TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

הפקולטה למדעים מדויקים

ע"ש ריימונד ובברלי סאקלר

בית הספר למדעי המחשב על שם בלבטניק

# ניתוח צורני דיסיונקטיבי חלקית

חיבור לשם קבלת תואר דוקטור לפילוסופיה מאת

**רומן מנביץ'**

עבודה זו נעשתה בהדרכתו של

פרופ' מולי שגיב

וייעוצו של

דר' גאנשן ראמלינגם

הוגש לסנאט של אוניברסיטת תל-אביב

פברואר 2009

# תמצית

תכניות מחשב מודרניות מסתמכות באופן משמעותי על מבני נתונים המוקצים באופן דינאמי. אלגוריתמי ניתוח צורני מנתחים באופן סטאטי תכניות, כדי לקבוע אינפורמציה על מבני הנתונים, כגון "האם משתנה x מצביע לרשימה לא-מעגלית?" ו-"האם ניתן להגיע לאובייקט ע"י מעקב מצביעים משני משתנים שונים?" אלגוריתמים אלו הינם קונסרבטיביים (נאותים), כלומר האינפורמציה שהם מגלים נכונה לכל קלט תוכנית, ולכן ניתן להשתמש בה לשימושים שונים, כגון אימות תוכנה, אופטימיזציה, מיקבול וכד'.

ניתוחים צורניים דיסיונקטיביים פועלים ע"י הפשטה של זיכרונות המחשב הקונקרטיים לגרפים צורניים (חסומים). בנקודות מיזוג של גרף הבקרה, הגרפים הצורניים ממוזגים ע"י שימוש בדיסיונקציה (איחוד קבוצות למעשה), דבר שלעיתים קרובות גורם לפיצוץ אקספוננציאלי במספר הגרפים הצורניים. בתכניות מקביליות הבעיה אף חמורה יותר בשל שילוב ביצועים של פתילים שונים.

אנו מציגים אנליזות צורניות "דיסיונקטיביות חלקית" המכוונות להקטין את גודל מרחב המצבים ע"י הפשטה של דיסיונקציות, וכמו-כן קירוב נאות של פקודות תוכנית. אנו מימשנו והפעלנו את האנליזות הללו כדי להוכיח תכונות של תכניות טוריות ותכניות עם מקביליות בגרעיניות עדינה. עלה ביכולתנו להוכיח תכונות מאתגרות הכוללות תכונות ניקיון, שמורות צורניות, וליניאריזציה של מימושי מבני נתונים מקביליים. האנליזות הצורניות החדשות מתאימות לתכניות גדולות יותר מאשר אנליזות צורניות דיסיונקטיביות ובדרך-כלל רצות מהר יותר בסדרי גודל, והן עדיין מסוגלות להוכיח את התכונות הרצויות.

# פרק 1 – הקדמה

בתזה זו, אנו מעוניינים בהסקה סטאטית של תכניות אשר משנות מבני נתונים מקושרים, אליה מתייחסים כ*ניתוח צורני* והיא מקרה פרטי של *פירוש מופשט* [CC77]. היישום העיקרי של אנליזה צורנית בתזה זו היא לצורך אימות תכונות בטיחות, הכוללות אי-ביצוע פנייה למחוון אפס, אי-ביצוע דליפות זיכרון, שמורות צורה, אי-ביצוע של חריגת שינוי מקבילי בג'אווה, ובדיקת *לינאריזציה* של מבני נתונים מקביליים.

