מבחן סוף סמסטר – מועד ב' טור 1

מרצה אחראי: דייר שחר יצחקי

מתרגלים: הילה לוי, תומר כהן, גיא ארבל, אנדריי בבין

: הוראות

- 1. בטופס המבחן 14 עמודים, מתוכם 5 דפי נוסחאות. בדקו שכל העמודים ברשותכם.
 - 2. משך המבחן שלוש שעות (180 דקות).
 - 3. כל חומר עזר חיצוני אסור לשימוש.
- 4. בשאלות הפתוחות, ניתן לציין לגבי סעיף או שאלה ״לא יודע/ת״ בלבד. תשובה זו תזכה ב-20% מהניקוד.תשובות שגויות לא יזכו בניקוד.
 - 5. קראו את כל המבחן לפני שאתם מתחילים לענות על השאלות.
 - 6. אין צורך להגיש את טופס מבחן זה בתום הבחינה.
- 7. את התשובות לשאלות הסגורות יש לסמן בטופס התשובות הנפרד בלבד. את התשובות לשאלות הפתוחות יש לכתוב במקומות המסומנים בבחינה.
 - 8. ודאו כי אתם מגישים טופס תשובות וטופס הבחינה בלבד (את מחברת הטיוטה ניתן לשמור אצלכם).

בהצלחה!

חלק א' - שאלות סגורות (45 נק')

שלבי קומפילציה (27 נקי)

:FanC נתונה התוכנית הבאה בשפת

```
int a = 15;
2. int b = 1;
3. int m = 1;
   while (a > 0) {
5.
        a = a - 1;
        b = b * 2;
6.
7.
        m = (m * a) - b;
8.
9.
              int c = 1;
10.
              m = m * b;
11.
        }
12. }
13. print("11");
```

בסעיפים הבאים (שאלות 1 עד 5) מוצגים שינויים (בלתי תלויים) לקוד של התוכנית. עבור כל שינוי ציינו את השלב המוקדם ביותר שבו נגלה את השגיאה :



<u>טאלון בוטבו ב.</u> (3 נקי) מוחקים את שורה 8

- אין שגיאה 💥
- ב. שגיאה בניתוח סמנטי
 - שגיאה בזמן ריצה 🌊
- שגיאה בניתוח תחבירי
- שגיאה בניתוח לקסיקלי
 - שגיאון בניונוון כקטיק
 - שגיאה בייצור קוד 👌

שאלה מספר 2:

c = b * 2;

(3 נקי) מחליפים את שורה 6 ב-

- א. שגיאה בזמן ריצה
- ב. שגיאה בניתוח לקסיקלי
 - ג. אין שגיאה
- ד. שגיאה בניתוח תחבירי
- ה שגיאה בניתוח סמנטי
 - שגיאה בייצור קוד

שאלה מספר 3: (3 נקי) מחליפים את שורה 13 בprint(11); א. שגיאה בזמן ריצה ב. אין שגיאה שגיאה בניתוח סמנטי ד. שגיאה בניתוח תחבירי ה. שגיאה בייצור קוד ו. שגיאה בניתוח לקסיקלי שאלה מספר 4: int a == 15; (3 נקי) מחליפים את שורה 1 ב-א. שגיאה בזמן ריצה שגיאה בניתוח הסמנטי ב. ג. אין שגיאה שגיאה בניתוח תחבירי ה. שגיאה בניתוח לקסיקלי ו. שגיאה בייצור קוד שאלה מספר 5: func(a > 0) { (3 נקי) מחליפים את שורה 4 ב-א. אין שגיאה שגיאה בניתוח לקסיקלי ב. שגיאה בזמן ריצה ډ. ד. שגיאה בייצור קוד

> ה שגיאה בניתוח תחבירי ו. שגיאה בניתוח סמנטי

שינויים ב-FanC- שאלות 6-7:

נרצה להוסיף לשפת FanC תמיכה בהוראה חדשה עבור קריאות לפונקציות ספרייה. בכל הצהרה (statement) שהיא קריאה לפונקציה, ניתן להוסיף @ לפני שם הפונקציה כדי להפוך את הקריאה ל-cALL בכל הצהרה היא שבמקום לבצע פקודת CALL בזמן ריצה, גוף הפונקציה מודבק לתוך הקוד.

