Specification and Verification of C3 linearization algorithm

Miguel Flor[1], António Ravara[1], and Mário Pereira[1]

NOVA LINCS, Nova School of Science and Technology
m.flor@campus.fct.unl.pt, {aravara, mjp.pereira}@fct.unl.pt

Abstract. Multiple hierarchical inheritance is common among many programming languages, and to produce a consistent and intuitive method resolution order (MRO) of the class hierarchy, the C3 linearization algorithm is one of the most widely used. However, the algorithm is not trivial, and because its widespread use, there is a need for a formal specification and verification of this algorithm. In this paper, we present a mathematical formal specification of the C3 algorithm, and its verification in OCaml code using Why3 and Cameleer.

1 Introduction

The need for object-oriented programming languages, the support for multiple inheritance and a predictable MRO is crucial for the development of large software systems. That's why the C3 linearization algorithm is widely used in many programming languages Python 2.3 [3], Perl [2], Solidity [1], and many others. And exactly because of its widespread use, there is a need for an implementation that is verified, in this paper the verification is done in OCaml using Why3 and Cameleer.

1.1 Why C3?

C3 algorithm is widely used for producing a consistent and predictable MRO when conflicts arise in multiple inheritance scenarios. The algorithm uses three core properties to ensure that the MRO is consistent (these properties are explained in section 2.2):

- Extended Precedence Graph (EPG)
- Local Precedence Order (LPO)
- Monotonicity

Consider the class hierarchy in Fig. 2, where class A inherits from both B and C. The C3 algorithm resolves the conflict ensuring that the MRO respects the order of inheritance (reading from left to right), resulting in an MRO of (A, B, C, D).

2 Background

We will first examine how multiple inheritance affects MRO algorithms, then examine the principles of the C3 linearization algorithm, followed by a brief overview of cameleer.

2.1 Influence of Multiple Inheritance on MRO Algorithms

MRO is a linear extension of some class hierarchy that induces a total order on the classes in the hierarchy. [6]

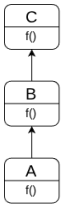

Fig. 1. Simple graph

For the example above (Fig. 1) a logical MRO could be (A, B, C), this means that when the function f is executed from the class A the implementation of A will be prioritized over B, and the implementation of B will be priorized over C. However, the MRO of a class hierarchy is not always that straightforward, and conflicts can start to arise when we introduce multiple inheritance. To demonstrate this, consider the following example, that represents a known problem called the diamond problem [7].

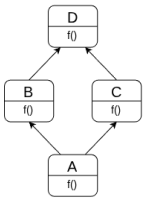

Fig. 2. Diamond problem graph

In this example (Fig. 2), a conflict emerges when determining the MRO of class A, since A inherits from both B and C. This raises the question: Which implementation, B's or C's, does A inherit from? This is where the C3 linearization algorithm comes into play, as it provides a consistent and predictable way to resolve this conflict, the algorithm says that we need to prioritize B over C,

because B comes first in the class hierarchy, and therefore the MRO of A will be (A, B, C, D).

2.2 C3 Linearization Algorithm

The C3 algorithm produces an MRO that is consistent with EPG, LPO, and monotonicity [4] (all of these properties are formally defined in section 3.2).

EPG To understand the EPG, consider the following graph (Fig. 3). For this graph instead of ordering the classes from left to right, they are arranged by the weights of the directed edges, from smallest to largest.

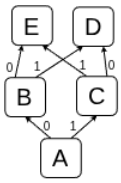

Fig. 3. Inconsistent EPG

This graph is inconsistent with the EPG, as the parent classes of B and C are ordered differently. A valid class hierarchy requires that for each class, the parent classes are ordered in the same way.

LPO Consider the graph in Fig. 2. An MRO that is consistent needs to respect the priority of the classes locally at each inheritance point; therefore, for this case, it cannot be (A, C, B, D).

Monotonicity If an MRO algorithm is consistency with monotonicity, it means that if a new class is added to the hierarchy the order of the previous MRO is preserved.

2.3 Cameleer

Cameleer is an automated tool for deductive verification in OCaml. It uses comments written in GOSPEL [5] (Generic OCaml Specification Language) to specify the properties of functions, and then leverages Why3 to verify these properties [8].

3 Formal Specification of C3 Linearization Algorithm

3.1 Ingredients

Let $\mathcal{C}$ be a finite set of symbols, ranged over by $C_0, C_1, \ldots, C_n$, possibly primed, which we refer to as *classes*, and let $(\mathcal{C}, <_c)$ be an ordered set of classes.

Consider $\mathcal{P} \subseteq (\mathcal{C}, <_c)$ and $\mathcal{D} \subseteq \mathcal{C} \times \wp(\mathcal{P})$.

The purpose of this document is to present the MRO algorithm, which maps each $\mathcal{C}$ with an ordered set of classes, $\mathcal{P}$. We show first the envisaged properties, then define it, and finally prove the definition ensures those properties.