**דרישות קדם ורקע תאורטי.** אחת הדרישות להבנת החומר בתזה היא היכרות עם תחום של אנליזה סטאטית של תכניות מחשב בדרך של פירוש מופשט (מבוא לפירוש מופשט ניתן ע"י נילסון ושות' [NNH99]), כולל הבנה של המושגים הבאים: *הפשטה, קונקרטיזציה, מרחבים מופשטים, מתמרים מופשטים*, והסקה של שמורות תוכנית ע"י *איטרציות כאוטיות* (חישוב נקודות שבת). אנו כוללים תזכורת קצרה למושגים אלו. אנו כוללים גם כן את החלקים הנדרשים מהתאוריה של אנליזה צורנית ע"י לוגיקה תלת-ערכית [SRW02] בפרק 2. במערכת זו, מצבים קונקרטיים ומצבים מופשטים מיוצגים ע"י מבנים לוגיים, אשר ניתן לחשוב עליהם כעל גרפים מכוונים עם ריבוי סוגים של קשתות ועם תכונות בוליאניות המשויכות לצמתים.

**פרטים עיקריים בפירוש מופשט.** בתזה זו, אנו מניחים שמרחב מופשט $A$ ניתן ע"י שריג שלם $D_A=(\sqsubseteq_A,\sqcup_A,\sqcap_A,\bot_A,\top_A)$ שבו $A$ הוא קבוצת אלמנטים; $\sqsubseteq_A$ הוא יחס סדר חלקי בין האלמנטים; $\sqcup_A$ הוא הגבול העליון הקטן ביותר, או אופרטור צירוף; $\sqcap_A$ הוא הגבול התחתון הגדול ביותר, או אופרטור חיתוך; $\bot_A$ הוא האלמנט המינימלי בשריג; ו-$\top_A$ הוא האלמנט המקסימלי בשריג. אנו נאמר שאלמנט $c_1$ מדויק יותר מאלמנט $c_2$ אם $c_1 \sqsubseteq_C c_2$

בפירוש מופשט [CC77], פונקצית הפשטה $\alpha^{C,A} : C \rightarrow A$ ממפה אלמנט של המרחב הקונקרטי C לאלמנט המדויק ביותר המייצג אותו *במרחב המופשט* A. המשמעות של אלמנט מופשט $a \in A$ ניתנת ע"י פונקצית *קונקרטיזציה* $\gamma^{A,C} : A \rightarrow C$. כלומר, אנו נאמר ש-$a \in A$ מייצג כל אלמנט $c \in C$, כך ש-$c \sqsubseteq \gamma^{A,C}(a)$. בנוסף לזאת, הזוג $(\gamma^{A,C}, \alpha^{C,A})$ יוצר *קשר גלואה*.

בהמשך, אנו נשמיט את הסימונים בכתב עילי ובכתב תחתי המציינים את המרחבים הסמנטיים כאשר הדבר אינו יוצר בלבול.

פונקציה סמנטית $F^{\#} : A \rightarrow A$ הינה קירוב מעל *נאות* של פונקציה סמנטית $F : C \rightarrow C$ כאשר התנאי הבא מתקיים:

$$F(\gamma(a)) \sqsubseteq \gamma(F^{\#}(a)) .$$

אנו נקרא לפונקציה $F$ המתמר הקונקרטי ולפונקציה $F^{\#}$ המתמר המופשט. בתזה זו אנו בדרך-כלל נהיה מעוניינים בקירוב עליון של המשמעות של פקודת תוכנית $[\![st]\!]$ מעל מרחב מופשט סופי $A$.

הסמנטיקה (או המשמעות) של תוכנית ניתנת במונחים של נקודת השבת הקטנה ביותר $lfp(F)$ והסמנטיקה המופשטת ניתנת ע"י $lfp(F^{\#})$[1]. ניתן להסיק תכונות באופן קונסרבטיבי ע"י התחלה מאלמנט תחילי $a_0$ והפעלה חוזרת של הפונקציה $F^{\#}$ עד לנקודת שבת. תהליך זה מובטח להסתיים כאשר גובה השריג $A$ סופי.

