修士 (工学) 学位論文
Master's Thesis of Engineering

GearsOS のメタ計算

2021 年 3 月
March 2021

清水 隆博
Takahiro Shimizu

琉球大学
大学院理工学研究科
情報工学専攻

Information Engineering Course
Graduate School of Engineering and Science
University of the Ryukyus

指導教員：教授 和田 知久
Supervisor: Prof. Tomohisa Wada

論文題目: GearsOS のメタ計算

氏　　名: 清水 隆博

本論文は、修士 (工学) の学位論文として適切であると認める。

論　文　審　査　会

印

(主　査)　　和田 知久

印

(副　査)　　山田 孝治

印

(副　査)　　當間 愛晃

印

(副　査)　　河野 真治

要旨

アプリケーションの信頼性を保証するには、土台となる OS の信頼性は高く保証されていなければならない。信頼性を保証する方法としてテストコードを使う手法が広く使われている。OS のソースコードは巨大であり、並列処理など実際に動かさないと発見できないバグが存在する。OS の機能をテストですべて検証するのは不可能である。

テストに頼らず定理証明やモデル検査などの形式手法を使用して、OS の信頼性を保証したい。証明を利用して信頼性を保証する定理証明は、 Agda や Coq などの定理証明支援系を利用することになる。支援系を利用する場合、各支援系で OS を実装しなければならない。証明そのものは可能であるが、 支援系で証明されたソースコードがそのまま OS として動作する訳ではない。このためには定理証明されたコードを等価な C 言語などに変換する処理系が必要となる。

信頼性を保証するほかの方法として、プログラムの可能な実行をすべて数え上げて仕様を満たしているかを確認するモデル検査がある。モデル検査は実際に動作しているプログラムに対して実行することが可能である。すでに実装したプログラムのコードに変化を加えずモデル検査を行いたい。

プログラムは本来やりたい計算であるノーマルレベルの計算と、その計算をするのに必要なメタレベルの計算に別けられる。メタレベルの計算では資源管理などを行うが、 モデル検査などの証明をメタレベルの計算で行いたい。

この実現にはノーマルレベル、メタレベルの計算の処理の切り分けと、メタレベルの計算をより柔軟に扱う OS、言語処理系が必要となる。両レベルを記述できる言語に Continuation Based (CbC) がある。CbC はスタック、あるいは環境を持たず継続によって次の処理を行う特徴がある。CbC を用いて、拡張性と信頼性を両立する OS である GearsOS を開発している。

GearsOS の開発ではノーマルレベルのコードとメタレベルのコードの両方が必要であり、メタレベルの計算の数は多岐にわたる。GearsOS の開発を進めていくには、メタレベルの計算を柔軟に扱う API や、 自動でメタレベルの計算を作製する GearsOS のビルドシステムが必須となる。本研究では GearsOS の信頼性と拡張性の保証につながる、メタ計算に関する API について考察し、言語機能などの拡張を行った。また、メタ計算を自動生成しているトランスコンパイラを改良し、従来の GearsOS のシステムよりさらに柔軟性が高いものを考案した。

Abstract

hogefuga

研究関連業績

- CbC を用いた Perl6 処理系 清水 隆博, 河野真治 第 60 回プログラミング・シンポジウム, Jan, 2019
- How to build traditional Perl interpreters. Takahiro SHIMIZU PerlCon2019 , Aug, 2019
- Perl6 Rakudo の内部構造について 清水 隆博 オープンソースカンファレンス 2019 Okinawa, Apr, 2020
- 継続を基本とした OS Gears OS 清水 隆博, 河野真治 第 61 回プログラミング・シンポジウム, Jan, 2020
- Perl6 のサーバを使った実行 福田 光希, 清水 隆博, 河野真治 第 61 回プログラミング・シンポジウム, Jan, 2020
- xv6 の構成要素の継続の分析 清水 隆博, 河野 真治 (琉球大学), 情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS), May, 2020

目 次

研究関連論文業績 **iii**

第 1 章　OS とアプリケーションの信頼性 **8**

第 2 章　Continuation Based C **11**
2.1 CodeGear 11
2.2 DataGear 12
2.3 CbC を使った例題 12
2.4 CbC を使ったシステムコールディスパッチの例題 16
2.5 メタ計算 16
2.6 MetaCodeGear 18
2.7 MetaDataGear 18

第 3 章　GearsOS **20**
3.1 GearsOS の構成 20
3.2 Context 21
3.3 Stub Code Gear 23
3.4 TaskManager 26
3.5 TaskQueue 27
3.6 Worker 27
3.7 union Data 型 27
3.8 Interface 28
3.8.1 Interface の定義 28
3.8.2 Interface の呼び出し 29
3.8.3 Interface のメタレベルの実装 30
3.8.4 Interface の Impl の実装 30
3.8.5 goto 時の Context と Interface の関係 31
3.9 GearsOS のビルドシステム 32
3.10 GearsOS の CbC から純粋な CbC への変換 33
3.11 generate_stub.pl 34

3.12 generate_context.pl 35
3.13 CbC xv6 36
3.14 ARM 用ビルドシステムの作製 37
3.15 Interface の取り扱い方法の検討 38

第 4 章 GearsOS の Interface の改良 40
4.1 GearsOS の Interface の構文の改良 40
4.2 Implement の型定義ファイルの導入 42
4.3 Implement の型をいれたことによる間違った Gears プログラミング 44
4.4 Interface のパーサーの構築 44
4.4.1 Gears::Interface の構成 45
4.5 Interface の実装の CbC ファイルへの構文の導入 45
4.6 GearsCbC の Interface の実装時の問題 45
4.7 Interface を満たすコード生成の他言語の対応状況 46
4.8 GearsOS での Interface を満たす CbC の雛形生成 46
4.8.1 雛形生成の手法 47
4.8.2 コンストラクタの自動生成 48
4.9 Interface の引数の検知 48
4.10 Interface の API の未実装の検知 51
4.11 par goto の Interface 経由の呼び出しの対応 51

第 5 章 トランスコンパイラによるメタ計算 52
5.1 トランスコンパイラ 52
5.2 トランスコンパイラによるメタレベルのコード生成 53
5.3 トランスコンパイラ用の Perl ライブラリ作製 53
5.4 context.h の自動生成 54
5.4.1 context.h の作製フロー 54
5.4.2 context.h のテンプレートファイル 55
5.5 メタ計算部分の入れ替え 55
5.6 コンパイルタイムでのコンストラクタの自動生成 56
5.7 Interface の API の自動保管 56
5.8 別 Interface からの書き出しを取得する必要がある CodeGear 56
5.9 別 Interface からの書き出しを取得する Stub の生成 60
5.9.1 初回 CbC ファイル読み込み時の処理 61
5.9.2 enum の差し替え処理 62
5.10 ジェネリクスのサポート 63

第 6 章　評価 **64**
6.1　GearsOS の構文作製 64
6.2　GearsOS のトランスコンパイラ 64
6.3　GearsOS のメタ計算 64

第 7 章　結論 **65**
7.1　今後の課題 65

謝辞 **66**

参考文献 **67**

付録 **69**

付 録 A 研究会業績 **70**
A-1　研究会発表資料 70

図 目 次

2.1 CbC と C の処理の差 14
2.2 CodeGear と MetaCodeGear 18

3.1 GearsOS の構成 21
3.2 Context の概要図 23
3.3 Context を参照した CodeGear のデータアクセス 26
3.4 GearsOS のビルドフロー 33
3.5 generate_sub.pl を使ったトランスコンパイル 35
3.6 generate_context.pl を使ったファイル生成 36
3.7 pmake.pl の処理フロー 38

4.1 impl2cbc の処理の流れ 47

5.1 stackTest1 の stub の概要 60

表 目 次

ソースコード目次

2.1 CbC の例題 12
2.2 ソースコード 2.1 の C での実装 13
2.3 ソースコード 2.1 のアセンブラの一部 14
2.4 ソースコード 2.2 のアセンブラの一部 15
2.5 CbC を利用したシステムコールのディスパッチ 16
3.1 context の定義 21
3.2 Gearef マクロ 24
3.3 enumData の定義 24
3.4 Stack に Push する CodeGear 24
3.5 3.4 の StubCodeGear 25
3.6 __code meta 25
3.7 CodeGear の番号である enumCode の定義 25
3.8 Queue の Interface 28
3.9 Interface の API の呼び出し 29
3.10 Queue の Interface に対応する構造体 30
3.11 SingleLinkedQueue の実装 30
3.12 take を呼び出す部分の変換後 32
3.13 CMakeList.txt 内でのプロジェクト定義 32
3.14 CMakeList.txt 内での Perl の実行部分 34
4.1 従来の Stack Interface 40
4.2 golang の interface 宣言 41
4.3 変更後の Stack Interface 41
4.4 cotnext.h に直接書かれた型定義 42
4.5 Java の Implement キーワード 43
4.6 SynchronizedQueue の定義ファイル 44
4.7 Perl レベルでの引数チェック 49
4.8 StackTestInterface の定義 50
4.9 StackTestInterface の API 呼び出し (引数不足) 50
4.10 Interface の API 呼び出し時の引数エラー 50

5.1 meta.pm 56
5.2 別 Interface からの書き出しを取得する CodeGear の例 57
5.3 SingleLinkedStack の pop2 57
5.4 SingleLinkedStack の pop2 のメタ計算 58
5.5 生成された Stub 58
5.6 goto 時に使用する interface の解析 61
5.7 Gearef のコード生成部分 62
5.8 enum の番号が差し替えられた CodeGear 63

第1章　OSとアプリケーションの信頼性

コンピューター上では様々なアプリケーションが常時動作している。動作しているアプリケーションは信頼性が保証されていてほしい。信頼性の保証には、 実行してほしい一連の挙動をまとめた仕様と、 それを満たしているかどうかの確認である検証が必要となる。アプリケーション開発では検証に関数や一連の動作をテストを行う方法や、デバッグを通して信頼性を保証する手法が広く使われている。

実際にアプリケーションを動作させる OS は、アプリケーションよりさらに高い信頼性が保証される必要がある。OS は CPU やメモリなどの資源管理と、 ユーザーにシステムコールなどの API を提供することで抽象化を行っている。OS の信頼性の保証もテストコードを用いて証明することも可能ではあるが、 アプリケーションと比較すると OS のコード量、 処理の量は膨大である。また OS は CPU 制御やメモリ制御、 並列・並行処理などを多用する。テストコードを用いて処理を検証する場合、テストコードとして特定の状況を作成する必要がある。実際に OS が動作する中でバグやエラーを発生する条件を、並列処理の状況などを踏まえてテストコードで表現するのは困難である。非決定的な処理を持つ OS の信頼性を保証するには、 テストコード以外の手法を用いる必要がある。

テストコード以外の方法として、 形式手法と呼ばれるアプローチがある。形式手法の具体的な検証方法の中で、 証明を用いる方法 [1][2][3] とモデル検査を用いる方法がある。証明を用いる方法では Agda[4] や Coq[5] などの定理証明支援系を利用し、 数式的にアルゴリズムを記述する。Curry-Howard 同型対応則により、型と論理式の命題が対応する。この型を導出するプログラムと実際の証明が対応する。証明には特定の型を入力として受け取り、証明したい型を生成する関数を作成する。整合性の確認は、記述した関数を元に定理証明支援系が検証する。証明を使う手法の場合、 実際の証明を行うのは定理証明支援系であるため、 定理証明支援系が理解できるプログラムで実装する必要がある。しかし Agda で証明ができても Agda のコードを直接 OS のソースコードとしてコンパイルすることはできない。検証されたアルゴリズムをもとに C で実装することは可能であるが、移植時にバグが入る可能性がある。検証ができているソースコードそのものを使って OS を動作させたい。

他の形式手法にモデル検査がある。モデル検査はプログラムの可能な実行をすべて数え上げて要求している使用を満たしているかどうかを調べる手法である。例えば Java のソースコードに対してモデル検査をする JavaPathFinder などがある。モデル検査を利用

する場合は、実際に動作するコード上で検証を行うことが出来る。OS のソースコードそのものをモデル検査すると、実際に検証された OS が動作可能となる。しかし OS の処理は膨大である。すべての存在可能な状態を数え上げるモデル検査では状態爆発が問題となる。状態を有限に制限したり抽象化を行う必要がある。また、モデル検査ができるモデル検査器は特定のプログラム形式でないと動かないものがある。例えば Spin は Promela 形式でないとモデル検査ができない。モデル検査ができる場合も、 モデル検査したコードと実際に動くコードを同一にしたい。また、モデル検査をする場合としない場合の切り替えを、より手軽に行いたい。

OS のシステムコールは、ユーザーから API 経由で呼び出され、 いくつかの処理を行う。その処理に着目すると OS は様々な状態を遷移して処理を行っていると考えることができる。OS を巨大な状態遷移マシンと考えると、 OS の処理の特定の状態の遷移まで範囲を絞ることができる。範囲が限られているため、有限時間でモデル検査などで検証することが可能である。この為には OS の処理を証明しやすくする表現で実装する必要がある。[6] 証明しやすい表現の例として、 状態遷移ベースでの実装がある。

証明を行う対象の計算は、 その意味が大きく別けられる。OS やプログラムの動作においては本来したい計算がまず存在する。これはプログラマが通常プログラミングするものである。これら本来行いたい処理のほかに、 CPU、メモリ、スレッドなどの資源管理なども必要となる。前者の計算をノーマルレベルの計算と呼び、後者をメタレベルの計算と呼ぶ。OS はメタ計算を担当していると言える。ユーザーレベルから見ると、データの読み込みなどは資源へのアクセスが必要であるため、システムコールを呼ぶ必要がある。システムコールを呼び出すと OS が管理する資源に対して何らかの副作用が発生するメタ計算と言える。副作用は関数型プログラムの見方からするとモナドと言え、 モナドもメタ計算ととらえることができる。OS 上で動くプログラムは CPU により並行実行される。この際の他のプロセスとの干渉もメタレベルの処理である。実装のソースコードはノーマルレベルであり検証用のソースコードはメタ計算だと考えると、OS そのものが検証を行ない、システム全体の信頼を高める機能を持つべきだと考える。ノーマルレベルの計算を確実に行う為には、メタレベルの計算が重要となる。

プログラムの整合性の検証はメタレベルの計算で行いたい。ユーザーが実装したノーマルレベルの計算に対応するメタレベルの計算を、自由にメタレベルの計算で証明したい。またメタレベルで検証ががすでにされたプログラムがあった場合、都度実行ユーザーの環境で検証が行われるとパフォーマンスに問題が発生する。この場合は検証を実行するメタ計算と、 検証をしないメタ計算を手軽に切り替える必要がある。さらに検証用とそうでない用で、動作させたいアルゴリズムの実装そのもののコードを変更したくない。これも検証をメタレベルで行い、実装をノーマルレベルで行い、各レベルを切り離すことで実現可能である。メタレベルの計算をノーマルレベルの計算と同等にプログラミングできると、動作するコードに対して様々なアプローチが掛けられる。ノーマルレベル、メタレベ

ル共にプログラミングできる言語と環境が必要となる。

プログラムのノーマルレベルの計算とメタレベルの計算を一貫して行う言語として、Continuation Based C(CbC) を用いる。CbC は基本 `goto` 文で `CodeGaar` というコードの単位を遷移する言語である。通常の関数呼び出しと異なり、スタックあるいは環境と呼ばれる隠れた状態を持たない。このため、計算のための情報は `CodeGear` の入力にすべてそろっている。そのうちのいくつかはメタ計算、つまり、OS が管理する資源であり、その他はアプリケーションを実行するためのデータ (`DataGear`) である。メタ計算とノーマルレベルの区別は入力のどこを扱うかの差に帰着される。CbC は C と互換性のある C の下位言語である。CbC は `GCC`[7][8] あるいは `LLVM`[9][10] 上で実装されていて、通常の C のアプリケーションやシステムプログラムをそのまま包含できる。C のコンパイルシステムを使える為に、 CbC のプログラムをコンパイルすることで動作可能なバイナリに変換が可能である。また CbC の基本文法は簡潔であるため、 Agda などの定理証明支援系 [11] との相互変換や、 CbC 自体でのモデル検査が可能であると考えられる。

CbC を用いてノーマルレベルとメタレベルの分離を行い、信頼性と拡張性を両立させることを目的として GearsOS を開発している。GeasOS では、 CbC の実行単位である CodeGear とデータの単位である DataGear を基本単位としている。GearsOS のメタ計算には MetaCodeGear と MetaDataGear を用いる。信頼性の保証は MetaCodeGear で行いたい。その為には GearsOS が柔軟にメタ計算を切り替えることが必要となる。また、GearsOS で実行されるメタ計算の数は膨大である。すべてをプログラミングするのではなく、いくつかのメタ計算は自動で生成されてほしい。GearsOS では拡張性の保証も重要な課題である。拡張性を保証するにはすべて純粋な CbC で実装すると、実装がきわめて煩雑である。その為には CbC とセマンティックが等しいより簡潔な GearsOS 独自のシンタックスなどが必要である。これらを踏まえて実装した GearsOS を動作させる際のビルドフローもより効率的なものにしたい。

本研究では GearsOS の信頼性と拡張性の保証につながる、メタ計算に関する API について考察する。GearsOS がメタ計算を自動生成しているトランスコンパイラで従来の GearsOS のシステムよりさらに拡張性の充実と、信頼性の保証を図る。

第2章　Continuation Based C

Continuation Based C(CbC) とはC 言語の下位言語であり、 関数呼び出しではなく継続を導入したプログラミング言語である。CbC では通常の関数呼び出しの他に、 関数呼び出し時のスタックの操作を行わず、次のコードブロックに `jmp` 命令で移動する継続が導入されている。この継続はScheme の call/cc などの環境を持つ継続とは異なり、 スタックを持たず環境を保存しない継続である為に軽量である事から軽量継続と呼べる。またCbC ではこの軽量継続を用いて `for` 文などのループの代わりに再起呼び出しを行う。これは関数型プログラミングでの Tail call スタイルでプログラミングすることに相当する。Agda よる関数型の CbC の記述も用意されている。実際の OS やアプリケーションを記述する場合には、GCC10[12] 及び LLVM10/clang 上 [13] の CbC 実装を用いる。

2.1　CodeGear

CbC では関数の代わりにCodeGear という単位でプログラミングを行う。CodeGear は通常のC の関数宣言の返り値の型の代わりに`__code` で宣言を行う。各CodeGear はDataGear と呼ばれるデータの単位で入力を受け取り、 その結果を別の DataGear に書き込む。入力の DataGear を InputDataGear と呼び、 出力の DataGear を OutputDataGear と呼ぶ。CodeGear がアクセスできる DataGear は、 InputDataGear と OutputDataGear に限定される。

CodeGear は関数呼び出し時のスタックを持たない為、一度ある CodeGear に遷移すると元の処理に戻ってこれない。しかし CodeGear を呼び出す直前のスタックは保存される。部分的に CbC を適用する場合はCodeGear を呼び出す `void` 型などの関数を経由することで呼び出しが可能となる。

この他に CbC から C へ復帰する為の API として、 環境付き goto がある。これは呼び出し元の関数を次の CodeGear の継続対象として設定するものである。これは GCC では内部コードを生成を行う。LLVM/clang では `setjmp` と `longjmp` を使い実装している。環境付き goto を使うと、通常の C の関数呼び出しの返り値を CodeGear から取得する事が可能となる。

2.2 DataGear

DataGear は CbC でのデータの単位である。CbC 上では構造体の形で表現される。各 CodeGear の入力として受ける DataGear を InputDataGear と呼ぶ。逆に次の継続に渡す DataGear を OutputDataGear と呼ぶ。

メタレベルでは DataGear はポインタを扱っているが、 ノーマルレベルの DataGear はポインタを扱っていないと仮定している。例えばリストの DataGear を考えると、 C の実装の場合はポインタを使った単方向リストが考えられる。リストのそれぞれの要素には、メタレベルでは次の要素を指し示すポインタが含まれている。ノーマルレベルの DataGear として見る場合は、 リストそのものや、 リストの中の値そのものとして判断するために、より抽象化された単位として見える。これは関数型プログラミングにおける末尾再起呼び出し時の値のやりとりに似た概念である。

2.3 CbC を使った例題

ソースコード 2.1 に CbC を使った例題を、 ソースコード 2.2 に通常の C で実装した例題を示す。この例では構造体 `struct test` を codegear1 に渡し、その次に codegear2 に継続している。codegear2 からは codegear3 に goto し、 最後に exit する。

ソースコード 2.1: CbC の例題

```
1  extern int printf(const char*,...);
2
3  typedef struct test {
4      int number;
5      char* string;
6  } TEST;
7
8
9  __code codegear1(TEST);
10 __code codegear2(TEST);
11 __code codegear3(TEST);
12
13 __code codegear1(TEST testin){
14     TEST testout;
15     testout.number = testin.number + 1;
16     testout.string = testin.string;
17     goto codegear2(testout);
18 }
19
20 __code codegear2(TEST testin){
21     TEST testout;
22     testout.number = testin.number;
23     testout.string = "Hello";
24     goto codegear3(testout);
25 }
```

```
26
27 __code codegear3(TEST testin){
28     printf("number = %d\t string= %s\n",testin.number,testin.string);
29     goto exit(0);
30 }
31
32 int main(){
33     TEST test = {0,0};
34     goto codegear1(test);
35 }
```

ソースコード 2.2: ソースコード 2.1 の C での実装

```
1  extern int printf(const char*,...);
2
3  typedef struct test {
4      int number;
5      char* string;
6  } TEST;
7
8
9  void codegear1(TEST);
10 void codegear2(TEST);
11 void codegear3(TEST);
12
13 void codegear1(TEST testin){
14     TEST testout;
15     testout.number = testin.number + 1;
16     testout.string = testin.string;
17     codegear2(testout);
18 }
19
20 void codegear2(TEST testin){
21     TEST testout;
22     testout.number = testin.number;
23     testout.string = "Hello";
24     codegear3(testout);
25 }
26
27 void codegear3(TEST testin){
28     printf("number = %d\t string= %s\n",testin.number,testin.string);
29     exit(0);
30 }
31
32 int main(){
33     TEST test = {0,0};
34     codegear1(test);
35 }
```

CbC の場合は継続で進んでいくが、 C 言語での実装は void 型の返り値を持つ関数呼び出しで表現される。codegear3 に遷移したタイミングで、 CbC は main 関数のスタックしか持たないが、 C 言語では codegear1、codegear2 のスタックをそれぞれ持つ違いがある。

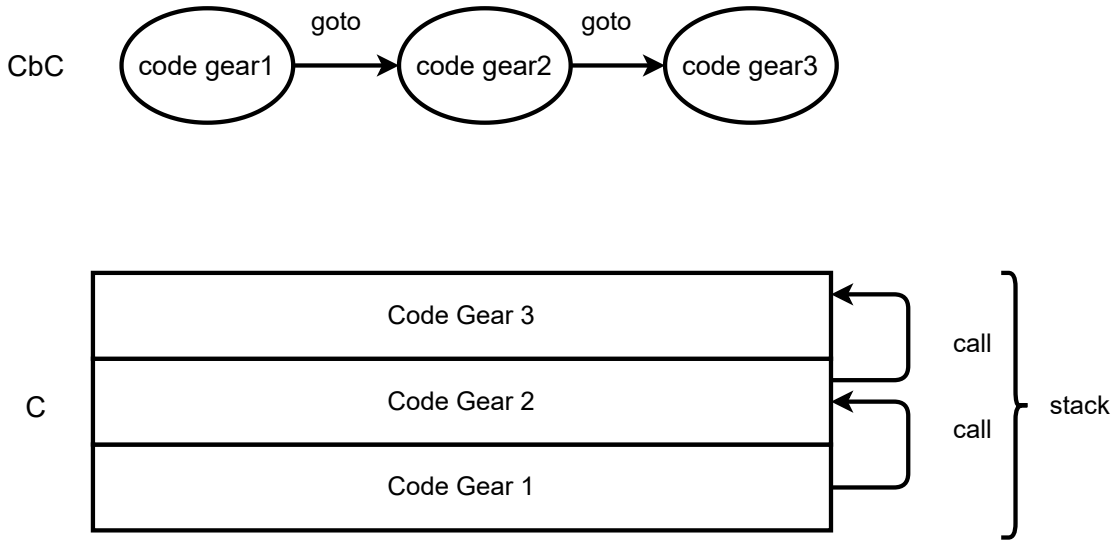

図 2.1: CbC と C の処理の差

(図 2.1)

