

# Capitolo 1

## CostCompiler

CostCompiler é un interprete per il linguaggio di programmazione definito nel capitolo precedente Grammatica ???. Una volta ricevuto il programma, CostCompiler procede alla verifica e correttezza sintattica e semantica del programma, successivamente si occupa della generazione dell'albero di sintassi astratta. Questo albero rappresenta una versione astratta del programma, che astrae i dettagli sintattici del codice e si concentra sulla sua struttura logica, associando ad ogni costrutto (eg. *if-then-else* un unico nodo, *IfNode* presente nel omonimo file in `src/ast`) i rispettivi sottonodi (nel caso di *IfNode* conterrà la guardia condizionale e i due statement). Dopo aver generato l'albero di sintassi astratta, CostCompiler si occupa della verifica Semantica ?? del linguaggio andando ad effettuare i controlli semantici e di tipo sul programma in input, andando a garantire alcune invarianti (eq. Le chiamate di funzioni devono rispettare i tipi di ritorno).

Una volta effettuata la verifica semantica, CostCompiler procede con la generazione delle equazioni di costo, andando a visitare l'AST secondo determinati criteri, ad ogni nodo figlio verrà passata una Mappa che contiene la mappatura di ogni variabile in una stringa che sarà la stessa stringa che compare nelle equazioni di costo.

Ogni nodo figlio invocato attraverso la funzione *toEquation()* ritorna una stringa, rappresentante l'equazione di costo del nodo figlio, e il padre va a concatenare le stringhe dei figli (anche in base al tipo di figlio da cui ricevere l'equazione), ad esempio la *return <EXP>* sarà diverso dal *return <function(Par) >*.

Il risultato finale di questo processo di concatenazione attraverso determinati nodi dell'AST é la generazione delle equazioni di costo. Una volta generate le equazioni di costo, CostCompiler le stampa in un file *equation.txt* inoltre lancia il risolutore PUBS 1.3 che va a calcolare gli upper bound del programma da stampare a video. Riportiamo un esempio di equazione di

costo generata da CostCompiler dato un programma scritto in HLCostLang:

```

1      struct Params {
2          address: array[int],
3          payload: any,
4          sender: string
5      }
6      service PremiumService : (string) -> void;
7      service BasicService : (any) -> void;
8      (isPremiumUser: bool, par: any) => {
9          if ( isPremiumUser ) {
10             call PremiumService("test");
11          } else {
12             call BasicService( par);
13          }
14      }

```

Listing 1.1: Listing8

Una volta preso in input Listing8, CostCompiler genera le seguenti equazioni di costo:

```

1 eq(main(P, ISPREMIUMUSER0, B), 0, [if9(ISPREMIUMUSER0, P, B)], []).
2 eq(if9(ISPREMIUMUSER0, P, B), nat(P), [], [ISPREMIUMUSER0=1]).
3 eq(if9(ISPREMIUMUSER0, P, B), nat(B), [], [ISPREMIUMUSER0=0]).

```

Listing 1.2: Equazioni di costo per Listing8

Andando a descriverle ci troveremo ad avere una equazione per la regola *init*, dove vediamo che *main* viene chiamata con costo 0 e verrà chiamata *if9* con parametri *ISPREMIUMUSER0, P, B*.

*P* e *B* sono il costo costante delle chiamate ai servizi *isPremiumUser* e *BasicService*, mentre *ISPREMIUMUSER0* sarà la valutazione del parametro *isPremiumUser* che sarà 1 se vero, 0 altrimenti; in altri termini *ISPREMIUMUSER0* sarà la valutazione della guardia del costrutto *if-then-else* e verrà eseguita la chiamata al servizio *PremiumService* se *ISPREMIUMUSER0* sarà 1 con costo *nat(P)*, altrimenti verrà eseguita la chiamata al servizio *BasicService* con costo *nat(B)*. Una volta avere generato l'equazioni di costo dal programma, lo stampiamo in un file *equation.txt*, così da poter eseguire PUBS(A Practical Upper Bounds Solver), per determinarci l'Upper Bound del programma. L'obiettivo di PUBS(che vedremo in seguito 1.3) è quello di ottenere automaticamente upper bound in forma chiusa per i sistemi di equazioni di costo, di conseguenza calcola i limiti superiori per la relazione di costo indicata come "Entry", oltre che per tutte le altre relazioni di costo di cui tale "Entry" dipende.