לבסוף, נציין ששני מרחבים מופשטים $A_1$ ו-$A_2$ עלולים להיות שקולים, כלומר איזומורפיים, ולהציע קידוד שונה של אותה אינפורמציה. כלומר, לכל אלמנט קונקרטי $c \in C$, התנאי הבא מתקיים:

$$\gamma^{A_1,C}(\alpha^{C,A_1}(c)) = \gamma^{A_2,C}(\alpha^{C,A_2}(c)).$$

[1] למעשה, נקודת השבת בדרך-כלל מחושבת לסדרת האיטרציות המוגדרת כך: $X_0=a_0$ ו-$X_{n+1}=X_n \sqcup F^{\#}(x_n)$.

**הפשטות דיסיונקטיביות לעומת הפשטות דיסיונקטיביות חלקית.** מרחב מופשט $A$ (וההפשטה המתאימה) נקראים *דיסיונקטיביים* כאשר התנאי הבא מתקיים לכל שני אלמנטים מופשטים $a_1, a_2 \in A$:

$$\gamma(a_1) \sqcup_C \gamma(a_2) = \gamma(a_1 \sqcup_A a_2) \ . \qquad (1.2)$$

אחרת, (כאשר $\gamma(a_1) \sqcup_C \gamma(a_2) \sqsubset \gamma(a_1 \sqcup_A a_2)$ אפשרי) נאמר שהמרחב *דיסיונקטיבי חלקית*.

כעת נשווה בין שתי צורות ההפשטה:

הפשטות דיסיונקטיביות, אשר מוגדרות ע"י חלוקה סופית של קבוצת המצבים הקונקרטיים, פופולאריות בקהילת בדיקת המודלים (model checking). הפשטות דיסיונקטיביות חלקית הן כלליות יותר, מכיוון שהן מאפשרות להגדיר הפשטה במונחים של כיסויים ע"י קבוצות מצבים חופפות (סגורות תחת צירוף וחיתוך).

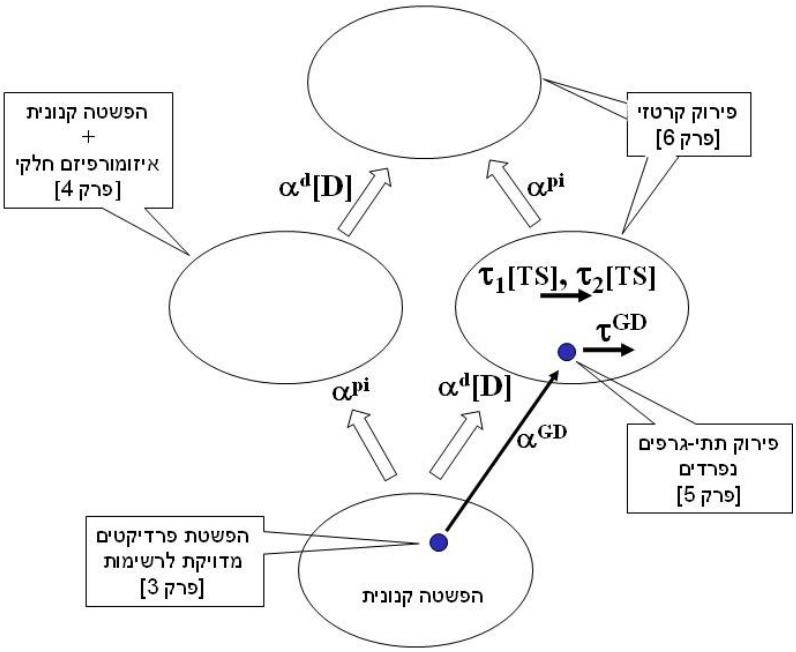
מרחבי הפשטות דיסיונקטיביות הן קבוצות חזקה של קבוצת מחלקות השקילות, ביחס ל-A, של המרחב הקונקרטי. לכן הן נחשבות יקרות למדי לאנליזות סטאטיות. לדוגמא, כאשר נתח סטאטי מפרש נקודות מפגש בגרף הבקרה, גודל האלמנט המצורף יכול להיות כפול מכל אחד מהאלמנטים אשר בענפים. בניגוד לזאת, מרחב דיסיונקטיבי חלקית יכול לקרב את האלמנטים משני הענפים בדרך שהגודל נשאר סביר (במיוחד במונחים של הייצוג במחשב) וחישוב נקודת השבת מסתיים מהר יותר. כמובן שהאלמנט המחושב בדרך זו יכול להיות מדויק פחות ולכן פחות שימושי, לדוגמא לצרכי אימות תוכנה. לכן, יש לבחור באופן זהיר את הפשטות דיסיונקטיביות חלקית כדי לתמוך בשני צרכים מנוגדים – לרסן את העלות של האנליזה ולספק אינפורמציה מדויקת מספיק לצרכי האנליזה.