実際に軽量継続になっているかを、この例題をアセンブラに変換した結果を見比べて確認する。

ソースコード 2.3: ソースコード 2.1 のアセンブラの一部

```
1  codegear1:
2  .LFB0:
3      .cfi_startproc
4      pushq   %rbp
5      .cfi_def_cfa_offset 16
6      .cfi_offset 6, -16
7      movq    %rsp, %rbp
8      .cfi_def_cfa_register 6
9      movl    %edi, %eax
10     movq    %rsi, %rcx
11     movq    %rcx, %rdx
12     movq    %rax, -32(%rbp)
13     movq    %rdx, -24(%rbp)
14     movl    -32(%rbp), %eax
15     addl    $1, %eax
16     movl    %eax, -16(%rbp)
17     movq    -24(%rbp), %rax
18     movq    %rax, -8(%rbp)
19     movl    -16(%rbp), %edx
20     movq    -8(%rbp), %rax
21     movl    %edx, %edi
22     movq    %rax, %rsi
23     popq    %rbp
24     .cfi_def_cfa 7, 8
```

```
25      jmp codegear2
26      .cfi_endproc
27  .LFE0:
28      .size   codegear1, .-codegear1
29      .section    .rodata
30  .LC0:
31      .string "Hello"
32      .text
33      .globl  codegear2
34      .type   codegear2, @function
```

ソースコード 2.4: ソースコード 2.2 のアセンブラの一部

```
1       pushq   %rbp
2       .cfi_def_cfa_offset 16
3       .cfi_offset 6, -16
4       movq    %rsp, %rbp
5       .cfi_def_cfa_register 6
6       subq    $32, %rsp
7       movl    %edi, %eax
8       movq    %rsi, %rcx
9       movq    %rcx, %rdx
10      movq    %rax, -32(%rbp)
11      movq    %rdx, -24(%rbp)
12      movl    -32(%rbp), %eax
13      addl    $1, %eax
14      movl    %eax, -16(%rbp)
15      movq    -24(%rbp), %rax
16      movq    %rax, -8(%rbp)
17      movl    -16(%rbp), %edx
18      movq    -8(%rbp), %rax
19      movl    %edx, %edi
20      movq    %rax, %rsi
21      call    codegear2
22      nop
23      leave
24      .cfi_def_cfa 7, 8
25      ret
26      .cfi_endproc
27  .LFE0:
28      .size   codegear1, .-codegear1
29      .section    .rodata
30  .LC0:
31      .string "Hello"
32      .text
33      .globl  codegear2
34      .type   codegear2, @function
```

codegear1 から codegear2 への移動の際に、CbC と C で発行されるアセンブラの命令を比較する。CbC の例題の場合のアセンブラのソースコード 2.3 は codegear2 へ 25 行目で jmp 命令を使って遷移している。対して C 言語での実装の場合 (ソースコード 2.4) は 21 行目で callq を使っている。jmp 命令はプログラムカウンタを切り替えるのみの命令であり、

call は関数呼び出しの命令であるためにスタックの操作が行われる。CbC での goto 文はすべてこの jmp 命令に変換されるため、 関数呼び出しより軽量に実行することが可能である。

2.4 CbC を使ったシステムコールディスパッチの例題

CbC を用いて MIT が開発した教育用の OS である xv6[14] の書き換えを行った。CbC を利用したシステムコールのディスパッチ部分をソースコード 2.5 に示す。この例題では特定のシステムコールの場合、 CbC で実装された処理に `goto` 文をつかって継続する。例題では CodeGear へのアドレスが配列 `cbccodes` に格納されている。引数として渡している `cbc_ret` は、 システムコールの返り値の数値をレジスタに代入する CodeGear である。実際に `cbc_ret` に継続が行われるのは、 `read` などのシステムコールの一連の処理の継続が終わったタイミングである。

ソースコード 2.5: CbC を利用したシステムコールのディスパッチ

```
1 void syscall(void)
2 {
3     int num;
4     int ret;
5
6     if((num >= NELEM(syscalls)) && (num <= NELEM(cbccodes)) && cbccodes[
      num]) {
7         proc->cbc_arg.cbc_console_arg.num = num;
8         goto (cbccodes[num])(cbc_ret);
9     }
```

軽量継続を持つ CbC を利用して、 証明可能な OS を実装したい。その為には証明に使用される定理証明支援系や、 モデル検査機での表現に適した状態遷移単位での記述が求められる。CbC で使用する CodeGear は、 状態遷移モデルにおける状態そのものとして捉えることが可能である。CodeGear を元にプログラミングをするにつれて、 CodeGear の入出力の Data も重要であることが解ってきた。

2.5 メタ計算

メタ計算のメタとは、高次元などの意味を持つ言葉であり、特定の物の上位に位置するものである。メタ計算の場合は計算に必要な計算や、 計算を行うのに必要な計算を指す。GearsOS でのメタ計算は、 通常の計算を管理している OS レベルの計算などを指す。OS から見たメタ計算は、自分自身を検証する計算などになる。

ノーマルレベルの計算からすると、メタ計算は通常隠蔽される。これは UNIX のプログラムを実行する際に、 OS のスケジューラーのことを意識せずに実行可能であることな

どから分かる。新しいプロセスを作製する場合は `fork` システムコールを実行する必要がある。システムコールの先は OS が処理を行う。fork システムコールの処理を OS が計算するが、 この計算はユーザープログラムから見るとメタ計算である。システムコールの中で何が起こっているかはユーザーレベルでは判断できず、 返り値などの API を経由して情報を取得する。現状の UNIX ではメタ計算はこの様なシステムコールの形としても表現される。

メタデータなどはデータのデータであり、 データを扱う上で必要なデータ情報を意味する。プログラムの中でプログラムを生成するものをメタプログラミングなどと呼ぶ。メタ計算やメタプログラムは、プログラム自身の検証などにとって重要な機能である。しかしメタレベルの計算をノーマルレベルで自在に記述してしまうと、 ノーマルレベルでの信頼性に問題が発生する。メタレベルではポインタ操作や資源管理を行うため、メモリオーバーフローなどの問題を簡単に引き起こしてしまう。メタレベルの計算とノーマルレベルの計算を適切に分離しつつ、 ノーマルレベルから安全にメタレベルの計算を呼び出す手法が必要となる。

プログラミング言語からメタ計算を取り扱う場合、 言語の特性に応じて様々な手法が使われてきた。関数型プログラミングの見方では、 メタ計算はモナドの形で表現されていた。[15] OS の研究ではメタ計算の記述に型付きアセンブラを用いることもある。[16]

通常ユーザーがメタレベルのコードを扱う場合は特定の API を経由することになる。プログラム実行中のスタックの中には、 プログラムが現在実行している関数までのフレームや、各関数でアロケートされた変数などの情報が入る。これらを環境と呼ぶ。現状のプログラミング言語や OS では、 この環境を常に持ち歩かなければならない。ノーマルレベルとメタレベルを分離しようとすると、 環境の保存について考慮しなければならず、 結果的にシステムコールなどの API を使わなければならない。システムコールを利用しても、 保存されている環境が常に存在する。また関数単位での分離を行っても、 呼び出す関数の数が細かくなってしまい、スタックの容量を容易に消費してしまう。既存言語ではメタ計算の分離が困難である。

CbC では goto 文による軽量継続によって、 スタックを goto の度に捨てていく。そもそもスタックが存在しないため、 暗黙の環境も存在せずに自由にプログラミングが可能となる。また CodeGear をどれだけ呼び出しても、関数呼び出し時に伴うスタックの消費も存在しない。メタ計算の単位で細かく CodeGear を切り分けても、実行の問題が生じない。その為従来のプログラミング言語ではできなかった、ノーマルレベルとメタレベルのコードの分離が容易に行える。

CbC でのメタ計算は CodeGear、 DataGear の単位がそのまま使用できる。メタ計算を行う CodeGear や、 メタな情報を持つ DataGear が存在する。これらの単位はそれぞれ、MetaCodeGear、 MetaDataGear と呼ばれる。

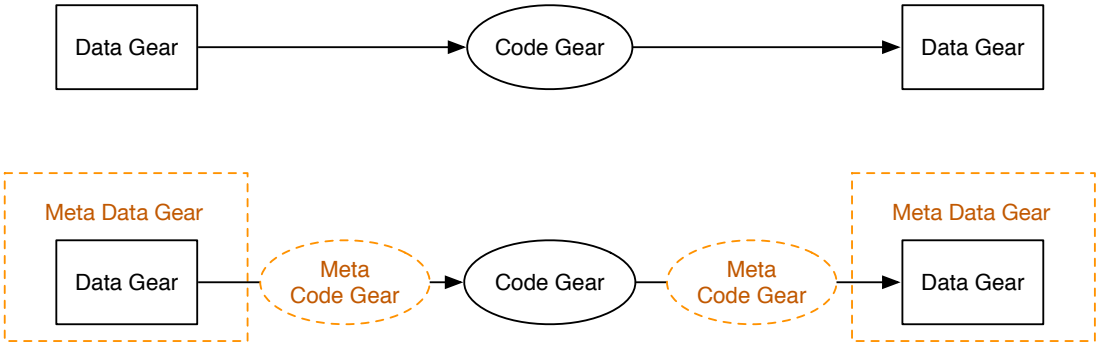

図 2.2: CodeGear と MetaCodeGear

2.6 MetaCodeGear

遷移する各 CodeGear の実行に必要なデータの整合性の確認などはメタ計算である。この計算は MetaCodeGear と呼ばれる各 CodeGear ごと実装されたメタな CodeGear で計算を行う。

特に対象の CodeGear の直前で実行される MetaCodeGear を StubCodeGear と呼ぶ。ユーザーからするとノーマルレベルの CodeGear 間の移動に見えるが、実際には StubCodeGear が挿入される。MetaCodeGear や MetaDataGear は、プログラマが直接実装せず、 Perl スクリプトによって GearsOS のビルド時に生成される。ただし Perl スクリプトはすでに書かれていた StubCodeGear は上書きしない。スクリプトに問題がある場合や、 細かな調整をしたい場合はプログラマが直接実装可能である。CodeGear から別の CodeGear に遷移する際の DataGear などの関係性を、図 2.2 に示す。

通常のコード中では入力の DataGear を受け取り CodeGear を実行、 結果を DataGear に書き込んだ上で別の CodeGear に継続する様に見える。この流れを図 2.2 の上段に示す。しかし実際は CodeGear の実行の前後に実行される MetaCodeGear や入出力の DataGear を MetaDataGear から取り出すなどのメタ計算が加わる。これは図 2.2 の下段に対応する。

2.7 MetaDataGear

基本は C 言語の構造体そのものであるが、 DataGear の場合はデータに付随するメタ情報も取り扱う。これはデータ自身がどういう型を持っているかなどの情報である。ほかに計算を実行する CPU、 GPU の情報や、 計算に必要なすべての DataGear の管理などの実行環境のメタデータも DataGear の形で表現される。このメタデータを扱う DataGear を MetaDataGear と呼ぶ。

また CbC はスタックを持たないため、 データを保存したい場合はスタック以外の場所に値を書き込む必要がある。このスタック以外の場所は DataGear であり、 メタなデータを扱っているために MetaDataGear と言える。具体的に MetaDataGear がどのように構成されているかは、CbC を扱うプロジェクトによって異なる。

第3章　GearsOS

GearsOS とは Continuation Based C を用いて信頼性と拡張性の両立を目指して実装している OS プロジェクトである。[17] CodeGear と DataGear を基本単位として実行する。CodeGear を基本単位としているため、 各 CodeGear は割り込みされず実行される必要がある。割り込みを完全に制御することは一般的には不可能であるが、 GearsOS のメタ計算部分でこれを保証したい。DataGear も基本単位であるため、各 CodeGear が DataGear をどのように扱っているか、書き込みをしたかは GearsOS 側で保証するとしている。

GearsOS は OS として実行する側面と、 CbC のシンタックスを拡張した言語フレームワークとしての側面がある。GearsOS はノーマルレベルとメタレベルの分離を目指して構築している OS である。すべてをプログラマが純粋な CbC で記述してしまうと、 メタレベルの情報を実装しなければならず、ノーマルレベルとメタレベルの分離をした意味がなくなってしまう。GesrsOS ではユーザーが書いたノーマルレベルのコードの特定の記述や、シンタックスをもとに、メタレベルの情報を含む等価な CbC へとコンパイル時にコードを変換する。コード変換は Perl スクリプトで行われている。

現在の GearsOS は Unix システム上のアプリケーションとして実装されているものと、xv6 の置き換えとして実装されているもの [18] がある。

3.1　GearsOS の構成

GearsOS は様々な役割を持つ CodeGear と DataGear で構成されている。また CodeGear と DataGear のモジュール化の仕組みとして Interface が導入されている。GearsOS の構成図を図 3.1 に示す。中心となる MetaDataGear は以下の要素である。

- Context
- TaskManager
- TaskQueue
- Worker

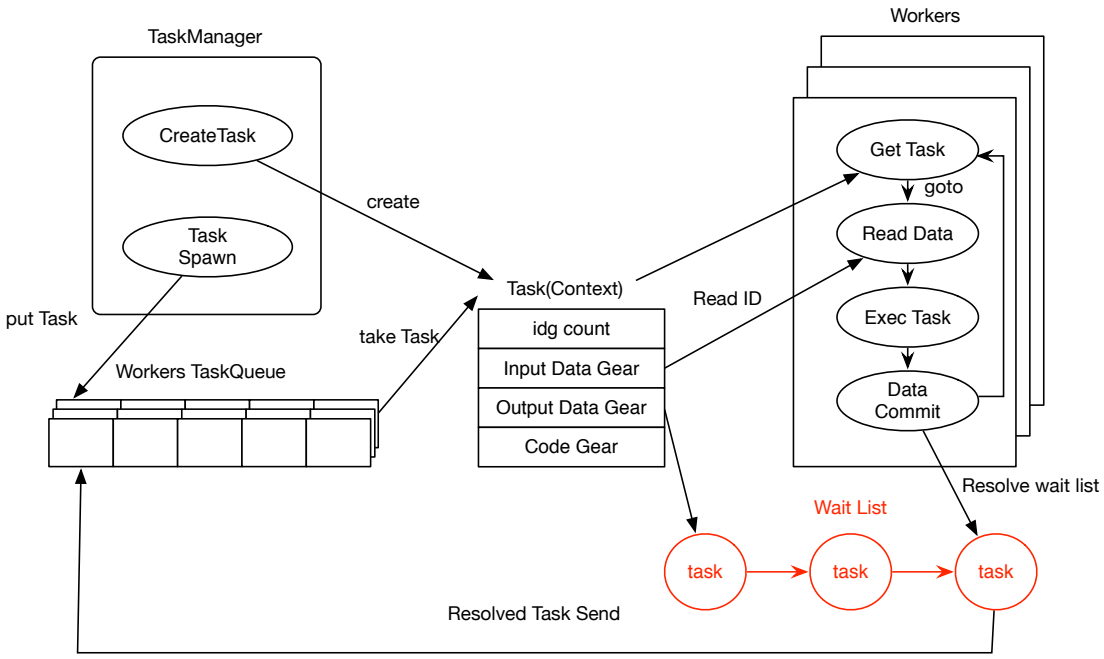

図 3.1: GearsOS の構成

3.2 Context

Context とは従来の OS のプロセスに相当する概念である。GearsOS でのデータの単位から見ると、 MetaDataGear に相当する。Context の概要図を図 3.2 に、実際の CbC 上での定義をソースコード 3.1 に示す。

Context は MetaDataGear である為に、 ノーマルレベルの CodeGear からは context は直接参照しない。context の操作をしてしまうと、メタレベルとノーマルレベルの分離をした意味がなくなってしまう為である。

Context はプロセスに相当するので、ユーザープログラムごとに Cotnext が存在する。この Context を User Context と呼ぶ。さらに実行している GPU や CPU ごとに Context が必要となる。これらは CPU Context と呼ばれる。GearsOS は OS であるため、全体を管理する Kernel の Context も必要となる。これは KernelContext や KContext と呼ばれる。KContext はすべての Context を参照する必要がある。OS が持たなければならない割り込みのフラグなどは KContext に置かれている。GearsOS のメタレベルのプログラミングでは、 今処理をしている Context が誰の Context であるかを強く意識する必要がある。

ソースコード 3.1: context の定義

```
1 struct Context {
2     enum Code next;
```

```
3     struct Worker* worker;
4     struct TaskManager* taskManager;
5     int codeNum;
6     __code (**code) (struct Context*);
7     union Data **data;
8     struct Meta **metaDataStart;
9     struct Meta **metaData;
10    void* heapStart;
11    void* heap;
12    long heapLimit;
13    int dataNum;
14
15    // task parameter
16    int idgCount; //number of waiting dataGear
17    int idg;
18    int maxIdg;
19    int odg;
20    int maxOdg;
21    int gpu; // GPU task
22    struct Context* task;
23    struct Element* taskList;
24 #ifdef USE_CUDAWorker
25    int num_exec;
26    CUmodule module;
27    CUfunction function;
28 #endif
29    /* multi dimension parameter */
30    int iterate;
31    struct Iterator* iterator;
32 };
```

Context は GearsOS の計算で使用されるすべての DataGear と CodeGear を持つ。つまり GearsOS で使われる CodeGear と DataGear は、誰かの Context に必ず書き込まれている。各 CodeGear、DataGear は Context はそれぞれ配列形式で Context にデータを格納する場所が用意されている。CodeGear が保存されている配列はソースコード 3.1 の 6 行目で定義している `code` である。StubCodeGear は Context のみを引数で持つため、 `__code stub(struct Context*)` の様な CodeGear の関数ポインタのポインタ、つまり CodeGear の配列としての定義されている。これは前述した StubCodeGear の関数ポインタが格納されており、 `__code meta` でのディスパッチに利用される。

DataGear が保存されている配列は 7 行目で定義している `data` である 。すべての DataGear は GearsOS 上では `union Data` 型として取り扱えるので、 `union Data` のポインタの配列として宣言されている。ただし GearsOS で使うすべての DataGear がこの Context に保存されている訳ではない。Interface を利用した goto 時の値の保存場所として、この配列に DataGear ごと割り振られた場所に DtaGear を保存する用途で利用している。CodeGear で利用している配列と同様に、 この配列の添え字も DataGear の番号に対応している。

DataGear は配列形式のデータ格納場所のほかに、Context が持つヒープに保存することも可能である。計算で必要な DataGear は、 CbC の中でアロケーションした場合は

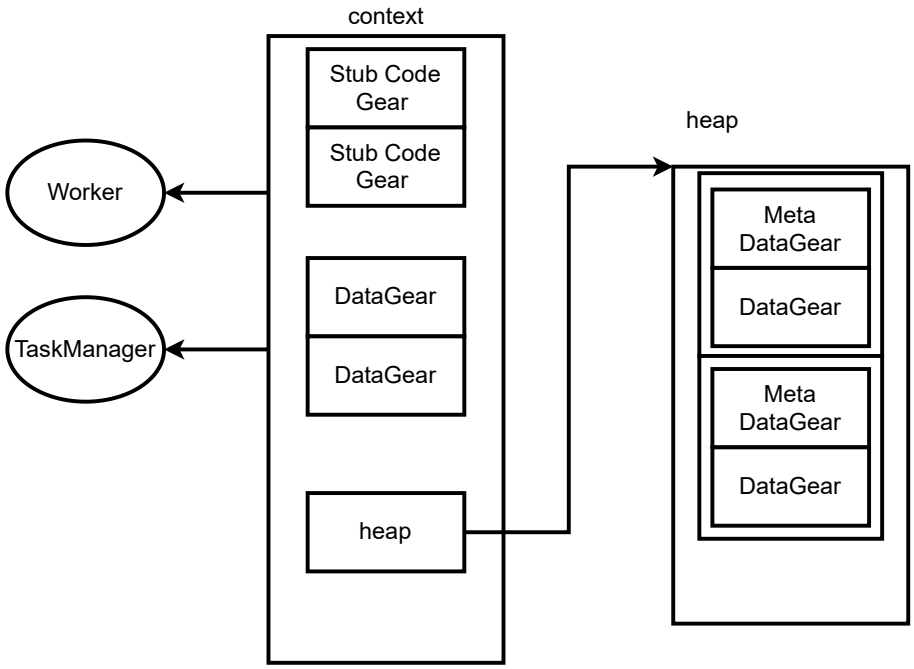

図 3.2: Context の概要図

Context にヒープに書き込まれる。ヒープには DataGear と、書き込んだ DataGear のメタ情報が記載されている MetaDataGear で構成されている。

3.3 Stub Code Gear

次の CodeGear に継続する際、ノーマルレベルから見ると次の CodeGear を直接指定しているように見える。さらに次の CodeGear に引数などを直接渡しているようにも見える。しかしノーマルレベルから次の CodeGear に継続する場合は関数ポインタなどが必要になるが、 これらはメタ計算に含まれる。その為純粋にノーマルレベルから CodeGear 間を自由に継続させてしまうと、 ノーマルレベルとメタレベルの分離ができなくなってしまう。ノーマルレベルとメタレベルの分離の為に、 次の CodeGear には直接継続させず、間に MetaCodeGear をはさむようにする必要がある。またポインタをノーマルレベルには持たせず、 継続先の CodeGear は番号を使って指定する。CodeGear 間の継続は GearsOS のビルド時に Perl スクリプトによって書き換えが行われ、MetaCodeGear を経由するように変更される。

GearsOS では DataGear はすべて Context を経由してやり取りをする。次の継続に DataGear を渡す場合、 継続する前に一度 Context に DataGear を書き込み、 継続先で Context

から DataGear を取り出す。Context は MetaDataGear であるために、 ノーマルレベルの CodeGear ではなく MetaCodeGear で扱う必要がある。各 CodeGear の計算で必要な DataGear を Context から取り出す MetaCodeGear は、 実行したい CodeGear の直前で実行される必要がある。この CodeGear を特に StubCodeGear と呼ぶ。StubCodeGear はすべての CodeGear に対して実装しなければならず、 手で実装するのは煩雑である。StubCodeGear も GearsOS のビルド時に Perl スクリプトによって自動生成される。

ソースコード 3.4 に示すノーマルレベルで記述した CodeGear を、Perl スクリプトによって変換した結果をソースコード 3.5 に示す。常に自分自身の Context を CodeGear は入力の形で受け取る為、変換後の `pushSingleLinkedStack` は、 第 1 引数に Context が加わっている。pushSingleLinkedStack は引数は 3 つ要求していた。これらの引数は生成された `pushSingleLinkedStack_stub` が Context の特定の場所から取り出す。この CodeGear は GearsOS の Interface を利用しており、Stack Interace の実装となっている。マクロ `Gearef` は、 context の Interface 用の DataGear の置き場所にアクセスするマクロであり、 Stack Interface の置き場所から、引数情報を取得している。マクロ `Gearef` の定義をソースコード 3.2 に示す。マクロ `Gearef` では引数で与えられた DataGear の名前を、enum を利用した番号に変換し、context から値を取り出している。DataGear は `enum Data` 型で各 DataGear の型ごとに番号が割り振られている。(ソースコード 3.3)

ソースコード 3.2: Gearef マクロ

```
1 #define Gearef(context, t) (&(context)->data[D_##t]->t)
```

