به نام خدا

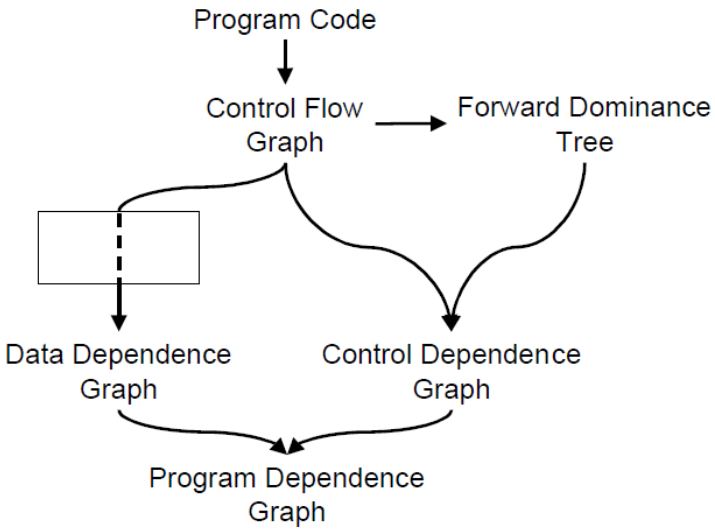
پروژه کارشناسی

گزارش فاز اول

سید محمد مهدی احمدپناه

در فاز قبل، سورس کد ورودی با استفاده از ابزارهای lex و yacc، گرفته می‌شد و در ادامه، در صورت وجود خطای نحوی، به کاربر اطلاع داده می‌شد.

حال در این مرحله، سورس کد ورودی معتبر و مطابق با گرامر زبان، گرفته می‌شود و گراف‌های مورد نیاز برای تحلیل ساخته می‌شود. هدف، تولید گراف وابستگی برنامه است که بدین منظور، فرایندهایی مطابق به شکل زیر انجام گرفت:



شکل ۱- نمودار کلی نحوه به دست آوردن گراف وابستگی برنامه

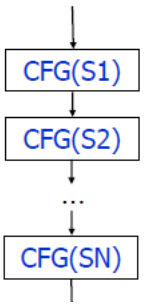
ابتدا گراف جریان کنترل[۱] یا به اختصار CFG به دست می‌آید. نحوه کار بدین شکل است که گزاره[۲]ها و عبارت[۳]های برنامه، به عنوان یگ گره[۴] در گراف در نظر گرفته می‌شود. هر بلوک پایه[۵]، شامل تعدادی گره است.

[1] Control Flow Graph
[2] Statement
[3] Expression
[4] Node
[5] Basic Block

هر بلوک پایه، فقط یک گره ورودی و یک گره خروجی خواهد داشت. برای این منظور، برای گزاره‌های ساده که اجرای آن‌ها مشروط نیست، به صورت دنباله پشت سر هم گره‌ها در نظر گرفته می‌شود.

CFG for Block Statement

CFG(S1; S2; ...; SN) =



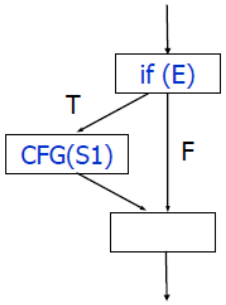
شکل ۲ - نحوه تولید زیرگراف بلوک پایه

پس برای این گزاره‌ها و عبارات، تنها ساختن یک گره جدید و متصل کردن آن‌ها به لیست پیوندی گراف کفایت می‌کند.

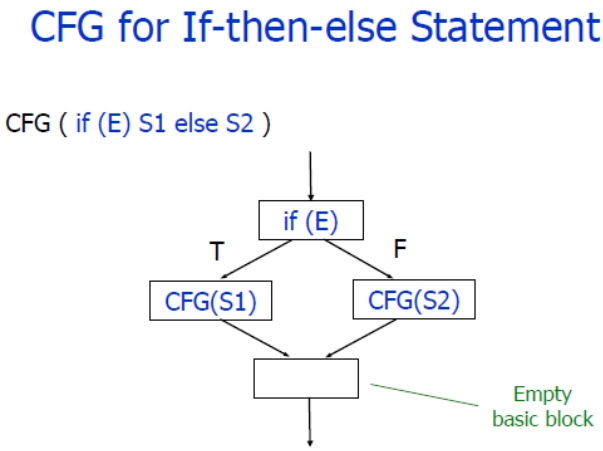
برای گزاره‌های شرطی، دو حالت زیر در نظر گرفته می‌شود:

