההמתנה כפונקציה של מקדמי ההשתנות, ע"פ שיטה ב.
 Graph 1.4.2.3 b: Percent of Waiting Time, as Function of CV's, according to the Second Method.

מסקנות:

- השפעה חזקה של המופע מתבטאת בעיקר עבור מקדם שרות נמוך וההשפעה הולכת וקטנה במקדמי שרות גבוהים.
- והמקרה ההפוך: השפעה גדולה של השרות מתבטא בעיקר עבור מקדם מופע נמוך, וההשפעה הולכת וקטנה במקדם מופע גבוה.
- במערכת דטרמיניסטית אחוז ההמתנה אפסי והוא מגיע ל- 90% עבור מקדמי מופע גבוהים.

להלן השפעת המופע והשרות על אחוז זמן הסנכרון:



גרף 1.4.2.3 ג: אחוז זמן הסנכרון כפונקציה של מקדמי ההשתנות, ע"פ שיטה ב.
 Graph 1.4.2.3 c: Percent of Synchronization Time, as Function of CV's, according to the Second Method.

מסקנות:

- השפעה קטנה של מקדם המופע על אחוז זמן הסנכרון (בד"כ במגמה להוריד את אחוז הסנכרון).
- השפעה חזקה של מקדמי השרות.

מסקנות כלליות.

- אחוז זמן הסנכרון אינו מושפע, כמעט, ממקדם המופע. ולכן, כחיזוק למסקנה הקודמת, ע"מ להוריד את אחוז הסנכרון כדאי להקטין מקדמי השתנות של השרות.
- לעומת זאת, עבור שונות שרות נמוכה, שונות המופע (והנצילות) משפיעים רבות על אחוז ההמתנה. עבור מערכת בעלת שונות שרות גבוהה, אין השפעה לשונות המופע, וכדאי להוריד נצילות.

1.4.3 מערכת לא הומוגנית.

נעשו סימולציות עבור מערכת לא הומוגנית, בה לפחות אחד מהשרתים (כלומר אחת התחנות כיוון שבכל תחנה שרת יחיד) בעל נצילות של 0.5 או 0.7 או 0.9 ושאר השרתים נשארים עם נצילות של 0.2.

נחלק את השרתים לשתי הקבוצות הנ"ל: קבוצה אחת בעלת נצילות 0.2, וקבוצה אחרת בעלת אחת מהנצילויות 0.5, 0.7, 0.9.

נסמן שרת בעל הנצילות הנמוכה ב-A, ושרת בעל הנצילות הגבוהה יותר, ב-B.

במערכת ישנם 4 שרתים, המסומנים כוקטור בעל 4 רכיבים:

שרת 1	שרת 2	שרת 3	שרת 4
-------	-------	-------	-------

מסתבר, מהסימולציות, כי זמן שהיה הממוצע של פרויקט במערכת יכול לקבל שני ערכים בלבד.

זמן שהיה נמוך, יחסית, יתקבל במקרים הבאים:

BAAA AABA ABAA AAAB

BBAA AAB

BAAB

זמן שהיה גבוה יותר יתקבל עבור שאר האפשרויות:

BABA ABAB ABBB BBBA

BABB BBAB ABBA BBBB

ננסה למצוא הסבר הגיוני לתוצאה שקיבלנו.

במערכת, ניתן לראות הקבלה מסוימת בין שרת 1 ושרת 2, והקבלה בין שרת 3 ושרת 4.

נגדיר, אינדוקטיבית, שלב עיבוד של שרת (תחנה) i : DIS_i

$$DIS(0) = 0$$

$$DIS(i) = \max_{j \in p(i)} (DIS(j)) + 1$$

פרוש אינטואיטיבי ל- $DIS(i)$ יהיה מספר שלבי העיבוד שצריך הפרויקט לעבור כדי להגיע לשרת

(מרחק מקסימלי בקשתות מצומת הכניסה -0).

נחלק את השרתים לפי שלבי העיבוד :

קבוצה בשלב 1: {1,2}.

קבוצה בשלב 2: {3,4}.

מתוצאות הסימולציה ניתן לראות, כי בכל המערכות שבהן השרתים בעלי ניצולת גבוהה שייכים כולם לאותו שלב (1 או 2 בדוגמא), זמני השהיה זהה. (ללא תלות במספר השרתים בעלי ניצולת גבוהה או בקבוצה שאליה הם משתייכים).
וכן אם השרתים בעלי ניצולת גבוהה ישתייכו ליותר מקבוצה אחת, גם זמן השהיה הממוצע יהיה זהה.

ההסבר לכך הוא פשוט והגיוני: שרת 1 ו-2 מקבילים, ולכן זה לא משנה אם אחד מהם בעל ניצולת גבוהה או שניהם, ההמתנה תעשה בשלב הראשון והעיבוד יוכל להמשיך רק אחרי ששניהם גמרו את עיבודם.

כנ"ל לגבי שרתים 3 ו-4, שם ההמתנה תעשה בשלב השני.
כאשר שרתים בעלי ניצולת גבוהה ממוקמים בקבוצות שונות, הרי ההמתנה נעשית בשני שלבי העיבוד ולכן ההמתנה ארוכה יותר.

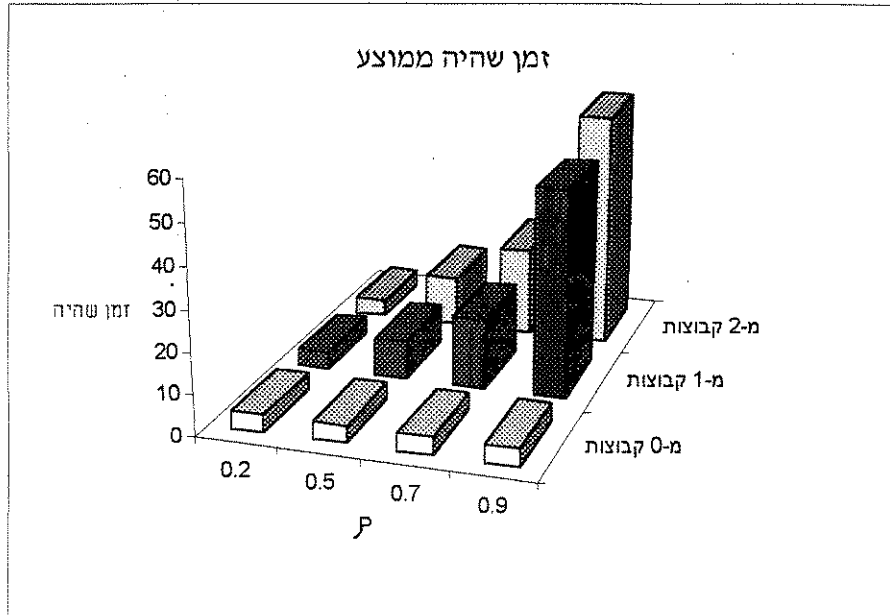
הכלל שנוסח למעלה היה מדויק אם המערכת הייתה סימטרית לחלוטין.
העובדה שהמערכת אינה סימטרית, מוסיפה לזמן השהיה הנמוך את המקרה BAAB (למרות ששרת 1 ושרת 4 אינם מאותה קבוצה).
גם לכך יש הסבר הגיוני: שרת 2 בעל ניצולת נמוכה, ולכן הפרויקט יכול להמשיך לשרת 4, למרות ששרת 1 עוד לא סיים. אבל שרת 4 בעל ניצולת גבוהה, וזמן השהיה הנגרם עקב כך לפרויקט חופף את זמן השהיה בשרת 1, ולכן בממוצע, הפרויקט ממתיין רק שלב אחד (ולכן הוא שייך לקבוצת זמן ההמתנה הנמוך יותר).

להלן זמן השהיה הממוצע של פרויקט במערכת לא הומוגנית, עבור שלוש האפשרויות :

א. כל השרתים בעלי ניצולת 0.2 (בגרף מסומן מ-0 קבוצות).

ב. השרתים בעלי ניצולת גבוהה יותר בעלי אותו שלב עיבוד ולשם הפשטות נוסיף גם את המקרה BAAB (בגרף מסומן מ-1 קבוצות).

ג. השרתים בעלי ניצולת גבוהה יותר בעלי שלבי עיבוד שונים (בגרף מסומן מ-2 קבוצות).



גרף 1.4.3 א: זמן שהיה ממוצע במערכת לא הומוגנית.
Graph 1.4.3 a: Average Throughput time in Nonhomogeneous System.

דוגמת סימולציה זו יכולה להוות מסגרת ומודל לחקירה אמפירית של רשת המבוססת על סימולציות תוך התייחסות לפרמטרים שונים, כגון: נצילות, מקדמי השתנות ופרופיל הזמן.

כמובן שניתוח אנליטי מדויק של הרשת היה רצוי, אך בד"כ לא ניתן לבצע ניתוח זה (אפילו עבור מבנה פרויקט פשוט הניתוח יתכן רק במקרים נדירים).

לכן, בפרק הבא (פרק 2) בחרנו לנתח אנליטית מבנה מיוחד, פשוט וסמטרי של פרויקט תוך התייחסות למושגים שהוזכרו: המתנה למשאבים וסינכרון.

הניתוח משתמש בחסמים אנליטיים לתוחלת זמן השהיה של פרויקט המערכת ומנסה לקבל קרובים אנליטיים לתוחלת זו בעזרת החסמים וסימולציות.

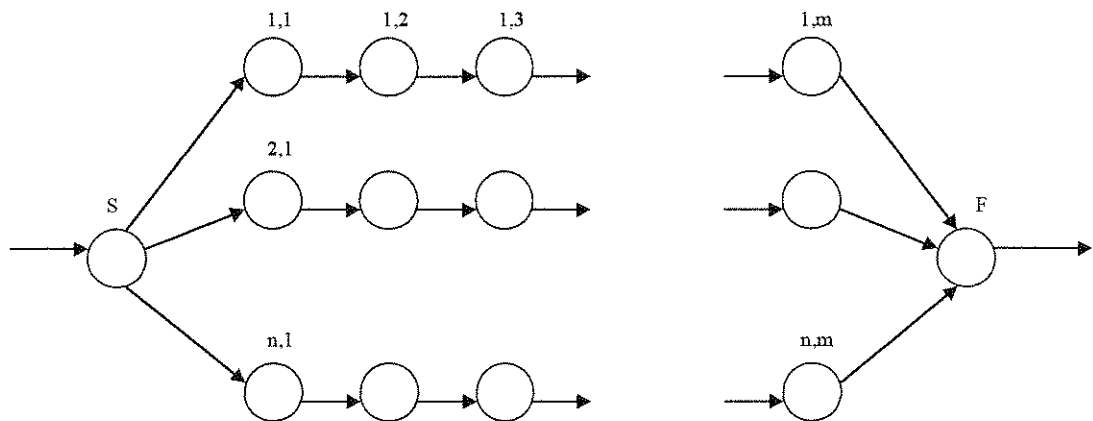
2. מערכת טורים במקביל.

2.1. תאור המערכת.

2.1.1. המודל.

נתונה מערכת המורכבת מפרויקטים ומשאבים.

להלן שרטוט רשת הפרויקט המתאר את המערכת.



שרטוט 2.1.1 א: רשת ה- PERT.

Figure 2.1.1 a: PERT Network.

כל פרויקט מכיל m קבוצות של משימות, המעובדות במקביל. כל טור i ($i=1, \dots, n$) מכילה m משימות

המעובדות בזה אחר זה. סה"כ מכיל הפרויקט $n \times m$ משימות.

נשתמש בסימון (i, j) לציון משימה j בקבוצה i (או טור i).

כמו כן, נסמן ב- S את התחנה הראשונה, ממנה מתפצל הפרויקט, ובאות F את התחנה הסופית המציינת

את סיום הפרויקט.

זמני ביצוע המשימות הם m ב"ת ש"ה מפולגים מעריכי עם קצב μ (ממוצע $1/\mu$), ומופע הפרויקטים

מהווה תהליך חידוש עם קצב λ . זימון המשימות מתרחש לפי מדיניות FCFS.

תנאי הכרחי ומספיק ליציבות המערכת הוא $\lambda < \mu$ (ראה Towsley, Baccelli, Massey [7]).

סימונים:

$s_k(i,j)$ - זמן ביצוע משימה (i,j) של הפרויקט ה- k (כפי שנאמר, מפולג מעריכי עם קצב μ).

$\tau_k, k=0,1,2,\dots$ - זמני מופע הפרויקטים למערכת.

$t_{k+1} = \tau_{k+1} - \tau_k$ - הזמן בין מופע k למופע $k+1$ (מ"מ ב"ת ש"ה).

$W_k(i,j)$ - זמן המתנה של הפרויקט ה- k בתור (i,j) (בהמתנה לביצוע משימה (i,j)).

$X_k(i,j) = W_k(i,j) + s_k(i,j)$ - זמן שהית פרויקט k בתחנה (i,j) , המתנה בתור + זמן עיבוד.

$$T_k(i) = \sum_{j=1}^m X_k(i,j) \quad \text{- זמן שהית פרויקט } k \text{ בטור } i.$$

$R_k(i,j) = X_k(i,1) + \dots + X_k(i,j)$ - הזמן שלוקח לפרויקט k , בטור i לסיים את משימה j .

2.1.2 משוואות Lindley (Lindley [23]).

משוואות Lindley מתייחסות למערכת $GI/G/1$. במקרה זה פרויקט מכיל משימה בודדת.

W_k - זמן המתנה בתור של הפרויקט ה- k .

X_k - זמן עיבוד הפרויקט ה- k .

t_k - הזמן הבין-מופעי בין הפרויקט ה- k והפרויקט ה- $k-1$.

$$W_{k+1} = \begin{cases} W_k + s_k - t_{k+1} & \text{if } t_{k+1} < W_k + s_k \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

בצורה פשוטה יותר: $W_{k+1} = [W_k + s_k - t_{k+1}]^+$

כאשר $[x]^+ = \max(x, 0)$.

2.1.3 הרחבת משוואות Lindley.

משוואות Lindley מתייחסות לתחנות שרות אחת.

ניתן להרחיב את המשוואות למספר תחנות המסודרות בטור.

הערה: זמן השהיה של פרויקט במערכת הוא מקסימום בין זמני השהיה של הפרויקט בטורים. כאשר פרויקט מגיע למערכת, הוא מתפצל מיד ל- n טורים במקביל, ואין השפעה של טור אחד על משנהו. לכן הרחבת המשוואות יכולה להתייחס רק למשימות המעובדות בטור.

המשוואה המורחבת:

לכל k ולכל j , מתקיים:

$$W_{k+1}(i, j) = \left[\sum_{l=1}^j X_k(i, l) - \sum_{l=1}^{j-1} X_{k+1}(i, l) - t_{k+1} \right]^+ \quad (2)$$

הוכחה:

נוכיח בעזרת אינדוקציה כפולה על מספר התחנות בטור (j) ועל מספר הפרויקטים (k).