ソースコード 3.3: enumData の定義

```
1 enum DataType {
2     D_Code,
3     D_Atomic,
4     D_AtomicReference,
5     D_CPUWorker,
6     D_Context,
7     D_Element,
8     ...
9 };
```

すべての引数を取得したのちに、`goto pushSingleLinkedStack` で、CodeGear に継続する。

ソースコード 3.4: Stack に Push する CodeGear

```
1 __code pushSingleLinkedStack(struct SingleLinkedStack* stack, union Data*
      data, __code next(...)) {
2     Element* element = new Element();
3     element->next = stack->top;
4     element->data = data;
5     stack->top = element;
6     goto next(...);
```

```
7 }
```

ソースコード 3.5: 3.4 の StubCodeGear

```
1  __code pushSingleLinkedStack(struct Context *context,struct
       SingleLinkedStack* stack, union Data* data, enum Code next) {
2      Element* element = &ALLOCATE(context, Element)->Element;
3      element->next = stack->top;
4      element->data = data;
5      stack->top = element;
6      goto meta(context, next);
7  }
8
9  __code pushSingleLinkedStack_stub(struct Context* context) {
10     SingleLinkedStack* stack = (SingleLinkedStack*)GearImpl(context,
       Stack, stack);
11     Data* data = Gearef(context, Stack)->data;
12     enum Code next = Gearef(context, Stack)->next;
13     goto pushSingleLinkedStack(context, stack, data, next);
14 }
```

Context と継続の関係性を図 3.3 に示す。StubCodeGear は GearsOS で定義されているノーマルレベルの CodeGear のすべてに生成される。

__code meta の定義をソースコード 3.6 に示す。

ソースコード 3.6: __code meta

```
1  __code meta(struct Context* context, enum Code next) {
2      goto (context->code[next])(context);
3  }
```

__code meta は Context に格納されている CodeGear の配列から CodeGear のアドレスを取得し継続する。この際に配列の要素を特定する際に使われる添え字は、各 CodeGear に割り振られた番号を利用している。この番号は C 言語の列挙体を使用した enum Code 型で定義されている。enum Code 型の定義をソースコード 3.7 に示す。命名規則は C_CodeGearName となっている。

ソースコード 3.7: CodeGear の番号である enumCode の定義

```
1  enum Code {
2      C_checkAndSetAtomicReference,
3      C_clearSingleLinkedStack,
4      C_clearSynchronizedQueue,
5      C_createTask,
6      C_decrementTaskCountTaskManagerImpl,
7      C_exit_code,
8      C_get2SingleLinkedStack,
9      ...
10 };
```

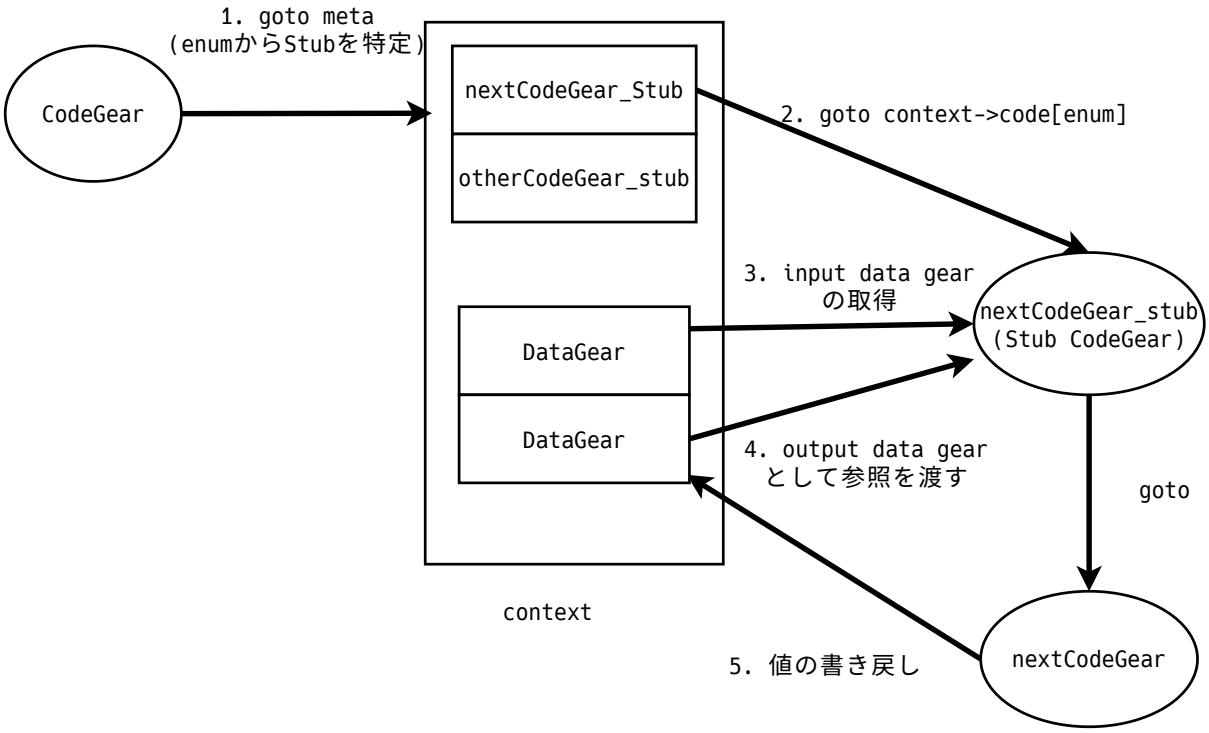

図 3.3: Context を参照した CodeGear のデータアクセス

`enum Code` 型は GearsOS のコンパイル時に利用されている CodeGear を数え上げて生成される。Context の code 配列には、各 CodeGear の StubCodeGear の関数ポインタが配置されている。よって`__code meta`から継続する先の CodeGear は、呼び出し先の CodeGear の直前に実行される StubCodeGear になる。

CodeGear から CodeGear への継続は、関数型プログラミングの継続先に渡す Data と Code の組の Closure となっている。シンタックスでは継続の際に引数 (`...`) を渡す。これは処理系では特に使用していないキーワードであるが、 この Closure を持ち歩いていることを意識するために導入されている。

3.4　TaskManager

TaskManager は GearsOS 上で実行される Task の管理を行う。GearsOS 上の Task とは Context のことであり、各 Context には自分の計算に必要な DataGear のカウンタなどが含まれている。TaskManager は、CodeGear の計算に必要な入力の DataGear(InputDataGear) が揃っているかの確認、揃っていなかったら待ち合わせを行う処理がある。すべての

DataGear が揃った場合、 Task を Worker の Queue に通知し Task を実行させる。この処理は GearsOS を並列実行させる場合に必要な機能となっている。

TaskManager は性質上シングルトンである。その為、複数 Worker を走らせた場合でも 1 全体で 1 つのみの値を持っていたいものは TaskManager が握っている必要がある。例えばモデル検査用の状態保存用のデータベース情報は、 TaskManager が所有している。

3.5　TaskQueue

GearsOS の TaskQueue は SynchronizedQueue で表現されている。TaskQueue は Worker が利用する Queue となっている。

Worker の Queue は、TaskManager に接続して Task を送信するスレッドと、 Task を実行する Worker 自身のスレッドで扱われる。さらに Worker が複数走る可能性もある。その為 SynchronizedQueue は、マルチスレッドでもデータの一貫性を保証する必要がある。GearsOS では CAS(Check and Set, Compare and Swap) を利用して実装が行われている。

3.6　Worker

Worker は Worker の初期化にスレッドを作る。GearsOS ではスレッドごとにそれぞれ Context が生成される。Worker はスレッド作製後に Context の初期化 API を呼び出し、自分のフィールドに Context のアドレスを書き込む。

スレッド作製後は TaskManager から Task を取得する。Task は Context の形で表現されているために、Worker の Context を Task に切り替え、 Task の次の継続に実行する。OutputDataGear がある場合は、Task 実行後に DataGear の書き出しが行われる。

Worker は CodeGear の前後で確実に呼び出される。この性質を利用すると、CodeGear の実行の前後での状態を記録することが可能である。つまりモデル検査が可能である為、モデル検査用の Worker を定義して入れ替えるとコードに変更を与えずに実行できる。Worker 自体は Interface で表現されているために、入れ替えは容易となっている。GearsOS では通常の Worker として CPUWorker を、GPU に関連した処理をする CUDAWorker、合間にモデル検査関連のメタ計算をはさむ MCWorker が定義されている。

3.7　union Data型

CbC/GearsOS では DataGear は構造体の形で表現されていた。すべての DataGear を管理する、Context は計算で使うすべての DataGear の型定義を持っている。各 DataGear は当然ではあるが別の型である。例えば Stack DataGear と Queue DataGear は、それぞ

れ struct Stack と struct Queue で表現されるが、 C 言語のシステム上別の型とみなされる。メタレベルで見れば、 この型定義は `union Data` 型にすべて書かれている。しかし Context はこれらの型をすべて DataGear として等しく扱う必要がある。この為に C 言語の共用体を使用し、汎用的な DataGear の型である union Data 型を定義している。共用体とは、構成するメンバ変数で最大の型のメモリサイズと同じメモリサイズになる特徴があり、複数の異なる型をまとめて管理することができる。構造体と違い、1 度に一つの型しか使うことができない。

実際にどの型が書き込まれているかは、 Context のどこに書き込まれているかによって確認の方法が異なる。Interface の入出力で利用している data 配列の場合は、enum の番号と data 配列の添え字が対応している。このため enum で指定した場所に入っている union Data の具体的な型は、 enum と対応する DataGear になる。context のヒープにアロケートされた DataGear の場合は、 型情報を取得できる MetaDataGear にアクセスすると、なんの型であったかが分かる。

Context から取り出してきた union Data から DataGear の型への変換はメタ計算で行われる。GearsOS の場合は、計算したい CodeGear の直前で実行される StubCodeGear で値のキャストが行われる。

3.8 Interface

GearsOS のモジュール化の仕組みとして Interface がある。Interface は CodeGear と各 CodeGear で使う入出力の DataGear の集合である。Interface に定義されている CodeGear は、各 Interface が満たすこと期待する API である。

Inteface は仕様 (Interface) と、実装 (Implement, Impl) を別けて記述する。Interface を呼び出す場合は、Interface に定義された API に沿ってプログラミングをすることで、Impl の内容を隠蔽することができる。これによってメタ計算部分で実装を入れ替えつつ Interface を使用したり、 ふるまいを変更することなく具体的な処理の内容のみを変更することが容易にできる。これは Java の Interface、Haskell の型クラスに相当する機能である。

3.8.1 Interface の定義

GearsOS に実装されている Queue の Interface の定義をソースコード 3.8 に示す。

ソースコード 3.8: Queue の Interface

```
1 typedef struct Queue<Impl>{
2         union Data* queue;
3         union Data* data;
4         __code whenEmpty(...);
5         __code clear(Impl* queue, __code next(...));
```

```
6         __code put(Impl* queue, union Data* data, __code next(...));
7         __code take(Impl* queue, __code next(union Data*, ...));
8         __code isEmpty(Impl* queue, __code next(...), __code whenEmpty
      (...));
9         __code next(...);
10 } Queue;
```

Interface の定義は、前半に入出力で利用する DataGear を列挙する。ここでは queue と data を利用する。4 行目からは API の宣言である。各 API は CodeGear であるので__code で宣言する。各 CodeGear の第 1 引数は Impl*型の変数になっている。これは Interface と対応する Implement の DataGear のポインタである。Java などのオブジェクト指向言語では self や this のキーワードで参照できるものとほぼ等しい。Interface 宣言時には具体的にどの型が来るかは不定であるため、 キーワードを利用している。Impl は Interface の API 呼び出し時に、メタレベルの処理である StubCodeGear で自動で入力される。その為ユーザーは Interface の API を呼び出す際は、 この Impl に対応する引数は設定しない。すなわち実際にいれるべき引数は、 Impl を抜いたものになる。

第 1 引数に Impl が来ない CodeGear として whenEmpty と next が Queue の例で存在している。これらは API の呼び出し時に継続として渡される CodeGear であるため、Interface の定義時には不定である。その為... を用いて、不定な CodeGear と DataGear の Closure が来ると仮定している。8 行目で定義している whenEmpty は Queue の状態を確認し、空でなければ next、空であれば whenEmpty に継続する。これらは呼び出し時に CodeGear を入力して与えることになる。

3.8.2 Interface の呼び出し

Interface で定義した API は interface->method の記法で呼びだすことが可能である。ソースコード 3.9 では、 Queue Interface の take API を呼び出している。take は__code next を要求しているので、 CodeGear の名前を引数として渡している。これはノーマルレベルでは enum の番号として処理される。take は出力を 1 つ出す CodeGear である為、継続で渡された odgCommitCUDAWorker4 は Stub でこの出力を受け取る。

ソースコード 3.9: Interface の API の呼び出し

```
1 __code odgCommitCUDAWorker3(struct CUDAWorker* worker, struct Context*
      task) {
2     int i = worker->loopCounter;
3     struct Queue* queue = GET_WAIT_LIST(task->data[task->odg+i]);
4     goto queue->take(odgCommitCUDAWorker4);
5 }
```

また、Interface を利用する場合はソースコード中に#interface "interfaceName.h"と記述する必要がある。例えば Queue を利用する場合は#interface "Queue.h"と記述しな

ければならない。#interface 構文は、一見すると C 言語のマクロの様に見える。実際にはマクロではなく、 Perl スクリプトによってメタレベルの情報を含む CbC ファイルに変換する際に、 Perl スクリプトに使っている Interface を教えるアノテーションの様な役割である。Perl スクリプトによって変換時に、 #interface の宣言は削除される。

3.8.3 Interface のメタレベルの実装

Interface 自身も DataGear であり、実際の定義は context の union Data 型に記述されている。メタレベルでは Inteface の DataGear の GearsOS 上の実装である構造体自身にアクセス可能である。Queue Interface に対応する構造体の定義をソースコード 3.10 に示す。

ソースコード 3.10: Queue の Interface に対応する構造体

```
1  struct Queue {
2      union Data* queue;
3      union Data* data;
4      enum Code whenEmpty;
5      enum Code clear;
6      enum Code put;
7      enum Code take;
8      enum Code isEmpty;
9      enum Code next;
10 } Queue;
```

Interface の実装は、この構造体に代入されている値で表現される。Interface の定義 (ソースコード 3.8) と、 実際の構造体 (ソースコード 3.10) を見比べると、 CodeGear は enum Code として表現し直されている。enum Code は GearsOS で使うすべての CodeGear に割り振られた番号である。Interface は API に対応する enum Code に、Impl 側の enumCode を代入することで、実装を表現している。Interface の Impl 側の DataGear は、各 Interface に存在する、Interface 名の最初の一文字が小文字になった union Data 型のポインタ経由で取得可能である。

3.8.4 Interface の Impl の実装

実際に Interface の初期化をしている箇所を確認する。Queue Interface に対応する SingleLinkedQueue の実装を 3.11 に示す。

ソースコード 3.11: SingleLinkedQueue の実装

```
1  #include "../context.h"
2  #include <stdio.h>
3  #interface "Queue.h"
4
5  Queue* createSingleLinkedQueue(struct Context* context) {
```

```
6      struct Queue* queue = new Queue();
7      struct SingleLinkedQueue* singleLinkedQueue = new SingleLinkedQueue()
       ;
8      queue->queue = (union Data*)singleLinkedQueue;
9      singleLinkedQueue->top  = new Element();
10     singleLinkedQueue->last = singleLinkedQueue->top;
11     queue->take  = C_takeSingleLinkedQueue;
12     queue->put   = C_putSingleLinkedQueue;
13     queue->isEmpty = C_isEmptySingleLinkedQueue;
14     queue->clear = C_clearSingleLinkedQueue;
15     return queue;
16 }
17
18 __code clearSingleLinkedQueue(struct SingleLinkedQueue* queue, __code
       next(...)) {
19     queue->top = NULL;
20     goto next(...);
21 }
22
23 __code putSingleLinkedQueue(struct SingleLinkedQueue* queue, union Data*
       data, __code next(...)) {
24     Element* element = new Element();
25     element->data = data;
26     element->next = NULL;
27     queue->last->next  = element;
28     queue->last = element;
29     goto next(...);
30 }
```

Interface の実装の場合も、Interface 呼び出しの API である`#interface "Queue.h"`を記述する必要がある。`createSingleLinkedQueue` は SingleLinkedQueue で実装した Queue Interface のコンストラクタである。これは関数呼び出しで実装されており、 返り値は Interface のポインタである。コンストラクタ内では Queue および SingleLinkedQueue のアロケーションを行っている。`new` 演算子が使われているが、 これは GearsOS で拡張されたシンタックスの 1 つである。new は GearsOS のビルド時に Perl スクリプトによって、context が持つ DataGear のヒープ領域の操作のマクロに切り替わる。ノーマルレベルでは context にアクセスできないので、 Java の様なアロケーションのシンタックスを導入している。

3.8.5 goto 時の Context と Interface の関係

Interface はモジュール化の仕組みとしてでなく、メタレベルでは一時的な変数の置き場所としても利用している。ソースコード 3.9 で呼び出している take は、 OutputDataGear がある API である。この OutputDataGear は、 context 内の DataGear の置き場所である data 配列の、 Interface のデータ格納場所に書き込まれる。OutputDataGear を取得する場合は、 継続先でなく、API の Interface から取得しないといけない。

また、goto 文で別の CodeGear に遷移する際も、引数情報を継続先の context の data 配列の場所に書き込む必要がある。この処理はメタレベルの計算であるため、 GearsOS のビルド時に Perl で変換される。ソースコード 3.12 にソースコード 3.9 の変換結果を示す。この例では StubCodeGear のディスパッチを行う `__code meta` への goto の前に、Gearef マクロを使った context への書き戻しが行われている。GearsOS は CbC を用いて実装している為、スタックを持っていない。その為都度データは Context に書き戻す必要があるが、データの一時保管場所としても利用されている。

ソースコード 3.12: take を呼び出す部分の変換後

```
1 __code odgCommitCPUWorker3(struct Context *context,struct CPUWorker*
     worker, struct Context* task) {
2     int i = worker->loopCounter;
3     struct Queue* queue = GET_WAIT_LIST(task->data[task->odg+i]);
4     Gearef(context, Queue)->queue = (union Data*) queue;
5     Gearef(context, Queue)->next = C_odgCommitCPUWorker4;
6     goto meta(context, queue->take);
7 }
```

この性質から Interface には、Interface を実装する DataGear が持っておきたい変数はいれてはいけない。例えば Queue の実装では先頭要素を指し示す情報が必要であるが、 これを Interface 側の DataGear にしてしまうと、 呼び出し時に毎回更新されてしまう。常に持っておきたい値は、 Impl 側の DataGear の要素として定義する必要がある。

3.9　GearsOS のビルドシステム

GearsOS ではビルドツールに CMake を利用している。ビルドフローを図 3.4 に示す。CMake は automake などの Make ファイルを作成するツールに相当するものである。GearsOS でプログラミングする際は、ビルドしたいプロジェクトで利用するソースコード群を CMake の設定ファイルである CMakeLists.txt に記述する。CMakeLists.txt では GearsOS のビルドに必要な一連の処理をマクロ `GearsCommand` で制御している。このマクロにプロジェクト名を `TARGET` として、 コンパイルしたいファイルを `SOURCES` に記述する。ソースコード 3.13 の例では `pop_and_push` が `TARGET` に指定されている。なおヘッダファイルは `SOURCES` に指定する必要はなく、 自動で解決される。

CMake 自身はコンパイルに必要なコマンドを実行することはなく、ビルドツールである make や ninja-build に処理を移譲している。CMake は make や ninja-build が実行時に必要とするファイルである Makefile、 build.ninja の生成までを担当する。

ソースコード 3.13: CMakeList.txt 内でのプロジェクト定義

```
1 GearsCommand(
2   TARGET
```

```
3   pop_and_push
4   SOURCES
5   examples/pop_and_push/main.cbc  examples/pop_and_push/StackTestImpl.cbc
      TaskManagerImpl.cbc CPUWorker.cbc SynchronizedQueue.cbc
    AtomicReference.cbc SingleLinkedStack.cbc examples/pop_and_push/
    StackTest2Impl.cbc examples/pop_and_push/StackTestImpl3.cbc
6 )
```

GearsOS のビルドでは直接 CbC コンパイラがソースコードをコンパイルすることはなく、 間に Perl スクリプトが 2 種類実行される。Perl スクリプトはビルド対象の GearsOS で拡張された CbC ファイルを、純粋な CbC ファイルに変換する。ほかに GearsOS で動作する例題ごとに必要な初期化関数なども生成する。Perl スクリプトで変換された CbC ファイルなどをもとに CbC コンパイラがコンパイルを行う。ビルドの処理は自動化されており、 CMake 経由で make や ninja コマンドを用いてビルドする。

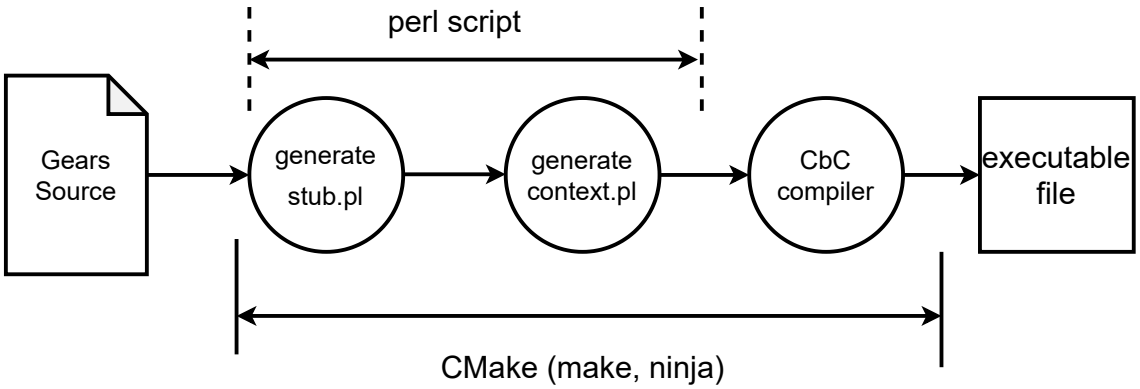

図 3.4: GearsOS のビルドフロー

3.10　GearsOS の CbC から純粋な CbC への変換

GearsOS は CbC を拡張した言語となっている。ただしこの拡張自体は CbC コンパイラである gcc、 llvm/clang には搭載されていない。その為 GearsOS の拡張部分を、等価な純粋な CbC の記述に変換する必要がある。現在の GearsOS では、 CMake によるコンパイル時に Perl で記述された `generate_stub.pl` と `generate_context.pl` の 2 種類のスクリプトで変換される。

これらの Perl スクリプトはプログラマが自分で動かすことはない。Perl スクリプトの実行手順は CMakeLists.txt に記述しており、 make や ninja-build でのビルド時に呼び出される。(ソースコード 3.14)

ソースコード 3.14: CMakeList.txt 内での Perl の実行部分

```
macro( GearsCommand )
    set( _OPTIONS_ARGS )
    set( _ONE_VALUE_ARGS TARGET )
    set( _MULTI_VALUE_ARGS SOURCES )
    cmake_parse_arguments( _Gears "${_OPTIONS_ARGS}" "${_ONE_VALUE_ARGS}"
    "${_MULTI_VALUE_ARGS}" ${ARGN} )

    set (_Gears_CSOURCES)
    foreach(i ${_Gears_SOURCES})
        if (${i} MATCHES "\\.cbc")
            string(REGEX REPLACE "(.*).cbc" "c/\\1.c" j ${i})
            add_custom_command (
                OUTPUT    ${j}
                DEPENDS   ${i}
                COMMAND  "perl" "generate_stub.pl" "-o" ${j} ${i}
            )
        elseif (${i} MATCHES "\\.cu")
            string(REGEX REPLACE "(.*).cu" "c/\\1.ptx" j ${i})
            add_custom_command (
                OUTPUT    ${j}
                DEPENDS   ${i}
                COMMAND  nvcc ${NVCCFLAG} -c -ptx -o ${j} ${i}
            )
        else()
            set(j ${i})
        endif()
        list(APPEND _Gears_CSOURCES ${j})
    endforeach(i)
```