CFG for If-then Statement

CFG(if (E) S)



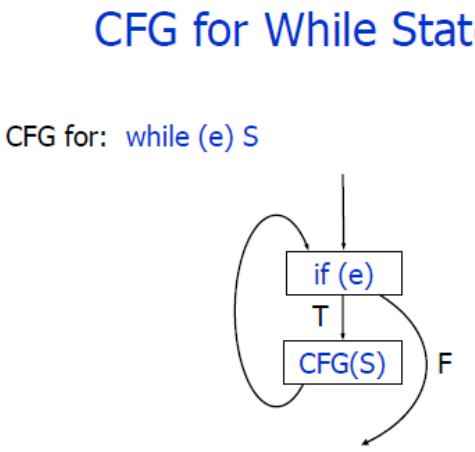
شکل ۳- نحوه تولید زیرگراف گزاره های شرطی - الف



شکل ۴- نحوه تولید زیرگراف گزاره‌های شرطی - ب

پس برای گره عبارت شرطی، دو پیوند برای برقراری و عدم برقراری شرط در نظر گرفته می‌شود. ضمنا به دلیل اینکه می‌خواهیم قالب کلی زیرگراف کنترل همه گزاره‌ها مشابه باشد، یک گره مجازی[6] در انتهای این‌گونه زیرگراف‌ها ایجاد می‌کنیم که مطمئن باشیم هر زیرگراف تنها یک مجرای خروجی دارد.

برای گزاره حلقه موجود در زبان، شکل زیر را در نظر می‌گیریم:



شکل ۵ - نحوه تولید زیرگراف گزاره‌های حلقه

مشابه قبل، یک گره مجازی به عنوان گره پایانی زیرگراف ساخته می‌شود.

[6] dummy

با توجه به زیرگراف‌های پایه‌ای فوق، در هر کدام از قاعده‌های موجود در گرامر زبان، حالت‌های بالا را ایجاد می‌کنیم و همزمان با ساخته شدن درخت تجزیه[۷]، CFG تولید می‌شود.

حال برای تولید گراف وابستگی کنترل[۸] یا CDG، نیاز به درخت غلبه رو به جلو[۹] یا درخت پس‌غلبه[۱۰] خواهیم داشت. برای تولید این درخت، الگوریتم ساخت درخت غلبه[۱۱] را بر روی معکوس CFG؛ یعنی همان گره‌ها ولی با جهت یال‌های معکوس شده، اعمال می‌کنیم.

در CFG، گره M بر گره N غلبه[۱۲] می‌کند، اگر و تنها اگر همه مسیرهای از گره شروع تا گره N، حتما و الزاماً از گره M بگذرند و گره M بر گره N اکیداً غلبه[۱۳] می‌کند، اگر و تنها اگر بر آن گره غلبه کند و M، همان گره N نباشد. واضح است که یک گره در CFG می‌تواند چندین غلبه‌کننده[۱۴] داشته باشد، اما برای تولید درخت غلبه، نزدیک‌ترین غلبه‌کننده که غلبه‌کننده بی‌درنگ[۱۵] اهمیت دارد. با استفاده از شبه‌کد زیر می‌توان غلبه‌کننده‌های یک گره در CFG را به دست آورد:

```
Compute Dominators(){
        For (each n ∈ NodeSet)
                Dom(n) = NodeSet
        WorkList = {StartNode}
        While (WorkList ≠ ϕ) {
                Remove any node Y from WorkList
                New = {Y} ∪ ∩x∈Pred(Y) Dom(X)
                If New ≠ Dom(Y) {
                        Dom(Y) = New
                For (each Z ∈ Succ(Y))
                WorkList = WorkList U {Z}
                }

        }
}
```

الگوریتم ۱ - محاسبه گره‌های غلبه‌کننده برای هر گره

[7] Parse tree
[8] Control Dependence Graph
[9] Forward Dominance Tree
[10] Post Dominance Tree
[11] Dominance Tree
[12] dominate
[13] Strictly dominate
[14] dominator
[15] Immediate dominator

سپس با توجه به مجموعه غلبه‌کنندگان به دست آمده از الگوریتم بالا و مقایسه با مجموعه غلبه‌کنندگان سایر گره‌ها، می‌توان گره غلبه‌کننده بی‌درنگ را یافت و درخت غلبه را تشکیل داد اما برای ما، درخت پس‌غلبه[16] کاربرد دارد.