למשל, נניח כי קיימת פונקציית ספרייה בשם double_print אשר מקבלת כפרמטר מחרוזת בשם str והגוף שלה הוא הקוד הבא:

```
    print(str);
    print(str);
    fanC יפלוט קוד זהה עבור שתי התוכניות הבאות:

            @double_print("A");
            if (11 > 22) {
            print("A");
            double_print("B");
            double_print("B");
            double_print("B");

    double_print("B");
    double_print("B");
    f (11 > 22) {
    double_print("B");
    f (11 > 22) {
```

Carv

שאלה מספר 6:

(6 נקי) על מנת לתמוך בפיציר החדש, מהו שלב הקומפילציה המוקדם ביותר מבין הרשומים כאן בו <u>לא</u> נבצע שינוי?

- א. ייצור קוד
- ב. ניתוח תחבירי
- ניתוח לקסיקלי
- . נצטרך לערוך שינויים בכל אחד משלבי הקומפילציה שרשומים כאן.

שאלה מספר 7:

 $\overline{(3)}$ נקי) בזמן שעבדת על מימוש הפיציר החדש, השותפה שלך לתרגילי הבית הוסיפה ל- $\overline{(3)}$ תמיכה ב- $\overline{(3)}$ נקי) בזמן שעבדת על מית הפיציר החדש, הוספת דגל "no-inline-י" אשר מכבה את פעולת ה- $\overline{(3)}$ (מתעלם מהתו $\overline{(3)}$). כאשר משווים ריצה בלי $\overline{(3)}$ לריצה עם $\overline{(3)}$ (בשתי הריצות יש $\overline{(3)}$), נסתכל על הטענות הבאות:

- .1 עבור כל תוכנית inline מגדיל את קוד הביניים שנוצר.
- 2. עבור כל תוכנית inline מקטין או לא משנה את זמן הריצה.

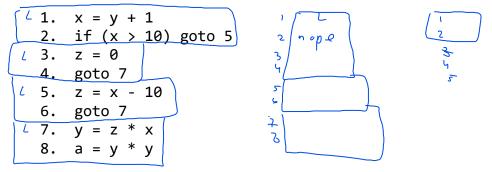
איזו מהטענות נכונה?

- א. רק הטענה השנייה נכונה.
 - שתי הטענות נכונות.
 - אף טענה אינה נכונה. 🦃
- . רק הטענה הראשונה נכונה.

אופטימיזציות (18 נקי)

שאלה מספר 8:

: (6 נקי) ליוסי נתון הקוד הבא



יוסי לא אוהב הרבה בלוקים בסיסיים, הוא הריץ את האלגוריתם שלמד בכיתה על מנת למצוא את מספר הבלוקים המינימלי של קטע הקוד שלו.

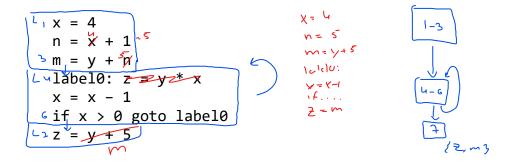
יוסי הופתע לרעה מכמות הבלוקים ולכן הוא מוכן לסכן את נכונות הקוד שלו על ידי מחיקת שורה בודדת! כאשר מוחקים שורה נשים במקומה את הפקודה NOPE אשר אינה עושה דבר.

מהו השינוי הגדול ביותר שיוסי יכול לעשות למספר הבלוקים הבסיסיים המינימלי על ידי מחיקה של שורה בודדת?

- א. קטן ב- 3
- ב קטן ב- 1
- ג. קטן ב- 2
- ד. מספר הבלוקים המינימלי לא משתנה לא משנה איזו שורה נמחק.