3.11 generate_stub.pl

generate_stub.pl は各 CbC ファイルごとに呼び出される。図 3.5 に generate_stub.pl を使った処理の概要を示す。ユーザーが記述した GearsOS の CbC ファイルは、 ノーマルレベルのコードである。generate_stub.pl は、 CbC ファイルにメタレベルの情報を付け加え、GearsOS の拡張構文を取り除いた結果の CbC ファイルを新たに生成する。返還前の GearsOS のファイルの拡張子は.cbc であるが、 generate_stub.pl によって変換されると、同名で拡張子のみ.c に切り替わったファイルが生成される。拡張子は.c であるが、中身は CbC で記述されている。generate_stub.pl はあるプログラムのソースコードから別のプログラムのソースコードを生成するトランスコンパイラとして見ることができる。

generate_stub.pl は GearsOS のソースコードを 2 回読む。1 度目の読み込みで、ソースコード中に登場する CodeGear と、CodeGear の入出力を検知する。この際に`#interface` 構文で Interface の利用が確認された場合、 Interface の定義ファイルを開き、実装されている CodeGear と DataGear の組を取得する。

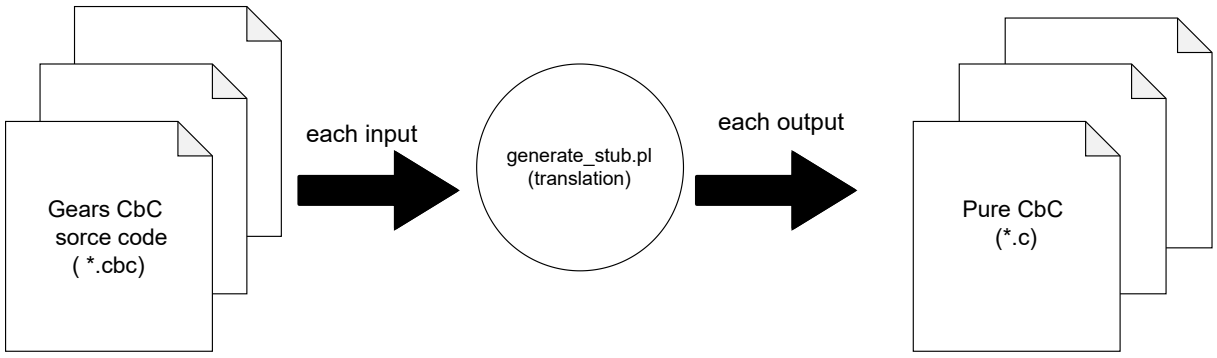

図 3.5: generate_sub.pl を使ったトランスコンパイル

1 度ファイルを完全に読み込み、 CodeGear、DataGear の情報を取得し終わると、以降はその情報をもとに変換したファイルを書き出す。ファイルを書き出す際は、 元の CbC ファイルを再読み込みし、 変換する必要があるキーワードが出現するまでは、変換後のファイルに転記を行う。例えば各 CodeGear の最後に実行される `goto` 文は、 GearsOS の場合は MetaCodeGear に継続するように、 対象を切り替える必要がある。この為に generate_stub.pl は、goto 文を検知すると context 経由で引数のやりとりをするメタ処理を付け加える。また、すべての CodeGear は context を入力として受け取る必要があるため、 引数を書き換えて Context を付け加えている。

generate_stub.pl は Perl で書かれたトランスコンパイラであり、 C 言語のコンパイラのように文字列を字句解析、構文解析をする訳ではない。いくつかあらかじめ定義した正規表現パターンに読み込んでいる CbC ファイルの行がパターンマッチされたら、 特定の処理をする様に実装されている。

CodeGear の入力を context から取り出す StubCodeGear の生成も generate_stub.pl で行う。なおすでに StubCodeGear が実装されていた場合は、 generate_stub.pl は StubCodeGear は生成しない。

3.12 generate_context.pl

generate_context.pl は、Context の初期化関連のファイルを生成する Perl スクリプトである。Context を初期化するためには、下記の処理をしなければならない。

- CodeGear のリストに StubCodeGear のアドレスの代入
- goto meta 時に引数を格納する data 配列のアロケーション

- 計算で使用するすべての DataGear、CodeGear に対して番号を割り振り、 enum を作製する
- コンストラクタ関数の extern の生成

これらの記述は煩雑であるものの、CbC ファイルと DataGear の情報が纏められた context.h を見れば、記述すべき内容は一意に決定でき為、自動化が可能である。generate_context.pl は、 context.h を読み、まず DataGear の取得を行う。CodeGear は、generate_stub.pl で変換された CbC ファイルを読み込み、 `__code` があるものを CodeGear として判断する。また、C の一般関数でも `create` が関数名に含まれており、ポインタ型を返す関数は Interface のコンストラクタとして判断する。これらの情報をもとに、 CodeGear、DataGear の番号を作製し、enumCode.h と enumData.h として書き出す。

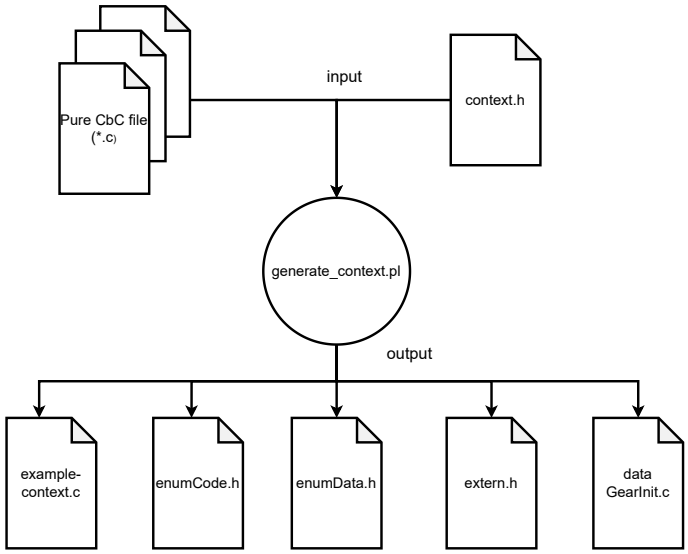

図 3.6: generate_context.pl を使ったファイル生成

3.13　CbC xv6

CbC xv6 は GearsOS のシステムを利用して xv6 OS の置き換えを目指しているプロジェクトである。[19] xv6 は v6 OS[20] を x86 アーキテクチャ用に MIT によって実装し直されたものである。Raspberry Pi 上での動作を目指しているため、 ARM アーキテクチャ用に改良されたバージョンを利用している。[21]

書き換えにおいてはビルドシステムは CMake を利用し、 Perl クロスコンパイラを導入してたりと GearsOS のビルドシステムとほぼ同じシステムを利用している。GearsOS を

使った比較的巨大な実用的なアプリケーションであるため、 xv6 の書き換えを進むに連れて様々な面で必要な機能や課題が生まれている。xv6 は UNIX OS である為プロセス単位で処理を行っていたが、 ここに部分的に Context を導入した。xv6 では割り込みのフラグなどを大域変数として使っていた。GearsOS で実装する場合は DataGear 単位になるため、 これらのフラグも DataGear の形で実装し直した。この DataGear は各プロセスに対応する Context ではなく、 中心的な Context がシングルトンで持っている必要がある。CbCxv6 の実装を通して Kernel の状況を記録しておく Context、つまり KernelContext が必要であることなどが判明した。

3.14　ARM 用ビルドシステムの作製

GearsOS をビルドする場合は、x86 アーキテクチャのマシンからビルドするのが殆どである。この場合ビルドしたバイナリは x86 向けのバイナリとなる。これはビルドをするホストマシンに導入されている CbC コンパイラが x86 アーキテクチャ向けにビルドされたものである為である。

CbC コンパイラは GCC と llvm/clang 上に構築した 2 種類が主力な処理系である。LVM/-clang の場合は LLVM 側でターゲットアーキテクチャを選択することが可能である。GCC の場合は最初から j ターゲットアーキテクチャを指定してコンパイラをビルドする必要がある。

時にマシンスペックの問題などから、別のアーキテクチャ向けのバイナリを生成したいケースがある。教育用マイコンボードである Raspberry Pi[22] は ARM アーキテクチャが搭載されている。Raspberry Pi 上で GearsOS のビルドをする場合、ARM 用にビルドされた CbC コンパイラが必要となる。Raspberry Pi 自体は非力なマシンであるため、GearsOS のビルドはもとより CbC コンパイラの構築を Raspberry Pi 上でするのは困難である。マシンスペックが高めの x86 マシンから ARM 用のバイナリをビルドして、 Raspberry Pi に転送し実行したい。ホストマシンのアーキテクチャ以外のアーキテクチャ向けにコンパイルすることをクロスコンパイルと呼ぶ。

GearsOS はビルドツールに CMake を利用しているので、 CMake でクロスコンパイル可能に工夫をしなければならない。ビルドに使用するコンパイラやリンカは CMake が自動探索し、 決定した上で Makefile や build.ninja ファイルを生成する。しかし CMake は今ビルドしようとしている対象が、自分が動作しているアーキテクチャかそうでないか、クロスコンパイラとして使えるかなどはチェックしない。つまり CMake が自動でクロスコンパイル対応の GCC コンパイラを探すことはない。その為そのままビルドすると x86 用のバイナリが生成されてしまう。

CMake を利用してクロスコンパイルする場合、CMake の実行時に引数でクロスコンパイラを明示的に指定する必要がある。この場合 x86 のマシンから ARM のバイナリを出力

する必要があり、 コンパイラやリンカーなどを ARM のクロスコンパイル対応のものに指定する必要がある。また、 xv6 の場合はリンク時に特定のリンカスクリプトを使う必要がある。これらのリンカスクリプトも CMake 側に、 CMake が提供しているリンカ用の特殊変数を使って自分で組み立てて渡す必要がある。CMake 側に使用したいコンパイラの情報を渡せれば、以降は CMake 側が自動的に適切なビルドスクリプトを生成してくれる。このような CMake の処理を手打ちで行うことは難しいので、 `pmake.pl` を作成した。`pmake.pl` の処理の概要を図 3.7 に示す。`pmake.pl` は Perl スクリプトで、 シェルコマンドを内部で実行しクロスコンパイル用のオプションを組み立てる。`pmake.pl` を経由して CMake を実行すると、 make コマンドに対応する Makefile、 ninja-build に対応する build.ninja が生成される。以降は cmake ではなく make などのビルドツールがビルドを行う。

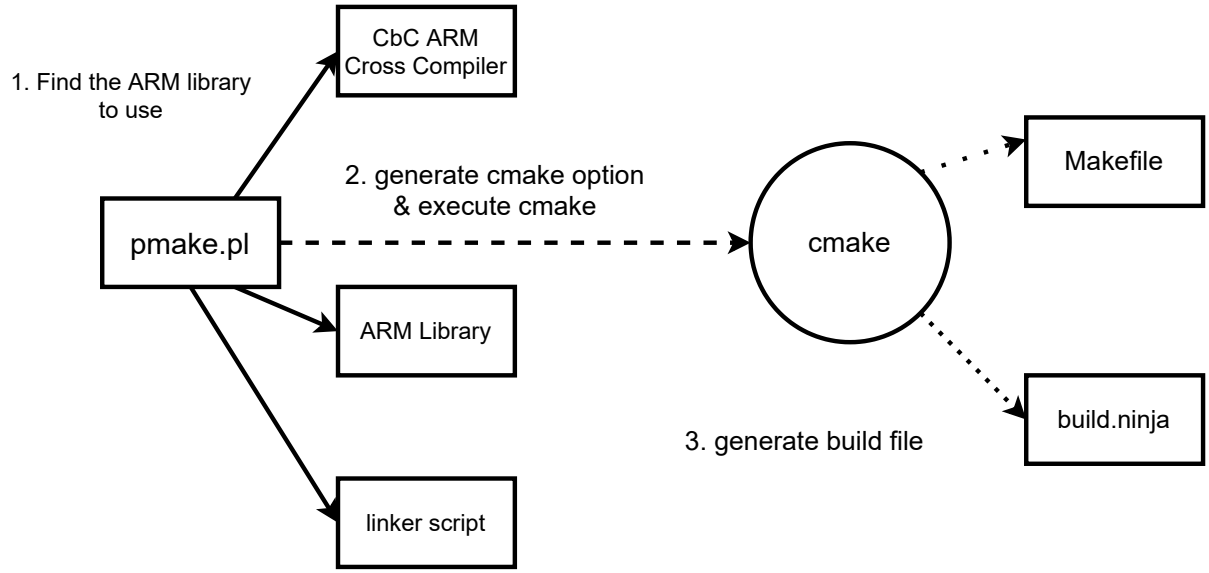

図 3.7: pmake.pl の処理フロー

3.15　Interface の取り扱い方法の検討

GearsOS の Interface はモジュール化の仕組みと `goto` 文での引数の一時保管場所としての機能を持っている。Interface の Implement のヘッダーファイルを実装したことで、GearsOS 上で Interface を実装する際に新たな方法での実装を検討した。Implement の CodeGear は今までは Interface で定義した CodeGear と 1 対 1 対応していた。Implement の CodeGear から goto する先は、入力として与えられた CodeGear か、Implement 内で独自

に定義した CodeGear に goto するケースとなっていた。後者の独自に定義した CodeGear に goto するケースも、 実装の CbC ファイルの中に記述されている CodeGear に遷移していた。

GearsOS を用いて xv6 OS を再実装した際に、 実装側の CodeGear を細かく別けて記述した。細分化によって 1 つの CbC ファイルあたりの CodeGear の記述量が増えてしまうという問題が発生した。見通しをよくする為に、 Interface で定義した CodeGear と直接対応する CodeGear の実装と、 それらから goto する CodeGear で実装ファイルを分離することを試みた。

第4章　GearsOSのInterfaceの改良

4.1　GearsOSのInterfaceの構文の改良

GearsOS の Interface では、 従来は DataGear と CodeGear を分離して記述していた。CodeGear の入出力を DataGear として列挙する必要があった。CodeGear の入出力として__code() の間に記述した DataGear の一覧と、Interface 上部で記述した DataGear の集合が一致している必要がある。ソースコード 4.1 は Stack の Interface の例である。

ソースコード 4.1: 従来の Stack Interface

```
1  typedef struct Stack<Type, Impl>{
2          union Data* stack;
3          union Data* data;
4          union Data* data1;
5          /* Type* stack; */
6          /* Type* data; */
7          /* Type* data1; */
8          __code whenEmpty(...);
9          __code clear(Impl* stack,__code next(...));
10         __code push(Impl* stack,Type* data, __code next(...));
11         __code pop(Impl* stack, __code next(Type* data, ...));
12         __code pop2(Impl* stack, __code next(Type* data, Type* data1,
       ...));
13         __code isEmpty(Impl* stack, __code next(...), __code whenEmpty
       (...));
14         __code get(Impl* stack, __code next(Type* data, ...));
15         __code get2(Impl* stack, __code next(Type* data, Type* data1,
       ...));
16         __code next(...);
17 } Stack;
```

従来の分離している記法の場合、 この DataGear の宣言が一致していないケースが多々発生した。また Interface の入力としての DataGear ではなく、 フィールド変数として DataGear を使うプログラミングスタイルを取るケースも見られた。GearsOS では、DataGear やフィールド変数をオブジェクトに格納したい場合、Interface 側ではなく Impl 側に変数を保存する必要がある。Interface 側に記述してしまう原因は複数考えられる。GearsOS のプログラミングスタイルに慣れていないことも考えられるが、構文によるところも考えられる。CodeGear と DataGear は Interface の場合は密接な関係性にあるが、分

離して記述してしまうと「DataGear の集合」と「CodeGear の集合」を別個で捉えてしまう。あくまで Interface で定義する CodeGear と DataGear は Interface の API である。これをユーザーに強く意識させる必要がある。

golang にも Interface の機能が実装されている。golang の場合は Interface は関数の宣言部分のみを記述するルールになっている。変数名は含まれていても含まなくても問題ない。

ソースコード 4.2: golang の interface 宣言

```
type geometry interface {
    area() float64
    perim() float64
}
```

GearsOS の Interface は入力と出力の API を定義するものであるので、golang の Interface のように、関数の API を並べて記述するほうが簡潔であると考えた。改良した Interface の構文で Stack を定義したものをソースコード 4.3 に示す。

ソースコード 4.3: 変更後の Stack Interface

```
typedef struct Stack<>{
        __code clear(Impl* stack,__code next(...));
        __code push(Impl* stack,union Data* data, __code next(...));
        __code pop(Impl* stack, __code next(union Data* data, ...));
        __code pop2(Impl* stack, __code next(union Data* data, union Data
   * data1, ...));
        __code isEmpty(Impl* stack, __code next(...), __code whenEmpty
   (...));
        __code get(Impl* stack, __code next(union Data* data, ...));
        __code get2(Impl* stack, __code next(union Data* data, union Data
   * data1, ...));
        __code next(...);
        __code whenEmpty(...);
} Stack;
```

従来の Interface では<Type, Impl>キーワードが含まれていた。これはジェネリクスの機能を意識して導入された構文である。Impl キーワードは実装自身の型を示す型変換として使われていた。しかし基本 Interface の定義を行う際に GearsOS のシステム上、CodeGear の第一引数は Impl 型のポインタが来る。これはオブジェクト指向言語で言う self に相当するものであり、 自分自身のインスタンスを示すポインタである。Impl キーワードは共通して使用されるために、 宣言部分からは取り外し、デフォルトの型キーワードとして定義した。Type キーワードは型変数としての利用を意識して導入されていたが、現在までの GearsOS の例題では導入されていなかった。ジェネリクスとしての型変数の利用の場合は T などの 1 文字変数がよく使われる。変更後の構文ではのちのジェネリクス導入のことを踏まえて、Type キーワードは削除した。

構文を変更するには、 GearsOS のビルドシステム上で Interface を利用している箇所を修正する必要がある。Interface は generate_stub.pl で読み込まれ、 CodeGear と入出力の DataGear の数え上げが行われる。この処理は Interface のパースに相当するものである。当然ではあるが、パース対象の Interface の構文は、変更前の構文にしか対応していない。

4.2　Implement の型定義ファイルの導入

Interface を使う言語では、 Interface が決まるとこれを実装するクラスや型が生まれる。GearsOS も Interface に対応する実装が存在する。例えば Stack Interface の実装は SingleLinkedStack であり、 Queue の実装は SingleLinkedQueue や SynchronizedQueue が存在する。

この SynchronizedQueue は GearsOS では DataGear として扱われる。Interface の定義と同等な型定義ファイルが、 実装の型については存在しなかった。従来は context.h の DataGear の宣言部分に、構造体の形式で表現したものを手で記述していた。(ソースコード 4.4)

ソースコード 4.4: cotnext.h に直接書かれた型定義

```
union Data {
   /* 略 */
    // Queue Interface
    struct Queue {
        union Data* queue;
        union Data* data;
        enum Code whenEmpty;
        enum Code clear;
        enum Code put;
        enum Code take;
        enum Code isEmpty;
        enum Code next;
    } Queue;
    struct SingleLinkedQueue {
        struct Element* top;
        struct Element* last;
    } SingleLinkedQueue;
    struct SynchronizedQueue {
        struct Element* top;
        struct Element* last;
        struct Atomic* atomic;
    } SynchronizedQueue;
   /* 略 */
};
```

CbC ファイルからは context.h をインクルードすることで問題なく型の使用は可能である。Perl のトランスコンパイラである generate_stub.pl は Interface の型定義ファイルをパースしていた。しかし型定義ファイルの存在の有無が Interface と実装で異なっている

為に、 generate_stub.pl で Implement の型に関する操作ができない。Implement の型も同様に定義ファイルを作製すれば、generate_stub.pl で型定義を用いた様々な処理が可能となり、ビルドシステムが柔軟な挙動が可能となる。また型定義は一貫して`*.h` に記述すれば良くなるため、 プログラマの見通しも良くなる。本研究では新たに Implement の型定義ファイルを考案する。

GearsOS ではすでに Interface の型定義ファイルを持っている。Implement の型定義ファイルも、 Interface の型定義ファイルと似たシンタックスにしたい。Implement の型定義ファイルで持たなければいけないのは、 どの Interface を実装しているかの情報である。この情報は他言語では Interface の実装を持つ型の宣言時に記述するケースと、型名の記述はせずに言語システムが実装しているかどうかを確認するケースが存在する。Java では `implements` キーワードを用いてどの Interface を実装しているかを記述する。[23] ソースコード 4.5 では、`Pig` クラスは `Animal` Interface を実装している。

ソースコード 4.5: Java の Implement キーワード

```
1  // interface
2  interface Animal {
3    public void animalSound(); // interface method (does not have a body)
4    public void sleep(); // interface method (does not have a body)
5  }
6
7  // Pig "implements" the Animal interface
8  class Pig implements Animal {
9    public void animalSound() {
10     // The body of animalSound() is provided here
11     System.out.println("The pig says: wee wee");
12   }
13   public void sleep() {
14     // The body of sleep() is provided here
15     System.out.println("Zzz");
16   }
17 }
```

golang では Interface の実装は特にキーワードを指定せずに、 その Interface で定義しているメソッドを、Implement に相当する構造体がすべて実装しているかどうかでチェックされる。これは golang はクラスを持たず、構造体を使って Interface の実装を行う為に、構造体の定義にどの Interface の実装であるかの情報をシンタックス上書けない為である。GearsOS では型定義ファイルを持つことができるために、 golang のような実行時チェックは行わず、 Java に近い形で表現したい。

導入した型定義で SynchronizedQueue を定義したものをソースコード 4.6 に示す。大まかな定義方法は Interface 定義のものと同様である。違いとして `impl` キーワードを導入した。これは Java の `implements` に相当する機能であり、 実装した Interface の名前を記述する。現状の GearsOS では Impl が持てる Interface は 1 つのみであるため、`impl` の後ろにはただ 1 つの型が書かれる。型定義の中では独自に定義した CodeGear を書いてもいい。

これは Java のプライベートメソッドに相当するものである。特にプライベートメソッドがない場合は、 実装側で所持したい変数定義を記述する。SynchronizedQueue の例では top などが実装側で所持している変数である。

ソースコード 4.6: SynchronizedQueue の定義ファイル

```
1 typedef struct SynchronizedQueue <> impl Queue {
2   struct Element* top;
3   struct Element* last;
4   struct Atomic* atomic;
5 } SynchronizedQueue;
```

従来 context.h に直接記述していたすべての DataGear の定義は、 スクリプトで機械的に Interface および Implement の型定義ファイルに変換している。

4.3 Implement の型をいれたことによる間違った Gears プログラミング

Implement の型を導入したが、 GearsOS のプログラミングをするにつれていくつかの間違ったパターンがあることがわかった。自動生成される StubCodeGear は、 goto meta から遷移するのが前提であるため、引数を Context から取り出す必要がある。Context から取り出す場合は、 実装している Interface に対応している置き場所からデータを取り出す。この置き場所は data 配列であり、 配列の添え字は enum Data と対応している。また各 CodeGear から goto する際に、 遷移先の Interface に値を書き込みに行く。

Interface で定義した CodeGear と対応している Implement の CodeGear の場合はこのデータの取り出し方で問題はない。しかし Implement の CodeGear から内部で goto する CodeGear の場合は事情が異なる。内部で goto する CodeGear は、 Java などのプライベートメソッドとして使用できる。この CodeGear のことを private CodeGear と呼ぶ。privateCodeGear に goto する場合、 goto 元の CodeGear からは goto meta 経由で遷移する。goto meta が発行されると Stub Code Gear に遷移するが、現在のシステムでは Interface から値を取得しに行く。private CodeGear の入力も Stub から取得したいと考え、 Implement を Interface のつもりで GearsOS のコードを記述した。