בסיס: $j=1$.

נציב $j=1$ במשוואה:

$$W_{k+1}(i, 1) = [X(i, 1) - t_{k+1}]^+ = [W_k(i, 1) + s(i, 1) - t_{k+1}]^+ \quad (3)$$

עבור מערכת בעלת משימה אחת בכל טור, המשוואה המתקבלת מקבילה למשוואת Lindley המקורית

ונכונה עבור כל k .

צעד: נניח נכונות המשוואה עבור j (לכל k), ונוכיח עבור $j+1$.

הוכחה: נוכיח באינדוקציה על מספר הפרויקטים k .

בסיס: $k=1$. (עבור משימה $j+1$).

$$W_1(i, j+1) = \left[\sum_{l=1}^{j+1} X_0(i, l) - \sum_{l=1}^j X_1(i, l) - t_1 \right]^+ = 0$$

זמני ההמתנה למשימות.

צעד: נניח נכונות עבור פרויקט k ונוכיח עבור פרויקט $k+1$.

הוכחה: ע"ס הנחת האינדוקציה, $W_k(i,1), W_k(i,2), \dots, W_k(i,j+1)$ זמני ההמתנה של הפרויקט ה- k בטור ה- i .

זמני השרות של הפרויקט התחנות הם $s_k(i,1), s_k(i,2), \dots, s_k(i,j+1)$.

מכאן, $\sum_{l=1}^{j+1} X_k(i,l) = W_k(i,1) + s_k(i,2) + \dots + W_k(i,j+1) + s_k(i,j+1)$ זהו הזמן שהפרויקט

ה- k גמר את עיבוד המשימה ה- $j+1$.

באותו אופן, $\sum_{l=1}^j X_{k+1}(i,l)$ - זמן שהפרויקט ה- $k+1$ גומר עיבוד משימה j .

פרויקט $k+1$ מתחיל t_{k+1} יח"ז לאחר פרויקט k , לכן אם $\sum_{l=1}^{j+1} X_k(i,l) < \sum_{l=1}^j X_{k+1}(i,l) + t_{k+1}$

אזי פרויקט $k+1$ ימצא את תחנה $j+1$ ריקה, ו- $W_{k+1}(i,j+1) = 0$, אחרת פרויקט $k+1$ ימתין עד פינוי התחנה.

כלומר: $W_{k+1}(i,j+1) = \left[\sum_{l=1}^{j+1} X_k(i,l) - \sum_{l=1}^j X_{k+1}(i,l) - t_{k+1} \right]^+$. מ.ש.ל.

2.2 תוצאות אנליטיות.

2.2.1 משתנים מקריים קשורים.

הגדרה.

משתנים מקריים $X = \{X_1, X_2, \dots, X_n\}$ נקראים "קשורים" (Associated random variables) אם $\text{cov}[f(X), g(X)] \geq 0$ עבור כל זוג פונקציות עולות $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^1$ (פונקציה עולה הכוונה לא יורדת).

תכונות מ"מ קשורים (ראה Baccelli, Makowski [4]).

(P1) קבוצה הנוצרת מפונקציות עולות של מ"מ קשורים - מהווה גם היא קבוצת מ"מ קשורים.

(P2) קבוצה הנוצרת מאיחוד שתי קבוצות בלתי תלויות של מ"מ קשורים מהווה גם היא קבוצת מ"מ קשורים.

(P3) מ"מ בלתי-תלויים - קשורים.

(P4) עבור $X = \{X_1, X_2, \dots, X_n\}$ מ"מ קשורים מתקיים האי שוויון הבא:

$$P \left[\max_{1 \leq i \leq n} X_i > x \right] \leq 1 - \prod_{i=1}^n (P[X_i \leq x]) \quad (4)$$

2.2.1.3 הוכחה על קשירות זמני שהיה.

משפט 1: $\{W_i(i,j), s_i(i,j), -t_i \mid i=1,2,\dots,n \ j=1,2,\dots,m \ l=1,2,\dots,k\}$ הם קבוצת מ"מ קשורים.

הוכחה: נוכיח באינדוקציה כפולה, על מספר המשימות (j), ומספר הפרויקטים (k). (בדומה להוכחה המופיעה ב- (6)).

בסיס: $j=1$ לכל k. הוכחה באינדוקציה על k:

בסיס: $k=1$ ($j=1$). $W_i(i,1) = 0$ לכל $i=1,2,\dots,n$. הקבוצה $\{s_i(i,1) \mid i=1,2,\dots,n\}$ היא קבוצת

מ"מ ב"ת ולכן ע"פ תכונה P3 קשורים. t_1 מ"מ ב"ת בשאר, ולכן ע"ס תכונה P2 הקבוצה

$\{W_i(i,1), s_i(i,1), -t_1 \mid i=1,2,\dots,n\}$ מ"מ קשורים.

צעד: נניח שקבוצת המ"מ $\{W_i(i,1), s_i(i,1), -t_1 \mid i=1,2,\dots,n \ l=1,2,\dots,k\}$ קשורים לכל k, ונוכיח עבור $k+1$.

הוכחה: ההוכחה נובעת מיידית ממאמרם של Nelson, Tantawi (1988) [25], נספח A וממאמרם

של Baccelli, Makowski (1988) [5] פרק III.4.2.

צעד: נניח שקבוצת המ"מ $\{W_i(i,r), s_i(i,r), -t_1 \mid i=1,2,\dots,n \ r=1,2,\dots,j \ l=1,2,\dots,k\}$ קשורים לכל k, ונוכיח עבור $j+1$.

הוכחה : באינדוקציה על k .

בסיס : $k=1$ (עבור $j+1$). עבור הפרויקט הראשון, $W_1(i,j+1)=0$ לכל i , ולכן הם ב"ת וע"ס תכונה

$P1, P2$ קבוצת המ"מ

$\{W_l(i,r), s_l(i,r), -t_l, W_l(i,j+1), s_l(i,j+1) \mid i=1,2,\dots,n \quad r=1,2,\dots,j \quad l=1,2,\dots,k\}$ קשורים לכל k .

צעד : נניח עבור $j+1$ ו- $l=1,2,\dots,k$ ונוכיח עבור $k+1$.

הוכחה : ע"ס משוואות Lindley המורחבות :

$$\begin{aligned} W_{k+1}(i,j+1) &= \left[\sum_{l=1}^{j+1} X_k(i,l) - \sum_{l=1}^j X_{k+1}(i,l) - t_{k+1} \right]^+ \\ &= \left[\sum_{l=1}^{j+1} (W_k(i,l) + s_k(i,l)) - \sum_{l=1}^j (W_{k+1}(i,l) + s_{k+1}(i,l)) - t_{k+1} \right]^+ \end{aligned}$$

מאחר והפונקציה $x \rightarrow [x]^+$ מונוטונית לא יורדת, וכל המ"מ הרלוונטים קשורים, ע"פ הנחת

אינדוקציה, הרי לפי תכונה $P3$ גם $W_{k+1}(i,j+1)$ קשורים.

עתה, הקבוצה $\{s_{k+1}(i,j+1) \mid i=1,2,\dots,n\}$ ב"ת ולכן לפי $P1$ קשורים, ולבסוף לפי $P2$ קבוצת

מ"מ $\{W_l(i,r), s_l(i,r), -t_l \mid i=1,2,\dots,n \quad r=1,2,\dots,j+1 \quad l=1,2,\dots,k+1\}$ לכל k . מ.ש.ל

משפט 2 : קבוצת המ"מ $\{T_k(i) \mid i=1,2,\dots,n\}$ לכל k , קשורים.

הוכחה : קבוצה חלקית של מ"מ קשורים גם קשורים. ע"ס הטענה הקודמת, ניקח את הקבוצה החלקית

$\{W_k(i,j), s_k(i,j) \mid i=1,2,\dots,n \quad j=1,2,\dots,m\}$.

$T_k(i) = \sum_{j=1}^m (W_k(i,j) + s_k(i,j))$ $i=1,2,\dots,n$ פונקציות לא יורדות של מ"מ קשורים ולכן גם קשורים.

2.2.1.3 תוצאות ומסקנות.

מסקנה 1: זמני השהיה של הפרויקט בקבוצות השונות הם מ"מ קשורים, ולכן ע"ס P4 מתקיים:

$$P \left[\max_{1 \leq i \leq n} T_k(i) > t \right] \leq 1 - \prod_{i=1}^n (P[T_k(i) \leq t]) \quad (5)$$

על מנת להבין את משמעות אי השוויון נשווה בין שתי מערכות: מערכת אחת המתוארת בסעיף 2.1.1 והמערכת השניה זהה לה מלבד מופע לא מסונכרן, כלומר μ מופעי חידוש ב"ת ש"ה (מופע i לטור i , $i=1,2,\dots,n$).

לצורך ההשוואה נשתמש במספר סימונים נוספים:

$T_k(n,m)$ - זמן השהיה של פרויקט k -י במערכת הראשונה, בעלת המופע המסונכרן.

$T_k^*(n,m)$ - זמן השהיה של פרויקט k -י במערכת השניה, בעלת μ מופעים ב"ת.

$T(n,m), T^*(n,m)$ - זמן השהיה של פרויקט במערכות במצב יציב, בהתאמה.

$$\begin{aligned} \lim_{k \rightarrow \infty} T_k(n,m) &= T(n,m) \\ \lim_{k \rightarrow \infty} T_k^*(n,m) &= T^*(n,m) \end{aligned} \quad (6)$$

כאשר הגבול נלקח בהתפלגות.

$T(i), X(i,j)$ - זמן השהיה של פרויקט בתחנה (i,j) ובטור (i) במצב היציב, בהתאמה.

$T^*(i), X^*(i,j)$ - זמן השהיה של פרויקט בתחנה (i,j) ובטור (i) במצב היציב, בהתאמה.

$$\begin{aligned} \lim_{k \rightarrow \infty} X_k(i,j) &= X(i,j) \\ \lim_{k \rightarrow \infty} T_k(i) &= T(i) \\ \lim_{k \rightarrow \infty} X_k^*(i,j) &= X^*(i,j) \\ \lim_{k \rightarrow \infty} T_k^*(i) &= T^*(i) \end{aligned} \quad (7)$$

$T^*(i), T(i); X^*(i,j), X(i,j); T_k^*(i), T_k(i)$;

$(X_k^*(i,j), X_k(i,j))$.

נסמן:

$\bar{T}(n, m)$, $\bar{T}^*(n, m)$ - תוחלת זמן שהיה במערכת במצב יציב, בהתאמה.

ע"ס מבנה המערכות:

$$\begin{aligned} T_k(n, m) &= \max_{1 \leq i \leq n} \{T_k(i)\} \\ T_k^*(n, m) &= \max_{1 \leq i \leq n} \{T_k^*(i)\} \end{aligned} \quad (8)$$

נסתכל עתה על אי השוויון (5).

אגף שמאל מתייחס למערכת המסונכרנת:

$$P[\max_{1 \leq i \leq n} T_k(i) > t] = P[T_k(n, m) > t]$$

נראה כי אגף ימין מתייחס למערכת הבלתי תלויה.

המערכת ב"ת בעלת מופעים ב"ת. לכן כל אחת מ- i הטורים ב"ת אחד בשני:

$$\begin{aligned} P(T_k^*(n, m) > t) &= P(\max_{1 \leq i \leq n} T_k^*(i) > t) = 1 - P(\max_{1 \leq i \leq n} T_k^*(i) \leq t) = \\ 1 - P(T_k^*(1) \leq t, T_k^*(2) \leq t, \dots, T_k^*(n) \leq t) &= \\ 1 - P(T_k^*(1) \leq t) \cdot P(T_k^*(2) \leq t) \dots P(T_k^*(n) \leq t) &= 1 - \prod_{i=1}^n P(T_k^*(i) \leq t) \end{aligned} \quad (9)$$

אבל $T^*(i), T(i)$ בעלי אותה התפלגות ולכן:

$$1 - \prod_{i=1}^n P(T_k^*(i) \leq t) = 1 - \prod_{i=1}^n P(T_k(i) \leq t) \quad (10)$$

המסקנה מ- (8), (9) ו- (10) היא:

$$P(T_k^*(n, m) > t) = 1 - \prod_{i=1}^n P(T_k(i) \leq t) \quad (11)$$

כלומר, אגף ימין של אי שוויון (5) מתייחס למערכת ב"ת.

מסקנה 2: כתוצאה משוויון (10) ואי שוויון (5)

$$P(T_k(n, m) > t) \leq P(T_k^*(n, m) > t), \quad t \geq 0 \quad (12)$$

כלומר, במערכת בעלת מופע מסונכרן, הסיכוי לזמן שהיה ארוך יותר - קטן מהסיכוי במערכת הבלתי תלויה. במילים אחרות, הסינכרון ההתחלתי מקטין סטוכסטית את משך זמן השהיה.

תוצאה.

$$P(T(n, m) > t) = P(\lim_{k \rightarrow \infty} T_k(n, m) > t) \leq P(\lim_{k \rightarrow \infty} T_k^*(n, m) > t) = P(T^*(n, m) > t) \quad (13)$$

נשתמש בסימון המקובל: $T(n, m) <_{st} T^*(n, m)$, זמן השהיה במערכת המסונכרנת קטן סטוכסטית מזמן השהיה במערכת הבלתי תלויה.

עתה, ע"י שימוש באי שוויון (12) נסיק לגבי התוחלות:

$$\bar{T}(n, m) = \int_0^{\infty} (P(T(n, m) > t)) dt \leq \int_0^{\infty} (P(T^*(n, m) > t)) dt = \bar{T}^*(n, m) \quad (14)$$

מסקנה 3: תוחלת זמן שהיה במערכת מסונכרנת קטנה מתוחלת זמן שהיה במערכת בלתי תלויה.

2.2.2 מופע דטרמיניסטי.

נניח שמופע הפרויקטים למערכת הוא דטרמיניסטי, כלומר $t_k = a \quad \forall k$, עבור מספר a כלשהוא.

המשפט הבא מכליל את ההוכחה במאמרם של Mandelbaum, Avi-Itzhak [24] למערכת D/M/1²-SPM.

משפט: לכל ערך קבוע של k , n קבוצות זמני ההמתנה $\{W_k(i,j), j=1,2,\dots,m\}$, $i=1,2,\dots,n$ בלתי תלויות.

הוכחה: נוכיח באינדוקציה על מספר הפרויקטים- k ועל המשימה j .

בסיס: $k=1$, עבור הפרויקט הראשון המערכת ריקה, $\forall j \quad W_1(i,j)=0$ ולכן זמני ההמתנה ב"ת.