## 1.1 Regole di Inferenza

I programmi di costo sono elenchi di equazioni che hanno termini:

$$f(\bar{x}) = e + \sum_{i \in 0..n} f_i(\bar{e}_i) \quad [\varphi]$$

Dove le variabili si presentano nel lato destro e in  $\varphi$  sono un sottoinsieme di  $\bar{x}$ ; mentre  $f$  e  $f_i$  sono i simboli delle funzioni. Ogni funzione ha un right-hand-side che é un'espressione aritmetica che può contenere:

- Un'espressione in Presburger aritmetica (PA):

$$e ::= x \quad | \quad q \quad | \quad e + e \quad | \quad e - e \quad | \quad q \cdot e \quad | \quad \max(e_1, \dots, e_n)$$

Dove  $x$  é una variabile,  $q$  é una costante intera,  $e_1, \dots, e_n$  sono espressioni aritmetiche e  $\max$  é un operatore che restituisce il massimo valore tra le sue espressioni.

- Un numero di invocazioni di funzioni di costo:  $f_i(\bar{e}_i)$ .
- La guardia *varphi* é un vincolo congiuntivo lineare nella forma:  $e_1 \geq e_2$  dove  $e_1$  e  $e_2$  sono espressioni aritmetiche di Presburger.

La soluzione di un equazione di costo é il calcolo dei limiti di un particolare simbolo di una funziona (generalmente la prima equazione) e i limiti sono parametrici nei parametri formali dei simboli della funzione. Definiamo un insieme di regole di inferenza che raccolgano frammenti di programmi di costo che vengono poi combinati in modo diretto dalla sintassi. Usiamo una variabile di ambiente  $\Gamma$  come dizionari:

- $\Gamma$  prende un servizio o un parametro e ritorna un espressione aritmetica di Presburger che di solito é una variabile.
- Quando scriviamo  $\Gamma + i : \text{Nat}$ , assumiamo che  $i$  non appartenga al dominio di  $\Gamma$ .

I giudizi hanno forma:

- $\Gamma \vdash e : \text{Nat}$  che significa che il valore di  $E$  in  $e$  é rappresentato dalla costEspression  $E$
- $\Gamma \vdash S : e; C; Q$  significa che il costo di  $S$  nell'ambiente  $\Gamma$  é  $e + C$  dato un insieme di equazioni  $Q$

$$\frac{[\text{MAIN}] \quad \Gamma \vdash S : e; C; Q \quad \bar{w} = \text{Var}(\bar{p}, e) \cup \text{Var}(C)}{\Gamma \vdash \bar{p} \rightarrow \{S\} : 0; \emptyset; Q'; C} \quad (1.1)$$

$$\frac{[\text{CALL}] \quad \Gamma \vdash S : e; C; Q}{\Gamma \vdash \text{call } h(\bar{E})S : e + e'; C; Q} \quad (1.2)$$

$$\frac{[\text{IF}] \quad \begin{array}{l} \Gamma \vdash E : \varphi \quad \Gamma \vdash S : e'; C; Q \quad \Gamma \vdash S : e''; C'; Q' \\ W = \text{Var}(e, e', e'') \cup \text{Var}(C) \quad Q'' = \left[ \begin{array}{l} \text{if}_l(\bar{w}) = e' + c[\varphi] \\ \text{if}_l(\bar{w}) = e'' + c[\neg\varphi] \end{array} \right] \end{array}}{\Gamma \vdash \text{if } E \text{ then } S_1 \text{ else } S_2 : 0; \text{if}_l(\bar{w}); Q; Q'; Q''} \quad (1.3)$$

$$\frac{[\text{LET}] \quad \Gamma + i : \text{Not} \quad \bar{w} = \text{Var}(E) \cup \text{Var}(c)}{\Gamma \vdash \text{let } i = e \text{ in } S : S; e; C; Q} \quad (1.4)$$

$$\frac{[\text{FOR}] \quad \begin{array}{l} \Gamma \vdash M : e \quad \Gamma \vdash i : \text{Not} \quad \Gamma \vdash S : e'; C; Q \\ \bar{w} = \text{Var}(e, e') \cup \text{Var}(C) \quad i \quad Q' = \left[ \begin{array}{l} \text{for}_l(i, \bar{w}) = e + c \quad [i' e] \\ \text{for}_l(i, \bar{w}) = 0 \quad [i \geq e] \end{array} \right] \end{array}}{\Gamma \vdash \text{for } i \text{ in } (0, M) \quad S : 0; \text{for}_l(0, \bar{w}); Q; Q'} \quad (1.5)$$

$$(1.6)$$

Riassumiamo le regole descritte in precedenza:

- Regola[call] gestisce l'invocazione di un servizio; il costo della call sarà il costo di  $S$  più il costo per l'accesso al servizio  $h$
- Regola[if] gestisce il costrutto condizionale; quando la guardi é un'espressione definita in aritmetica di Presburger e il costo verrà rappresentato da entrambi i rami con i due condizionali  $\varphi$  e  $\neg\varphi$ . Rappresentiamo a livello di equazione  $\text{if}_l$  dove  $l$  é la linea di codice dove inizia il costrutto.
- Regola [for] descritto all'interno del rispettivo frammento di codice come  $\text{for}_l$  per lo stesso motivo citato in precedenza; Definiamo  $i$  come Nat e verifichiamo che non sia presente nell'ambiente  $\Gamma$  e scriviamo il rispettivo  $S$  come caso base in cui  $e \geq i$  oppure  $i \geq e + 1$

- Regola [LetIn] Dove viene definita  $E$  nell'ambiente  $\Gamma$  con costo  $e$  (il costo per eseguire l'espressione  $e$ ). Andremo a valutare se in  $\Gamma$  è presente  $E$  e andiamo a valutare  $\Gamma \vdash S$  che ritornerà un'equazione  $Q'$  con costo  $C$ .

## 1.2 Generazione delle Equazioni di costo

La generazione delle equazioni di costo viene eseguita andando a implementare le regole di inferenza viste in precedenza. Ogni nodo all'interno del nostro AST contiene il metodo `toEquation()` che prende come argomento la variabile del nostro ambiente  $\Gamma$  e sarà appunto un dizionario. Questo dizionario di tipo `EnvVar` è un `HashMap` che contiene come chiave l'oggetto `Nodo` della variabile e come valore la stringa rappresentante. Abbiamo deciso di utilizzare questo approccio per focalizzarci sull'efficienza del farci restituire la variabile che mappa quel determinato `Nodo`, senza dover andare a cercare all'interno dell'`HashMap` la chiave che mappa quel valore, cosa che viene fatta all'inserimento di un `Nodo`. L'inserimento del nodo però non sempre è un'operazione onerosa per il fatto che abbiamo già il controllo semantico che ci garantisce che non ci saranno variabili non dichiarate oppure variabili non dichiarate prima del loro utilizzo.

Andiamo ad analizzare un esempio semplice, all'interno del `Nodo` di tipo `CallService.java` che rappresenta l'invocazione di un servizio, abbiamo il metodo `toEquation()` che prende come argomento l'ambiente  $\Gamma$  e restituisce una stringa che rappresenta l'equazione di costo del nodo, che sarà una stringa riportata all'interno dell'equazione di costo del padre.

```

1      @Override
2      public String toEquation(EnvVar e) {
3          return "nat("+e.get(this)+")" + (stm!= null ? "+"+stm
4              .toEquation(e) : "");
5      }

```

La sezione `e.get(this)` ritorna la variabile mappata per quel determinato nodo, che è una stringa che rappresenta la variabile all'interno dell'equazione di costo, mentre la sezione `stm.toEquation(e)` è la chiamata sul metodo `toEquation()` del nodo figlio, che restituirà la stringa che rappresenta l'equazione di costo del nodo, andando a richiamare il medesimo metodo sui sottonodi contenuti all'interno del nodo figlio.

Per avere una panoramica completa del processo di generazione delle equazioni di costo, riportiamo il frammento di codice di `toEquation()` del `programNode`:

```

1      public String toEquation(EnvVar e){

```

```

2
3     for (Node n : decServices){
4         n.checkVarEQ(e);
5     }
6     StringBuilder equ = new StringBuilder();
7     for(Node n : funDec){
8         equ.append(n.toEquation(e));
9     }
10    equ.append(main.toEquation(e));
11    return equ.toString();
12 }

```

Come possiamo vedere, prima di generare le equazioni di costo del programma, andiamo a controllare che le variabili dichiarate all'interno dei servizi siano presenti all'interno dell'ambiente  $\Gamma$  e le mappiamo con determinate stringhe che appariranno nelle equazioni. Successivamente andiamo a iterativamente all'interno delle singole funzioni le generiamo e le concateniamo alla stringa che rappresenta le equazioni di costo del programma. Infine ci occupiamo di generare le equazioni di costo della funzione main, che saranno concatenate anch'esse con la stringa che rappresenta le equazioni di costo del programma.