4.4 Interface のパーサーの構築

従来の GearsOS のトランスコンパイラでは、 generate_stub.pl が Interface ファイルを開き、情報を解析していた。この情報解析は getDataGear 関数で行われていた。しかしこの関数は、CbC ファイルの CodeGear、DataGear の解析で使用するルーチンと同じものである。従って、 Interface 特有のパースが出来ていなかった。

例えば開いたヘッダファイルが Interface のファイルでも、そうでない C のヘッダファイルでも同様の解析をしてしまう。Interface の定義ファイルの構文はすでに統一されたものを使用している。この構文で実装されていない Interface ファイルを読み込んだ場合は、エラーとして処理したい。また、Interface が満たすべき CodeGear の種類や InputDataGear の数の管理も行いたい。さらに Interface ではなく、Implement の定義ファイルも同様にパースし、情報を解析したい。

これらを実現するには、最初から Interface に特化したパーサーが必要となる。本研究では Gears::Interface モジュールとして実装した。

4.4.1 Gears::Interface の構成

4.5 Interface の実装の CbC ファイルへの構文の導入

4.6 GearsCbC の Interface の実装時の問題

Interface とそれを実装する Impl の型が決定すると、最低限満たすべき CodeGear の API は一意に決定する。ここで満たすべき CodeGear は、Interface で定義した CodeGear と、Impl 側で定義した private な CodeGear となる。例えば Stack Interface の実装を考えると、各 Impl で `pop`, push, shift, isEmpty などを実装する必要がある。

従来はプログラマが手作業でヘッダーファイルの定義を参照しながら `.cbc` ファイルを作成していた。手作業での実装のため、 コンパイル時に下記の問題点が多発した。

- CodeGear の入力のフォーマットの不一致
- Interface の実装の CodeGear の命名規則の不一致
- 実装を忘れている CodeGear の発生

特に GearsOS の場合は Perl スクリプトによって純粋な CbC に一度変換されてからコンパイルが行われる。実装の状況とトランスコンパイラの組み合わせによっては、 CbC コンパイラレベルでコンパイルエラーを発生させないケースがある。この場合は実際に動作させながら、gdb, lldb などの C デバッガを用いてデバッグをする必要がある。また CbC コンパイラレベルで検知できても、すでに変換されたコード側でエラーが出る。このため、トランスコンパイラの挙動をトレースしながらデバッグをする必要がある。Interface の実装が不十分であることのエラーは、 GearsOS レベル、最低でも CbC コンパイラのレベルで完全に検知したい。

4.7　Interfaceを満たすコード生成の他言語の対応状況

Interface を機能として所持している言語の場合、Interface を完全に見たいしているかどうかはコンパイルレベルか実行時レベルで検知される。例えば Java の場合は Interface を満たしていない場合はコンパイルエラーになる。

Interface の API を完全に実装するのを促す仕組みとして、Interface の定義からエディタやツールが満たすべき関数と引数の組を自動生成するツールがある。

Java では様々な手法でこのツールを実装している。Microsoft が提唱している IDE とプログラミング言語のコンパイラをつなぐプロトコルに Language Server がある。Language Server はコーディング中のソースコードをコンパイラ自身でパースし、 型推論やエラーの内容などを IDE 側に通知するプロトコルである。主要な Java の Language Server の実装である eclipse.jdt.ls[24] では、 LanguageServer の機能として未実装のメソッドを検知する機能が実装されている。[25] この機能を応用して vscode 上から未実装のメソッドを特定し、 雛形を生成する機能がある。他にも IntelliJ IDE などの商用 IDE では、 IDE が独自に未実装のメソッドを検知、雛形を生成する機能を実装している。

golang の場合は主に `josharian/impl`[26] が使われている。これはインストールすると `impl` コマンドが使用可能になり、 実装したい Interface の型と、 Interface を実装する Impl の型 (レシーバ) を与えることで雛形が生成される。主要なエディタである vscode の golang の公式パッケージである `vscode-go`[27] でも導入されており、 vscode から呼び出すことが可能である。vscode 以外にも vim などのエディタからの呼び出しや、 シェル上で呼び出して標準出力の結果を利用することが可能である。

4.8　GearsOSでのInterfaceを満たすCbCの雛形生成

GearsOS でも同様の Interface の定義から実装する CodeGear の雛形を生成したい。LanguageServer の導入も考えられるが、 今回の場合は C 言語の LanguageServer を CbC 用にまず改良し、 さらに GearsOS 用に書き換える必要がある。現状の GearsOS が持つシンタックスは CbC のシンタックスを拡張しているものではあるが、これは CbC コンパイラ側には組み込まれていない。LanguageServer を GearsOS に対応する場合、CbC コンパイラ側に GearsOS の拡張シンタックスを導入する必要がある。CbC コンパイラ側への機能の実装は、 比較的難易度が高いと考えらる。CbC コンパイラ側に手をつけず、 Interface の入出力の検査は既存の GearsOS のビルドシステム上に組み込みたい。

対して golang の `impl` コマンドのように、 シェルから呼び出し標準出力に結果を書き込む形式も考えられる。この場合は実装が比較的容易かつ、 コマンドを呼び出して標準出力の結果を使えるシェルやエディタなどの各プラットフォームで使用可能となる。先

行事例を参考に、コマンドを実行して雛形ファイルを生成するコマンド `impl2cbc.pl` を GearsOS に導入した。`impl2cbc.pl` の処理の概要を図 4.1 に示す。

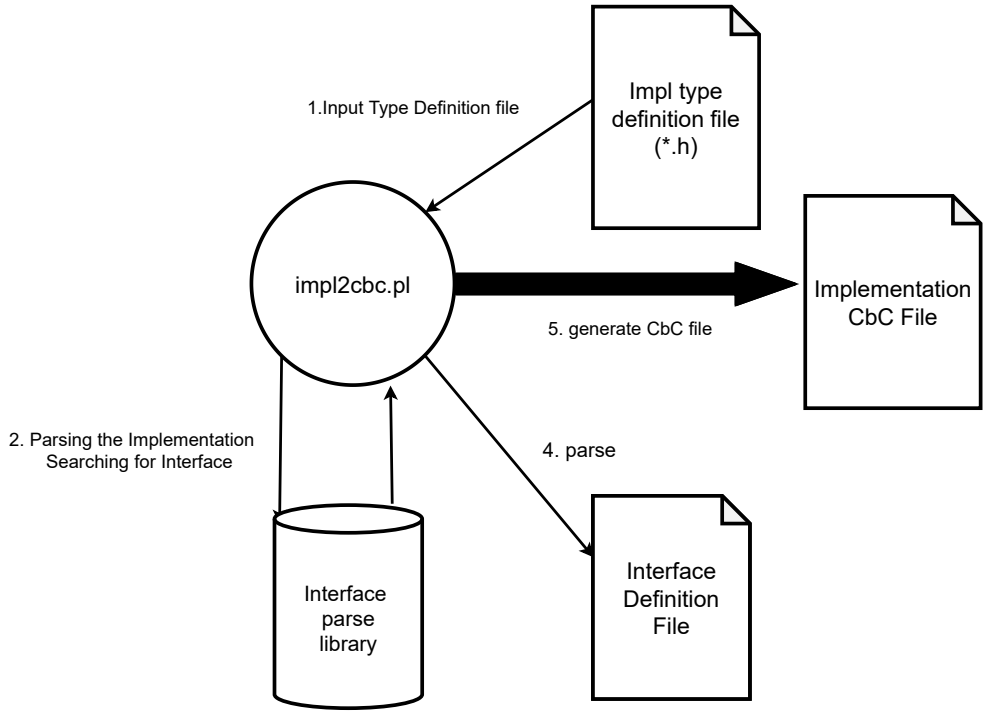

図 4.1: impl2cbc の処理の流れ

4.8.1　雛形生成の手法

Interface では入力の引数が Impl と揃っている必要があるが、 第一引数は実装自身のインスタンスがくる制約となっている。実装自身の型は、Interface 定義時には不定である。その為、 GearsOS では Interface の API の宣言時にデフォルト型変数 `Impl` を実装の型として利用する。デフォルト型 `Impl` を各実装の型に置換することで自動生成が可能となる。

実装すべき CodeGear は Interface と Impl 側の型を見れば定義されている。`__code` で宣言されているものを逐次生成すればよいが、 継続として呼び出される CodeGear は具体的な実装を持たない。GearsOS で使われている Interface には概ね次の継続である `next` が登録されている。`next` そのものは Interface を呼び出す際に、入力として与える。その為各 Interface に入力として与えられた `next` を保存する場所は存在するが、 next そのものの独自実装は各 Interface は所持しない。したがってこれを Interface の実装側で明示的に

実装することはできない。雛形生成の際に、入力として与えられる CodeGear を生成してしまうと、プログラマに混乱をもたらしてしまう。

入力として与えられている CodeGear は、Interface に定義されている CodeGear の引数として表現されている。コードに示す例では、`whenEmpty` は入力して与えられている CodeGear である。雛形を生成する場合は、入力として与えられた CodeGear を除外して出力を行う。順序は Interface をまず出力した後に、 Impl 側を出力する。

4.8.2　コンストラクタの自動生成

雛形生成では他にコンストラクタの生成も行う。GearsOS の Interface のコンストラクタは、 メモリの確保及び各変数の初期化を行う。メモリ上に確保するのは主に Interface と Impl のそれぞれが基本となっている。Interface によっては別の DataGear を内包しているものがある。その場合は別の DataGear の初期化もコンストラクタ内で行う必要があるが、 自動生成コマンドではそこまでの解析は行わない。

コンストラクタのメンバ変数はデフォルトでは変数は 0、ポインタの場合は NULL で初期化するように生成する。このスクリプトで生成されたコンストラクタを使う場合、CbC ファイルから該当する部分を削除すると、`generate_stub.pl` 内でも自動的に生成される。自動生成機能を作成すると 1CbC ファイルあたりの記述量が減る利点がある。

明示的にコンストラクタが書かれていた場合は、Perl スクリプト内での自動生成は実行しないように実装した。これはオブジェクト指向言語のオーバーライドに相当する機能と言える。現状の GearsOS で使われているコンストラクタは、 基本は `struct Context*` 型の変数のみを引数で要求している。しかしオブジェクトを識別するために ID を実装側に埋め込みたい場合など、 コンストラクタ経由で値を代入したいケースが存在する。この場合はコンストラクタの引数を増やす必要や、 受け取った値をインスタンスのメンバに書き込む必要がある。具体的にどの値を書き込めば良いのかまでは Perl スクリプトでは判定することができない。このような細かな調整をする場合は、 generate_stub.pl 側での自動生成はせずに、 雛形生成されたコンストラクタを変更すれば良い。あくまで雛形生成スクリプトはプログラマ支援であるため、 いくつかの手動での実装は許容している。

4.9　Interface の引数の検知

GearsOS のノーマルレベルでは、Interface の API の呼び出しは `interface->method(arg)` の呼び出し方であった。`arg` は引数であり、これは Interface で定義した API の引数の一致している必要がある。Interface の定義の引数は、Impl の実装自身が第一引数でくる制約があった。この制約の為に、厳密には Interface の定義ファイルに書かれている CodeGear の引数と、Interface の呼び出しの引数は数が揃ってはいない。generate_stub.pl は第一引

数が実装自身の型であるので、union Data 型にキャストし、Context の引数保存場所に書き込むようになっている。問題が第 1 引数以外の引数が揃っていない場合である。generate_stub.pl を通すと、次の継続は goto meta に変換されてしまい、引数情報が抜けてしまう。その為引数はすべて適切に context に書き込まれている必要があるが、 一部引数が足りず書き込みが出来なかったケースでも、 CbC コンパイラレベルでは引数関係のエラーが発生しない。また上手く Interface の入力の数を取得できなかった場合も、generate_stub.pl は止まらずにマクロを生成してしまう。Gearef を通して context に書き込む右辺値が抜けているコードなどがよく発生した。この場合は原因を.c ファイルと.cbc ファイル、Interface ファイル、context ファイルのすべてを確認しなければならず、デバッグが非常に困難だった。Interface の API 呼び出し時の引数検知は、 Interface の型定義ファイルから CodeGear の入力の数の取得が不十分であるのが主な原因であった。

この問題は Perl スクリプトレベルで引数のチェックを十分に行う必要がある。すでに Interface のパーサーは実装している為、 パーサー経由で呼び出している API を持つ Interface の情報を取得する。パースした結果の情報に、 各 CodeGear の引数情報と引数の数を取得できれば、それらと API 呼び出し時に与えられている引数を比較すればチェックが可能である。現状は引数の数が揃っているかどうかで確認をしている。

Inteface の引数を確認し、Gearef マクロを生成している generate_stub.pl の箇所に、引数の確認処理を実装した。(ソースコード 4.7) ここで API 呼び出し時の引数は、`$tmpArgs` に代入されている。CbC の関数呼び出しの引数はカンマで区切るので、2 行目でカンマで文字列を分割し、引数を配列`@args` に変換している。

generate_stub.pl はローカル変数のすべての型を記録しているので、6 行目で API 呼び出しをしているインスタンスの名前から Interface を特定する。特定後、ヘッダファイルの場所を取得し、8 行目で Interface のパーサーを呼び出している。パーサーから取得した情報から、メソッドの引数の数を 14 行目で取得し、 引数が格納されている配列`@args` の要素数と比較している。

ソースコード 4.7: Perl レベルでの引数チェック

```
1  #$tmpArgs =~ s/\(.*\)/\(\)/;
2  my @args = split(/,/,$tmpArgs);
3
4  #....
5
6  my $nextType           = $currentCodeGearInfo->{localVar}->{$next}  //
       $currentCodeGearInfo->{arg}->{$next};
7  my $nextTypePath       = $headerNameToInfo->{$nextType}->{path};
8  my $parsedNextTypePath = Gears::Interface->detailed_parse($nextTypePath);
9
10 unless (exists $parsedNextTypePath->{codeName}->{$method}) {
11   die "[ERROR] not found $next definition at $_ in $filename\n";
12 }
13 my $nextMethodInfo = $parsedNextTypePath->{codeName}->{$method};
14 my $nextMethodWantArgc = $nextMethodInfo->{argc};
```

```
15
16 if ($nextMethodWantArgc != scalar(@args)) {
17   die "[ERROR] invalid arg $line  you shoud impl $nextMethodInfo->{args}\
       n";
18 }
```

Perl スクリプトでエラーを検知すると、エラーで終了する。ソースコード 4.8 の Interface の insertTest1 を呼び出す例題でエラーを発生さる。

ソースコード 4.8: StackTestInterface の定義

```
1 typedef struct StackTest <> {
2   __code insertTest1(Impl* stackTest, struct Stack* stack, __code next
      (...));
3   __code next(...);
4 } StackTest;
```

ソースコード 4.9 で API を呼び出しているが、この呼び出し方法では stack が引数にない。

ソースコード 4.9: StackTestInterface の API 呼び出し (引数不足)

```
1 __code gmain(){
2     Stack* stack = createSingleLinkedStack(context);
3     StackTest* stackTest = createStackTestImpl3(context);
4     goto stackTest->insertTest1(shutdown);
5 }
```

GearsOS のビルドを行うと、ソースコード 4.10 のエラーが発生し、以降のビルドが停止する。Cmake はエラーを検知するとビルドを止めるように Makefile を作製するため、GearsOS の拡張構文のレベルで停止ができる。

ソースコード 4.10: Interface の API 呼び出し時の引数エラー

```
1 [ 12%] Generating c/examples/pop_and_push/main.c
2 [ERROR] invalid arg     goto stackTest->insertTest1(shutdown);
3   you shoud impl Impl* stackTest, struct Stack* stack, __code next(...)
4 make[3]: *** [CMakeFiles/pop_and_push.dir/build.make:81: c/examples/
      pop_and_push/main.c] Error 25
5 make[3]: *** Deleting file 'c/examples/pop_and_push/main.c'
```

generate_stub.pl 側で、出てきたローカル変数と型の組はすべて保存している。Interface 側の CodeGear の定義にも当然引数の型と名前は書かれている。このローカル変数の型と、CodeGear の定義の引数の型が、完全に一致しているかどうかのチェックを行うと、さらに強固な引数チェックが可能となる。ただし引数で渡す際に、例えば int 型の値の加算処理などを行っていると、その処理の結果が int 型になっているかどうかを Perl レベルでチェックする必要が出てしまう。

4.10 Interface の API の未実装の検知

InterfaceAPI 呼び出し時に引数の数以外に、そもそも実装していない API を呼び出してしまうことがある。この場合は CbC が Perl スクリプトによって変換された後でエラーが出る。内容は CbC コンパイラのコンパイル時に Interface の構造体に、API に対応するフィールドがないエラーである。コンパイル時に発覚できるので問題ないが、 これも変換する前に発見したほうがデバッグが容易である。

API 呼び出し時の処理は、ソースコード 4.7 の処理そのものであるため、この処理の中に未実装の API を検知する様にした。呼び出し元の Interface の情報パースした結果、ヘッダファイルに API の定義がなかった場合は 11 行目の `unless` に処理が落ち、 エラー終了する。これによって Interface 呼び出しの問題が、 Perl スクリプトによって変換する前に検知可能になった。

4.11 par goto の Interface 経由の呼び出しの対応

第5章　トランスコンパイラによるメタ計算

GearsOS は CbC で実装を行う。CbC は C 言語よりアセンブラに近い言語である。すべてを純粋な CbC で記述すると記述量が膨大になる。またノーマルレベルの計算とメタレベルの計算を、全てプログラマが記述する必要がでる。メタ計算では値の取り出しなどを行うが、 これはノーマルレベルの CodeGear の API が決まれば一意に決定される。したがってノーマルレベルのみ記述すれば、 機械的にメタ部分の処理は概ね生成可能となる。また、メタレベルのみ切り替えたいなどの状況が存在する。ノーマルレベル、メタレベル共に同じコードの場合は記述の変更量が膨大であるが、 メタレベルの作成を分離するとこの問題は解消される。

GearsOS ではメタレベルの処理の作成に Perl スクリプトを用いており、ノーマルレベルで記述された CbC から、メタ部分を含む CbC へと変換する。変換前の CbC を GearsCbC と呼ぶ。

5.1　トランスコンパイラ

プログラミング言語から実行可能ファイルやアセンブラを生成する処理系のことを、一般的にコンパイラと呼ぶ。特定のプログラミング言語から別のプログラミング言語に変換するコンパイラのことを、 トランスコンパイラと呼ぶ。トランスコンパイラとしては JavaScript を古い規格の JavaScript に変換する Babel[28] がある。

またトランスコンパイラは、変換先の言語を拡張した言語の実装としても使われる。JavaScript に強い型制約をつけた拡張言語である TypeScript は、 TypeScript から純粋な JavaScript に変換を行うトランスコンパイラである。すべての TypeScript のコードは JavaScript にコンパイル可能である。JavaScript に静的型の機能を取り込みたい場合に使われる言語であり、 JavaScript の上位の言語と言える。

GearsOS は CbC にノーマルレベル、メタレベルの書き別けの機能などを追加した拡張言語であると言える。コンパイル時に CMake によって呼び出される 2 種類の Perl スクリプトで等価な純粋な CbC に変換される。これらの Perl スクリプトは GearsOS の CbC から純粋な CbC へと変換している為に一種のトランスコンパイラと言える。

5.2　トランスコンパイラによるメタレベルのコード生成

トランスコンパイラはノーマルレベルで記述された GearsOS を、メタレベルを含む CbC へと変換する役割である。変換時に様々なメタ情報を CbC のファイルに書き出すことが可能である。従来は Stub の生成や、 引数の変更などを行っていたが、 さらにメタレベルのコードをトランスコンパイラで作製したい。トランスコンパイラ上でメタレベルのコードを作製することによって、 GearsOS 上でのアプリケーションの記述が容易になり、かつメタレベルのコードを柔軟に扱うことができる。本研究では様々なメタレベルのコードを、トランスコンパイラで生成することを検討した。

5.3　トランスコンパイラ用の Perl ライブラリ作製

従来の Perl トランスコンパイラは generate_stub.pl と generate_context.pl の 2 種類のスクリプトで構築されていた。これらのスクリプトはそれぞれ独立した処理を行っていた。

しかし本研究を進めるにつれて、Interface のパーサーやメタ計算部分の操作を行う API など、Perl スクリプトで共通した実装が見られた。さらに generate_stub.pl ら Perl スクリプトの行数や処理の複雑度が上がり、 適切に処理をモジュール化する必要が生じた。この為新しく実装した Perl トランスコンパイラが利用する API は、 Perl のモジュール機能を利用しモジュールの形で実装した。以下に実装したモジュールファイルと、その概要を示す。

- `Gears::Context`
  - context.h の自動生成時に呼び出されるモジュール
  - 変換後の CbC のコードを解析し、使用されている DataGear の数え上げを行う
- `Gears::Interface`
  - Interface および Implement のパーサー
- `Gears::Template` Gears::Template 以下は Perl スクリプトが生成する際に、テンプレートとして呼び出すファイルの定義などがある
  - `Gears::Template::Context`
    - context.h のテンプレート
  - `Gears::Template::Context::XV6`
    - CbC Xv6 専用の context.h のテンプレート

  - `Gears::Template::Gmain`
    - GearsOS Main 関数のテンプレート

- `Gears::Stub`
  - Stub Code Gear 生成時に呼び出されるモジュール

これらは generate_stub.pl および generate_context.pl および、本研究で作製した Perl のツールセットからも呼び出される。

5.4　context.h の自動生成

GearsOS の Context の定義は context.h にある。Context は GearsOS の計算で使用されるすべての CodeGear、DataGear の情報を持っている。context.h では DataGear に対応する `union Data` 型の定義も行っている。Data 型は C の共用体であり、 Data を構成する要素として各 DataGear がある。各 DataGear は構造体の形で表現されている。各 DataGear 自体の定義も context.h の union Data の定義の中で行われている。

DataGear の定義は Interface ファイルで行っていた。Interface ファイルは GearsOS 用に拡張されたシンタックスのヘッダファイルを使っており、 直接 CbC からロードすることができない。その為従来はプログラマが静的に Interface ファイルを CbC の文脈に変換し、 context.h に構造体に変換したものを書いていた。この手法では手書きでの構築のために自由度は高かったが、 GearsOS の例題によっては使わない DataGear も、 context.h から削除しない限り context に含んでしまう問題があった。さらに Interface ファイルで定義した型を context.h に転記し、それをもとに Impl の型を考えて CbC ファイルを作製する必要があった。これらをすべてユーザーが行うと、ファイルごとに微妙な差異が発生したりとかなり煩雑な実装を要求されてしまう。DataGear の定義は Interface ファイルを作製した段階で決まり、 使用している DataGear、CodeGear はコンパイル時に確定する。使用している各 Gear がコンパイル時に確定するならば、 コンパイルの直前に実行される Perl トランスコンパイラでも Gear の確定ができるはずである。ここから context.h をコンパイルタイミングで Perl スクリプト経由で生成する手法を考案した。

5.4.1　context.h の作製フロー

GearsCbC からメタ計算を含む CbC ファイルに変換する generate_stub.pl は各 CbC ファイルを 1 つ 1 つ呼び出していた。context.h を生成しようとする場合、 プロジェクトで利用する全 CbC ファイルを扱う必要がある。

Context の初期化ルーチンを作製する generate_context.pl は、その特性上すべての CbC ファイルをロードしていた。したがって context.h を作製する場合はこのスクリプトで行うと現状の CMake に手をつけずに変更ができる。

5.4.2　context.h のテンプレートファイル

Perl のモジュールとして `Gears::Template::Context` を作製した。xv6 プロジェクトの場合は一部ヘッダファイルに含める情報が異なる。

派生モジュールとして `Gears::Template::Context::XV6` も実装した。これらのテンプレートモジュールは generate_context.pl の実行時のオプションで選択可能である。