<u>שאלות 10-9:</u>

: נתון קטע הקוד הבא



שאלה מספר 9:

(6 נקי) ציירו את ה-CFG של קוד הביניים הנתון. מה מספר הצמתים והקשתות בגרף!

- א. 3 קשתות, 4 צמתים ב. 3 קשתות, 3 צמתים
- ל קטונות, 4 צמתים 1. אף תשובה אינה נכונה 1. אף תשובה אינה נכונה
 - אן אנטובווא זיין אנטור א קשתות, 3 צמתים

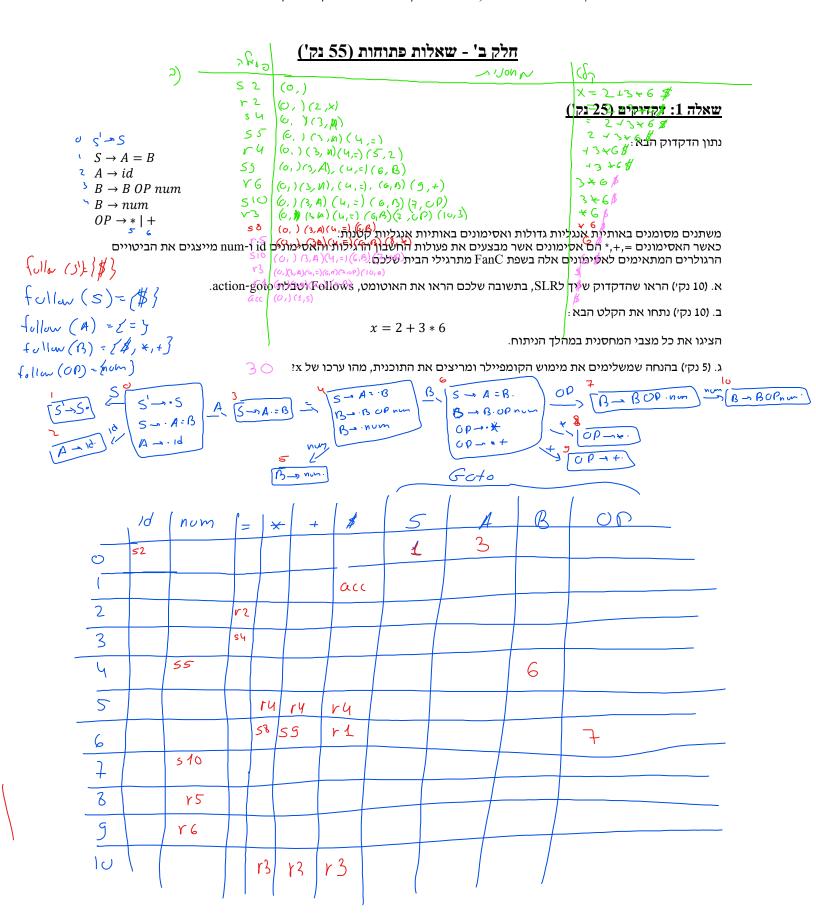
שאלה מספר 10:

היים אות קטע קוד הביניים הכולל שמכיל את קטע זה הופעלה תחילה אנליזת חיות אשר דיווחה כי לאחר קטע קוד זה חיים zו-z בלבד.

כעת עליך להריץ את כל האופטימיזציות השונות שנלמדו בכיתה עד לקבלת קוד אופטימלי.

איזו אופטימיזציה **אינה** מתבצעת!

- Useless code elimination
 - Constant propagation 🛣
- Commonsub-expression elimination **
 - Constant folding
- כל האופטימיזציות שמצויינות בתשובות מתבצעות.