**מטרות.** המטרות שלנו בתזה זו הן כדלהלן: (1) למצוא הפשטות חדשות ומדויקות למבני נתונים מקושרים, ו-(2) למצוא הפשטות דיסיונקטיביות חלקית ומתמרים יעילים אשר ניתן להשתמש בהם לניתוח תכניות משנות-זכרון ברמת דיוק דומה לזו אשר מושגת בעזרת אנליזות דיסיונקטיביות, אבל עם ביצועים טובים יותר. בפרט, אנו מחפשים הפשטות דיסיונקטיביות חלקית אשר מאפשרות למתכנן האנליזה להפשיט במתכוון אינפורמציה אשר אותה הוא מחשיב כלא-רלבנטית להוכחת תכונה מסוימת (לדוגמא, הפשטה של המתאם בין תכונות של רשימות נפרדות להוכחת אי-ביצוע פנייה למחוון אפס).

## 1.1 מבט-על

החומר בתזה זו מבוסס בעיקר על ארבעה מאמרים [MSRF04,MYRS05,MBC+07,MLAS+08], אשר לכל אחד מהם פרק טכני מתאים. פרק זה מכיל מבט-על לא פורמאלי, אשר מציג את התרומות של כל אחד מהמאמרים ואת הקשרים ביניהם.

איור 1.1 ממחיש את ההפשטות השונות והאלגוריתמים למתמרים השונים אשר פותחו בתזה. המרחב הקונקרטי, אשר לא נראה באיור, הוא קבוצת החזקה של מבנים לוגיים 2-ערכיים (בפרקים 3,4, ו-6) או קבוצת החזקה של גרפים צורניים (בפרק 5).




**איור 1 1 : הפשטות ומתמרים אשר פותחו בתזה**

## 1.1.1 הפשטות מדויקת לרשימות חד-מקושרות

בפרק 3 אנו מציגים הפשטה דיסיונקטיבית מדויקת למדי לתכניות המכילות מספר סופי של רשימות (עם אפשרות למעגליות ובעלות אורך לא חסום). הרעיונות הללו אומצו והורחבו ע"י חוקרים אחרים בעבודות מאוחרות יותר [APV06,APV08,LAIS06a].

**קידוד מרחבים מופשטים.** אנו מראים כיצד לקודד את האלמנטים המופשטים בעזרת הפשטת פרדיקטים [GS97] ובעזרת הפשטה קנונית [SRW02] ומוכיחים שהקידודים שקולים במובן שהוגדר למעלה. (בפרק 5 אנו משתמשים בשיטה מותאמת וישירה יותר לקודד את האלמנטים המופשטים בעזרת גרפים צורניים.)

**יישום.** אנו משתמשים בהפשטה כדי להוכיח תכונות בטיחות בסיסיות ושמורות צורניות בפרוצדורות סטנדרטיות הפועלות על רשימות.

**תרומות עיקריות:** (1) אנו מגדירים דוגמא של הפשטה קנונית [SRW02] אשר מפשיטה רשימות מעגליות בדיוק רב יותר מאשר דוגמאות קיימות המבוססות על הפשטה קנונית; (2) אנו משווים הפשטת פרדיקטים והפשטה קנונית במונחים של מספר הפרדיקטים הנדרש לקודד הפשטה זהה ומראים שקילות להפשטת הרשימות; ו-(3) אנו מדווחים על הערכה אמפירית של אנליזה המבוססת על ההפשטה החדשה על קבוצת תכניות בדיקה.