呼び出しには Perl の動的モジュールロード機能を利用している。各モジュールに共通の API を記述しており、 テンプレートに限らず共通して呼び出すことが可能である。

5.5　メタ計算部分の入れ替え

GearsOS では次の CodeGear に移行する前の MetaCodeGear として、 デフォルトでは `__code meta` が使われている。`__code meta` は context に含まれている CodeGear の関数ポインタを、 enum からディスパッチして次の Stub CodeGear に継続するものである。

例えばモデル検査を GearsOS で実行する場合、 通常の Stub CodeGear のほかに状態の保存などを行う必要がある。この状態の保存に関する一連の処理は明らかにメタ計算であるので、 ノーマルレベルの CodeGear ではない箇所で行いたい。ノーマルレベル以外の CodeGear で実行する場合は、 通常のコード生成だと StubCodeGear の中で行うことになる。StubCodeGear は自動生成されてしまうため、 値の取り出し以外のことを行う場合は自分で実装する必要がある。しかしモデル検査に関する処理は様々な CodeGear の後に行う必要があるため、 すべての CodeGear の Stub を静的に実装するのは煩雑である。これを避けるには、 Stub 以外の Meta Code Gear をユーザーが自由に定義できる必要がある。

ノーマルレベルの CodeGear の処理の後に、StubCodeGear 以外の Meta Code Gear を実行したい。Stub Code Gear に直ちに遷移してしまう `__code meta` 以外の Meta CodeGear に、 特定の CodeGear の計算が終わったら遷移したい。このためには、特定の CodeGear の遷移先の MetaCodeGear をユーザーが定義できる API が必要となる。この API を実装すると、ユーザーが柔軟にメタ計算を選択することが可能となる。これはいわゆるリフレクション処理に該当する。

GearsOS のビルドシステムの API として `meta.pm` を作製した。これは Perl のモジュールファイルとして実装した。meta.pm は Perl で実装された GearsOS のトランスコンパイラである generate_stub.pl から呼び出される。meta.pm の中のサブルーチンである `replaceMeta` に変更対象の CodeGear と変更先の MetaCodeGear への goto を記述する。ユーザーは

meta.pm の Perl ファイルを API として GearsOS のトランスコンパイラにアクセスすることが可能となる。

具体的な使用例をコード 5.1 に示す。meta.pm はサブルーチン `replaceMeta` が返すリストの中に、特定のパターンで配列を設定する。各配列の 0 番目には、goto meta を置換したい CodeGear の名前を示す Perl 正規表現リテラルを入れる。コード 5.1 の例では、`PhilsImpl` が名前に含まれる CodeGear を指定している。すべての CodeGear の goto の先を切り替える場合は `qr/.*/`などの正規表現を指定する。

ソースコード 5.1: meta.pm

```
1  package meta;
2  use strict;
3  use warnings;
4
5  sub replaceMeta {
6    return (
7      [qr/PhilsImpl/ => \&generateMcMeta],
8    );
9  }
10
11 sub generateMcMeta {
12   my ($context, $next) = @_;
13   return "goto mcMeta($context, $next);";
14 }
15
16 1;
```

generate_stub.pl は Gears CbC ファイルの変換時に、 CbC ファイルがあるディレクトリに meta.pm があるかを確認する。meta.pm がある場合はモジュールロードを行う。meta.pm がない場合は meta Code Gear に goto するものをデフォルト設定として使う。これらの処理は Perl のクロージャの形で表現しており、 トランスコンパイラ側では共通の API で呼び出すことが可能である。各 Gode Gear が `goto` 文を呼び出したタイミングで replaceMeta を呼び出し、 ルールにしたがって goto 文を書き換える。変換する CodeGear がルールになかった場合は、 デフォルト設定が呼び出される。

5.6　コンパイルタイムでのコンストラクタの自動生成

5.7　Interface の API の自動保管

5.8　別 Interface からの書き出しを取得する必要がある CodeGear

従来の MetaCodeGear の生成では、 別の Interface からの入力を受け取る CodeGear の Stub の生成に問題があった。具体的なこの問題が発生する例題をソースコード 5.2 に示す。

ソースコード 5.2: 別 Interface からの書き出しを取得する CodeGear の例

```
#interface "String.h"
#interface "Stack.h"

#impl "StackTest.h" for "StackTestImpl3.h"

/* 略 */

__code pop2Test(struct StackTestImpl3* stackTest, struct Stack* stack,
    __code next(...)) {
    goto stack->pop2(pop2Test1);
}


__code pop2Test1(struct StackTestImpl3* stackTest, union Data* data,
    union Data* data1, struct Stack* stack, __code next(...)) {
    String* str = (String*)data;
    String* str2 = (String*)data1;

    printf("%d\n", str->size);
    printf("%d\n", str2->size);
    goto next(...);
}
```

この例では pop2TestCode Gear から stack->pop2 を呼び出し、 継続として pop2Test1 を渡している。pop2Test 自体は StackTest Interface であり、 stack->pop2 の stack は Stack Interface である。例題では Stack Interface の実装は SingleLinkedStack である。SingleLinkedStack の pop2 の実装をソースコード 5.3 に示す。

ソースコード 5.3: SingleLinkedStack の pop2

```
__code pop2SingleLinkedStack(struct SingleLinkedStack* stack, __code next
    (union Data* data, union Data* data1, ...)) {
    if (stack->top) {
        data = stack->top->data;
        stack->top = stack->top->next;
    } else {
        data = NULL;
    }
    if (stack->top) {
        data1 = stack->top->data;
        stack->top = stack->top->next;
    } else {
        data1 = NULL;
    }
    goto next(data, data1, ...);
}
```

pop2 はスタックから値を 2 つ取得する API である。pop2 の継続は next であり、 継続先に data と data1 を渡している。data、 data1 は引数で受けている union Data*型の変数であり、 それぞれ stack の中の値のポインタを代入している。この操作で stack から値を

2 つ取得している。

このコードを generate_stub.pl 経由でメタ計算を含むコードに変換する。変換した先のコードを 5.4 に示す。

ソースコード 5.4: SingleLinkedStack の pop2 のメタ計算

```
1  __code pop2SingleLinkedStack(struct Context *context,struct
       SingleLinkedStack* stack, enum Code next,union Data **O_data,union
       Data **O_data1) {
2    Data* data  __attribute__((unused))  = *O_data;
3    Data* data1  __attribute__((unused))  = *O_data1;
4      if (stack->top) {
5          data = stack->top->data;
6          stack->top = stack->top->next;
7      } else {
8          data = NULL;
9      }
10     if (stack->top) {
11         data1 = stack->top->data;
12         stack->top = stack->top->next;
13     } else {
14         data1 = NULL;
15     }
16   *O_data = data;
17   *O_data1 =  data1;
18     goto meta(context, next);
19 }
20
21
22 __code pop2SingleLinkedStack_stub(struct Context* context) {
23   SingleLinkedStack* stack = (SingleLinkedStack*)GearImpl(context, Stack,
       stack);
24   enum Code next = Gearef(context, Stack)->next;
25   Data** O_data = &Gearef(context, Stack)->data;
26   Data** O_data1 = &Gearef(context, Stack)->data1;
27   goto pop2SingleLinkedStack(context, stack, next, O_data, O_data1);
28 }
```

実際は next は goto meta に変換されてしまう。data、data1 は goto meta の前にポインタ変数 O_data が指す値にそれぞれ書き込まれる。O_data は pop2 の Stub CodeGear である pop2SingleLinkedStack_stub で作製している。つまり O_data は context 中に含まれている Stack Interface のデータ保管場所にある変数 data のアドレスである。pop2 の API を呼び出すと、 Stack Interface 中の data に Stack に保存されていたデータのアドレスが書き込まれる。

当初 Perl スクリプトが生成した pop2Test1 の stub CodeGear はソースコード 5.5 のものである。CodeGear 間で処理されるデータの流れの概要図を図 5.1 に示す。

ソースコード 5.5: 生成された Stub

```
1  __code pop2Test1StackTestImpl3_stub(struct Context* context) {
```

```
2   StackTestImpl3* stackTest = (StackTestImpl3*)GearImpl(context,
      StackTest, stackTest);
3   Data* data = Gearef(context, StackTest)->data;
4   Data* data1 = Gearef(context, StackTest)->data1;
5   Stack* stack = Gearef(context, StackTest)->stack;
6   enum Code next = Gearef(context, StackTest)->next;
7   goto pop2Test1StackTestImpl3(context, stackTest, data, data1, stack,
      next);
8 }
```

__code pop2Test で遷移する先の CodeGear は StackInterface であり、呼び出している API は pop2 である。pop2 で取り出したデータは、上記で確認した通り Context 中の Stack Interface のデータ格納場所に書き込まれる。しかしソースコード 5.5 の例では Gearef(context, StackTest) で Context 中の StackTest Interface の data の置き場所から値を取得している。これは Interface の Impl の CodeGear は、Interface から値を取得するという GearsOS のルールの為である。現状では pop2 でせっかく取り出した値を StubCodeGear で取得できない。

ここで必要となってくるのは、 実装している Interface 以外の呼び出し元の Interface からの値の取得である。今回の例では StackTest Interface ではなく Stack Interface から data、 data1 を取得したい。どの Interface から呼び出されているかは、 コンパイルタイムには確定できるので Perl のトランスコンパイラで Stub Code を生成したい。

別 Interface から値を取得するには別の出力がある CodeGear の継続で渡された CodeGear をまず確定させる。今回の例では pop2Test1 が該当する。この CodeGear の入力の値と、出力がある CodeGear の出力を見比べ、 出力をマッピングすれば良い。Stack Interface の pop2 は data と data1 に値を書き込む。pop2Test1 の引数は data, data1, stack であるので、前 2 つに pop2 の出力を代入したい。

Context から値を取り出すのはメタ計算である Stub CodeGear で行われる。別 Interface から値を取り出そうとする場合、 すでに Perl トランスコンパイラが生成している Stub を書き換える方法も取れる。しかし StubCodeGear そのものを、 別 Interface から値を取り出すように書き換えてはいけない。これは別 Interface の継続として渡されるケースと、次の goto 先として遷移するケースがあるためである。前者のみの場合は書き換えで問題ないが、 後者のケースで書き換えを行ってしまうと Stub で値を取り出す先が異なってしまう。どのような呼び出し方をしても対応できるようにするには、 Stub を別に別ける必要がある。

GearsOS では継続として渡す場合や、 次の goto 文で遷移する先の CodeGear はノーマルレベルでは enum の番号として表現されていた。enum が降られる CodeGear は、厳密には CodeGear そのものではなく Stub CodeGear に対して降られる。StubCodeGear を実装した分だけ enum の番号が降られるため、 goto meta で遷移する際に enum の番号さえ合わせれば独自定義の Stub に継続させることが可能である。別 Interface から値を取り出

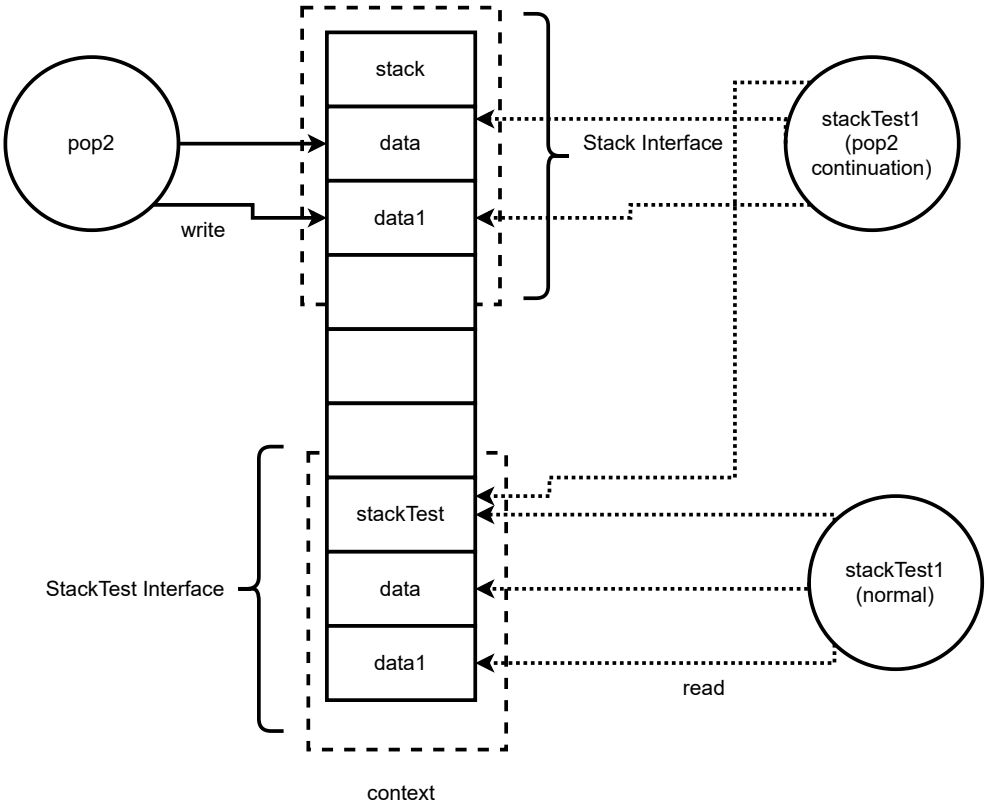

図 5.1: stackTest1 の stub の概要

したいケースの場合、 取り出してくる先の Interface と呼び出し元の CodeGear が確定したタイミングで別の StubCodeGear を生成する。呼び出し元の CodeGear が継続として渡す StubCodeGear の enum を、独自定義した enum に差し替えることでこの問題は解決する。この機能を Perl のトランスコンパイラである `generate_stub.pl` に導入した。

5.9　別 Interface からの書き出しを取得する Stub の生成

別 Interface からの書き出しを取得する場合、 generate_stub.pl では次の点をサポートする機能をいれれば実現可能である。

- goto 先の CodeGear が出力を持つ Interface でかつ継続で渡している CodeGear が別 Interface の場合の検知
  - この場合は goto している箇所で渡している継続の enum を、新たに作製した stub の enum に差し替える

- 継続で実行された場合に別に Interface から値をとってこないといけない CodeGear 自身
  - Stub を別の Interface から値をとる実装のものを別に作製する

generate_stub.pl 内では変換対象の CbC のソースコードを 2 度読み込む。最初の読み込み時に継続の状況を確認し、 2 度目の読み込み時に状況を踏まえてコードを生成すれば良い。初回の読み込み時に Interface 経由の goto 文があった場合に、別 Interface からの出力があるかなどの情報を確認したい。

5.9.1　初回 CbC ファイル読み込み時の処理

Interface 経由での goto 文は goto interface->method() の形式で呼び出される。ソースコード 5.6 はこの形式で来ていた行を読み込んだタイミングで実行される処理である。

ソースコード 5.6: goto 時に使用する interface の解析

```
1  } elsif (/^(.*)goto (\w+)\-\>(\w+)\((.*)\);/) {
2    debug_print("getDataGear",__LINE__, $_) if $opt_debug;
3    # handling goto statement
4    # determine the interface you are using, and in the case of a goto
     CodeGear with output, create a special stub flag
5    my $prev = $1;
6    my $instance = $2;
7    my $method = $3;
8    my $tmpArgs = $4;
9    my $typeName = $codeGearInfo->{$currentCodeGear}->{arg}->{$instance};
10   my $nextOutPutArgs = findExistsOutputDataGear($typeName, $method);
11   my $outputStubElem = { modifyEnumCode => $currentCodeGear,
     createStubName => $tmpArgs };
12
13   if ($nextOutPutArgs) {
14     my $tmpArgHash = {};
15     for my $vname (@$nextOutPutArgs) {
16       $tmpArgHash->{$vname} = $typeName;
17     }
18
19     $outputStubElem->{args} = $tmpArgHash;
20
21     #We're assuming that $tmpArgs only contains the name of the next
     CodeGear.
22     #Eventually we need to parse the contents of the argument. (eg.
     @parsedArgs)
23     my @parsedArgs = split /,/ , $tmpArgs; #
24
25     $generateHaveOutputStub->{counter}->{$tmpArgs}++;
26     $outputStubElem->{counter} = $generateHaveOutputStub->{counter}->{
     $tmpArgs};
```

```
27        $generateHaveOutputStub->{list}->{$currentCodeGear} =
      $outputStubElem;
28     }
```

1 行目の正規表現は Interface 経由での goto 文の正規表現パターンである。変数$instance は Interface のインスタンスである。正規表現パターンでは interface->method の->の前に来ている変数名に紐づけられる。変数$method は goto 先の Interface の API である。正規表現パターンでは interface->method の->の後に来ている API 名である。ソースコード 5.2 の pop2Test では、 stack->pop2 の呼び出しをしているため、 stack がインスタンスであり、 pop2 が API である。現在解析している goto 文が含まれている CodeGear の名前は、変数$currentCodeGear で別途保存している。連想配列である$codeGearInfo の中には、 各 CodeGear で使われている変数と変数の型などの情報が格納されている。ソースコード 5.6 の 9 行目では、 $codeGearInfo 経由で Interface のインスタンスから、具体的にどの型が呼ばれているかを取得する。pop2Test では、 インスタンス stack に対応する型名は Stack と解析される。

ソースコード 5.6 の 10 行目で実行されている findExistsOutputDataGear は generate_stub.pl 内の関数である。これは Interface の名前とメソッド名を与えると、 Interface の定義ファイルのパース結果から出力の有無を確認する動きをする。出力がある場合は出力している変数名の一覧を返す。ソースコード 5.2 の例では pop2 は data と data1 を出力している為、 これらがリストとして関数から返される。出力がない場合は偽値を返すために 13 行目からの if 文から先は動かない。出力があった場合は generate_stub.pl の内部変数に出力する変数名と、 Interface の名前の登録を行う。生成する Stub は命名規則は、Stub の本来の CodeGear の名前の末尾に_に続けて数値をいれる。__code CodeGearStub の場合は、 __code CodeGearStub_1 となる。この数値は変換した回数となるため、 この回数の計算を行う。

27 行目で$generateHaveOutputStub の list 要素に現在の CodeGear の名前と、 出力に関する情報を代入している。現在の CodeGear の名前を保存しているのは、この後のコード生成部分で enum の番号を切り替える必要があるためである。ソースコード 5.2 の例では pop2Test が使う enum を書き換える必要がある為、 ここの$currentCodeGear は pop2Test となる。ここで作製した$outputStubElem は、返還後の CbC コードを生成しているフェーズで呼びされる。

5.9.2　enum の差し替え処理

ソースコード 5.7 の箇所は遷移先の enum を Perl スクリプトで生成し、 GearsOS が実行中に enum を context に書き込むコードを生成するフェーズである。

ソースコード 5.7: Gearef のコード生成部分

```
1 if ($outputStubElem && !$stub{$outputStubElem->{createStubName}."_stub
    "}->{static}) {
2   my $pick_next = "$outputStubElem->{createStubName}_$outputStubElem->{
    counter}";
3   $return_line .= "${indent}Gearef(${context_name}, $ntype)->$pName =
    C_$pick_next;\n";
4   $i++;
5   next;
6 }
```

if 文で条件判定をしているが、前者は出力があるケースかどうかのチェックである。続く条件式は GearsOS のビルドルールとして静的に書いた stub の場合は変更を加えない為に、静的に書いているかどうかの確認をしている。変数$pick_next で継続先の CodeGear の名前を作製している。CodeGear の名前は一度目の解析で確認した継続先に_とカウント数をつけている。ここで作製した CodeGear の名前を、3 行目で context に書き込む CbC コードとして生成している。

実際に生成された例題をソースコード 5.8 に示す。

ソースコード 5.8: enum の番号が差し替えられた CodeGear

```
1 __code pop2TestStackTestImpl3(struct Context *context,struct
    StackTestImpl3* stackTest, struct Stack* stack, enum Code next) {
2     Gearef(context, Stack)->stack = (union Data*) stack;
3     Gearef(context, Stack)->next = C_pop2Test1StackTestImpl3_1;
4     goto meta(context, stack->pop2);
5 }
```

5.10　ジェネリクスのサポート

第6章　評価

6.1　GearsOSの構文作製

GearsOSで使われるInterface、およびそのImplementの型定義ファイルを導入した。GearsOSでプログラミングする際に通常のC言語やJavaなどの言語の様に、まず型を作成してからプログラミングすることが可能になった。

ただし現状のGearsOSでは1ファイルに1つの型定義しかできない。アプリケーションとしてGearsOSを動かす現在の例題ではそこまで問題になっていない。しかし、CbC xv6などの実用的なアプリケーションを実装する場合は、ファイルの数が莫大になる可能性がある。1ファイル内で様々な型が定義可能になれば、より見通しの良いプログラミングが可能であると考えられる。

6.2　GearsOSのトランスコンパイラ

6.3　GearsOSのメタ計算

第7章　結論

7.1　今後の課題

謝辞

ホゲ様，フガ様ありがとうございます

参考文献

[1] Gerwin Klein, Kevin Elphinstone, Gernot Heiser, June Andronick, David Cock, Philip Derrin, Dhammika Elkaduwe, Kai Engelhardt, Rafal Kolanski, Michael Norrish, Thomas Sewell, Harvey Tuch, and Simon Winwood. sel4: Formal verification of an os kernel, 2009.

[2] Helgi Sigurbjarnarson, James Bornholt, Emina Torlak, and Xi Wang. Push-button verification of file systems via crash refinement. pp. 1–16, 2016.

[3] Haogang Chen, Daniel Ziegler, Tej Chajed, Adam Chlipala, M. Frans Kaashoek, and Nickolai Zeldovich. Using crash hoare logic for certifying the fscq file system. pp. 18–37, 2015.

[4] Ulf Norell. Dependently typed programming in agda. pp. 1–2, 2009.

[5] the coq proof assistant. *https://coq.inria.fr/*.

[6] Luke Nelson, Helgi Sigurbjarnarson, Kaiyuan Zhang, Dylan Johnson, James Bornholt, Emina Torlak, and Xi Wang. Hyperkernel: Push-button verification of an os kernel. *Proceedings of the 26th Symposium on Operating Systems Principles*, 2017.

[7] GNU Compiler Collection (GCC) Internals. *http://gcc.gnu.org/onlinedocs/gccint/*.

[8] 大城信康, 河野真治. Continuationbasedc の gcc4.6 上の実装について. 第 53 回プログラミング・シンポジウム予稿集, Vol. 2012, pp. 69–78, jan 2012.

[9] Chris Lattner and Vikram Adve. LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation. In *Proceedings of the 2004 International Symposium on Code Generation and Optimization (CGO'04)*, Palo Alto, California, Mar 2004.

[10] Kaito TOKKMORI and Shinji KONO. Implementing continuation based language in llvm and clang. *LOLA 2015*, July 2015.

[11] 外間政尊, 河野真治. Gearsos の hoare logic をベースにした検証手法. ソフトウェアサイエンス研究会, Jan 2019.

[12] 並列信頼研究室. Cbc_gcc. http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_gcc/. Accessed: 2021-02-02.

[13] 並列信頼研究室. Cbc_llvm. http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_llvm/. Accessed: 2021-02-02.

[14] Russ Cox, Frans Kaashoek, Robert Morris. xv6 a simple, unix-like teaching operating system. https://pdos.csail.mit.edu/6.828/2018/xv6/book-rev11.pdf.

[15] Eugenio Moggi. Notions of computation and monads, July 1991.

[16] Jean Yang and Chris Hawblitzel. Safe to the last instruction: Automated verification of a type-safe operating system, 2010.

[17] 河野真治, 伊波立樹, 東恩納琢偉. Code gear、data gear に基づく os のプロトタイプ. 情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS), May 2016.

[18] 坂本昂弘, 桃原優, 河野真治. 継続を用いた x.v6 kernel の書き換え. 情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS), No. 4, may 2019.