שאלה 2: אנליזה סטטית (30 נק')

טונטו חזר משירות מילואים ונזכר שלא סיים את מימוש הקומפיילר שעליו הוא עבד בסמסטר שעבר. הוא עצר בשלב שבו יש לו frontend אשר הפלט שלו הוא קוד ביניים בצורה של 3AC. הנה כמה דוגמאות של תוכניות שנוצרו, מחולקות לבלוקים בסיסיים.

absur		signur	sum_squr
0: 1: 2:	w0 := readi() if w0 > 0 goto 2 w1 := -w0 t0 := readi() t1 := 3 * t0 goto 3	0: w0 := readi() if w0 > 0 goto 5 1: if w0 = 0 goto 3 2: x0 := 45 goto 4 3: x1 := 48	0: $n := readi()$ i0 := 1 sum0 := 0 1: $i1 := \varphi(i0, i2)$ if i1 >= n goto 4 2: $t0 := i1 * i1$
3:	$w2 := \varphi(w0, w1)$ t2 := w1 + t1 printi(t2)	 4: print("nisse") goto 6 5: x2 := 43 6: x4 := φ(x0, x1, x2) printi(x4) 	$sum1 := \varphi(sum0, sum2) \\ sum2 := sum1 + t0 \\ goto 3$ 3: $i2 := i1 + 1 \\ goto 1$ 4: $sum4 := \varphi(sum0, sum2) \\ printi(sum4)$

כעת הוא חייב להשלים בהקדם את ה-backend ולייצר קוד אסמבלי. לצורך כך הוא מעוניין להמיר את קוד הביניים שלו לתצורה של בעת הוא חייב להשלים בהקדם את הבעיה הבעיה המרכזית LLVM וכך לסיים את המשימה מהר יותר. טונטו נזכר מה הייתה הבעיה המרכזית עד בגינה לא הספיק להשלים את המשימה לפני תקופת מועדי א': בתשתית LLVM, האופרטור φ (פִּי) צריך לקבל, בנוסף למשתנים המייצגים את הערכים, גם את הבלוק שממנו מגיעה ההגדרה של כל אחד מהארגומנטים.

יתר על כן, כל ציון בלוק המוצמד לארגומנט חייב להיות בלוק קודם ישיר (direct predecessor) של הבלוק שבו מופיע השימוש באופרטור ϕ כלומר חייבת להיות קשת אחת בדיוק המקשרת ביניהם, ואסורים מסלולים באורך גדול מ-1.

בתור שלב ראשון, טונטו מריץ על הקוד אנליזה בשם Elvish Definitions, שהיא גרסה של Reaching Definitions שהותאמה על-ידי אלפור לקוד SSA באופן הבא:

 $L = P(Vars); \sqsubseteq = \subseteq$

Statement at l	kill(ℓ)	gen(l)
$v := \varphi(u_1, \ldots, u_k)$	$\{u_1,\ldots,u_k\}$	{ v }
$v := expr$ (that is not φ)	Ø	{ v }
anything else	Ø	Ø

שימו לב: מכיוון שב-SSA לכל משתנה יש בדיוק הגדרה אחת, אלפור החליט שמספיק לשמור רק את שמות המשתנים באנליזה (Vars). כמו כן הוסכם כי הגדרה של משתנה מפסיקה להיות תקפה ברגע שהמשתנה משמש כארגומנט לאופרטור au.

א. (10 נקי) ציירו את גרף בקרת הזרימה של כל אחת מהתוכניות שבטבלה בראש העמוד, והריצו את אנליזת Elvish Definitions על הגרפים שהתקבלו. יש לרשום תוצאות סופיות בלבד (out ו-out של כל בלוק).

(הבלוקים באנליזה הם הבלוקים הבסיסיים של התוכנית בשפת הביניים ; כאשר בבלוק יש מספר השמות ברצף, מופעלות, כרגיל, פונקציות המעבר ברצף בזו אחר זו.)