به این صورت که، گره Z، گره Y را پس‌غلبه می‌کند اگر و تنها اگر همه مسیرهای از Y تا گره پایانی، حتما و الزاماً از Z عبور کنند. حال، در صورتی که این الگوریتم را برای معکوس CFG اعمال کنیم، درخت پس‌غلبه تولید می‌شود.

پس از این، برای ساخت CDG، به مرز پس‌غلبه[17] نیاز داریم. مرز پس‌غلبه گره X، مجموعه‌ای از گره‌هایی هستند که توسط X اکیدا پس‌غلبه نمی‌شوند اما گره‌های ما بَعدی[18] دارند که توسط X، پس‌غلبه می‌شوند. تعریف ریاضی این گره‌ها بدین شرح است:

$PDF(X) = \{y \mid (\exists z \in Succ(y) \text{ such that } x \text{ post-dominates } z) \text{ and } x \text{ does not strictly post-dominate } y\}$

که این مجموعه بیانگر نزدیک‌ترین نقاط انشعابی[19] است که به گره X منجر می‌شوند.

طبق قضیه زیر، می‌توان وابستگی‌های کنترلی برنامه را برای هر گره موجود در CFG، به دست آورد:

قضیه – گره y تعلق دارد به مجموعه PDF(X) اگر و تنها اگر X به Y وابستگی کنترلی داشته باشد.

حال با استفاده از الگوریتم زیر، می‌توان مجموعه PDF هر گره را به دست آورد که بیانگر وابستگی‌های کنترلی خواهد بود.

```
For each x in the bottom-up traversal of the postdominator tree do
    PDF(X) = ϕ
    Step 1: For each y in Predecessor(X) do
        If X is not immediate post-dominator of y then
            PDF(X) ← PDF(X) ∪ {y}
    Step 2: For each z that x immediately post-dominates, do
        For each y ∈ PDF(Z) do
            If X is not immediate post-dominator of y then
                PDF(X) ← PDF(X) ∪ {y}
```

الگوریتم ۲ - محاسبه گره‌های پس‌غلبه‌کننده مرزی برای هر گره

[16] Post-dominance tree
[17] Post-dominance Frontier
[18] successor
[19] Diverging points

پس از این محاسبات، وابستگی‌های کنترلی برنامه برای هر گره – گزاره یا عبارت برنامه ورودی – به دست آمده است. در نتیجه، تا این مرحله، گراف وابستگی کنترلی برنامه یا CDG تولید شده است.

حال نوبت تولید گراف وابستگی داده[20] یا به اختصار DDG است. وابستگی‌های داده مختلفی وجود دارد اما برای کار ما و طبق مقاله اصلی پروژه، ارتباط بین گره‌هایی که شامل مقداردهی یک متغیر و استفاده از آن متغیر هستند، اهمیت دارد؛ یعنی گره X به گره Y وابستگی داده‌ای دارد اگر و تنها اگر در گره Y متغیری وجود داشته باشد که در گره X مقداردهی شده باشد. پس با توجه به همین تعریف، مطابق با قواعد، متغیری که به آن مقداری نسبت داده شده یا استفاده شده است، نگه‌داری می‌شوند. حال برای به دست آوردن وابستگی‌های داده‌ای، در صورتی که در یک گره، از متغیری که در گره دیگری مقداردهی شده است، استفاده شود، یک وابستگی داده‌ای در نظر گرفته می‌شود. برای دقیق‌تر بودن و عدم محافظه‌کارانه بودن وابستگی‌ها، تنها نزدیک‌ترین گزاره‌ای که آن متغیر در آن مقداردهی شده است، وابستگی را خواهد داشت؛ و نه همه گزاره‌هایی که آن متغیر را مقداردهی کردند، که این کار با پیمایش CFG امکان‌پذیر است و انجام شده است.

پس از این مرحله، گراف وابستگی داده‌ای برنامه نیز آماده است. تنها کار باقی‌مانده، ترکیب این دو گراف که دارای گره‌های یکسان هستند، خواهد بود تا گراف وابستگی برنامه تولید شود.

برای نمایش گرافیکی این گراف‌ها، از ابزار GraphViz استفاده شده که سورس کد آن نیز به پروژه اضافه شده است و بهترین نحوه نمایش ممکن برای گراف را در یک فایل تصویری، به نمایش درمی‌آورد.

در فاز بعدی، الگوریتم بازنویس برای حالت غیرحساس به پیشرفت[21] پیاده‌سازی خواهد شد.