[19] 並列信頼研究室. Cbc_xv6. http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_xv6/. Accessed: 2021-02-02.

[20] J. Lions. *Lions' Commentary on UNIX 6th Edition with Source Code*. Computer classics revisited. Peer-to-Peer Communications, 1996.

[21] Zhiyi Wang. xv6-rpi. https://code.google.com/archive/p/xv6-rpi/, 2013.

[22] Raspberry Pi. https://www.raspberrypi.org.

[23] Java implements keyword. https://www.w3schools.com/java/ref_keyword_implements.asp.

[24] Eclipse jdt language server. https://github.com/eclipse/eclipse.jdt.ls.

[25] yaohaizh. Add unimplemented methods code action.

[26] josharian/impl. https://github.com/josharian/impl.

[27] golang. golang/vscode-go.

[28] Babel. `https://babeljs.io/`.

付 録A　研究会業績

A-1　研究会発表資料

- CbC を用いた Perl6 処理系 清水 隆博, 河野真治 第 60 回プログラミング・シンポジウム, Jan, 2019
- 継続を基本とした OS Gears OS 清水 隆博, 河野真治 第 61 回プログラミング・シンポジウム, Jan, 2020
- xv6 の構成要素の継続の分析 清水 隆博，河野 真治 (琉球大学), 情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS), May, 2020

CbCを用いたPerl6処理系

清水 隆博[1,a)]　河野 真治[1,b)]

概要：スクリプト言語である Perl5 の後継言語として Perl6 が現在開発されている．Perl6 は設計と実装が区分されており様々な処理系が開発されている．現在主流な Perl6 は Rakudo と言われるプロジェクトである．Rakudo では Perl6 自体を NQP(NotQuitPerl) と言われる Perl6 のサブセットで記述し, NQP を VM が解釈するという処理の流れになっている．この VM は任意の VM が選択できるようになっており, 主に利用されている VM に C で書かれた MoarVM が存在する．MoarVM は JIT コンパイルなどをサポートしているが, 全体的な起動時間及び処理速度が Perl5 と比較し非常に低速である．この問題を解決するために Continuation based C (CbC) という言語を一部用いて MoarVM の書き換えを行う．CbC は C よりも細かな単位で記述が可能である為, 言語処理系の実装に適していると考えられる．CbC に関する今までの研究においては, 言語処理系に CbC を利用した事例が少ない．その為, 本稿では CbC を言語処理系に用いた場合の利点やデバッグ手法などについても述べる．

キーワード：プログラミング言語, コンパイラ, CbC, Perl6, MoarVM

1. はじめに

当研究室では Continuation Based C(以下 CbC) という言語を開発している．CbC は C よりきめ細やかな単位で実装する事が可能である為, 言語処理系に応用すれば効率的な開発，実行が出来ると期待される．現在活発に開発が進んでいる言語に Perl6 がある．Perl6 は MoarVM と呼ばれる VM を中心とした Rakudo と呼ばれる実装が現在の主流となっている．Rakudo は処理速度が他のプログラミング言語と比較しても非常に低速である．その為, 現在日本国内では Perl6 を実務として利用するケースは概ね存在しない．Perl6 の持つ言語機能や型システムは非常に柔軟かつ強力であるため, 実用的な処理速度に達すれば, 言語の利用件数が向上することが期待される．その為本稿では, CbC を用いた言語処理系の実装の一例として MoarVM を CbC で書き換えた CbCMoarVM を提案する．

CbC を MoarVM の実装として利用した場合, CbC の持つ機能によって MoarVM の高速化を中心とした改良に有益な効果があると推測出来る．また, 現在までの CbC を用いた研究においては言語処理系への応用例が少ない．従って, 本稿は CbC をスクリプト言語の実装に適応した場合, どのような利点やプログラミング上の問題点に遭遇するか, CbC の応用としての側面でも行う．この際に CbC を用いた言語処理系のデバッグを行う際には, CbC を使わずに記述されたオリジナルの言語処理系との並列デバッグが必要となる．従って MoarVM に CbC を適応した場合, どのようにすれば並列デバッグが行えるかについても述べる．本稿ではまず CbC, Perl6 の特徴及び現在の実装に

[1] 琉球大学工学部情報工学科
[a)] anatofuz@cr.ie.u-ryukyu.ac.jp
[b)] kono@ie.u-ryukyu.ac.jp

ついて述べ, CbC で書き換えた MoarVM についてデバッグ手法も含め解説する. 研究にあたり, 得られた CbC を言語処理系に適応した場合の利点と欠点について述べ, 今後の展望について記載する.

2. CbC

2.1 CbC の概要

CbC は当研究室で開発しているプログラミング言語である．C レベルでのプログラミングを行う場合, 本来プログラマが行いたい処理の他に malloc などを利用したメモリのアロケートやエラーハンドリングなどを記述する必要がある．これらの処理を meta computation と呼ぶ．これら meta computation と通常の処理を分離することでバグの原因が meta computation 側にあるか処理側にあるかの分離などが可能となる．しかし C 言語などを用いたプログラミングで meta computation の分離を行おうとすると, それぞれ事細かに関数やクラスを分割せねばならず容易ではない．CbC では関数より meta computation を細かく記述する為に CodeGear という単位を導入した．また CodeGear の実行に必要なデータを DataGear という単位で受け渡す．CbC では CodeGear, DataGear を基本単位として記述するプログラミングスタイルを取る.

2.2 CodeGear と DataGear

CbC では C の関数の代わりに CodeGear を導入する．CodeGear は C の関数宣言の型名の代わりに__code と書くことで宣言できる．__code は CbC コンパイラの扱いは void と同じ型であるが, CbC プログラミングでは CodeGear である事を示す識別子としての意味で利用する．CodeGear 間の移動は goto 文によって記述する.

```
extern int printf(const char*,...);

int main (){
    int data = 0;
    goto cg1(&data);
}

__code cg1(int *datap){
    (*datap)++;
    goto cg2(datap);
}

__code cg2(int *datap){
    (*datap)++;
    printf("%d\n",*datap);
}
```

Code 1: cbc_example.cbc

Code1 に示す CbC のコードでは main 関数から cg1, cg2 に遷移し, 最終的に data の値が 2 となる．CodeGear 間の入出力の受け渡しは引数を利用し行う.

ある CodeGear の実行に必要なデータを, DataGear と呼ぶ．DataGear には CodeGear で実行される関数や変数などの情報を含む．Code1 に示す例では, CodeGear に渡す引数 datap が, 一種の DataGear と言える.

2.3 軽量継続

CbC では次の CodeGear に移行する際, C の goto 文を利用する．通常の C の関数呼び出しの場合, スタックポインタを操作しローカル変数などをスタックに保存する．CbC の場合スタックフレームを操作せず, レジスタの値を変更せずそのまま次の CodeGear に遷移する事が可能である．通常 Secheme の call/cc などの継続は現在の位置までの情報を環境として所持した状態で遷移する．対して CbC は環境を持たず遷移する為, 通常の継続と比較して軽量であることから軽量継続であると言える．CbC は軽量継続を利用するためレジスタレベルでのきめ細やかな実装が可能となっている.

2.4 現在の実装

CbC は現在主要な C コンパイラである gcc 及び llvm をバックエンドとした clang 上の 2 種類の実装が存在する．gcc はバージョン 9.0.0 に, clang は 7.0.0 に対応している.

2.5 CbC と C の互換性

CbC コンパイラはコンパイル対象のソースコードが CbC であるかどうかを判断する．この際に

CodeGear を利用していない場合は通常の C プログラムとして動作する．その為今回検証する MoarVM のビルドにおいても CbC で書き換えたソースコードがある MoarVM と, 手を加えていないオリジナルの MoarVM の 2 種類を同一の CbC コンパイラでビルドする事が可能である．

また C から CbC への遷移時に, 再び C の関数に戻るように実装したい場合がある．その際は環境付き goto と呼ばれる手法を取る．これは_CbC_return 及び_CbC_environment という変数を使用する．この変数は_CbC_return が元の環境に戻る際に利用する CodeGear を指し, _CbC_environment は復帰時に戻す元の環境である．復帰する場合, 呼び出した位置には帰らず, 呼び出した関数の終了する位置に帰る．

```
__code cg(__code (*ret)(int,void *),void *env
    ){
  goto ret(1,env);
}

int c_func(){
  goto cg(_CbC_return,_CbC_environment);
  return -1;
}

int main(){
  int test;
  test = c_func();
  printf("%d\n",test);
  return 0;
}
```

Code 2: 環境付き継続の例

Code2 に示す例では c_func から環境付き継続で cg に継続している．通常 c_func の返り値は-1 であるが, cg から環境付き継続で main に帰る為に cg から渡される 1 が test の値となる．

2.6 言語処理系における CbC の応用

CbC を言語処理系, 特にスクリプト言語に応用すると幾つかの箇所に置いて利点がある．CodeGear はコンパイラの基本ブロックに相当する．その為従来のスクリプト言語では主に case 文で記述していた命令コードディスパッチの箇所を CodeGear の遷移として記述する事が可能である．通常の言語処理系では命令コードディスパッチ部分は巨大な case 文となり, この部分を実装した C ファイルが巨大化してしまう．CodeGear を導入することで巨大な case 文を CodeGear として分割する事が可能となり, 処理のモジュール化が可能となる．また, CodeGear と CodeGear 間の遷移は軽量継続で行われる為, レジスタレベルでの最適化も可能となる．

CbC は状態を単位として記述が可能であるため, 命令コードなどにおける状態を利用するスクリプト言語の実装は応用例として適していると考えられる．

3. Perl6 の概要

この章では現在までの Perl6 の遍歴及び Perl6 の言語的な特徴について記載する．

3.1 Perl6 の構想

Perl6 は 2002 年に LarryWall が Perl を置き換える言語として設計を開始した．Perl5 の言語的な問題点であるオブジェクト指向機能の強力なサポートなどを取り入れた言語として設計された．Perl5 は設計と実装が同一であり, Larry らによって書かれた C 実装のみだった．Perl6 は設計と実装が分離している．言語的な特徴としては, 独自に Perl6 の文法を拡張可能な Grammar, Perl5 と比較した場合のオブジェクト指向言語としての進化も見られる．また Perl6 は漸進的型付け言語である．従来の Perl の様に変数に代入する対象の型や, 文脈に応じて型を変更する動的型言語としての側面を持ちつつ, 独自に定義した型を始めとする様々な型に, 静的に変数の型を設定する事が可能である．

Perl6 は言語仕様及び処理実装が Perl5 と大幅に異なっており, 言語的な互換性が存在しない．従って現在では Perl6 と Perl5 は別言語としての開発方針になっている．Perl6 は現在有力な処理系である Rakudo から名前を取り Raku という別名がつけられている．

3.2 Rakudo

Rakudo とは NQP, NQP に基づく Perl6 を基に

したプロジェクトである．NQP とは, 以前の Perl6 処理系である Parrot[4] で, 構想に上がった Perl6 のサブセットである．Rakudo が Perl6 のコンパイラかつインタプリタであると考えても良い．Rakudo は図 1 に示す構成になっている．Rakudo におけるコンパイラとは厳密には 2 種類存在する．まず第 1 のものが Perl6, もしくは NQP を MoarVM, JVM のバイトコードに変換する NQP コンパイラである．次にその NQP が出力したバイトコードをネイティブコードに変換する VM の 2 種類である．この VM は現在 MoarVM, JavaVM を選択可能である．Rakudo 及び NQP project ではこの NQP コンパイラの部分をフロントエンド, VM の部分をバックエンド [13] と呼称している．NQP で主に書かれ, MoarVM など NQP が動作する環境で動く Perl6 のことを Rakudo と呼ぶ．Perl6 は NQP 以外にも NQP を拡張した Perl6 自身で書かれている箇所が存在し, これは NQP コンパイラ側で MoarVM が解釈可能な形へ変換を行う．

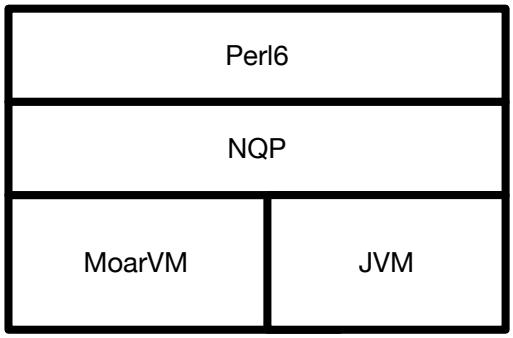

図 1: Rakudo の構成

3.3 MoarVM

MoarVM とは Rakudo で主に開発が進められている VM である．Perl6 及び NQP の専用処理系であり, レジスタマシンである．MoarVM は NQP から与えられた MoarVM のバイトコードを実行する．

MoarVM のバイトコードインタプリタは src/core/interp.c で定義されている．この中の関数 MVM_interp_run で命令に応じた処理を実行する．関数内では命令列が保存されている cur_op, 現在と次の命令を指し示す op, Thread の環境が保存されている Threadcontext などの変数を利用する．命令実行は大きく二種類の動作があり, C の goto が利用できる場合は Code3 に示す MVM_CGOTO フラグが立ちラベル遷移を利用する．goto 文が利用できない場合は巨大な case 文として命令を実行する．

ラベル遷移を利用する場合は Code4 に示すラベルテーブル LABELS にアクセスし, テーブルに登録されているアドレスを取得し, マクロ NEXT で遷移する．Code5 に示す no_op は何もせず次の命令に移動する為, goto NEXT; のみ記述されている．

```
#define NEXT_OP (op = *(MVMuint16 *)(cur_op),
      cur_op += 2, op)

#if MVM_CGOTO
#define DISPATCH(op)
#define OP(name) OP_ ## name
#define NEXT *LABELS[NEXT_OP]
#else
#define DISPATCH(op) switch (op)
#define OP(name) case MVM_OP_ ## name
#define NEXT runloop
#endif
```

Code 3: interp.c のマクロ部分

```
static const void * const LABELS[] = {
    &&OP_no_op,
    &&OP_const_i8,
    &&OP_const_i16,
    &&OP_const_i32,
    &&OP_const_i64,
    &&OP_const_n32,
    &&OP_const_n64,
    &&OP_const_s,
    &&OP_set,
    &&OP_extend_u8,
    &&OP_extend_u16,
    &&OP_extend_u32,
    &&OP_extend_i8,
    &&OP_extend_i16,
```

Code 4: ラベルテーブルの一部分

```
        DISPATCH(NEXT_OP) {
            OP(no_op):
```

```
            goto NEXT;
        OP(const_i8):
        OP(const_i16):
        OP(const_i32):
            MVM_exception_throw_adhoc(tc, "
                const_iX␣NYI");
        OP(const_i64):
            GET_REG(cur_op, 0).i64 =
                MVM_BC_get_I64(cur_op, 2);
            cur_op += 10;
            goto NEXT;
        OP(pushcompsc): {
            MVMObject * const sc = GET_REG(
                cur_op, 0).o;
            if (REPR(sc)->ID !=
                MVM_REPR_ID_SCRef)
                MVM_exception_throw_adhoc(
                    tc, "Can␣only␣push␣an␣
                    SCRef␣with␣pushcompsc")
                    ;
            if (MVM_is_null(tc, tc->
                compiling_scs)) {
                MVMROOT(tc, sc, {
                    tc->compiling_scs =
                        MVM_repr_alloc_init
                        (tc, tc->instance->
                        boot_types.
                        BOOTArray);
                });
            }
            MVM_repr_unshift_o(tc, tc->
                compiling_scs, sc);
            cur_op += 2;
            goto NEXT;
        }
    }
```

Code 5: オリジナル版 MoarVM のバイトコードディスパッチ

この為 MoarVM 内の命令コードに対応する処理は, 命令ディスパッチが書かれている C ソースファイルの, 特定の場所のみに記述せざるを得ない．その為命令コードのモジュール化などが行えず, 1 ファイル辺りの記述量が膨大になってしまう．また各命令コードに対応する処理は, ラベルジャンプもしくは switch 文に展開されてしまう為, Threaded Code の実装を考えた場合, 大幅なコードの改修が要求される．デバッグ時には, C レベルでのデバッグ時にはアドレスと実際に呼ばれる箇所を確認する事に手間がかかる．

3.4 NQP

Rakudo における NQP[6] は現在 MoarVM, JVM 上で動作する．NQP は Perl6 のサブセットであるため, 主な文法などは Perl6 に準拠しているが幾つか異なる点が存在する．NQP は最終的には NQP 自身でブートストラップする言語であるが, ビルドの最初にはすでに書かれた MoarVM のバイトコードを必要とする．この MoarVM のバイトコードの状態を Stage0 と言う.Perl6 の一部は NQP を拡張したもので書かれている為, Rakudo を動作させる為には MoarVM などの VM, VM に対応させる様にビルドした NQP がそれぞれ必要となる．現在の NQP では MoarVM, JVM に対応する Stage0 はそれぞれ MoarVM のバイトコード, jar ファイルが用意されている．MoarVM の ModuleLoader は Stage0 にある MoarVM のバイトコードで書かれた一連のファイルが該当する．

Stage0 にあるファイルを MoarVM に与えることで, NQP のインタプリタが実行される様になっている．これは Stage0 の一連のファイルは, MoarVM のバイトコードなどで記述された NQP コンパイラのモジュールである為である．NQP のインタプリタはセルフビルドが完了すると, nqp というシェルスクリプトとして提供される．このシェルスクリプトは, ライブラリパスなどを設定して MoarVM の実行バイナリである moar を起動するものである．

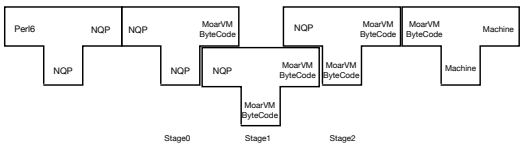

図 2: NQP のビルドフロー

NQP のビルドフローを図 2 に示す．Rakudo による Perl6 に処理系は NQP における nqp と同様に, moar にライブラリパスなどを設定した perl6 というシェルスクリプトである．この perl6 を動かすためには self build した NQP コンパイラが必要となる．その為に Stage0 を利用して Stage1 をビルドし NQP コンパイラを作成する．Stage1 は

中間的な出力であり, 生成された NQP ファイルは Stage2 と同一であるが, MoarVM のバイトコードが異なる．Perl6 では完全なセルフコンパイルを実行した NQP が要求される為, Stage1 を利用してもう一度ビルドを行い Stage2 を作成する．

Perl6 のテストスイートである Roast[10] やドキュメントなどによって設計が定まっている Perl6 とは異なり NQP 自身の設計は今後も変更になる可能性が開発者から公表されている．現在の公表されている NQP のオペコードは NQP のリポジトリ [7] に記述されているものである．

3.5 処理速度

現在の Perl6 が他のプログラミング言語と比較した場合どのような違いがでるのか計測した．macOS の/var/log/system.log ファイルから正規表現でログ中のプログラムが書き込んだ回数を個別に数え上げるというものである．今回はファイルを 231K と 3GB の二種類用意し, どの様な違いが出るのか測定した．

測定した環境は次の通りである．今回は現在広く使用されているスクリプト言語である Perl5 を計測対象に追加した．また Rakudo の処理系による処理時間の差を計測する為に MoarVM, JVM に構築した Perl6 の処理速度を計測を行った．JVM 自体の処理時間と Rakudo を構築した JVM の速度の差を見るために, 同様のプログラムを Java10 でも行った．

- Perl6 (MoarVM) ver.2018.04.01
- Perl6 (JVM) 2018.06-163-g612d071b8 built on JVM
- Java 10
- Perl5

測定した結果を表 1 に示す． 測定結果の単位は秒である．

FileSize	MoarVM	Perl6 on JVM	Java	Perl5
231K	0.86	21.48	0.27	0.04
3G	2331.08	1665.56	48.85	41.35

表 1: ログファイル処理時間の計測結果

計測結果からファイルサイズが小さい場合は MoarVM より JVM に乗せた Perl6 が低速であるが, ファイルサイズが大きい場合は Java の JIT が働くため MoarVM より高速に動いていると推測できる．

4. CbC による MoarVM

この章では改良を行った Perl6 処理系である MoarVM について述べる．今回改良を行った MoarVM は 2018.04.01 であり, 利用した NQP は 2018.04-3-g45ab6e3 バージョンである．

4.1 方針

MoarVM の中心は, バイトコードを解釈する, バイトコードインタプリタ部分である．その為 CbC を用いて, MoarVM のバイトコードインタプリタ部分を記述し直し, CbCMoarVM として実装する．CbC の CodeGear はコンパイラの基本ブロックに該当する．従って MoarVM における基本ブロックの箇所を CodeGear に書き換える事が可能である．

4.2 MoarVM のバイトコードのディスパッチ

interp.c では命令コードのディスパッチはマクロを利用した cur_op の計算及びラベルの遷移, もしくはマクロ DISPATCH が展開する switch 文で行われていた．このディスパッチ方法では, ラベルジャンプや巨大な case 文として記述する必要があり, ファイルが冗長になるなどの問題が生じる．

CbCMoarVM ではこの問題を解決するために, それぞれの命令に対応する CodeGear を作成し, 各 CodeGear の名前を要素として持つ CbC の CodeGear のテーブルを作成した．この CodeGear のテーブルを参照する CodeGear は cbc_next であり, この中のマクロ NEXT は interp.c のマクロ NEXT を CbC 用に書き直したものである．

```
#define NEXT_OP(i) (i->op = *(MVMuint16 *)(i
    ->cur_op), i->cur_op += 2, i->op)

#define DISPATCH(op) {goto (CODES[op])(i);}
#define OP(name) OP_ ## name
#define NEXT(i) CODES[NEXT_OP(i)](i)
static int tracing_enabled = 0;
```

```
__code cbc_next(INTERP i){
    goto NEXT(i);
}
```

Code 6: CbCMoarVM のバイトコードディスパッチ

Code6 に示す変更例では, マクロ NEXT などの引数に変数 i を導入している．この i とは, バイトコードインタプリタ内で利用する MoarVM のレジスタ情報などが, 格納された, 構造体へのポインタである．i が示す構造体 INTER, 及び i の型であるポインタ INTERP は Code7 に示すように宣言している．これはマクロ内部で現在の命令を示す op や命令列 cur_op などにアクセスする必要があるが, CbC の CodeGear を適応した場合に元のマクロの記述方法ではアクセスできない為に導入したものである．

```
typedef struct interp {
    MVMuint16 op;
    /* Points to the place in the bytecode
        right after the current opcode. */
    /* See the NEXT_OP macro for making sense
         of this */
    MVMuint8 *cur_op;

    /* The current frame's bytecode start. */
    MVMuint8 *bytecode_start;

    /* Points to the base of the current
        register set for the frame we
     * are presently in. */
    MVMRegister *reg_base;

    /* Points to the current compilation unit
        . */
    MVMCompUnit *cu;

    /* The current call site we're
        constructing. */
    MVMCallsite *cur_callsite;

    MVMThreadContext *tc;
} INTER,*INTERP;
```

Code 7: MoarVM の情報を格納した構造体 INTER

4.3 命令実行箇所の CodeGear への変換

ラベルテーブルや case 文の switch 相当の命令実行箇所を CbC に変換し, CodeGear の遷移として利用する．interp.c は Code5 に示す様にマクロ OP を利用して記述されている．

OP(.*) の.*に該当する箇所はバイトコードの名前である．通常このブロックには LABEL から遷移, または swtich-case によって分岐する為, バイトコードの名前は配列 LABELS の添字に変換されている．そのため対象となる CodeGear を LABELS の並びと対応させ, Code8 に示す CodeGear の配列 CODES として設定すれば CodeGear の名前は問わない．今回は CodeGear である事を示す為に接頭辞として cbc_をつける．

```
__code (* CODES[])(INTERP) = {
  cbc_no_op,
  cbc_const_i8,
  cbc_const_i16,
  cbc_const_i32,
  cbc_const_i64,
  cbc_const_n32,
  cbc_const_n64,
  cbc_const_s,
  cbc_set,
  cbc_extend_u8,
  cbc_extend_u16,
```

Code 8: CodeGear 配列の一部分

Code9 に示す命令の実行処理で MoarVM のレジスタである reg_base や, 命令列 cur_op などの情報を利用しているが, これらは MVM_interp_run 内のローカル変数として利用している．ラベルを利用しているオリジナル版では同一関数内であるためアクセス可能であるが, CodeGear 間の移動で命令を表現する CbC ではアクセスできない．その為 Code7 に示す様に, インタプリタの情報を集約した構造体 inter を定義する．この構造体へのポインタである INTERP 型の変数 i を CodeGear の入出力として与える．CodeGear 内では INTERP を経由することでインタプリタの各種情報にアクセスする．CodeGear 間の遷移ではレジスタの値の調整は行われない為, 入力引数を使ってレジスタマッピングを管理できる．INTERP のメンバである MoarVM のレジスタそのものをアーキテクチャのレジスタ上に乗せる事が可能となる．

命令実行中の CodeGear の遷移を図 3 に示す．

この中で実線で書かれている部分は CbC の goto 文で遷移し, 波線の箇所は通常の C の関数呼び出しとなっている.