: ב. (20 נקי) כעת על טונטו להחליף את הביטויים מסוג $\varphi(u_1,\ldots,u_k)$ לביטויים מסוג שמתאים ל כעת על טונטו להחליף את הביטויים מסוג $\varphi(u_1,\ldots,u_k)$ ב. $\varphi([u_1,\ell_1],\ldots,[u_k,\ell_k])$

שבו כל זוג מורכב משם של משתנה u_i ומתווית של בלוק הקודם לבלוק שבו נמצאת ההשמה, כך שבכל פעם שהריצה מגיעה אל השבו כל זוג מורכב משם של משתנה v_i ומתווית של הערך של u_i .

,absur טונטו שם לב, כי ההגדרה של u_i לא בהכרח מגיעה מבלוק קודם ישיר, ולכן לא ניתן להשתמש בתווית הזאת. למשל: בתוכנית הוא לא יכול להחליף את ההשמה בבלוק u_i באופן הבא:

 $w2 := \varphi([w0, 0], [w1, 1]) \leftarrow (3 בכלוק בבלוק) אינו חוקי$

יתרה מכך, המסלול שמוביל מהגדרות של u_i שונים עשוי לעבור דרך **אותו** בלוק ישיר קודם. למשל: בתוכנית absur, ההגדרות של w_i שונים עשוי לעבור דרך אותו בלוק ישיר קודם. למשל: בתוכנית w_i את ההשמה אזיעות שתיהן לבלוק 3 דרך בלוק 2 (שהוא הבלוק הקודם היחיד של 3). כלומר, טונטו גם לא יכול להחליף את ההשמה w_i את ההשמה בלוק 3 באופן הבא:

$$w2 := \varphi([w0, 2], [w1, 2]) \leftarrow$$
הביטוי הזה אינו חוקי

הפתרון היחיד במקרה זה הוא להוסיף משתנה זמני חדש בתחילת בלוק 2:

$$t_w0_w1 := \varphi([w0, 0], [w1, 1])$$

ולשנות את ההשמה בתחילת בלוק 3 ל-

$$t2 := t_w0_w1$$

העיקרון המנחה הוא שכל פקודת ϕ תופיע במיקום הראשון שבו ההגדרות (של w0 ו-w1, במקרה זה) נפגשות.

 φ עזרו לטונטו לכתוב אנליזה סטטית שבעזרתה יוכל הקומפיילר שלו להוסיף את פקודות φ החדשות וכן לשנות את כל פקודות הקיימות בהתאם. טונטו גילה בפתקית נייר שנשכחה בין כריות של ספה ישנה במעונות כי הדומיין של האנליזה עשוי להיות

$$L = P(Lab \times P(Vars)); \sqsubseteq = \subseteq$$

כאשר Vars = קבוצת המשתנים, בוצת התוויות.

הגדירו את פונקציות המעברים. הקפידו על הגדרות מתמטיות מדויקות, תוך שימוש בסימונים הבאים:

$\operatorname{pred}(\ell)$	ℓ קבוצת הבלוקים (תוויות) שהם הקודמים הישירים של הבלוק בעל התווית	
$\operatorname{Ein}(\ell)$, $\operatorname{Eout}(\ell)$	ℓ על הבלוק, Elvish Definitions, על הבלוק, על האנליזה	
С	קבוצה של קבוצות של משתנים המתארת את כל המופעים של $arphi$ (בתוכנית המקורית) ; כל איבר ב-C הוא קבוצה של משתנים המופיעים יחד כאופרנדים בפקודת $arphi$ בודדת כלשהי	

מותר להניח כי:

- התוכנית היא תקינה ובצורת SSA.
- כל השימושים במשתנים בתוכנית מכבדים את ההגדרה של Elvish Definitions, כלומר יש הגדרה של המשתנה שמגיעה לאותו שימוש לפי האנליזה של אלפור.
 - דרגת הכניסה של כל בלוק בסיסי בגרף בקרת הזרימה היא לכל היותר 2.

הריצו את האנליזה על שלושת הגרפים מסעיף א. יש לרשום תוצאות סופיות בלבד (in ו-out של כל בלוק).