[20] Data Dependence Graph
[21] Progress insensitive

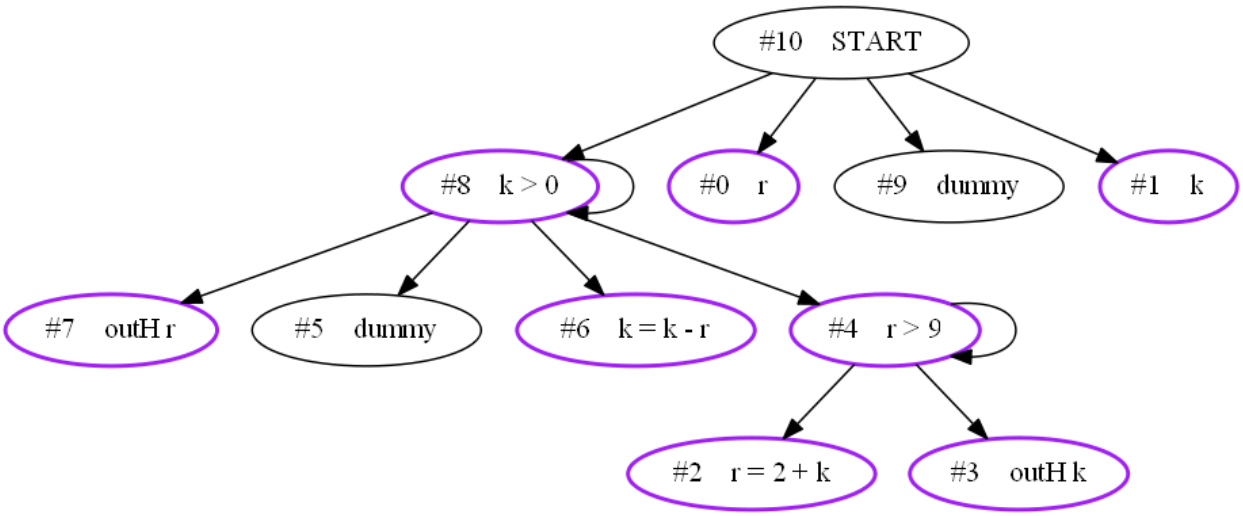
در ادامه نمونه‌ای از یک برنامه مورد آزمون[۲۲] آمده است (سایر موارد آزمون از طریق آدرس پروژه داده شده بر روی گیت‌هاب قابل مشاهده است):

فایل سورس کد ورودی:

```
program;
inH k , r;
while k>0 do
        while r>9 do
                r = 2 + k;
                outH k
        done;
        k = k - r;
        outH r
done
```

برنامه ۱ - یکی از موارد آزمون بررسی‌شده

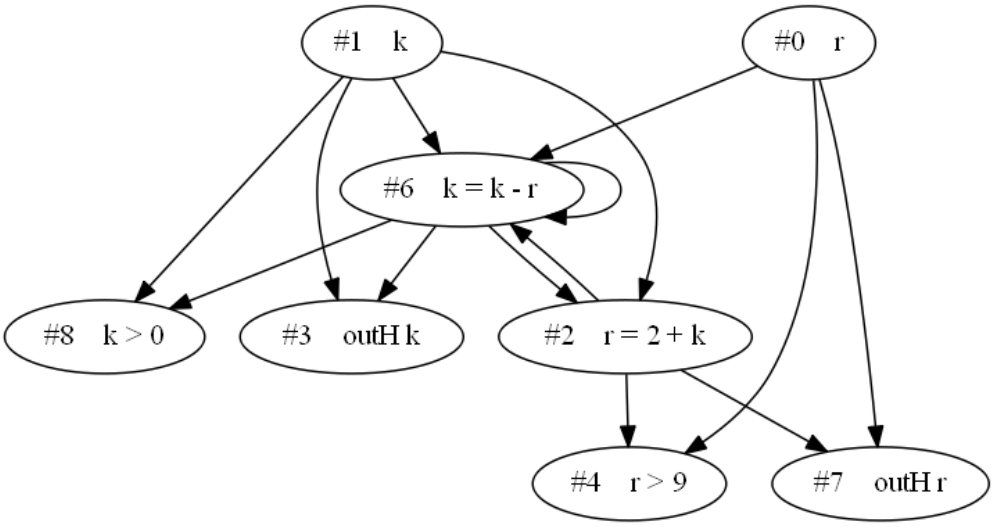
گراف وابستگی کنترل:



شکل ۶ - گراف وابستگی کنترل برای برنامه ۱

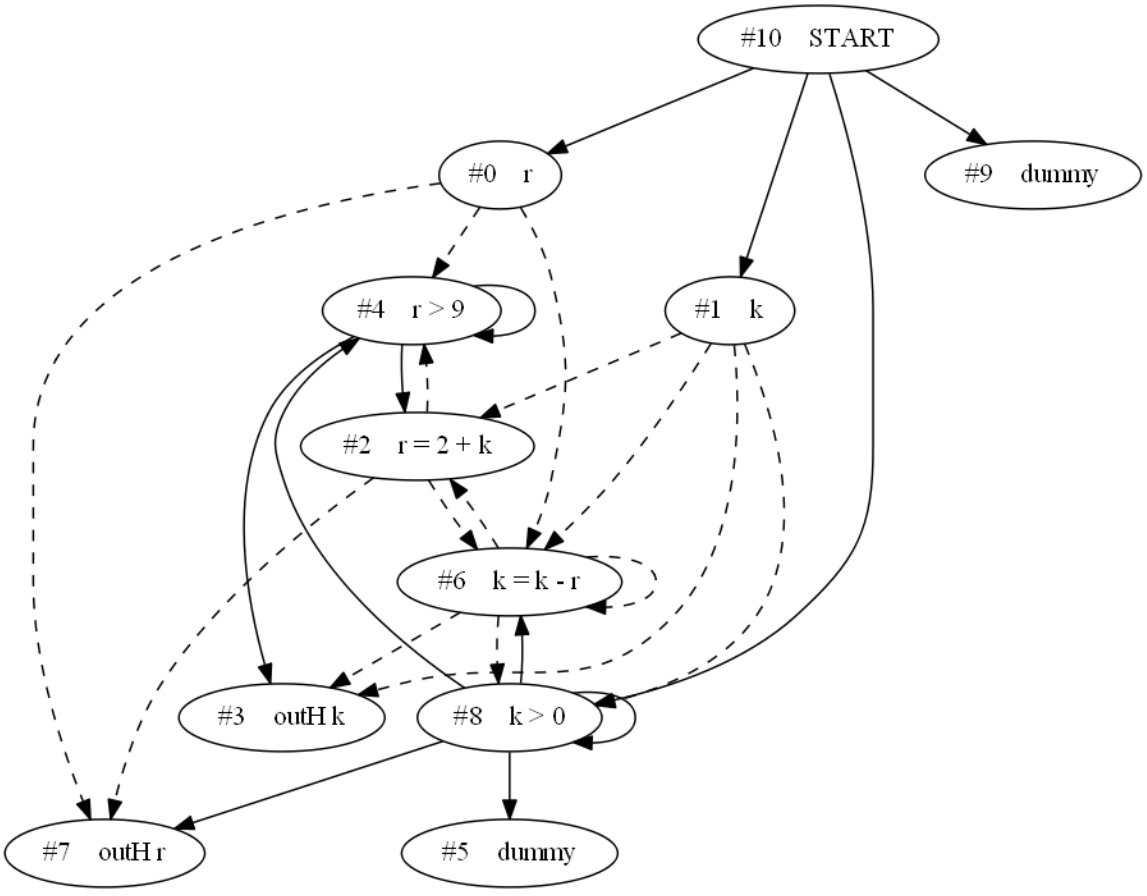
[22] Test case

گراف وابستگی داده:



شکل ۷ - گراف وابستگی داده برای برنامه ۱

گراف وابستگی برنامه:



شکل ۸ - گراف وابستگی برنامه برای برنامه ۱

مراجع مورد استفاده برای این مرحله:

- J. Ferrante et al. “The Program Dependence Graph and Its Use in Optimization”. ACM Transactions on Programming Languages and Systems, vol. 9, No. 3, pp. 319-349, July 1987.

- R. Cytron et al. “An Efficient Method of Computing Static Single Assignment Form”. Proceedings of the 16th ACM SIGPLAN-SIGACT symposium on Principles of programming languages, 1989, pp. 25-35.

- M. Wolfe and U. Banerjee. “Data Dependence and Its Application to Parallel Processing”. International Journal of Parallel Programming, vol. 16, No. 2, pp. 137-178, October 1987.

- E. R. Gansner and S. C. North. “An Open Graph Visualization and Its Application to Software Engineering”. Software – Practice and Experience Journal, vol. 30, No. 11, pp. 1203-1233, 2000.

مراحل انجام پروژه از طریق گیت‌هاب به آدرس https://github.com/smahmadpanah/BScProject قابل مشاهده و پیگیری است.

گزارش فازهای بعدی نیز به تدریج ارائه خواهد شد.