צעד: נניח נכונות עבור כל הפרויקטים עד פרויקט k , כלומר הקבוצה $\{W_i(i,r) \quad r=1,2,\dots,j\} \quad i=1,2,\dots,n$

$k+1, i=1,2,\dots,n$ ב"ת. נוכיח עבור פרויקט $k+1$.

הוכחה: נוכיח באינדוקציה על מספר המשימות j .

בסיס: $j=1$, הוכחה ניתן למצוא במאמרם של Mandelbaum, Avi-Itzhak (1967) [24].

צעד: נניח נכונות עד משימה j ונוכיח עבור משימה $j+1$ (לפרויקט $k+1$).

הוכחה: ע"ס הרחבת משוואת Lindley:

$$W_{k+1}(i, j+1) = \left[\sum_{l=1}^{j+1} X_k(i, l) - \sum_{l=1}^j X_{k+1}(i, l) - a \right]^+$$

$s_k(i, l) - i=1,2,\dots,n$ מ"מ ב"ת, ע"ס הגדרת המודל.

$W_k(i, l) - i=1,2,\dots,n$ מ"מ ב"ת, ע"י האינדוקציה.

$X_k(i, l) = W_k(i, l) + s_k(i, l)$ - סכום מ"מ, ב"ת לפי משתנה i .

באותו אופן ניתן להוכיח כי $X_{k+1}(i, l)$ $i=1,2,\dots,n$ ב"ת לפי משתנה i .

לכן, $W_{k+1}(i, j+1)$ $i=1,2,\dots,n$ ב"ת לפי משתנה i וניתן להוסיפו לקבוצה. מכאן הוכחנו שעבור כל k

קבוע הקבוצות $\{W_{k+1}(i, r) \quad r=1,2,\dots,j+1\}$ ב"ת עבור $i=1,2,\dots,n$. מ.ש.ל.

מסקנות:

$$R_k(i, j) = \sum_{l=1}^j (W_k(i, j) + s_k(i, j)) - \text{הזמן הלוךח לפרויקט לסיים משימה } j \text{ בטור } i, \text{ גם ב"ת לפי משתנה } i$$

יכן $T_k(i) = \sum_{j=1}^m (W_k(i, j) + s_k(i, j))$, זמן השהיה בכל טור, $i=1,2,\dots,m$, בלתי תלוי במשתנה i , כלומר יש לנו פה m מערכות ב"ת.

2.3 חסמים.

2.3.1 מופע פואסון.

בסעיף זה נתייחס למערכת המתוארת בסעיף 2.1.1, המכילה m טורים של משימות המעובדות במקביל, כל

טור מכיל m משימות המעובדות בזה אחר זה. מופע הפרויקטים פואסוני עם קצב λ , וזמני עיבוד

המשימות מעריכי עם קצב μ .

ע"ס משפט Burke (1956) [8] למערכת $M/M/1$, מופע היציאה מתחנת השרות אף הוא פואסוני עם קצב

λ , ולכן במערכת הנוכחית מופע הכניסה לכל תחנה אף הוא פואסוני וניתן להתייחס אל כל תחנה כאל

מודל $M/M/1$ (לצורכי חישובי תוחלות).

ידוע כי זמן שהיה במערכת $M/M/1$ מפולג מעריכי עם תוחלת $1/(\mu-\lambda)$.

ב- [32] מוכיח J. Walrand כי m התחנות ב"ת תלויות זו בזו, ולכן זמן השהיה בתחנה ב"ת תלוי בזמן

השהיה בתחנה אחרת.

$$\sum_{j=1}^m X(i,j) = T(i) \sim \text{Erlang}(m, \mu-\lambda) \text{ ולכן } X(i,j) \sim \text{Exp}(\mu-\lambda)$$

עתה, ע"י שימוש באי שוויון (5) ו- (14) והצגת תוצאת התפלגות Erlang נקבל **חסם עליון** :

(15)

$$\bar{T}(n, m) \leq \bar{T}^*(n, m) = \int_{t=0}^{\infty} \left[1 - \left(1 - \sum_{j=0}^{m-1} \frac{e^{-(\mu-\lambda) \cdot t} \cdot ((\mu-\lambda) \cdot t)^j}{j!} \right)^n \right] dt$$

ניתן לקבל שני חסמים תחתונים.

חסם ראשון:

נסתכל על מערכת סטטית הנוצרת מהמערכת מסעיף 2.1.1 אך ללא מופע פרויקטים (ולכן לא נוצרים תורי המתנה למשאבים). כלומר הפרויקט עובר m משימות ב"ת מפולגות מעריכי בכל טור והטורים לא תלויים אחד בשני.

נסמן:

$T_*(n, m)$ - זמן השהיה של פרויקט במערכת הסטטית.

$T_*(i)$ - זמן השהיה בטור i .

$\bar{T}_*(n, m)$ - תוחלת זמן השהיה במערכת הסטטית.

$$\begin{aligned} T_*(i) &\approx \text{Erlang}(m, \mu) \\ T_*(n, m) &= \max_{1 \leq i \leq n} \{ T_*(i) \} \end{aligned} \quad (16)$$

$$\bar{T}_*(n, m) = \int_{t=0}^{\infty} [1 - (1 - \sum_{j=0}^{m-1} \frac{e^{-\mu \cdot t} \cdot (\mu \cdot t)^j}{j!})^n] dt \quad (17)$$

ברור שתוחלת זמן השהיה במערכת סטטית קטנה מתוחלת זמן השהיה במערכת הדינמית, ולכן נקבל:

$$\bar{T}(n, m) \geq \bar{T}_*(n, m) = \int_{t=0}^{\infty} [1 - (1 - \sum_{j=0}^{m-1} \frac{e^{-\mu \cdot t} \cdot (\mu \cdot t)^j}{j!})^n] dt \quad (18)$$

חסם שני:

התיחסות רק לטור אחד. עבור $n=1$ התוצאה ידועה:

$$\bar{T}(1, m) = \bar{T}^*(1, m) = \frac{m}{\mu - \lambda}$$

$$\begin{aligned} \bar{T}(n, m) &= ET(n, m) = E[\text{Max}(T(1), T(2), \dots, T(n))] \geq E[T(1)] = \bar{T}(1, m) \\ &\Downarrow \\ \bar{T}(n, m) &\geq \bar{T}(1, m) = \frac{m}{\mu - \lambda} \end{aligned} \quad (19)$$

טיב החסמים תלוי בנצילות. עבור נצילות נמוכה, החסם הראשון טוב יותר. קל לראות כי עבור נצילות שואפת ל-0, שני החסמים (החסם העליון והחסם התחתון הראשון) שואפים להתלכד (ולפי משפט הסנדוויץ גם תוחלת זמן השהיה במערכת).
עבור נצילות גבוהה, זמן השהיה בתורים משמעותי. התעלמות מהם (כפי שעושה החסם הראשון) מעוות את התוצאה. לעומת זאת הזמן שנוסף כתוצאה מסנכרון מספר הטורים במקביל אינו משמעותי ולכן התעלמות ממנו (כפי שעושה החסם השני) אינו מעוות את התוצאה. לכן עבור נצילות גבוהה החסם השני יעיל יותר.

בפרק 2.4 נדגים את ההבדל בין החסמים בעזרת סימולציות.

ii גדול (מספר טורים גדול).

אנו נעסוק עתה במערכת טורים במקביל, בעלת מופע פואסוני כאשר ii גדול ו-iii קטן (יחסית ל-ii).
עבור אותה מערכת רק ii גדול ו-iii קטן יחסית, ניתן להשתמש במאמרם של Kang, Serfozo [17].
מודל 1 בסעיף 2 במאמר עוסק בהתפלגות מקסימום בין משימות ב"ת מפולגות ארלנג, המעובדות במקביל.

כזכור, ניתן להתייחס לשלוש מערכות:

- א. מערכת סטטית (התעלמות מהתורים) תוחלת זמן השהיה בה מסומנת ב- $\bar{T}_*(n, m)$.
- במערכת זו זמני השהיה בטורים ב"ת ומפולג ארלנג $(T_*(i) \sim \text{Erlang}(m, \mu))$.
- ב. מערכת המתוארת בסעיף 2.1.1 תוחלת זמן השהיה בה מסומנת ב- $\bar{T}(n, m)$. זמני השהיה בטורים תלוי.
- ג. מערכת בעלת ii מופעים ב"ת ותוחלת זמן השהיה בה מסומנת ב- $\bar{T}^*(n, m)$.
- במערכת זו זמני השהיה בטורים ב"ת ומפולג ארלנג $(T^*(i) \sim \text{Erlang}(m, \mu - \lambda))$.

עתה, ע"ס מאמרים של Kang, Serfozo [17] :

$$\begin{aligned} \frac{T^*(n, m) - b_n^*}{a_n^*} &\xrightarrow[n \uparrow \infty]{} X, \quad X \approx \text{Gumbel} \\ \frac{T_*(n, m) - b_{*n}}{a_{*n}} &\xrightarrow[n \uparrow \infty]{} X, \quad X \approx \text{Gumbel} \end{aligned} \quad (20)$$

כאשר פרמטרי הנרמול המתאימים הם :

$$a_n^* = 1/(\mu - \lambda) \quad b_n^* = (\mu - \lambda)^{-1} [\ln n + (m-1) \ln \ln n - \ln(m-1)!] \quad (21)$$

$$a_{*n} = 1/\mu \quad b_{*n} = (\mu)^{-1} [\ln n + (m-1) \ln \ln n - \ln(m-1)!] \quad (22)$$

בסעיף 2.3.1 הראינו כי לכל n, m מתקיים :

$$ET_*(n, m) \leq \bar{T}(n, m) \leq ET^*(n, m) \quad (23)$$

נתייחס לאי השוויון הימני :

$$\begin{aligned} \frac{\bar{T}(n, m) - b_n^*}{a_n^*} &\leq E\left[\frac{T^*(n, m) - b_n^*}{a_n^*}\right] \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_n^*}{a_n^*} &\leq \lim_{n \uparrow \infty} E\left[\frac{T^*(n, m) - b_n^*}{a_n^*}\right] \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_n^*}{a_n^*} &\leq \gamma \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_n^*}{1/(\mu - \lambda)} &\leq \gamma \end{aligned} \quad (24)$$

לצורכי חישוב, לכל m ולכל n גדול וקבוע :

$$\begin{aligned} E\left[\frac{T^*(n, m) - b_n^*}{a_n^*}\right] &\cong \gamma \\ \bar{T}(n, m) &\leq b_n^* + \frac{\gamma}{\mu - \lambda} \end{aligned} \quad (25)$$

אותו תהליך נעשה עבור אי שוויון שמאלי :

$$\begin{aligned} \frac{\bar{T}(n, m) - b_{*n}}{a_{*n}} &\geq E\left[\frac{T_*(n, m) - b_{*n}}{a_{*n}}\right] \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_{*n}}{a_{*n}} &\geq \lim_{n \uparrow \infty} E\left[\frac{T_*(n, m) - b_{*n}}{a_{*n}}\right] \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_{*n}}{a_{*n}} &\geq \gamma \\ \lim_{n \uparrow \infty} \frac{\bar{T}(n, m) - b_{*n}}{1/\mu} &\geq \gamma \end{aligned} \quad (26)$$

לצורכי חישוב, לכל m ו- n גדול וקבוע:

$$\begin{aligned} E\left[\frac{T_*(n, m) - b_{*n}}{a_{*n}}\right] &\cong \gamma \\ \bar{T}(n, m) &\geq b_{*n} + \frac{\gamma}{\mu} \end{aligned} \quad (27)$$

ולכן נקבל:

$$b_{*n} + \frac{\gamma}{\mu} \leq \bar{T}(n, m) \leq b_n^* + \frac{\gamma}{\mu - \lambda} \quad (28)$$

נסמן : $C(n, m) = \ln n + (m-1) \ln \ln n - \ln(m-1)!$ נציב בחסמים ונקבל:

$$\frac{\gamma + C(n, m)}{\mu} \leq \bar{T}(n, m) \leq \frac{\gamma + C(n, m)}{\mu - \lambda} \quad (29)$$

הערה : בחישובים הנ"ל השתמשנו במאמרו של J. Pickands [28] המוכיח כי עבור Z_n , $G(x)$ המקיימים:

$$\lim_{n \rightarrow \infty} P\left(\frac{Z_n - b_n}{a_n} \leq x\right) = G(x) \quad (30)$$

מתקיים, בתנאים מסויימים (כגון : Z_n מ"מ חיובי), גם התכנסות התוחלות:

$$\lim_{n \rightarrow \infty} a_n^{-m} E(Z_n - b_n)^m = \int_0^\infty x^m dG(x) \quad (31)$$

כמו כן נקבל קצב התכנסות לוגריתמי כפונקציה של n וליניארי כפונקציה של m .

$$\begin{aligned} \lim_{n \uparrow \infty} \frac{\bar{T}(n, m)}{\ln n} &= 1 & \forall m \\ \lim_{m \uparrow \infty} \frac{\bar{T}(n, m)}{m} &= 1 & \forall n \end{aligned} \quad (32)$$

2.3.2 מופע דטרמיניסטי.

נשנה את מופע הפרויקטים למערכת בעלת מופע דטרמיניסטי.

ע"ס המשפט במאמרם של Baccelli, Makowski [4], מופע דטרמיניסטי ממוזער זמן שהיה ממוצע של פרויקט במערכת.

נסמן $\bar{T}(n, m)^D$ - תוחלת זמן שהיה של פרויקט במערכת בעלת מופע דטרמיניסטי.

חסם עליון יתקבל ע"י התוחלת של המערכת בעלת מופע פואסוני.

$$\bar{T}(n, m)^D \leq \bar{T}(n, m) \quad (33)$$

בנוסף, נסינו להעזר במאמרם של Suresh, Whitt [30]. במאמר הם נותנים נוסחה מקורבת רקורסיבית

למקדם השתנות מופע הפרויקטים ברשת טורית:

$$c_{a,i+1}^2 = \rho_i^2 \cdot c_{s,i}^2 + (1 - \rho_i^2) \cdot c_{a,i}^2, \quad i \geq 1 \quad (34)$$

כאשר:

$c_{a,i}^2$ - מקדם השתנות המופע של תחנה i . $c_{s,i}^2$ - מקדם השתנות השרות של תחנה i .

ρ_i - נצילות תחנה i .

חישוב החסם העליון התבצע בשלושה שלבים:

א. חישוב מקדמי ההשתנות של כל תחנה בהתאם לנוסחה ($c_{a,1}^2 = 0$, כיון שהמופע דטרמיניסטי).

ב. קירוב זמן שהיה בכל תחנה למעריכי עם תוחלת שחושבה בעזרת נוסחת חיצוין-פולצ'ק.