הסבירו כיצד הקומפיילר ישתמש בתוצאות האנליזה על מנת להמיר את התוכנית לצורה הרצויה. הדגימו זאת על התוכניות signur signur ו-sum_squr

נוסחאות ואלגוריתמים

G = (V, T, P, S) כל ההגדרות מתייחסות לדקדוק

Top Down

```
\begin{split} & \text{first}(\alpha) = \big\{ \ t \in T \mid \alpha \Rightarrow^* t\beta \land \beta \in (V \cup T)^* \ \big\} \\ & \text{follow}(A) = \big\{ \ t \in T \cup \{\$\} \mid S\$ \Rightarrow^* \alpha A t\beta \land \alpha \in (V \cup T)^* \ \land \beta \in (V \cup T)^*(\epsilon |\$) \ \big\} \\ & \text{select}(A \rightarrow \alpha) = \left\{ \begin{array}{l} & \text{first}(\alpha) \cup \text{follow}(A) & \alpha \Rightarrow^* \epsilon \\ \\ & \text{first}(\alpha) & \text{otherwise} \end{array} \right. \end{split}
```

:LL(1) עבור דקדוק $M: V \times (T \cup \{\$\}) \rightarrow P \cup \{error\}$ עבור עבלת מעברים

```
M[A , t] = \begin{cases} A \to \alpha & t \in select(A \to \alpha) \\ error & t \notin select(A \to \alpha) \text{ for all } A \to \alpha \in P \end{cases}
```

:LL(1) אלגוריתם מנתח

```
Q.push(S)
while !Q.empty() do
    X = Q.pop()
    t = next token
    if X ∈ T then
        if X = t then MATCH
        else ERROR
    else // X ∈ V
        if M[X , t] = error then ERROR
        else PREDICT(X , t)
    end if
end while
t = next token
if t = $ then ACCEPT
else ERROR
```

Bottom Up

```
t\in T\cup \{\$\}\ ,A\to\alpha\beta\in P\ \text{Casur}\ (A\to\alpha\bullet\beta\ ,\ t)\ \text{The}\ LR(1) סגור (closure) על קבוצת פריטים I מוגדר באופן אינדוקטיבי: closure(I) = I מוגדר באופן אינדוקטיבי:  A\to\alpha\bullet\beta\beta\ ,\ t)\in \text{closure}(I)=I צעד: אם (A \to\alpha\bullet\beta או לכל I מוגדר באופן אינדוקטיבי:  A\to\alpha\bullet\beta\beta\ ,\ t)\in \text{closure}(I) צעד: אם (B \to\bullet\gamma\ ,\ x)\in \text{closure}(I) (B \to\bullet\gamma\ ,\ x)\in \text{closure}(I) פונקציית המעברים של האוטומט:  \delta(I\ ,\ X)=\bigcup \Big\{\ \text{closure}(A\to\alpha X\bullet\beta\ ,\ t)\mid (A\to\alpha\bullet X\beta\ ,\ t)\in I \Big\}
```

:SLR למנתח מבלת action הגדרת טבלת

$$\begin{aligned} \text{action}[i \ , \, t] = & \begin{cases} \text{SHIFT}_j & \delta(I_i \ , t) = I_j \\ \text{REDUCE}_k & \text{rule k is } A \to \alpha, \, (A \to \alpha \bullet) \in I_i \text{ and } t \in \text{follow}(A) \\ \text{ACCEPT} & (S' \to S \bullet) \in I_i \text{ and } t = \$ \\ \text{ERROR} & \text{otherwise} \end{cases}$$

:LR(1) למנתח action הגדרת טבלת

$$\text{action[i , t] = } \begin{cases} \text{SHIFT}_j & \delta(I_i \text{ ,t}) = I_j \\ \text{REDUCE}_k & \text{rule k is } A \rightarrow \alpha \text{ and } (A \rightarrow \alpha \bullet \text{ , t}) \in I_i \\ \text{ACCEPT} & (S' \rightarrow S \bullet \text{ , \$}) \in I_i \text{ and } t = \$ \\ \text{ERROR} & \text{otherwise} \end{cases}$$