## 2.1.1 הפשטת איזומורפיזם חלקי

בפרק 4 אנו מציגים הפשטה דיסיונקטיבית חלקית הפועלת מעל הפשטה קנונית. הרעיון הוא להשתמש ביחס שקילות על מבנים, ע"י *איזומורפיזם חלקי*, כדי להגדיר מושג של דמיון ביניהם. אנו ממזגים מבנים דומים למבנה יחיד, ובכך מפחיתים את מספר המבנים לאחר צירוף. אנו לא ממזגים מבנים אשר אינם דומים, מכיוון שאנו מחשיבים אותם כנפרדים ע"י תכונות אשר חשובות לאנליזה.

**קידוד מרחב מופשט.** אנו בונים על גבי התיאוריה של ניתוח צורני 3-ערכי ומקודדים את האלמנטים המופשטים ע"י מבנים 3-ערכיים. זה מאפשר לנו להגדיר הפשטה מאוד כללית ולעשות שימוש חוזר במתמרים אשר זמינים במערכת זו.

**יישום.** הפעלנו את האנליזה למגוון רב של תכניות בדיקה, אשר הוגדרו במשך השנים ע"י משתמשי מערכת ה-TVLA [LAS00], אשר כוללות תכניות סדרתיות ומקביליות, בדיקות של שגרות הפועלות על רשימות ועצים ובדיקה של חריגות שינוי מקביל בג'אווה.

**תרומות עיקריות:** (1) הפשטה מאוד כללית (וריאציות של רעיון זה אומצו ע" חוקרים אחרים [YLB+08]); (2) מימוש יציב בתוך TVLA; ו-(3) הערכה אמפירית על מגוון רב של תכניות בדיקה אשר מראות האצות דרמטיות לביצועי TVLA.

## 3.1.1 פירוק תתי-גרפים נפרדים

בפרק 5 אנו מגדירים הפשטה דיסיונקטיבית חלקית לניצול תלויות חלשות בין מבני נתונים שונים, נפרדים. הפשטה זו מאפשרת להפחית גורמים אקספוננציאליים בניתוח תכניות המנהלות מספר רב של מבני נתונים נפרדים.

**קידוד מרחב מופשט.** אנו מבטאים את המצבים הקונקרטיים ואת המצבים המופשטים בעזרת גרפים צורניים מותאמים. אנו עושים זאת לשם פשטות ההצגה, כאשר למעשה ניתן לעשות הצגה מחדש של התוצאות במונחים של מבנים לוגיים בצורה ישירה.

**יישום.** אנו מימשנו והפעלנו את האנליזה לתכניות המשנות מספר מרובה של רשימות חד-מקושרות מעגליות, כולל תכניות המעוצבות בדומה למנהלי התקנים במערכת ההפעלה חלונות.

**תרומות עיקריות.** (1) אנו מציגים סוג חדש של הפשטה צורנית אשר משתמשת בקיום מבני נתונים נפרדים כדי להפחית את גובה המרחב המופשט; (2) אנו חוקרים את סיבוכיות המתמרים המופשטים ומראים שהמתמרים המדויקים ביותר הם NP-complete; (3) אנו מציעים מתמרים ($\tau^{GD}$) פולינומיאליים ויעילים אשר אינם המדויקים ביותר, אבל בדרך-כלל טובים באופן מעשי, כלומר האנליזה המתקבלת מדויקת דיה כדי להוכיח תכונות כמו האנליזה המבוססת על הפשטה דיסיונקטיבית; ו-(4) אנו מימשנו והראינו האצות נכבדות של האנליזה על קבוצת תכניות בדיקה הפועלות על רשימות.