ג. קבלת חסם עליון שהוא המקסימום בין μ ארלנגים.

לאחר ביצוע סימולציות, ראינו כי עבור מערכת בעלת μ גדול, ו- m קטן יחסית ל- μ החסם הנ"ל גרוע

מהחסם המתקבל בעזרת מאמרם של Kang, Sorfozo [16] (ראה נוסחה (29)).

עבור מערכת בעלת m גדול, ו- μ קטן יחסית ל- m אין הבדל משמעותי בין החסם הנ"ל לבין החסם הרגיל

(המתקבל ממערכת בעלת מופע פואסון- נוסחה (15)). לכן החלטנו לא להשתמש בחסם זה בעבודתנו.

2.4 קרובים בעזרת שימוש בסימולציות ובחסמים.

בסעיף זה נחזור לעסוק במערכת המתוארת בסעיף 2.1.1.

אנו נבדוק את טיב החסמים שהוצעו בסעיפים הקודמים ונסה לקבל קרובים אנליטיים לממוצע זמן השהיה של פרויקט במערכת. ע"י כך נבחן את השפעת הפרמטרים השונים, כגון: נצילות המערכת, השפעת המופע המסונכרן, השפעת הסנכרון הסופי והיחס בין מספר הטורים לאורך הטור במערכת. נראה כי אורך הטור בעל השפעה רבה יותר על תוחלת זמן השהיה, מאשר מספר הטורים. ככל שאורך הטור גדל (iii) כך תוחלת זמן השהיה גדל בקצב ליניארי. לעומת זאת, ככל שמספר הטורים גדל (iii) התוחלת גדלה בקצב לוגריתמי.

במערכת קיימים שני סנכרונים עיקריים:

- הסנכרון ההתחלתי, הנוצר עקב מופע משותף של פרויקטים לטורים השונים.
 - הסנכרון הסופי, הנוצר עקב האילוץ שהפרויקט מסיים את עיבודו בסיום כל משימותיו.
- אנו נווכח כי השפעת אורך הטור או מספר הטורים זהה על שני סנכרונים אך הנצילות משפיעה באופן הפוך.

2.4.1 מספר טורים ואורך טור סופי וקטן.

סעיף זה יעסוק במערכת המתוארת בסעיף 2.1 כאשר m, n קטנים.

עבור מערכת קטנה זו ניתן בקלות לחשב את החסמים (חסם עליון ושני חסמים תחתונים) שהובאו בסעיף 2.3.1.

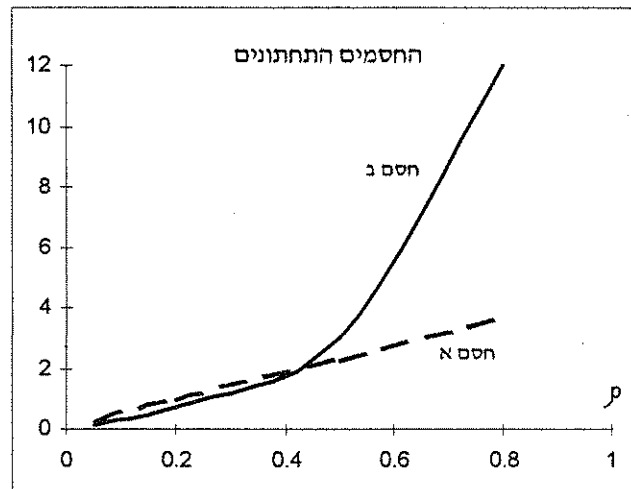
שני החסמים התחתונים:

א. חסם המתקבל ע"י הזנחת התורים (שסומן $(\bar{T}^*(n, m))$).

ב. חסם המתקבל ע"י $n=1$ $(\bar{T}(1, m))$.

החסם הראשון מזניח את השפעת התורים במערכת. לעומתו, החסם השני מזניח את השפעת הטורים במערכת והסנכרון הסופי שנגרם כתוצאה ממספר טורים.

כפי שהוסבר, טיב החסמים תלוי בנצילות כפי שניתן לראות מגרף 2.4.1 א (התוצאות התקבלו עבור מערכת $n=m=3$).



גרף 2.4.1 א: חסמים תחתונים עבור n, m קטנים.
Graph 2.4.1 a: Lower Bounds for Small m, n .

בנצילות שונות, השפעת התורים והשפעת הטורים שונות.
בנצילות נמוכות, המערכת כמעט ריקה, תורים מעטים. זמני השהיה במערכת עם תורים או בלי תורים קרובים והחסם הראשון יהיה קרוב יותר לתוצאות הסימולציה ולכן עדיף.
לעומת זאת, במערכת עמוסה, זמני שהיה בתורים מהווים חלק ארי מזמן השהיה במערכת והזנחת התורים מעוותת את התוצאה, לכן החסם הראשון איננו טוב והחסם השני עדיף.

חסם עליון נקבל ע"י מערכת בעלת מופעי פואסון ב"ת. כזכור סימנו תוחלת זמן שהיה במערכת

$$\bar{T}^*(n, m)$$

גרף 2.4.1 ב מציג תוחלת זמן שהיה אופייני והחסמים (חסם עליון והחסם התחתון הראשון) המתאימים כפונקציה של m .



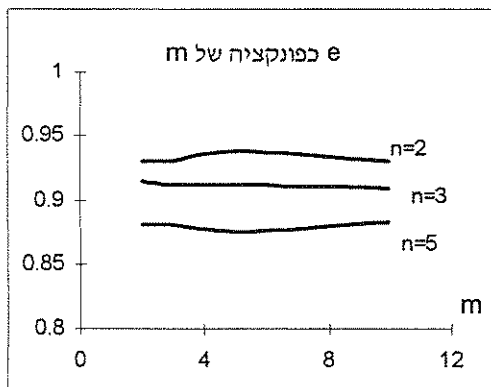
גרף 2.4.1 ב: תוחלת זמן שהיה במערכת m, n קטנים, החסם העליון והתחתון.
Graph 2.4.1 b: Expected Throughput Time and Bounds for Small m, n .

ניתן לראות כי החסם העליון (מופעי פואסון ב"ת) יכול להוות קרוב טוב.

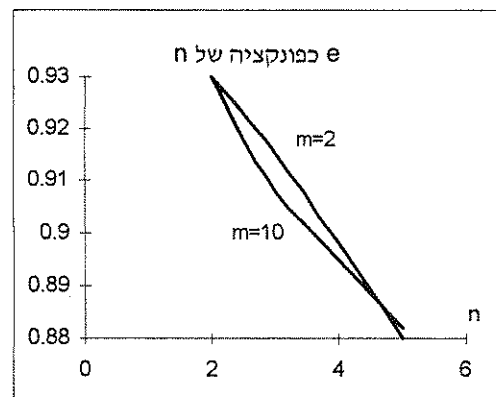
נגדיר $e = \frac{\bar{T}(n, m)}{\bar{T}^*(n, m)}$, היחס בין תוחלת זמן שהיה במערכת, לבין החסם העליון. ברור כי $0 < e \leq 1$.

ככל ש- e קרוב ל-1, הרי החסם העליון טוב יותר.

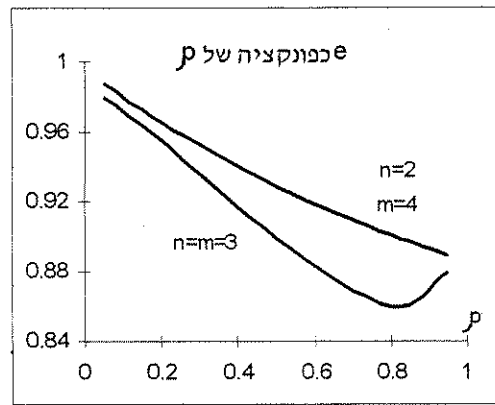
גרפים 2.4.1 ג, ד, ה מציגים התנהגות אופיינית של e כפונקציה של m, n .



גרף 2.4.1 ד: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של אורך הטור.
Graph 2.4.1 d: e as function of m .



גרף 2.4.1 ג: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של מספר הטורים.
Graph 2.4.1 c: e as function of n .



גרף 2.4.1 ה: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של הנצילות.

Graph 2.4.1 e: e as function of ρ .

הורצו סימולציה עבור $m=2,3,5,10$ ו- $n=2,3,5$. לאחר עיון בתוצאות הסימולציות, נוכל לסכם את

התנהגות e :

1. היחס e נשאר קבוע כפונקציה של אורך הטור m (ראה שרטוט 2.4.1 ג).

2. היחס e יורד כפונקציה של n (ראה שרטוט 2.4.1 ד).

3. היחס e יורד כפונקציה של ρ (ראה שרטוט 2.4.1 ה).

הסברים להתנהגות היחס יובאו בהמשך, עבור המקרים הגבוליים ($n \gg m, m \gg n$).

באופן כללי, החסם העליון יכול להוות קרוב טוב יחסית. טבלה 5.1.1 מביאה את תוצאות ההרצות,

החסם העליון והטעויות.

בטבלה רואים כי ככל ש- n גדל או ρ גדל, טיב הקרוב יורד.

מעבר לכך, במאמרם (שתואר בסקר הספרות) מציעים Cooper, Thompson, Charnes [9] דרך למציאת

התפלגות זמן שהיה במערכת סטטית, המתבססת על שני סוגי חישובים בלבד :

א. חישוב התפלגות s+t, סכום שני מ"מ (קונוולוצית ההתפלגויות).

ב. חישוב התפלגות $\max(s, t)$, מקסימום שני מ"מ (מכפלת ההתפלגויות).

נשתמש בשיטה זו כדי לחשב את התפלגות החסם העליון.

עבור מערכת בעלת מופעי פואסון ב"ת, הראנו בסעיף 2.3.1 כי זמני השהיה בכל תחנה ב"ת מפולגים

מעריכי. עתה, בעזרת שני סוגי חישובים אלו, בדרך רקורסיבית ניתן למצוא את התפלגות החסם העליון

של זמן שהיה במערכת. עבור n, m קטנים, זמן השהיה במערכת קרוב לזמן השהיה במערכת הב"ת, ולכן התפלגות החסם תהיה קרובה להתפלגות זמן השהיה האמיתי במערכת. לדוגמא, החישוב עבור מערכת 2×2 (4 תחנות שרות):

$$X^*(1,1), X^*(1,2), X^*(2,1), X^*(2,2) \sim \text{Exp}(\mu - \lambda)$$

$$T^*(1) = X^*(1,1) + X^*(1,2) \sim \text{Erlang}(2, \mu - \lambda) \quad (\text{חישוב א})$$

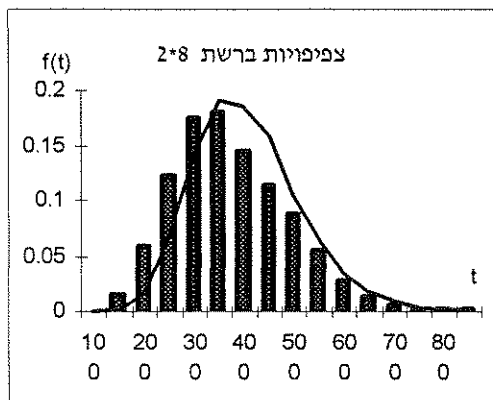
$$T^*(2) = X^*(2,1) + X^*(2,2) \sim \text{Erlang}(2, \mu - \lambda) \quad (\text{חישוב א})$$

$$T^*(2,2) = \text{Max}(T^*(1), T^*(2)) \quad (\text{חישוב ב})$$

$$P(T^*(2,2) \leq t) = P(T^*(1) \leq t) \cdot P(T^*(2) \leq t) = \left(1 - \sum_{j=0}^1 \frac{e^{-(\mu-\lambda) \cdot t} ((\mu-\lambda) \cdot t)^j}{j!}\right)^2$$

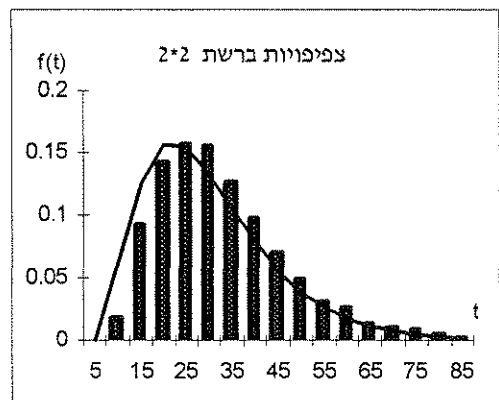
עתה, לאחר גזירת ההתפלגות מקבלים צפיפות של החסם העליון.

לדוגמא, גרף 2.4.1 ו, ז מציג צפיפויות עבור רשתות 2×2 , 2×8 שהתקבלו ע"ס החישוב שתואר למעלה, והסטוגרמת זמן השהיה של המערכות המקוריות (בעלות המופע המשותף). בטבלה 5.1.2 מרוכזים התפלגות זמן השהיה במערכות כפי שמוערך ע"י הסימולציה והתפלגות החסם.



גרף 2.4.1 ז: השוואת צפיפויות ברשת 8×2
Graph 2.4.1 g: Throughput Time Density,
 8×2 Network.

נתוני סימולציה: ממוצע- 349 ס.תקן- 114.32
נתוני החסם: ממוצע- 381.5 ס.תקן- 106.6



גרף 2.4.1 ו: השוואת צפיפויות ברשת 2×2
Graph 2.4.1 f: Throughput Time Density,
 2×2 Network

נתוני סימולציה: ממוצע- 25.5 ס.תקן- 14.3
נתוני החסם: ממוצע- 27.5 ס.תקן- 14.7

ע"ס הגרפים, ניתן להסיק כי עבור n, m קטנים, התפלגות החסם העליון מהווה קרוב טוב להתפלגות זמן השהיה במערכת.

2.4.2 טורים ארוכים.

סעיף זה יעסוק במערכת המתוארת בסעיף 2.1.1 כאשר $n \gg m$, כלומר מספר המשימות המעובדות

בקבוצה גדול יחסית למספר הקבוצות במערכת.

כפי שהוזכר, ניתן להשוות את תוחלת הזמן במערכת לשתי מערכות בהן התוצאה ידועה:

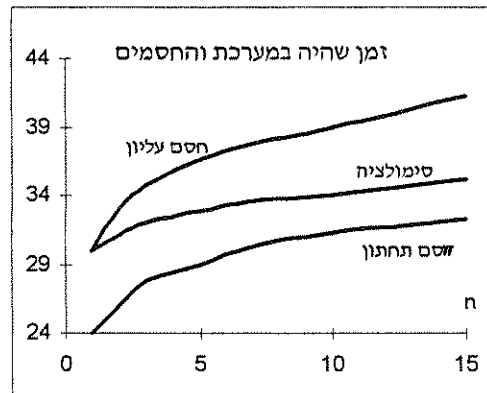
א. מערכת ב"ת.