:LR(1) -ו SLR למנתח goto הגדרת טבלת

$$\label{eq:goto} \text{goto}[i \ , X] = \left\{ \begin{array}{ll} j & \delta(I_i \ , X) = I_j \\ \\ \text{error} & \text{otherwise} \end{array} \right.$$

:shift/reduce אלגוריתם מנתח

קוד ביניים

סוגי פקודות בשפת הביניים:

```
    x := y op z
    x := op y
    x משפטי השמה עם פעולה בינארית
    x := y
    a משפטי העתקה
    a משפטי בלתי מותנה
    a קפיצה בלתי מותנה
    c קפיצה מותנה
    d קפיצה מותנה
    d הדפסה
```

Data-Flow Analysis

G = (V, E) מהצורה CFG- מהייחסות מתייחסות

הצורה הכללית של המשוואות בחישוב סריקה קדמית:

```
in(B) = \bigcup_{(S,B)\in E} out(S)
out(B) = f_B(in(B))
```

הצורה הכללית של המשוואות בחישוב סריקה אחורית:

```
\operatorname{in}(B) = \bigcup_{(B,D) \in E} \operatorname{out}(D)
\operatorname{out}(B) = f_B(\operatorname{in}(B))
```

שפת FanC

אסימונים:

תבנית	אסימון אסימון
int	INT
byte	BYTE
b	В
bool	BOOL
and	AND
or	OR
not	NOT
true	TRUE
false	FALSE
return	RETURN
if	IF
else	ELSE
while	WHILE
break	BREAK
continue	CONTINUE
•	SC
(LPAREN
)	RPAREN
{	LBRACE
}	RBRACE
=	ASSIGN
== != < > <= >=	RELOP
+ - * /	BINOP
[a-zA-Z][a-zA-Z0-9]*	ID
0 [1-9][0-9]*	NUM
"([^\n\r\"\\] \\[rnt"\\])+"	STRING

דקדוק:

- 1. $Program \rightarrow Statements$
- 2. $Statements \rightarrow Statement$
- 3. $Statements \rightarrow Statements Statement$
- 4. Statement \rightarrow LBRACE Statements RBRACE
- 5. $Statement \rightarrow Type\ ID\ SC$
- 6. Statement \rightarrow Type ID ASSIGN Exp SC
- 7. Statement \rightarrow ID ASSIGN Exp SC
- 8. $Statement \rightarrow Call SC$
- 9. Statement \rightarrow RETURN SC
- 10. Statement \rightarrow IF LPAREN Exp RPAREN Statement
- 11. Statement \rightarrow IF LPAREN Exp RPAREN Statement ELSE Statement
- 12. $Statement \rightarrow WHILE\ LPAREN\ Exp\ RPAREN\ Statement$
- 13. $Statement \rightarrow BREAKSC$
- 14. $Statement \rightarrow CONTINUE\ SC$
- 15. $Call \rightarrow ID LPAREN Exp RPAREN$
- 16. $Type \rightarrow INT$
- 17. $Type \rightarrow BYTE$
- 18. $Type \rightarrow BOOL$
- 19. $Exp \rightarrow LPAREN Exp RPAREN$
- 20. $Exp \rightarrow Exp \ BINOP \ Exp$
- 21. $Exp \rightarrow ID$
- 22. $Exp \rightarrow Call$
- 23. $Exp \rightarrow NUM$
- 24. $Exp \rightarrow NUM B$
- 25. $Exp \rightarrow STRING$
- 26. $Exp \rightarrow TRUE$
- 27. $Exp \rightarrow FALSE$
- 28. $Exp \rightarrow NOT Exp$
- 29. $Exp \rightarrow Exp \ AND \ Exp$
- 30. $Exp \rightarrow Exp \ OR \ Exp$
- 31. $Exp \rightarrow Exp \ RELOP \ Exp$
- 32. $Exp \rightarrow LPAREN Type RPAREN Exp$