3.2 Properties

Let $D = \langle C, \{P_1, \ldots, P_n\}\rangle \in \mathcal{D}$, with $n \in \mathbb{N}_0$.
Let $\texttt{MRO}(C) = \{C, C_1 \ldots, C_m\}$, with $m \in \mathbb{N}_0$.

Consistency with the EPG This property requires that $\{P_1, \ldots, P_n\} \subseteq \{C_1, \ldots, C_m\}$.

Consistency with the LPO For $m, n \geq 2$.

$$\forall\, i, j \ \big(0 \leq i < j \leq n \implies \exists\, p, q \ \big(0 \leq p < q \leq m \wedge C_p = P_i \wedge C_q = P_j\big)\big)$$

Consistency with monotonicity Let $D', D'' \in \mathcal{D}$
If $C \in \pi_2(D') \setminus \pi_2(D'')$, then
$\exists p, q.\ 0 < p < q \leq m \ \wedge\ C_p = C \ \wedge\ C_q = \pi_1(D'')$.

3.3 Functions

Let $(\mathcal{C}, <_c)^*$ be a sequence of ordered sets of classes.
Let $L = (L_1, \ldots, L_n)$, $L \in (\mathcal{C}, <_c)^*$
Let $C \in \mathcal{C}$

Remove $\texttt{remove} : (\mathcal{C}, <_c)^* \times \mathcal{C} \Rightarrow (\mathcal{C}, <_c)^*$

$$\texttt{remove}((\,), C) = (\,)$$

$$\texttt{remove}(l :: L, C) = l \setminus \{C\} :: \texttt{remove}(L, C)$$

Merge $\texttt{merge} : (\mathcal{C}, <_c)^* \Rightarrow (\mathcal{C}, <_c)$

$$\texttt{merge}(L) = \begin{cases} \{C\} \cup \texttt{merge}(\texttt{remove}(L, C)), & \text{if (1), (2), (3)} \\ \text{fail}, & \text{otherwise} \end{cases}$$

where:

$$\begin{aligned} &(1) \quad \exists k \in [\![1, n]\!], L_k \neq \emptyset \wedge C = \texttt{head}(L_k) \\ &(2) \quad \forall j < k, C \neq \texttt{head}(L_j) \\ &(3) \quad \forall i \in [\![1, n]\!], C \notin \texttt{tail}(L_i) \end{aligned}$$

C3 Linearization $\texttt{c3linearization} : \mathcal{D} \Rightarrow (\mathcal{C}, <_c)$
Let $D = \langle C, P \rangle$ where $D \in \mathcal{D}$
Let $D' = (D_1, D_2, \ldots, D_{|P|})$, such that
$\forall P_i \in P,\ \exists D_i \in \mathcal{D}$ where $D_i = \langle P_i, P' \rangle$ where
$i \in [\![1, |P|]\!]$.
Let `C3linearization` be denoted as `cl` for brevity.

$$\texttt{cl}(D) = \begin{cases} \{C\} & \text{if } P = \emptyset \\ \{C\} \cup \texttt{merge}\left((\texttt{cl}(D_i))_{D_i \in D'}\ ,\ P\right) & \text{otherwise} \end{cases}$$

References

1. Language Influences — Solidity 0.8.31 documentation. https://docs.soliditylang.org/en/latest/language-influences.html
2. Mro - Method Resolution Order - Perldoc Browser. https://perldoc.perl.org/mro
3. The Python 2.3 Method Resolution Order. https://docs.python.org/3/howto/mro.html
4. Barrett, K., Cassels, B., Haahr, P., Moon, D.A., Playford, K., Withington, P.T.: A monotonic superclass linearization for Dylan. In: Proceedings of the 11th ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications. pp. 69–82. OOPSLA '96, Association for Computing Machinery, New York, NY, USA (Oct 1996). https://doi.org/10.1145/236337.236343
5. Charguéraud, A., Filliâtre, J.C., Lourenço, C., Pereira, M.: GOSPEL—Providing OCaml with a Formal Specification Language. In: ter Beek, M.H., McIver, A., Oliveira, J.N. (eds.) Formal Methods – The Next 30 Years. pp. 484–501. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-30942-8_29
6. Hivert, F., Thiéry, N.M.: Controlling the C3 Super Class Linearization Algorithm for Large Hierarchies of Classes. Order **41**(1), 83–98 (Apr 2024). https://doi.org/10.1007/s11083-022-09607-5
7. Monday Eze, Charles Okunbor, Umoke Chukwudum: Studies in object-oriented programming backbone implementations. Global Journal of Engineering and Technology Advances **8**(3), 020–031 (Sep 2021). https://doi.org/10.30574/gjeta.2021.8.3.0119
8. Pereira, M., Ravara, A.: Cameleer: A Deductive Verification Tool for OCaml. In: Silva, A., Leino, K.R.M. (eds.) Computer Aided Verification. pp. 677–689. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-81688-9_31