ב. מערכת בעלת $n=1$ (טור יחיד).

א.השפעת הסנכרון ההתחלתי.

כפי שנאמר, חסם עליון יתקבל מהמערכת בעלת n מופעי פואסון ב"ת.

גרף 2.4.2 א אופייני לחסם ולסימולציה.



גרף 2.4.2 א: זמן שהיה במערכת, החסם העליון והתחתון.

Graph 2.4.2 a: Throughput Time and Bounds.

נגדיר: $a = \frac{\bar{T}(n, m)}{\bar{T}^*(n, m)}$, היחס בין תוחלת זמן שהיה במערכת (כפי שמוערכת ע"י תוצאות הסימולציה)

לבין תוחלת זמן שהיה במערכת הב"ת. ברור כי $0 < a \leq 1$.

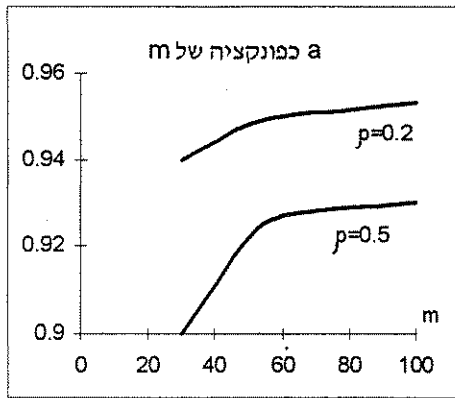
היחס a מבטא את השפעת הסנכרון ההתחלתי (לעומת מערכת בלתי תלויה ללא סנכרון). ככל ש- a יותר

קרוב ל-1, השפעת הסנכרון קטנה יותר, ולכן תוחלת זמן שהיה בשתי המערכות דומה. ככל ש- a קטן,

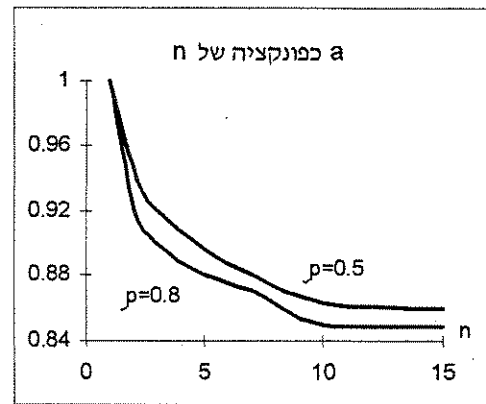
השפעת הסנכרון גדולה, וההבדל בין התוחלות משמעותי.

גרפים 2.4.2 ב, ג, ד מראים התנהגות אופיינית של a כפונקציה של m, n, ρ . כל גרף נבנה ע"י הרצת ערכים

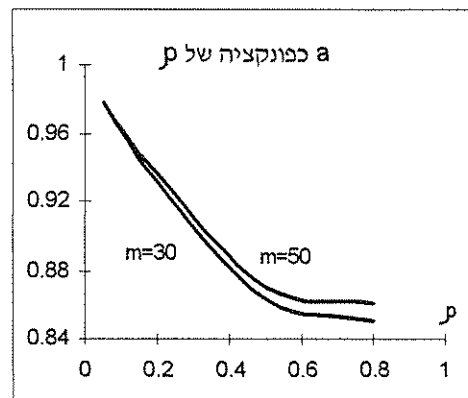
לפרמטר יחיד, כאשר שאר שני הפרמטרים קבועים.



גרף 2.4.2 ג: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של אורך הטור.
Graph 2.4.2 c: a as function of m.



גרף 2.4.2 ב: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של מספר הטורים.
Graph 2.4.2 b: a as function of n.



גרף 2.4.2 ד: היחס $\bar{T}(n, m)/\bar{T}^*(n, m)$ כפונקציה של נצילות.
Graph 2.4.2 d: a as function of p.

הורצו סימולציות עבור $n=1,2,3,5,7,10,15$, $m=30,50,60,100$ ו- $p=0.05,0.2,0.5,0.8,0.95$.

לאחר עיון בתוצאות הסימולציה, נוכל לסכם את התנהגות היחס a :

1. היחס a הוא פונקציה של מספר הטורים n, אורך הטור m, והנצילות p.
2. ככל ש-n גדל, a קטן אך קצב הירידה הולך ופוחת ולכן היחס הולך ומתייצב לפונקציה שנשמך אותה $g(m, p)$.
3. ככל ש-m גדל, היחס a גדל.
4. ככל ש-p גדל, היחס a קטן.

הסבר:

2. ככל שמספר הטורים גדל, יש השפעה גדולה יותר למופע המסונכרן אשר מקטין את זמן השהיה הממוצע ולכן היחס קטן.

מקרה קצה הוא $n=1$ שבו $\bar{T}(1, m) = \bar{T}^*(1, m)$, כיוון שעבור טור אחד אין הבדל בין המופעים השונים (כפי שניתן לראות בגרף 2.4.2 א).

3. ככל שאורך הטור גדל, השפעת הסנכרון ההתחלתי קטן.

מקרה קצה יהיה $m \rightarrow \infty$, שבו הסנכרון ההתחלתי לא ישפיע כמעט על התחנות המרוחקות ממנו ולכן

$$\bar{T}(n, m) \rightarrow \bar{T}^*(n, m), \text{ כלומר } a \rightarrow 1.$$

ככל שהמערכת עמוסה יותר, גודל התורים וזמן השהיה בהם עולה, ולכן חשיבות הסנכרון גדלה ומשפיע יותר. כדי להבין את השפעת הנצילות בצורה אינטואיטיבית יותר, כדאי להסתכל על מקרה הקצה בו $\rho \rightarrow 0$. כדי שהנצילות תשאף ל-0 קצב השרות μ צריך להיות גדול יחסית לקצב המופע λ , עד ש- $\mu \rightarrow \lambda$. עתה נסתכל על החסמים מסעיף 2.3.1 (15), (18) - כאשר קצב השרות μ גדול יחסית לקצב המופע λ החסמים (החסם העליון והתחתון) מתלכדים ולכן $a \rightarrow 1$.

בעקבות המאפיינים הנ"ל, ובהתאם לתוצאות הסימולציה התאמנו את הנוסחה הבאה:

$$a = g(m, \rho) + (1 - g(m, \rho)) \cdot e^{-(n-1)/4} \quad (1)$$

$$g(m, \rho) = 1 - \frac{\ln m + 2 \cdot \rho}{m}$$

הנוסחה הסופית:

$$\bar{T}(n, m) = \left[\left(1 - \frac{\ln m + 2 \cdot \rho}{m} \right) + \left(\frac{\ln m + 2 \cdot \rho}{m} \right) \cdot e^{-(n-1)/4} \right] \cdot \bar{T}^*(n, m) \quad (2)$$

סיכום השוואה:

מתוך 42 סימולציות שהורצו, 31 מקרים (73.8%) היו במרחק של עד 3% מהקרוב, ו-39 מקרים (92.8%) היו במרחק של עד 5% מהקרוב. סיכומי התוצאות ראה בטבלה 5.1.3.

ב. השפעת הסנכרון הסופי.

חסם תוחתון מתקבל מהמערכת בעלת $n=1$.

נגדיר: $b = \frac{\bar{T}(n, m)}{\bar{T}(1, m)}$, היחס בין תוחלת זמן השהיה במערכת (כפי שמוערכת ע"י תוצאות הסימולציה)

לבין תוחלת זמן השהיה במערכת בעלת טור אחד $(m/(\mu-\lambda))$. ברור כי $b \geq 1$.

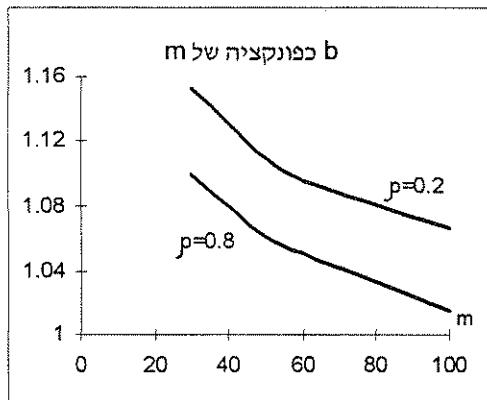
היחס b מבטא את השפעת הסנכרון הסופי, כיוון שבמערכת בעלת טור אחד אין סנכרון סופי.

ככל ש- b גדול יותר, השפעת הסנכרון הסופי גדולה, כלומר תוחלת זמן השהיה במערכת בעלת הסנכרון

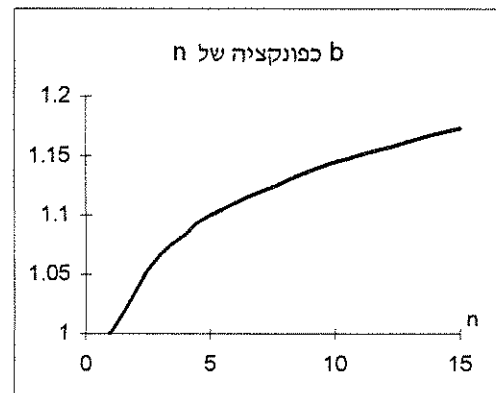
הסופי גדולה יותר ממערכת בעלת טור אחד (בלי הסנכרון הסופי).

גרפים 2.4.2 ה, ו, ז מראים התנהגות אופיינית של b כפונקציה של m, n, p . כל גרף נבנה ע"י הרצת ערכים

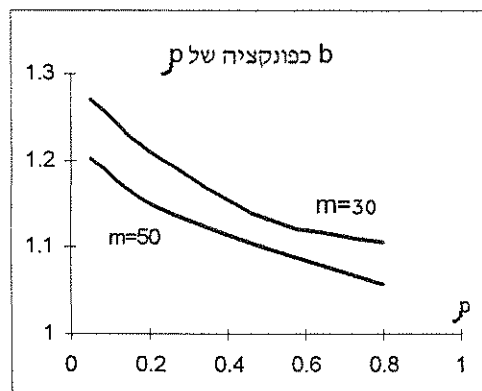
לפרמטר יחיד, כאשר שאר שני הפרמטרים קבועים.



גרף 2.4.2 ו: היחס $\bar{T}(n, m)/\bar{T}(1, m)$ כפונקציה של אורך הטור.
Graph 2.4.2 f: b as function of m .



גרף 2.4.2 ה: היחס $\bar{T}(n, m)/\bar{T}(1, m)$ כפונקציה של מספר הטורים.
Graph 2.4.2 e: b as function of n .



גרף 2.4.2 ז: היחס $\bar{T}(n, m)/\bar{T}(1, m)$ כפונקציה של הנצילות.
Graph 2.4.2 g: b as function of p .

הורצו סימולציות עבור $m=30,50,60,100$ ו- $p=0.05,0.2,0.5,0.8,0.95$, $n=2,3,5,7,10,15$

לאחר עיון בתוצאות הסימולציה, נוכל לסכם את התנהגות היחס b :

1. היחס b הוא פונקציה של מספר הטורים m , אורך הטור m , והנצילות p .

2. ככל ש- n גדל, b גדל בקצב הנראה לוגריתמי.

3. ככל ש- m גדל, היחס b קטן.

4. ככל ש- p גדל, היחס b קטן.

הסבר :

2. ככל שמספר הטורים גדל, יש השפעה גדולה יותר לסנכרון הסופי כיוון שיש יותר טורים שהפרויקט

צריך לעבד במקביל. אך ברור שהתרומה השולית של תוספת טור תלך ותקטן. את העלייה

הלוגריתמית ניתן לצפות בעקבות מאמרים קודמים, כפי שצוין בסקר הספרות.

מקרה קצה הוא $n=1$ שבו $\bar{T}(n, m) = \bar{T}(1, m)$.

3. ככל שאורך הטור גדל, המערכת נעשית יותר דטרמיניסטית (עבור זמן שהיה בטור המפולג

$Erlang(m, \mu - \lambda)$, מקדם השונות יהיה $CV = \frac{\sigma}{E} = \frac{\sqrt{m}/(\mu - \lambda)}{m/(\mu - \lambda)} = \frac{1}{\sqrt{m}}$ כלומר, השפעת סטיות

התקן על התוחלת הולכת וקטנה, המערכת נעשית יותר ויותר דטרמיניסטית לכן השפעת הסנכרון

הסופי קטנה.

מקרה קצה יהיה $m \rightarrow \infty$, שבו זמן שהיה בכל קבוצה ישאף לדטרמיניסטי ($CV \rightarrow 0$) ואז לא יהיה

הבדל בין n זמני שהיה בטורים השונים ולכן $\bar{T}(n, m) \rightarrow \bar{T}(1, m)$, כלומר $b \rightarrow 1$.

4. ככל שהמערכת עמוסה יותר, גודל התורים וזמן שהיה בהם עולה, ולכן השפעת הסנכרון הסופי

(הגורם להוצרות תורי סנכרון) זעומה.

כדי להבין את השפעת הנצילות בצורה אינטואיטיבית יותר, כדאי להסתכל על מקרה הקצה בו

הנצילות שואפת ל-0. במערכת כזו לא יהיו כמעט תורי המתנה למשאבים. אז יישארו תורי הסנכרון,

שלא מופיעים כלל במערכת בעלת טור אחד, ולכן היחס b יקבל את ערכו המקסימלי.

בעקבות המאפיינים הנ"ל, ובהתאם לתוצאות הסימולציה התאמנו את הנוסחה הבאה:

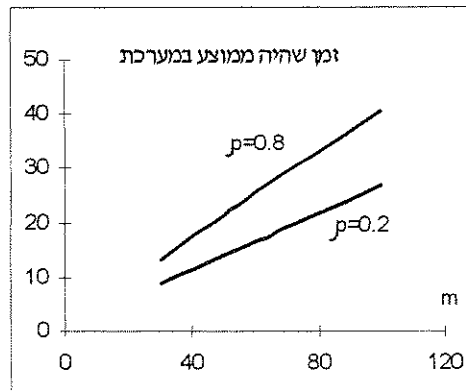
$$b = 1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2}$$

$$\bar{T}(n,m) = b \cdot \bar{T}(1,m) = \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \cdot \frac{m}{\mu - \lambda} \quad (3)$$

סיכום השוואה:

מתוך 42 סימולציות שהורצו, 29 מקרים (69%) היו במרחק של עד 2% מהקרוב, וכל המקרים היו במרחק של עד 4% מהקרוב. סיכומי התוצאות ראה בטבלה 5.1.4.

הערה: מהנוסחה ניתן להסיק שקצב העלייה לפי m ליניארי, והמסקנה מאושרת גם בגרף 2.4.2 ת.



גרף 2.4.2 ת: תוחלת זמן שהיה כפי שמוערך ע"י הסימולציה.