## 4.1.1 פירוק קרטזי של תתי-ערמות

בפרק 6 אנו מציגים מערכת לבניית הפשטות דיסיונקטיביות חלקית המבוססת על הרעיון של פירוק מבנים לוגיים לתתי-מבנים ושימוש בהפשטה קרטזית. משתמש של המערכת יכול לפרט את הפירוקים השונים והמתמרים השונים ולהיות מובטח בקבלת אנליזה ניאותה.

**קידוד מרחב מופשט.** אנו מציגים את הרעיונות ע"י שימוש בערמות קונקרטיות. האלגוריתמים משולבים לתוך TVLA ומשתמשים במבנים לוגיים. דבר זה מאפשר מערכת כללית ביותר.

**יישום.** הפעלנו את הרעיונות לניתוח תכניות מקביליות בגרעיניות עדינה הפועלות על מבני נתונים של רשימות כדי להוכיח תכונות בטיחות בסיסיות (לדוגמא, אי-ביצוע של פניות למחוון אפס) וליניאריזציה [HW90], ע"י בנייה מעל האנליזה של עמית ושות' [ARR+07].

**תרומות עיקריות.** (1) אנו מגדירים את המושג של פירוק ערמות ומראים כיצד ניתן להשתמש בו יחד עם הפשטה קרטזית; (2) אנו מפתחים טכניקות למתמרים ניאותים ויעילים ($\tau_1[TS], \tau_2[TS]$) אשר מתכנן האנליזה יכול לקבוע להם את הפרמטרים; (3) האלגוריתמים משולבים לתוך TVLA, כאשר מתכנן האנליזה יכול לפרט פרמטר פירוק $D$ ומפרמטר מתמר $TS$ ולקבל, באופן אוטומטי, אנליזה ניאותה, אשר איתה הוא יכול להתנסות; ו-(4) אנו מראים את השימושיות של המערכת לניתוח תכניות מקביליות בגרעיניות עדינה ובדיקת לינאריזציה. הראינו שהשימוש בטכניקות הללו הוא חיוני לבדיקה יעילה של לינאריזציה בתכניות עם מספר לא-חסום של פתילים [BLAM+08].

בעזרת הפשטת פירוק קרטזי ניתן לתפוס את הפשטת הפירוק לתתי-גרפים, שהוצגה בפרק 5, כולל המתמרים שהוצגו שם.

הפשטת האיזומורפיזם החלקי והפשטת הפירוק הקרטזי ניתנות לשילוב. אנו משתמשים בשתי ההפשטות כדי לקבל את האנליזות והתוצאות המדווחות בפרק 6 ובעבודה עוקבת [BLAM+08].

## 2.1 מבנה התזה

שאר התזה מאורגנת באופן הבא:

- בפרק 2 אנו כוללים חומרי רקע על הפשטה קנונית [SRW02];
- בפרק 3 אנו מציגים אבסטרקציה סופית לזיכרונות המכילים מספר חסום של רשימות חד-מקושרות (בעלות אורך לא חסום) ומתארים את הקידוד של ההפשטה בפורמליזמים שונים;
- בפרק 4 אנו מציגים הפשטה דיסיונקטיבית חלקית מעל הפשטה קנונית המבוססת על מיזוג מצבים מופשטים דומים ואת יישומה למגוון רב של אנליזות;
- בפרק 5 אנו מציגים הפשטה דיסיונקטיבית חלקית המבוססת על פירוק של ערמות (מופשטות) לתתי-ערמות נפרדות ואת יישומה לניתוח תכניות סדרתיות;
- בפרק 6 אנו מציגים מערכת פרמטרית להפשטות דיסיוקנטיביות חלקית המבוססת על פירוק ערמות (מופשטות) לקבוצות של ערמות – לא-בהכרח נפרדות – ואת יישומה לניתוח תכניות סדרתיות ומקביליות;
- פרק 7 מסכם את התזה ודן בכיווני מחקר עתידיים.