## 1.3 PUBS

Pubs é un risolutore di vincoli di costo, che prende in input un file di equazioni di costo e restituisce un file con i limiti superiori per ogni relazione di costo. PUBS (Practical Upper Bounds Solver) ha l'obiettivo di ottenere automaticamente upper bound in forma chiusa per i sistemi di equazioni di costo, di conseguenza calcola i limiti superiori per la relazione di costo indicata come "Entry", oltre che per tutte le altre relazioni di costo di cui tale "Entry" dipende. Nell'output di pubs vengono mostrati anche i passaggi intermedi eseguiti che coinvolgono il calcolo delle funzioni di classificazione e degli invarianti di ciclo.

```

CRS $pubs_aux_entry$(A,B,C) -- THE MAIN ENTRY

* Non Asymptotic Upper Bound: max([nat(A),nat(B)])

* LOOPS $pubs_aux_entry$(D,E,F) -> $pubs_aux_entry$(G,H,I)

* Ranking function: N/A

* Invariants $pubs_aux_entry$(A,B,C) -> $pubs_aux_entry$(D,E,F)

  entry  : []
  non-rec: [A=D,B=E,C=F]
  rec    : [0=1]
  inv    : [1*A+ -1*D=0,1*B+ -1*E=0,1*C+ -1*F=0]

CRS main(A,B,C)

* Non Asymptotic Upper Bound: max([nat(A),nat(B)])

* LOOPS main(D,E,F) -> main(G,H,I)

* Ranking function: N/A

* Invariants main(A,B,C) -> main(D,E,F)

  entry  : []
  non-rec: [A=D,B=E,C=F]
  rec    : [0=1]
  inv    : [1*A+ -1*D=0,1*B+ -1*E=0,1*C+ -1*F=0]

```

Figura 1.1: Esempio di output PUBS su Listing 1

Come vediamo nell’immagine sopra, PUBS ci restituisce un’analisi dell’intera equazione (*pub\_aux\_entry*) e delle singole funzioni da cui essa dipende, in questo caso *pubs\_aux\_if9* e *pubs\_aux\_main*. In questo caso con “Listing1” abbiamo un Upper Bound non Asintotico di  $\max(\text{Nat}(A), \text{Nat}(B))$  che ci determina che il costo del programma dipende dalle variabili A e B, e che il costo del programma sarà il massimo tra i due. Inoltre non abbiamo la presenza di cicli, quindi PUBS non ci restituisce nessun invariante di ciclo.

In PUBS però abbiamo una grammatica da rispettare, tenuta in considerazione da ogni equazione di costo generata da CostCompiler, che é la seguente:

```

1  <equation> ::= eq(Head,costExpression,[listOfCall],[
2  ListOfSizeRelation]).
3  <Head> ::= Name | Name(Par).
4  <costExpression> ::= nat(<variable>)
5  | <costExpression> + <costExpression>
6  | <costExpression> - <costExpression>
7  | <costExpression> * <costExpression>
8  | max(<costExpression>,<costExpression>).
9
10 <listOfCall> ::= [] | <call> <listOfCall>.
    <call> ::= <function>(<listOfParameters>).

```

```

11  <listOfParameters> ::= [] | <variable> <listOfParameters
    >.

```

Listing 1.3: Grammatica PUBS

Dove  $\langle \text{Head} \rangle$  é il nome della entry che andremo ad analizzare insieme ai suoi parametri.  $\text{CostExpression}$  é l'espressione di costo che rappresenta il costo della entry e rispetta la grammatica della aritmetica di Presburger.  $\text{ListOfCall}$  é la lista delle chiamate alle altre entry, che sono rappresentate come  $\langle \text{call} \rangle$  e  $\langle \text{listOfCall} \rangle$ , la lista di queste chiamate; in questo modo PUBS riesce a costruire un grafo delle dipendenze tra le entry. Infine abbiamo  $\langle \text{listOfSizeRelation} \rangle$  che sarà la lista delle relazioni di costo che dipendono dalla entry che stiamo analizzando, e che PUBS andrà a calcolare. Riportiamo Listing1 come esempio di equazione di costo generata da CostCompiler, dove troviamo l'analisi alla pagine precedente:

```

1  eq(main(B),1,[svc(B)],[]).
2  eq(svc(B),0,[for3(0, B)],[] ).
3  eq(for3(M, B) ,nat(B),[for3(M+1, B)], [10>= M]).
4  eq(for3(M, B) ,0,[],[M >= 10+ 1]).

```

Listing 1.4: Equazione di costo PUBS per Listing6

Nella prima riga troviamo l'entry *main* che prende in input B, con costo 1, chiama la funzione *svc*. Quest'ultima andrà a chiamare la funzione *for3* inserendo un'ulteriore parametro che sarà il counter del ciclo con parametro 0 e B, che avrà costo 0 in caso  $M \geq 10 + 1$  altrimenti avrà costo  $\text{nat}(B)$ .