Graph 2.4.2 h: Expected Throughput Time (from simulations).

2.4.3 מספר טורים גדול.

סעיף זה יעסוק במערכת המתוארת בסעיף 2.1, כאשר $m \gg n$ כלומר מספר הטורים גדול יחסית למספר המשימות המעובדות בטור.

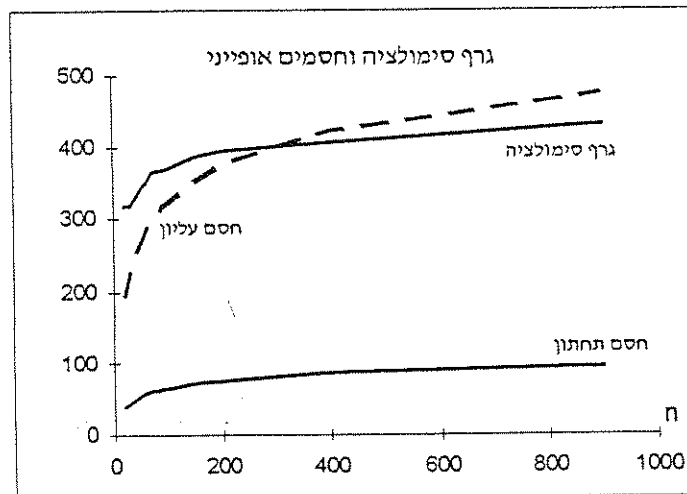
לשם קבלת החסמים אנו נשתמש במאמרם של Kang, Serfozo [16], בהוכחות האנליטיות, כפי שהבאנו בסעיף 2.3.1 ואי שוויון (30), עבור n גדול.

כזכור, סימנו: $C(n, m) = \ln n + (m-1) \ln \ln n - \ln(m-1)!$ וקבלנו שני החסמים:

$$\frac{\gamma + C(n, m)}{\mu} \leq \bar{T}(n, m) \leq \frac{\gamma + C(n, m)}{\mu - \lambda} \quad (4)$$

הורצו סימולציות עבור $m=3,4,5$ $n=50-200$, $\rho=0.05, 0.2, 0.5, 0.8$.

גרף 2.4.3 א מתאר תוחלת זמן שהיה אופייני במערכת כפי שהתקבל מהערכות הסימולציה, וחסמים.



גרף 2.4.3 א: זמן שהיה אופייני במערכת, וחסמים המתאימים.
Graph 2.4.3 a: Throughput Time and Bounds.

ניתן להבחין שני מאפיינים חשובים:

א. גרף הסימולציה מקביל לחסם התחתון.

ב. גרף הסימולציה נפגש בחסם העליון בערך n גדול (שממנו החסם מתקיים) ואז המרחק בין

החסם והסימולציה הולך וגדל (הסיבה, כפי שהוזכרה, היא שככל שמספר הטורים גדל, השפעת

הסנכרון ההתחלתי גדולה).

בעקבות המאפיינים הוצעה הנוסחה הבאה:

$$\bar{T}(n, m) = \frac{\gamma + C(n, m)}{\mu} + D(m, \mu, \rho) \quad (5)$$

כאשר $D(m, \mu, \rho)$ זהו ההפרש בין גרף הסימולציה לחסם התחתון שנשאר קבוע (כפונקציה של n).

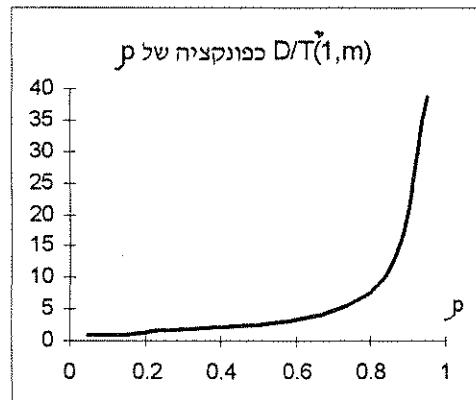
על פי הצעתנו, ניתן לבטא את ההפרש כמכפלת שני מרכיבים:

1. מרכיב התלוי בזמן השהיה הבסיס במערכת סטטית עבור $n=1$ (m/μ).

2. מרכיב התלוי בנצילות.

בגרף 2.4.3 ב נראה ההפרש המנורמל (מחולק בערך $\bar{T}^*(1, m)$) כפונקציה של הנצילות, כאשר שאר שני

הפרמטרים n, m נשארו קבועים.



גרף 2.4.3 ב: ההפרש המנורמל כפונקציה של הנצילות.

Graph 2.4.3 b: The Normalized Difference, as Function of Utilization.

המסקנה היא שהמרכיב התלוי בנצילות מתנהג כמו עקומת צפף (עולה ככל שהנצילות עולה).

הסבר:

1. בגלל שאנו מחפשים את ההפרש האבסולוטי בין שני הגרפים, ברור שהוא יהיה תלוי בזמן השרות ולא רק בקצבו, ולכן נוח לקחת כמדד תוצאה ידועה, שהיא זמן השהיה במערכת סטטית בעלת טור אחד.

2. ככל שהנצילות גדלה, ההפרש בין שני החסמים גדל, ולכן גם מרחק גרף הסימולציה והחסם התחתון גדל. עבור נצילות שואפת לאפס, שני החסמים מתאחדים (וכן גם ממוצע זמן השהיה במערכת).

בעקבות המאפיינים הנ"ל, ובהתאם לתוצאות הסימולציה התאמנו את הנוסחה הבאה:

$$\begin{aligned} D(m, \mu, \rho) &= \frac{m}{\mu} \cdot \frac{\rho}{1-\rho} \cdot (2.8 - \rho) \\ \bar{T}(n, m) &= \frac{\gamma + C(n, m)}{\mu} + \frac{m}{\mu} \cdot \frac{\rho}{1-\rho} \cdot (2.8 - \rho) \end{aligned} \quad (6)$$

סיכום השוואה:

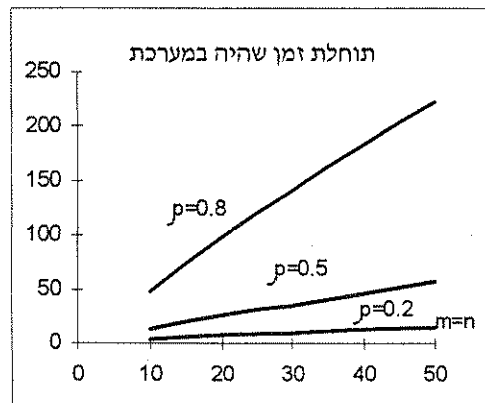
מתוך 34 סימולציות שהורצו, 26 מקרים (76.4%) היו במרחק של עד 3% מהקרוב, ו-32 מקרים (94.1%) היו במרחק של עד 4% מהקרוב. סיכומי התוצאות ראה בטבלה 5.1.5.

2.4.4 מספר טורים ואורך טור גדולים.

סעיף זה יעסוק במערכת המתוארת בסעיף 2.1.1, כאשר m, n מאותו סדר גודל וגדולים.

הורצו סימולציות עבור המקרים $\rho=0.05, 0.2, 0.5, 0.8$, $m=n=10-60$.

גרף 2.4.4 א מתאר תוחלת זמן שהיה אופייני במערכת כפי שהתקבל מהערכות הסימולציה.



גרף 2.4.4 א: תוחלת זמן שהיה במערכת, כפי שמוערך ע"י הסימולציה.

Graph 2.4.4a: Throughput Time.

תוצאות הסימולציה מרמזות כי זמן שהיה הוא פונקציה ליניארית של m או n , ובדיקה קפדנית יותר

הובילה למסקנה התלות הליניארית היא דרך m (בהתאמה לנוסחה (3)).

לאחר השוואה, נמצא כי נוסחה (3) מסעיף 2.4.2 מתאימה גם במקרה שלנו.

כלומר:

$$\bar{T}(n, m) = \frac{m}{\mu - \lambda} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \quad (7)$$

סיכום השוואה:

מתוך 21 סימולציות שהורצו, 18 מקרים (85.7%) היו במרחק של עד 3% מהקרוב, ו-20 מקרים (95.2%)

היו במרחק של עד 5% מהקרוב. סיכומי התוצאות ראה בטבלה 5.1.6.

2.4.5 מופע דטרמיניסטי

סעיף זה יעסוק במערכת המתוארת בסעיף 2.1.1, כאשר הזמן הבינומפעי של פרויקטים למערכת דטרמיניסטי.

גרפים 2.4.5 א, ב מתארים תוחלת זמן שהיה אופייני במערכת הדטרמיניסטית לעומת הפואסונית

כפונקציה של הפרמטרים n, m כפי שהתקבל מהערכות הסימולציה.



גרף 2.4.5 ב: תוחלת זמן שהיה במערכות כפונקציה של אורך הטור.

Graph 2.4.5 b: Throughput Time in systems, as function of m .



גרף 2.4.5 א: תוחלת זמן שהיה במערכות כפונקציה של מספר הטורים.

Graph 2.4.5 a: Throughput Time in systems, as function of n .

נגדיר $d = \frac{\bar{T}(n, m)^D}{\bar{T}(n, m)}$, היחס בין תוחלת זמן שהיה במערכת בעלת מופע דטרמיניסטי למערכת בעלת

מופע פואסוני (כפי שמוערך ע"י הסימולציה). ברור כי $0 < d \leq 1$.

היחס d מבטא את השפעת המופע על תוחלת זמן השהיה במערכת. ככל ש- d מתקרב ל-1, הרי אין השפעה גדולה של המופע. ככל ש- d קטן, השפעת המופע גדלה. הורצו סימולציות עבור כל המקרים הקודמים (בהם המופע הוא פואסון). לאחר עיון בתוצאות הסימולציה, נוכל לסכם את התנהגות היחס d :

1. היחס d הוא פונקציה של אורך הטור m והנצילות ρ . ניתן לראות מגרף 2.4.5 א כי היחס נשמר כפונקציה של n .

2. ככל ש- m גדל, d גדל (ניתן לראות זאת בגרף 2.4.5 ב).

3. ככל ש- ρ גדל, היחס d קטן.

הסבר:

2. כפי שהוסבר קודם, ככל שאורך הטור גדל, השפעת המופע קטן נצפה כי במקרה הקצה בו $m \rightarrow \infty$

המופע ההתחלתי לא ישפיע, ולכן $\bar{T}(n, m)^D \rightarrow \bar{T}(n, m)$, כלומר $d \rightarrow 1$.

3. כפי שהוסבר, ככל שהמערכת יותר עמוסה (בנצילות גבוהות) השפעת המופע ההתחלתי גדול. לעומת זאת, בנצילות נמוכות המערכת כמעט ריקה, ולכן המופע ההתחלתי פחות משפיע.

כדי לראות זאת אנליטית, ניתן להסתכל במערכת $m=n=1$ (כלומר $D/M/1$ לעומת $M/M/1$). עבור

מערכות אלו ניתן לחשב את תוחלת זמן השהיה ע"ס נוסחת חניצין-פולצ'ק:

$$\bar{T} = \frac{1}{\mu} \cdot \left(1 + \frac{\rho}{1-\rho} \cdot \frac{C_a^2 + C_s^2}{2} \right) \quad (8)$$

עבור מופע דטרמיניסטי: $C_a=0, C_s=1$. עבור מופע פואסוני: $C_a=1, C_s=1$.

בהצבת הפרמטרים בנוסחה מקבלים:

$$d = \frac{\frac{1}{\mu} \cdot \left(1 + \frac{\rho}{2(1-\rho)} \right)}{\frac{1}{\mu} \cdot \left(1 + \frac{\rho}{1-\rho} \right)} = \frac{2-\rho}{2} \quad (9)$$

כלומר, ככל שהנצילות יותר נמוכה, היחס d יותר גדול.

בעקבות המאפיינים הנ"ל, ובהתאם לתוצאות הסימולציה התאמנו את הנוסחה הבאה:

$$\begin{aligned} d &= 1 - \frac{\rho}{3 \cdot \ln m} \\ \bar{T}(n, m)^D &= \left(1 - \frac{\rho}{3 \cdot \ln m}\right) \cdot \bar{T}(n, m) \end{aligned} \quad (10)$$

כאשר $\bar{T}(n, m)$ - תוחלת זמן השהיה במערכת המקבילה בעלת מופע פואסוני, כפי שהתקבל בקרוב במתאים.

סיכום השוואה:

מתוך 65 סימולציות שהורצו, 42 מקרים (64.6%) היו במרחק של עד 5% מהקרוב, ו- 62 מקרים (95.3%) היו במרחק של עד 10% מהקרוב. זה אומנם לא קרוב מעולה, אך חלק מהטעויות נובעות מהעובדה שזה קרוב ע"ג קרוב (הנוסחה משתמשת בקרוב קודם), ולכן רמת הדיוק יורדת.

באופן כללי, ניתן לראות מהתוצאות כי עבור אורך טור גדול (30 ומעלה) הקרוב די טוב.

סיכומי התוצאות ראה בטבלה 5.1.7.

2.5 סיכום, תוצאות ומסקנות עיקריות.

בסעיף זה התייחסנו למערכת מסוימת, המכילה n טורים במקביל, כל טור מורכב מ- m שרתים. מופע הפרויקטים למערכת מהווה תהליך חידוש וזמני ביצוע המשימות הם m ב"ת מפולגים מעריכי.

בנינו שלושה חסמים לתוחלת זמן שהיה של פרויקט במערכת:

- חסם עליון המתקבל ע"י מערכת בעלת n מופעים ב"ת של פרויקטים.
- חסם תחתון המתקבל ע"י מערכת ללא מופע פרויקטים, מערכת סטוכסטית סטטית.
- חסם תחתון המתקבל ע"י מערכת בעלת טור אחד ($n=1$).

בעזרת שימוש בחסמים הנ"ל ובתוצאות סימולציות קיבלנו קרובים אנליטיים לתוחלת זמן שהיה של פרויקט במערכת.

כמו כן, בדקנו את השפעת הפרמטרים השונים על התוחלת, כגון: נצילות המשאבים, אורך הטור, מספר הטורים והשתנות המופע.

הבחנו בין שני סוגי סנכרון הקיימים במערכת:

- סנכרון התחלתי, הנוצר עקב מופע משותף של פרויקטים למערכת.
- סנכרון סופי, הנוצר עקב האילוץ שהפרויקט מסיים את עיבודו בסיום עיבוד כל משימותיו.

תוצאות ומסקנות עיקריות.

- ככל שאורך הטור גדול יותר, השפעת הסנכרון ההתחלתי והסופי קטנה.
- ככל שמספר הטורים גדול יותר, השפעת הסנכרון ההתחלתי והסופי גדלה.
- ככל שהמערכת עמוסה יותר, השפעת הסנכרון ההתחלתי גדלה אך השפעת הסנכרון הסופי קטנה.
- הקרוב לתוחלת זמן שהיה פרויקט במערכת בו אנו ממליצים להשתמש (מתאים ל- $m > 10$), ולכל

(n):

$$\begin{aligned} \bar{T}(n, m) &= \frac{m}{\mu - \lambda} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \\ \frac{\bar{T}(n, m)}{1/\mu} &= \frac{m}{1 - \rho} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \end{aligned} \quad (1)$$

קרוב זה נותן, על פי רוב, טעות של עד 5% מהתוחלת כפי שחושבה ע"י הסימולציות.

- כפי שניתן לראות מהקרוב, תוחלת זמן ההחיה גדל בקצב ליניארי כפונקציה של m , ובקצב לוגריתמי כפונקציה של n .

- קיימים שני סוגי תורים במערכת: תורי המתנות עקב משאבים מוגבלים ותורי סנכרון עקב אילוצי קדימות ואי תלות.

באופן גס ניתן לשייך את אורך הטור (iii) ומספר התורים (ii) לשני סוגי התורים באופן הבא: את אורך הטור נשייך לתורי המתנות עקב משאבים מוגבלים, כיוון שלפני כל שרת-משאב יש תור המתנה למשאבים.

את מספר התורים נשייך לתורי סנכרון, כיוון שתורי הסנכרון נוצרים עקב מספר טורים במקביל.

המסקנה הנובעת מכך היא, שאם נרצה לתת משקל לשני סוגי העכובים: פער הסנכרון ומשאבים מוגבלים, הרי פער הסנכרון משפיע בקצב לוגריתמי בעוד משאבים מוגבלים משפיע בקצב ליניארי. לכן, בקרוב מסדר ראשון, כדאי להשקיע יותר במשאבים מוגבלים, ע"מ לשפר את המערכת. (מסקנה מותאמת למסקנה מסעיף 1.4.2.2, גם שם ראינו כי יש משקל רב להמתנה למשאבים ע"פ שתי השיטות וחלק מזמן הסינכרון מושפע בעקיפין מזמן המתנה ולכן כדאי להשקיע במשאבים).

3. סיכום והרחבות.

בעבודה זו התמקדנו בחקר ביצועים של רשתות PERT/CPM סטוכסטיות דינמיות. ניסינו להתאים לרשתות אלו מושגים שהוגדרו עבור רשתות דטרמיניסטיות וסטוכסטיות: מסלול קריטי ומשימה קריטית. כדי לישם את ההגדרות פתחנו אלגוריתם המבוסס על סימולציות. בנוסף, החלטנו כי ניתן לחלק את פרופיל השהיה של פרויקט במערכת לשלושה סוגי זמנים: זמן עיבוד, זמן המתנה למשאבים וזמן סנכרון בין משימות. הצענו שתי שיטות להגדרת זמנים אלו: שיטה א': שחרור המערכת מאילוצי משאבים ותזמון תוך מדידת הרווח ברמת הביצועים. שיטה זו כוללת ארבעה שלבים הניתנים לביצוע בסדר שרירותי. שיטה ב': התייחסות לכל משימה בנפרד והגדרת מדד ביצועים השווה לסכום זמני שהיית המשימות בתחנות. הדגמנו את ישום השיטות הנ"ל ובחינת פרמטרים נוספים במערכת קטנה (אך מעניינת מספיק), ובכך סיפקנו מסגרת לחקירה אמפירית המבוססת על סימולציות של רשתות תורים. לבסוף, התמקדנו במערכת פשוטה וסימטרית, בעלת $m \times m$ משימות: m טורים במקביל, m משימות בכל טור. זמני השרות ב"ת מפולגים מעריכית ומופע פרויקטים למערכת מהווה תהליך חידוש. מצאנו חסמים אנליטיים לתוחלת זמן השהיה של פרויקט במערכת וע"י שימוש בחסמים אלו ובסימולציות התאמנו נוסחאות לתוחלת זמן השהיה עבור מערכת בעלת מופע פואסוני. בניתוח התייחסנו למושגים קודמים, כגון: משאבים מוגבלים ופער הסנכרון.

הרחבות.

- במחקר זה התמקדנו במערכת הומוגנית של טורים במקביל שבה מופע הפרויקטים מהווה תהליך פואסון או דטרמיניסטי. כדאי היה לבדוק מופעי פרויקטים נוספים, מערכת הומוגנית בעלת זמני שרות לא מעריכיים או אפילו מערכת לא הומוגנית ולבדוק את היחס בין השפעת אורך הטור (iii) למספר הטורים (ii).
- עבור מערכת GI/G/1 ידועה נוסחת חיצ"ק-פולצ'ק לתוחלת זמן השהיה במערכת. מעניין להרחיב נוסחה זו למערכת $k \times GI/G/1$, כלומר פרויקט מגיע, מתפצל ל- k משימות המעובדות

- במקביל. הפרויקט מסיים את עיבודו לאחר סיום עיבוד כל משימותיו. מאמרם של Nelson, Tantawi [24] מקרב נוסחה לתוחלת זמן השהיה במערכת $k \times M/M/1$. מאמרם של Avi-Itzhak, Mandelbourn [23] מקבל נוסחה לתוחלת זמן השהיה במערכת $k \times D/M/1$. ניתן להעזר במאמרים אלו ע"י הוספת גורם המתחשב במקדמי ההשתנות (כגון $(C_a^2 + C_s^2)/2$ הקיים בנוסחת חיצ"ק-פולצ'ק).
- ראינו, עבור מערכת טורים במקביל, כי נוסחת הקרוב לתוחלת זמן השהיה מכילה שלושה פרמטרים: ρ, η, θ . כדאי לבדוק עבור מערכות לא סימטריות ופשוטות, אולי ניתן לקרב או לחסום את התוחלת רק ע"י שימוש באותם פרמטרים (ρ - נצילות המערכת, θ - עומק המערכת המקסימלי, כלומר המספר המקסימלי של תחנות בטור, η - פריסה מקסימלית של המערכת, כלומר המספר המקסימלי של תחנות במקביל).
 - כדאי לישים את המושגים שנחקרו בעבודה זו (כגון: מסלול קריטי, משימה קריטית, פרופיל שהיה וכד') במערכת אמיתית, כגון: ניהול תחנות משנה בחברת החשמל (ירון למברג [33]).
 - פיתוח מסגרת מתמטית מדויקת להגדרות במחקר (כגון: הגדרת מסלול קריטי ומשימה קריטית).

4. רשימת מקורות:

- [1] P.S.Adler, A.Mandelbaum, V.Nguyen, E.Schwerer, :”From Project to Process Management: An Empirically-Based Framework for Analyzing Product Development Time”, Management Science 41, 458-484 (1995).
- [2] N.J.Aquilano, D.E.Smith:”A Formal Set of Algorithms for Project Scheduling with Critical Path Scheduling / Material Requirements Planning”, Journal of Operations Management 1, 57-67 (1980).
- [3] F.Baccelli, C.A.Courcoubetis, M.I.Reiman, “Construction of the Stationary Regime of Queues with Locking”, Stochastic Processes and their Applications 26, 257-265 (1987).
- [4] F.Baccelli, A.M.Makowski, “Multi Dimensional Stochastic Ordering and Associated Random Variables”, Operations Research 37, 478-487 (1989).
- [5] F.Baccelli, A.M.Makowski, “Queuing Models For Systems With Synchronization Constraints”, Proceedings of the IEEE 77, 138-161 (1988).
- [6] F.Baccelli, A.M.Makowski, A.Shwartz, “The Fork-Join Queue and Related Systems with Synchronization Constraints : Stochastic Ordering and Computable Bounds”, Advances in Applied Probability. 21, 629-660 (1989).
- [7] F.Baccelli, W.A.Massey, D.Towsley , “Acyclic Fork Join Queuing Networks”, J.Assoc. Comput. Mach. 36, 615-642 (1989).
- [8] P.J.Burke :”The Output of a Queueing System”, Operations Research 4, 699-704 (1956).
- [9] A.H.Buss, M.J.Rosenblatt :” Activity Delay in Stochastic Project Networks”, Working Paper, (1995).

- [10] A.Charnes, W.W.Cooper, G.L.Thompson :” Critical Path Analyses Via Chance Constrained and Stochastic Programming”, Operations Research 12, 460-470 (1963).
- [11] C.A.Courcoubetis, M.I.Reiman, “Optimal Control of a Queuing System with Simultaneous Service Requirements”, IEEE Transactions On Automatic Control 32, 717- 727 (1987).
- [12] C.A.Courcoubetis, M.I.Reiman, B.Simon, “Stability of a Queuing System with Concurrent Service and Locking”, SIAM J. Comput 16, 169-178 (1987).
- [13] R.H.Doersch, J.H.Patterson, ” Scheduling a Project to Maximize its Present Value: a Zero-One Programming Approach”, Management Science 23, 882-889 (1977).
- [14] Golenko-Ginzburg, ” PERT Assumptions Revisited”, OMEGA 17, 393-396 (1989).
- [15] J.M.Harrison , “Simulation Studies of Open Network Performance”, Teaching Notes,Stanford University, (1992)..
- [16] M.Harrison , “Assembly-Like Queues” , Journal of Applied Probability. 10, 354-367 (1973).
- [17] S.Kang, R.F.Serfozo :” Parallel-Processing Times: Extreme Values of Phase-Type and Mixed Random Variables”, Technical Report, Georgia Tech (1995-6).
- [18] L.Kleinrock, “Computer Application”,Queuing Systems, Volume 11, 271-291 (1975).
- [19] V.G.Kulkarni, V.G.Adlakha :“Markov and Markov- Regenerative PERT Networks” , Operations Research 34, 769-781 (1986).
- [20] I.S.Kurtulus, S.C.Narula :”Multi Project Scheduling: Analysis of Project Performance”, IIE Transactions 17, 58-65 (1985).
- [21] R.C.Larson , “Improving the New-York City Arrest to Arraignment System”, Interfaces 23, 76-96 (1993).
- [22] C.A.Leonhard, J.S.Davis, “Job Shop Development Model: A Case Study”, IEEE Software, 86-92 (1995).

- [23] D.V.Lindley, "The Theory of Queues with a Single Server", *Proc. Camb. Phil. Soc* 48, 277-289 (1952).
- [24] M.Mandelbaum, B.Avi-Itzhak, "Introduction to Queuing with Splitting and Matching ", *Israel Journal of Technology* 6 , 376-382 (1967).
- [25] R.Nelson, A.N.Tantawi, "Approximate Analysis of Fork/Join Synchronization in Parallel Queues", *IEEE Transactions on Computers* 37, 739-743 (1988).
- [26] R.Nelson, D.Towsley, A.N.Tantawi , "Performance Analysis of Parallel Processing Systems", *IEEE Transactions on Software Engineering* 14, 532-540 (1988).
- [27] V.Nguyen, "Processing Networks with Parallel and Sequential Tasks: Heavy Traffic Analysis and Brownian Limits", *The Annals of Applied Probability* 3 , 28-55 (1993).
- [28] J.Pickands III, "Moment Convergence of Sample Extremes", *The Annals of Mathematical Statistics* 39, 881-889 (1967).
- [29] M.J.Rosenblatt, Y.Roll, " A Future Value Approach to Determining Project Duration, *IIE Transactions* 17, 164-167 (1985).
- [30] C.E.Sigal, A.A.B.Pritsker, J.J.Solberg, " The Use of Cutsets in Monte Carlo Analysis Stochastic Networks", *Mathematical and Computer in Simulation* XXI, 376-384 (1979).
- [31] D.E.Smith-Daniels, V.L.Smith-Daniels, " Maximizing the Net Project Value of a Project Subject to Materials and Capital Constraints", *Journal of Operation Management* 7, 33-45 (1987).
- [31] S.Suresh, W.Whitt, "The Heavy Traffic Bottleneck Phenomenon in Open Queueing Networks", *AT&T Bell Laboratories, Murray Hill, New Jersey*, (1989).
- [32] J.Walrand, "Sojourn Time and Flows", in *An Introduction To Queueing Networks* , Chapter 4, Prentic-Hall (1988).

[33] R.Weber, "Scheduling and Interchangeability in Tandem Queues", Scheduling Theory and its Application, Chapter 15, Edited by Chre'tiene, P. et.al., Wiley (1995).

[34] י.למברג, "שיטות תכנון ובקרה במערכת מרובת פרויקטים עם יסומים לנושא ניהול פרויקטים תחמ"ש בחברת חשמל", עבודת מסטר, הפקולטה להנדסת תעו"נ, טכניון (1995).

5. נספחים.

5.1 טבלאות.

בכל הטבלאות שלהלן רווחי הסמך (רמת בטחון 95%) לא צוינו מפורשות בגלל היותם קטנים.

לדוגמא, הבאנו את רווחי הסמך בטבלה מספר 3.

5.1.1 טבלה מספר 1:

מערכת בעלת אורך טור ומספר טורים קטן יחסית.

השוואת תוחלת זמן ההחלפה במערכת בעלת מופע משותף, כפי שהוערכו ע"י הסימולציות

והתוחלת כפי שחושבה ע"י החסם העליון, פרק 2.3 נוסחה (18):

$$\bar{T}^*(n, m) = \int_{t=0}^{\infty} \left[1 - \left(1 - \sum_{j=0}^{m-1} \frac{e^{-(\mu-\lambda) \cdot t} \cdot ((\mu-\lambda) \cdot t)^j}{j!} \right)^n \right] dt \quad (1)$$

טעות (באחוזים)	חסם עליון	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
1.17	0.145	0.1433	2	0.05	2
3.10	0.688	0.66664	2	0.2	2
7.05	2.75	2.556	2	0.5	2
10.36	11	9.86	2	0.8	2
19.97	52.25	41.811	2	0.95	2
1.11	0.268	0.265	2	0.05	4
3.37	1.273	1.23	2	0.2	4
7.14	5.094	4.73	2	0.5	4
9.25	20.375	18.49	2	0.8	4
11.025	96.78	86.11	2	0.95	4
2.10	0.237	0.232	3	0.05	3
4.44	1.124	1.074	3	0.2	3

טעות (באחוזים)	חסם עליון	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
10.16	4.496	4.039	3	0.5	3
14.029	17.983	15.46	3	0.8	3
12.12	85.419	75.06	3	0.95	3

5.1.2 טבלה מספר 2:

מערכת בעלת $n=m=2$, $\mu=0.2$, $\lambda=0.1$,

השוואה בין התפלגות זמן שהיה במערכת ב"ת (המהווה קרוב טוב) לבין התפלגות זמן שהיה של המערכת בעלת מופע משותף, כפי שמוערך מהסימולציה.

התפלגות זמן שהיה במערכת הב"ת:

$$P(T^*(n,m) \leq t) = (1 - \sum_{j=0}^1 \frac{e^{-(\mu-\lambda)} \cdot ((\mu-\lambda) \cdot t)^j}{j!})^2 \quad (2)$$

חסם עליון	סימולציה	התחום
0	0	0-5
0.062	0.0173	5-10
0.126	0.0915	10-15
0.157	0.1418	15-20
0.155	0.1569	20-25
0.133	0.1556	25-30
0.105	0.1265	30-35
0.079	0.0961	35-40
0.056	0.0694	40-45
0.039	0.049	45-50
0.027	0.0309	50-55

התחום	סימולציה	חסם עליון
55-60	0.0251	0.018
60-65	0.0136	0.012
65-70	0.0098	0.008
70-75	0.0083	0.005
75-80	0.0043	0.003
80-85	0.0023	0.002
85-	0.0005	0.001

5.1.3 טבלה מספר 3:

מערכת בעלת טורים ארוכים.

השוואת תוחלת זמן השהיה כפי שהוערכו ע"י הסימולציות והתוחלת כפי שחושבה ע"י נוסחת

הקרוב בפרק 2.4.2 נוסחה (2):

$$\bar{T}(n, m) = \left[\left(1 - \frac{\ln m + 2 \cdot \rho}{m} \right) + \left(\frac{\ln m + 2 \cdot \rho}{m} \right) \cdot e^{-(n-1)/4} \right] \cdot \bar{T}^*(n, m) \quad (3)$$

טעות (באחוזים)	קרוב ע"פ הנוסחה	חצי רווח סמך (95%)	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
0.93	1.703	0.0042	1.72	2	0.05	30
8.47	1.83	0.0018	2.007	10	0.05	30
4.89	8.41	0.0329	8.642	5	0.2	30
2.56	8.63	0.024	9.082	10	0.2	30
2.72	31.93	0.48	31.09	2	0.5	30
1.53	32.7	0.33	32.21	3	0.5	30
1.23	33.21	0.14	32.81	5	0.5	30
0.43	33.47	0.17	33.625	7	0.5	30
0.42	33.85	0.20	34.0	10	0.5	30
1.91	34.52	0.215	35.2	15	0.5	30
5.15	127.42	2.34	121.17	2	0.8	30
3.83	129.7	1.89	124.93	3	0.8	30
0.63	131.00	1.97	131.85	5	0.8	30
0.55	131.56	1.21	132.3	7	0.8	30
0.01	132.62	1.46	132.6	10	0.8	30
1.21	134.97	1.47	136.63	15	0.8	30
2.94	623.37	3.96	642.3	10	0.95	30

טעות (באחוזים)	קרוב ע"פ הנוסחה	חצי רווח סמך (95%)	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
5.03	3.00	0.0049	3.15	10	0.05	50
1.28	13.23	0.067	13.07	2	0.2	50
0.81	13.53	0.054	13.43	3	0.2	50
0.06	13.82	0.057	13.83	5	0.2	50
1.11	13.98	0.085	14.14	7	0.2	50
1.44	14.17	0.067	14.38	10	0.2	50
1.37	14.43	0.048	14.64	15	0.2	50
2.04	56.63	0.87	54.91	10	0.5	50
2.93	217.58	2.34	211.39	5	0.8	50
3.79	221.48	2.76	213.39	10	0.8	50
3.27	999.06	6.58	967.9	2	0.95	50
0.83	1062.06	7.24	1071.0	15	0.95	50
0.26	17.216.49	0.042	16.45	5	0.2	60
0.39	17.2	0.04	17.27	15	0.2	60
3.12	252.03	2.12	244.4	2	0.8	60
4.64	256.58	2.54	245.2	3	0.8	60
1.38	260.4	2.90	256.86	5	0.8	60
1.98	262.54	2.01	257.44	7	0.8	60
2.84	265.25	2.78	257.93	10	0.8	60
4.37	269.47	2.88	258.18	15	0.8	60
1.85	26.11	0.08	25.64	2	0.2	100
1.59	27.08	0.08	26.66	5	0.2	100
1.79	27.68	0.08	27.2	10	0.2	100
1.42	28.09	0.04	27.7	15	0.2	100
5.9	429.97	3.11	406.01	5	0.8	100

5.1.4 טבלה מספר 4:

מערכת בעלת טורים ארוכים.

השוואת תוחלת זמן השהיה כפי שהוערכו ע"י הסימולציות והתוחלת כפי שחושבה ע"י נוסחת

הקרוב בפרק 2.4.2, נוסחה (3).

$$\bar{T}(n,m) = \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \cdot \frac{m}{\mu - \lambda} \quad (4)$$

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
1.41	1.69	1.72	2	0.05	30
2.017	1.96	2.007	10	0.05	30
1.55	8.5	8.642	5	0.2	30
1.53	8.94	9.082	10	0.2	30
1.05	31.41	31.09	2	0.5	30
0.12	32.24	32.21	3	0.5	30
1.47	33.29	32.81	5	0.5	30
1.06	33.98	33.625	7	0.5	30
2.09	34.71	34.0	10	0.5	30
0.97	35.54	35.2	15	0.5	30
3.18	125.02	121.17	2	0.8	30
2.42	127.96	124.93	3	0.8	30
0.13	131.66	131.85	5	0.8	30
1.36	134.1	132.3	7	0.8	30
3.08	136.69	132.6	10	0.8	30
2.19	139.62	136.63	15	0.8	30
0.5	645.52	642.3	10	0.95	30

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
3.47	3.05	3.15	10	0.05	50
0.7	12.97	13.07	2	0.2	50
1.27	13.25	13.43	3	0.2	50
1.58	13.61	13.83	5	0.2	50
2.1	13.84	14.14	7	0.2	50
2.02	14.08	14.38	10	0.2	50
1.85	14.64	14.64	15	0.2	50
0.71	55.30	54.91	10	0.5	50
0.9	213.31	211.39	5	0.8	50
2.64	219.04	213.39	10	0.8	50
0.84	976.09	967.9	2	0.95	50
1.77	1051.96	1071.0	15	0.95	50
1.96	16.93	16.45	5	0.2	60
1.83	16.14	17.27	15	0.2	60
0.64	245.98	244.4	2	0.8	60
1.74	249.48	245.2	3	0.8	60
1.15	253.89	256.86	5	0.8	60
0.24	256.8	257.44	7	0.8	60
0.757	259.88	257.93	10	0.8	60
2.015	263.38	258.18	15	0.8	60
0.39	25.53	25.64	2	0.2	100
1.53	26.25	26.66	5	0.2	100
1.51	26.78	27.2	10	0.2	100
2.15	27.103	27.7	15	0.2	100
2.34	415.54	406.01	5	0.8	100

5.1.5 טבלה מספר 5:

מערכת בעלת מספר רב של טורים.

השוואת תוחלת זמן השהיה כפי שהוערכו ע"י הסימולציות והתוחלת כפי שחושבה ע"י נוסחת

הקרוב בפרק 2.4.3, נוסחה (5):

$$C(n, m) = \ln n + (m - 1) \cdot \ln \ln n - \ln(m - 1)!$$

$$\bar{T}(n, m) = \frac{\gamma + C(n, m)}{\mu} + \frac{m}{\mu} \cdot \frac{\rho}{1 - \rho} \cdot (2.8 - \rho) \quad (5)$$

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
3.97	16.94	17.65	50	0.2	3
3.23	17.49	18.08	60	0.2	3
2.38	17.95	18.39	70	0.2	3
1.38	18.34	18.6	80	0.2	3
0.617	18.68	18.8	90	0.2	3
1.108	18.98	19.2	100	0.2	3
1.184	20.13	19.9	150	0.2	3
1.622	20.93	20.6	200	0.2	3
0.655	25.21	25.09	1000	0.2	3
4.62	244.19	233.4	50	0.8	3
3.77	246.38	237.41	60	0.8	3
0.31	248.2	249.0	70	0.8	3
2.35	249.76	255.41	80	0.8	3
1.67	251.13	255.8	90	0.8	3
2.936	252.34	260.83	100	0.8	3
5.29	256.94	271.31	150	0.8	3
5.57	260.13	275.48	200	0.8	3
1.7	81.54	80.18	60	0.5	4

טעות (באחוזים)	קרוֹב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
0.71	82.86	82.27	70	0.5	4
2.102	84.00	82.28	80	0.5	4
1.012	84.98	84.13	90	0.5	4
0.551	85.86	85.39	100	0.5	4
1.01	89.15	88.26	150	0.5	4
1.42	91.42	91.3	200	0.5	4
0.0005	99.74	99.23	600	0.5	4
4.07	377.05	362.29	60	0.8	5
3.19	379.47	367.72	70	0.8	5
3.11	381.52	370.01	80	0.8	5
2.89	383.32	372.55	90	0.8	5
2.84	384.9	374.26	100	0.8	5
0.80	390.84	387.72	150	0.8	5
0.03	394.93	395.05	200	0.8	5
0.53	404.41	406.6	400	0.8	5
3.81	414.96	431.43	900	0.8	5

5.1.6 טבלה מספר 6:

מערכת בעלת מספר טורים ואורך טור גדולים.

השוואת תוחלת זמן השהיה כפי שהוערכו ע"י הסימולציות והתוחלת כפי שחושבה ע"י נוסחת

הקרוב בפרק 2.4.4, נוסחה (6):

$$\bar{T}(n, m) = \frac{m}{\mu - \lambda} \cdot \left(1 + \frac{\ln n}{m} \cdot \frac{\ln m - \ln \rho}{2} \right) \quad (6)$$

אורך הטור ומספר טורים (m=n)	נצילות (ρ)	סימולציה	קרוב ע"פ הנוסחה	טעות (באחוזים)
10	0.05	0.79	0.847	7.26
20	0.05	1.49	1.52	2.34
30	0.05	2.16	2.15	0.39
40	0.05	2.8	2.75	1.63
50	0.05	3.44	3.34	2.82
10	0.2	3.56	3.62	1.85
20	0.2	6.72	6.72	0.066
30	0.2	9.63	9.63	0.71
40	0.2	12.44	12.44	1.16
50	0.2	15.2	15.2	1.49
10	0.5	13.26	13.44	1.42
20	0.5	25.02	25.52	2.02
30	0.5	36.04	36.96	2.56
40	0.5	46.7	48.08	2.96
50	0.5	57.66	59.00	2.33
60	0.5	68.61	69.8	1.73
10	0.8	50.16	51.63	2.93
20	0.8	99.67	99.28	0.38

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	נצילות (ρ)	אורך הטור ומספר טורים ($n=m$)
2.86	144.65	140.62	0.8	30
3.03	188.86	183.3	0.8	40
4.47	232.35	222.41	0.8	50

5.1.7 טבלה מספר 7:

מערכת בעלת מופע דטרמיניסטי.

השוואת תוחלת זמן שהיה כפי שהוערכו ע"י הסימולציות והתוחלת כפי שחושבה ע"י נוסחת

הקרוב בפרק 2.4.5, נוסחה (7):

$$\bar{T}(n,m)^D = \left(1 - \frac{\rho}{3 \cdot \ln m}\right) \cdot \bar{T}(n,m) \quad (7)$$

כאשר $\bar{T}(n, m)$ - תוחלת זמן שהיה במערכת המקבילה, בעלת מופע פואסוני, כפי שהתקבל

בקרב המתאים.

עבור מערכת בעלת טורים ארוכים, השתמשנו בקרב השני.

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
5.58	8.34	7.9	5	0.2	30
4.49	8.7	8.39	10	0.2	30
4.83	29.87	28.5	2	0.5	30
3.96	30.66	29.5	3	0.5	30
4.49	31.66	30.3	5	0.5	30
4.05	32.31	31.6	7	0.5	30
4.56	33.01	31.57	10	0.5	30
4.48	33.8	32.35	15	0.5	30
1.57	115.22	113.44	2	0.8	30
2.41	117.93	115.15	3	0.8	30
1.64	121.34	119.38	5	0.8	30
1.45	123.59	121.82	7	0.8	30
0.89	125.97	124.86	10	0.8	30
1.99	128.68	126.17	15	0.8	30

טעות (באחוזים)	קרב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
0.58	585.42	582	10	0.95	30
1.93	897.08	914.8	2	0.95	50
4.37	966.8	1011	15	0.95	50
2.54	15.85	15.49	5	0.2	60
2.11	16.65	16.31	15	0.2	60
0.046	229.96	230.07	2	0.8	60
0.16	233.23	232.85	3	0.8	60
0.21	237.36	236.86	5	0.8	60
0.199	240.07	239.6	7	0.8	60
0.38	242.95	242.03	10	0.8	60
0.59	246.22	247.7	15	0.8	60
2.47	25.16	24.56	2	0.2	100
1.45	25.87	25.5	5	0.2	100
2.03	15.91	16.25	50	0.2	3
1.35	16.43	16.66	60	0.2	3
0.811	16.86	17	70	0.2	3
0.409	17.22	17.3	80	0.2	3
0.169	17.55	17.58	90	0.2	3
0.253	17.83	17.79	100	0.2	3
1.305	18.91	18.67	150	0.2	3
2.044	19.66	19.27	200	0.2	3
3.69	184.92	192.01	50	0.8	3
6.60	186.57	198.5	60	0.8	3
6.57	187.95	201.18	70	0.8	3
8.47	189.14	206.65	80	0.8	3
8.74	190.17	208.4	90	0.8	3

טעות (באחוזים)	קרוב ע"פ הנוסחה	סימולציה	מספר טורים (n)	נצילות (ρ)	אורך הטור (m)
9.5	191.09	211.16	100	0.8	3
11.99	194.57	221.1	150	0.8	3
13.18	196.99	226.9	200	0.8	3
8.22	71.73	64.29	60	0.5	4
8.79	72.9	67.01	70	0.5	4
8.04	73.9	68.4	80	0.5	4
8.34	74.76	69.01	90	0.5	4
7.49	75.53	69.87	100	0.5	4
7.74	78.48	72.8	150	0.5	4
7.93	80.43	74.9	200	0.5	4
7.58	314.58	292.4	60	0.8	5
7.86	316.59	293.5	70	0.8	5
7.06	318.31	297.3	80	0.8	5
6.00	319.8	301.7	90	0.8	5
5.66	321.12	303.9	100	0.8	5
3.1	326.08	316.28	150	0.8	5
1.13	329.49	325.8	200	0.8	5
11.07	3.52	3.17	10	0.2	10
7.25	6.57	6.13	20	0.2	20
5.25	9.44	8.97	30	0.2	30
4.7	45.65	43.6	10	0.8	10
3.48	90.44	87.4	20	0.8	20
1.99	133.31	130.7	30	0.8	30
1.98	175.2	171.8	40	0.8	40
1.66	216.51	212.96	50	0.8	50