現在の CbCMoarVM は次の命令セットのディスパッチを cbc_next が行っている. cbc_next から命令コードに対応する CodeGear に継続し, CodeGear から cbc_next に継続するサイクルが基本の流れである. CodeGear 内から C の関数は問題なく呼ぶ事が可能であるため, C の関数を利用する処理は変更せず記述する事ができる. また変換対象は switch 文であるため, break せず次の case に移行した場合に対応するように別の CodeGear に継続し, その後 cbc_next に継続するパターンも存在する.

```
__code cbc_no_op(INTERP i){
    goto cbc_next(i);
}
__code cbc_const_i8(INTERP i){
    goto cbc_const_i16(i);
}
__code cbc_const_i16(INTERP i){
    goto cbc_const_i32(i);
}
__code cbc_const_i32(INTERP i){
    MVM_exception_throw_adhoc(i->tc, "const_iX
        ␣NYI");
   goto cbc_const_i64(i);
}
__code cbc_const_i64(INTERP i){
    GET_REG(i->cur_op, 0,i).i64 =
         MVM_BC_get_I64(i->cur_op, 2);
    i->cur_op += 10;
    goto cbc_next(i);
}
__code cbc_pushcompsc(INTERP i){
    static MVMObject * sc;
    sc = GET_REG(i->cur_op, 0,i).o;
    if (REPR(sc)->ID != MVM_REPR_ID_SCRef)
        MVM_exception_throw_adhoc(i->tc, "Can␣
            only␣push␣an␣SCRef␣with␣pushcompsc
            ");
    if (MVM_is_null(i->tc, i->tc->
         compiling_scs)) {
        MVMROOT(i->tc, sc, {
            i->tc->compiling_scs =
                 MVM_repr_alloc_init(i->tc, i->
                 tc->instance->boot_types.
                 BOOTArray);
        });
    }
    MVM_repr_unshift_o(i->tc, i->tc->
         compiling_scs, sc);
    i->cur_op += 2;
    goto cbc_next(i);
}
```

Code 9: CbCMoarVM のバイトコードに対応する CodeGear

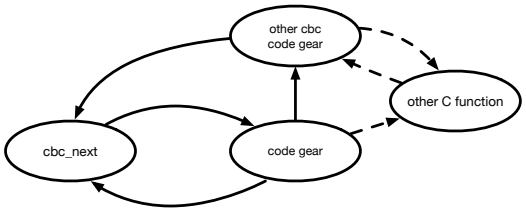

図 3: CbC における MoarVM バイトコードインタプリタ内の状態遷移

バイトコードの数は膨大である為, すべてを手作業で変換する事は望ましくない. 従って PerlScript を用いて interp.c から CbC の CodeGear を自動生成するスクリプトを作成した. このスクリプトでは以下の修正手続きを実行する.

- OP(.*) の.*部分を CodeGear の名前として, 先頭に cbc_をつけた上で設定する.
- cur_op など構造体 INTER のメンバ変数はポインタ i から参照するように修正する
- GC 対策のためマクロ MVMROOT を使い局所変数のポインタをスタックに積む箇所は, 局所変数を static 化する
- 末尾の goto NEXT を goto cbc_next(i) に修正する
- case 文で下の case 文に落ちている箇所は, case 文に対応する CodeGear に遷移する様に goto 文を付け加える

上記 Code9 では cbc_const_i8 などが case 文の下の case 部分に該当する cbc_const_i64 に遷移する様に変換されている. また cbc_pushcompsc では MVMROOT に局所変数 sc を渡している為, これを static で宣言し直している.

現在 CbC で記述された OS である GearsOS には Interface が導入されている. これは Java の interface, Haskell の型クラスに該当する概念であり,

次の CodeGear に Interface 経由で継続する事が可能である．Interface は現在の CbCMoarVM の実装には用いていないが, 今後 ThreadedCode の実装を行うにあたり命令コードディスパッチ箇所に導入を検討している．

5. MoarVM のデバッグ

MoarVM 自体のデバッグは MoarVM のリポジトリにテストコードが付随していない為単体では実行不可能である．また, CbC を用いて言語処理系の改良時を行う際に, 処理系のデバッグを行う場合は, CbC を用いないオリジナルの処理系との並列デバッグが必要となる．MoarVM 自体にはデバッグを支援するツールが存在しない為, MoarVM 自体のデバッグ方法や, CbC を用いた処理系との並列デバッグについて独自の手法を考案する必要がある．本稿では MoarVM のデバッグにおける C デバッガの使用方法と MoarVM のテスト方法についても示す．

5.1 MoarVM のバイトコードのデバッグ

MoarVM の実行バイナリである moar に対して, MoarVM のバイトコードを dump オプションを付けて読み込ませると, バイトコードが MoarVM によるアセンブリコードとして出力される．しかしこれは MoarVM が実行したバイトコードのトレースではなく, MoarVM のバイトコードを変換したものに過ぎない．また, 明らかに異なる挙動を示すオリジナルの MoarVM と, CbC で書き換えた CbCMoarVM 両者の moar を利用しても同じ結果が返ってきてしまう．そのため今回の MoarVM のバイトコードインタプリタの実装のデバッグにはこの方法は適さない．従って実際に実行した命令を確認するには gdb などの C デバッガを利用して MoarVM を直接トレースする必要がある．

CbC 側は Code10 に示す様に cbc_next に break point を設定する．オリジナル側は次のオペコードの設定のマクロにダミーの関数を呼び出すように修正し, そこに break point を設定する．CbC 側では CodeGear の名前をデバッガ上で直接確認できるが, オリジナル版は LABLE の配列の添え字から自分でどのオペコードに対応しているかをデバッガの外で探す必要がある．

添字を確認するためには Code11 に示すようにオリジナルの MoarVM の場合 cur_op の値を MVMuint16 のポインタでキャストし, これが指す値を出力する．break point を掛けているダミー関数では cur_op にアクセスする事が出来ない為, スタックフレームを一つ up する必要がある．

```
(gdb) b cbc_next
Breakpoint 2 at 0x7ffff7560288: file src/core
    /cbc-interp.cbc, line 61.
(gdb) command 2
Type commands for breakpoint(s) 2, one per
    line.
End with a line saying just "end".
>p CODES[*(MVMuint16 *)i->cur_op]
>p *(MVMuint16 *)i->cur_op
>c
>end
```

Code 10: CbCMoarVM に対しての break point 設定

```
dalmore gdb --args ../../MoarVM_Original/
    MoarVM/moar --libpath=src/vm/moar/stage0
    gen/moar/stage1/nqp
(gdb) b dummy
Function "dummy" not defined.
Make breakpoint pending on future shared
    library load? (y or [n]) y
Breakpoint 1 (dummy) pending.
(gdb) command 1
Type commands for breakpoint(s) 1, one per
    line.
End with a line saying just "end".
>up
>p *(MVMuint16 *)(cur_op)
>c
>end
```

Code 11: オリジナル版 MoarVM に対しての break point 設定

5.2 MoarVM の並列デバッグ手法

しかし MoarVM が実行する命令は膨大な数がある．その為 gdb などの C デバッガで, オリジナルの MoarVM と, 一部を CbC で記述した CbCMoarVM の並列デバッグを手動で全て行うことは困難である．Perl などのスクリプトを用いて自

動的に解析したいため, ログを残す為に script コマンドを実行した状態で gdb を起動する．トレースでは実行した命令名のみ取得できれば良い為, Code10, 11 で break point に command として設定している様に，設定された cur_op の値を出力し続けるのみの動きを導入する．

実際に実行したログ・ファイルの一部をそれぞれ Code12, 13 に示す．

```
Breakpoint 1, dummy () at src/core/interp.c
     :46
46 }
#1 0x00007ffff75608fe in MVM_interp_run (tc=0
     x604a20,
    initial_invoke=0x7ffff76c7168 <
        toplevel_initial_invoke>, invoke_data
        =0x67ff10)
    at src/core/interp.c:119
119 goto NEXT;
$1 = 159

Breakpoint 1, dummy () at src/core/interp.c
     :46
46 }
#1 0x00007ffff75689da in MVM_interp_run (tc=0
     x604a20,
    initial_invoke=0x7ffff76c7168 <
        toplevel_initial_invoke>, invoke_data
        =0x67ff10)
    at src/core/interp.c:1169
1169 goto NEXT;
$2 = 162
```

Code 12: オリジナル版 MoarVM のバイトコードのトレース

```
Breakpoint 2, cbc_next (i=0x7fffffffdc30) at
     src/core/cbc-interp.cbc:61
61 goto NEXT(i);
$1 = (void (*)(INTERP)) 0x7ffff7566f53 <
     cbc_takeclosure>
$2 = 162

Breakpoint 2, cbc_next (i=0x7fffffffdc30) at
     src/core/cbc-interp.cbc:61
61 goto NEXT(i);
$3 = (void (*)(INTERP)) 0x7ffff7565f86 <
     cbc_checkarity>
$4 = 140

Breakpoint 2, cbc_next (i=0x7fffffffdc30) at
     src/core/cbc-interp.cbc:61
61 goto NEXT(i);
$5 = (void (*)(INTERP)) 0x7ffff7579d06 <
     cbc_paramnamesused>
$6 = 558
```

Code 13: CbCMoarVM のバイトコードのトレース

オリジナル版では実際に実行する命令処理はラベルに変換されてしまう為名前をデバッガ上では出力できないが, CbC では出力する事が可能である．CbC とオリジナルの CODES, LABEL の添字は対応している為, ログの解析を行う際はそれぞれの添字を抽出し違いが発生している箇所を探索する．これらは script コマンドが作成したログを元に異なる箇所を発見するスクリプトを用意し自動化する．(Code 14)

```
131 : 131
139 : 139
140 : 140
144 : 144
558 : 558
391 : 391
749 : 749
53 : 53
*54 : 8
```

Code 14: バイトコードの差分検知の一部分

違いが生じている箇所が発見できた場合, その前後の CodeGear 及びディスパッチ部分に break point をかけ, それぞれの変数の挙動を比較する．主に cbc_return 系の命令が実行されている場合は, その直前で命令を切り替える cbc_invoke 系統の命令が呼ばれているが, この周辺で何かしらの違いが発生している可能性が高い．また主に次の CodeGear に遷移する際に CbC コンパイラのバグが生じている可能性もある為, アセンブラレベルの命令を確認しながらデバッグを進めることとなる．

5.3 MoarVM のテスト方法

MoarVM は単体で実行する事が不可能である．また NQP のリポジトリに付随するテストは NQP で書かれている．従って NQP を解釈可能な, NQP のセルフビルド時に生成されるシェルスクリプト nqp が必要である．その為, シェルスクリプト nqp を生成出来ない場合, MoarVM のテストを行

う事が出来ない．CbCMoarVM は NQP のセルフビルドが現時点では達成出来ない為, 通常ではテストが実行出来ない．しかし, MoarVM のバイナリ moar は MoarVM のバイトコードを読み込むことは NQP をセルフビルドしなくとも可能である．

その為, 正常に動作している MoarVM のバイナリ moar と nqp を用意し, この nqp 側から MoarVM のバイトコードに NQP で記述されたテストを変換する．変換された MoarVM のバイトコードはバイナリ moar に渡す事で実行可能であり, テストを行う事が出来る．

6. CbCMoarVM の利点と欠点

MoarVM の様な巨大なスクリプト言語処理系に CbC を適応した所現在までに複数の利点と欠点が発見された．本章ではまず利点を述べ, 次に現段階での CbC を適応した場合の欠点について考察する．

オリジナルの MoarVM では命令コードに対応する箇所はラベルジャンプ, もしくは switch 文で実装されていた．その為同じ C ファイルに命令コードの実行の定義が存在しなければならない．今後 MoarVM に新たなバイトコードが導入されていく事を考えると interp.c が巨大になる可能性がある．関数単位での処理の比重が偏る事に加え, switch 文中に書かれている処理は他の関数から呼ぶ事が出来ないため, 余計な手間がかかる箇所が発生すると考えられる．

CbCMoarVM の場合, CodeGear として基本ブロックを記述出来る為オリジナルの MoarVM の様に swtich 文のブロック中に書く必要性が無くなる．その為類似する命令系をコード分割し, モジュール化する事が可能である．また CbC は goto 文で遷移する以外は通常の C の関数と同じ扱いをする事が可能である．従って CodeGear 内部の処理を別の箇所から使用する事も可能となる為再利用性も向上する．

ThrededCode を実装する場合, 通常命令ディスパッチの箇所と, 実際に実行される命令処理を大幅に変更しなければならない．CbC を用いた実装の場合, 命令処理はただの CodeGear の集合である．その為 CodeGear を ThrededCode に対応した並びとして選択する事ができれば命令処理部分の修正をほぼせずに ThrededCode を実現する事が可能である．

また CodeGear はバイトコードレベルと同じ扱いができるため, ThrededCode そのものを分離して最適化をかける事が可能である．これも CodeGear が関数単位として分離できる事からの利点である．

MoarVM のバイトコードインタプリタの箇所はオリジナルの実装ではラベルジャンプを用いて実装されている．その為, 直接ラベルに break point をかける事が出来ない．作業者がデバッガが読み込んでいる C ソースコードの位置を把握し, 行番号を指定して break point を設定する必要があった．

CbCMoarVM の場合, CodeGear 単位でバイトコードの処理単位を記述している為, 通常の関数と同じく直接 CodeGear に break point をかける事が可能である．これは C プログラミングの関数に対してのデバッグで, 状態ごとに break point をかける事が出来ることを意味する．通常の C 言語で言語処理系を実装した場合と比較して扱いやすくなっていると言える．さらにラベルテーブルでの管理の場合, 次のバイトコード箇所は数値でしか確認できず, 実際にどこに飛ぶのかはラベルテーブル内と数値を作業者が手作業で確認する必要があった．スクリプトなどを組めば効率化は出来るがデバッガ上で完結しない為手間がかかる．CbC 実装では CODES テーブル内は次の CodeGear の名前が入っている為, 数値から CodeGear の名前をデバッガ上で確認する事が出来る．

現在 MoarVM は LuaJit[3] を搭載し JIT コンパイルを行っている．LuaJIT そのものを CbC に適応させるわけではないが, CbC の ABI に JIT されたコードを合わせる事が可能であると推測できる．

しかし, 言語処理系は広く使われる為に著名な OSS などを利用して開発するのが望ましいが, CbC プロジェクトの認知度が低いという現状がある．

また, CbC コンパイラが現在非常にバグを発生させやすい状態になっている．CbC コンパイラは gcc と llvm/clang 上に実装している為, これらのアップデートに追従する必要がある．しかしコン

パイラのバージョンに応じて CbC で利用するコンパイラ内の API が異なる場合が多く, API の変更に伴う修正作業などを行う必要がある．

CbCMoarVM では C から CodeGear へ, CodeGear から C への遷移などが複数回繰り返されているが, この処理中の CodeGear での tail call の強制が非常に難関である．tail call の強制には関数定義の箇所や引数, スタック領域のサイズ修正などを行う必要がある．現在の CbC コンパイラのバグでは CodeGear 内部での不要なスタック操作命令を完全に排除しきれていない．

また CodeGear から C に帰る場合, 環境付き継続を行う必要がある．C の関数の末尾で CodeGear を呼び出している場合など環境付き継続を使用しなくても良いケースは存在するが, 頻繁に C と CbC を行き来する場合記述が冗長になる可能性はある．

7. Threaded Code

現在の MoarVM は次の命令をバイトコードからディスパッチし決定後, ラベルジャンプを利用し実行している．この処理ではディスパッチの箇所にコストが掛かってしまう．CbC を MoarVM に導入することで, バイトコード列を直接サブルーチンコールの列に置き換えてしまう事が可能である．これは CbC が基本ブロックの単位と対応している為である．CbC では現在ディスパッチを行う CodeGear である cbc_next を利用しているが, Threaded Code を実装するにあたり, cbc_next と次の CodeGear に直接遷移する cbc_fixt_next の実装を予定している．

また段階的に現在 8 バイト列を 1 命令コードとして使用しているが, これを 16 バイトなどに拡張し 2 命令を同時に扱えるように実装する事なども検討している．

Perl5 においては perlcc というモジュールが開発されている．これは Perl5 内部で利用している Perl バイトコードを, Perl の C API である XS 言語の様な C のソースファイルに埋め込み, それを C コンパイルでコンパイルするというものである．perlcc を利用することで Perl インタプリタが無い状況でも可動するバイナリファイルを作成する事が可能である．しかし perlcc は Perl スクリプトが複雑になるほど正確に C に移植を行う事が出来ず, 現在では Perl のコアモジュールから外されている．perlcc は Perl のバイトコードを C への変換のみ行う為, C で実装されている Perl 経由で実行した場合と処理速度はほぼ変わらない．また perlcc で生成された C のソースコードは難解であり, これをデバッグするのが困難でもある．MoarVM で threaded code を実現出来た場合, その箇所のみ CbC プログラムとして切り出す事が可能である為 perlcc と似たツールを作成することも可能である．C 言語でも perlcc の様に内部構造を C の関数化すれば ThrededCode の様な物を構築できるが, CbC と比較して処理の単位が明確ではない為高速化は見込めない．また, CbC の CodeGear は基本ブロックそのものである為, CbC プログラムとして切り出す場合, CodeGear をそのまま出力すればよく, 生成された CbC プログラム自体も perlcc と比較して扱いやすい．CbC を用いた ThrededCode で perlcc の様なツールを作成した場合, CodeGear の単位が正常に機能すれば CbC の CodeGear が ThrededCode をより効率化出来ると推測できる．

CbC の CodeGear は goto 文で遷移するため, 次の CodeGear が一意に決定している場合 C コンパイラ側でインライン展開する事が可能である．CodeGear がインライン展開される限界については別途研究する必要があるが, CbC を利用した場合 CodeGear 単位でインライン展開が可能である．その為, ThrededCode を実装する場合に決定した次の CodeGear をインライン展開する事が可能である．従って ThreadedCode を実現するにあたり新たな処理系を開発する必要がなく, 既存の資源を利用して ThreadedCode が実現出来る．これを繰り返す事で perlcc などと比較してより高速化した ThrededCode が実現できる．

CbC を使わずにバイトコードディスパッチの箇所を改良する際に, 関数ポインタを利用する場合も考えられる．この場合は, 関数ポインタの配列を作成し, 次の命令コードに対応する関数をポインタ経由で実行する．C の関数ポインタを利用した場合, CbC と同様に処理のモジュール化は可能である．

しかし, CbC とは違い軽量継続ではなく関数呼び出しで処理をする為, C のスタックフレームが非常に巨大になる．C の関数呼び出しのコストから, 通常の case 文やラベルジャンプを利用した場合と速度差的に優位にならない．また, ThreadedCode の観点では, 命令列に対応した関数を ThreadedCode 用に大幅に修正する必要がある．その為, CbC の様に関数そのものの並びで ThreadedCode に対応させることは出来ない．

8. まとめ

本稿では CbC によって Perl6 の処理系である MoarVM インタプリタの一部改良とその手法を示した．CbCMoarVM ではオリジナルの MoarVM と比較して以下の様な利点が見られた．

- CodeGear 単位で命令処理を記述する事が可能となり, モジュール化が可能となった．
- ThreadedCode を実装する際に効率的に実装ができる見込みが立った．
- CodeGear を導入した命令単位での最適化が可能となった．
- break point を命令の処理単位でかける事が可能となった．
- 現在は命令処理部分を CodeGear に書き換えたのみである為, ラベルを利用した場合と比較して速度としては同等である．

今後 CbC での開発をより深く行っていくにあたり, CbC コンパイラそのものの信頼性を向上させる必要がある．MoarVM の開発を行うにあたり新たに発見された複数のバグを修正し, より安定するコンパイラにする為に改良を行う．

現在 CbCMoarVM で直接バイトコードを入力した場合の NQP のテストは JVM などのテストを除く中で 80%パスする．また数値の計算と出力などの簡単な NQP の例題を作成し, オリジナルの NQP,MoarVM でバイトコード化したものを入力した際も正常に動作している．しかし NQP のセルフビルドは現在オブジェクトの生成に一部失敗している為成功していない．今後はさらに複雑な例題や NQP のセルフビルド, Perl6 の動作を行っていく．

MoarVM では GC からオブジェクトを守る為に MVMROOT というマクロを利用し, 局所変数のポインタをスタックに登録する処理を行っている．GC の制御を効率的に行えば本来は必要ない処理であり, 実行すると CodeGear の優位性が損なわれてしまう．従って MoarVM の GC の最適化を行う．

また高速化という面では, Perl の特徴である正規表現に着目し, 正規表現の表現のみ高速で動く最適化の導入なども検討している．他に rakudo のコンパイラ系統から CbC のコードを直接生成させ, それを llvm でコンパイルすることによって LLVM の最適化フェーズを得て高速化することも可能であると推測できる．

Perl6 の開発は非常に活発に行われている為, CbCMoarVM の最新版の追従も課題となっている．現在は interp.c から Perl スクリプトを用いて自動で CbC の CodeGear を生成している．今後の開発領域の拡大と共により効率的に CbC コードへの自動変換も複数の C コードに対応する様に開発を行っていく．

参考文献

[1] Bell, J. R.: Threaded Code, *Commun. ACM*, Vol. 16, No. 6, pp. 370–372 (online), DOI: 10.1145/362248.362270 (1973).

[2] Ertl, A.: Threaded Code, Technische Universitat Wien (online), available from ⟨`https://www.complang.tuwien.ac.at/forth/threaded-code.html`⟩ (accessed 2018-11-21).

[3] Pall, M.: The LuaJIT Project, luajit.org (online), available from ⟨`http://luajit.org/`⟩ (accessed 2018-11-21).

[4] ParrotFoundation: Parrot, ParrotFoundation (online), available from ⟨`http://parrot.org/`⟩ (accessed 2018-11-21).

[5] Piumarta, I. and Riccardi, F.: Optimizing Direct Threaded Code by Selective Inlining, *Proceedings of the ACM SIGPLAN 1998 Conference on Programming Language Design and Implementation*, PLDI '98, New York, NY, USA, ACM, pp. 291–300 (online), DOI: 10.1145/277650.277743 (1998).

[6] ThePerlFoundation: NQP - Not Quite Perl (6), GitHub (online), available from ⟨`https://github.com/perl6/nqp`⟩ (accessed 2018-11-

21).

[7] ThePerlFoundation: NQP Opcode List, GitHub (online), available from ⟨https://github.com/perl6/nqp/blob/master/docs/ops.markdown⟩ (accessed 2018-11-21).

[8] ThePerlFoundation: Perl 6 Design Documents, ThePerlFoundation (online), available from ⟨https://design.perl6.org/⟩ (accessed 2018-11-21).

[9] ThePerlFoundation: Perl6 Documentation, ThePerlFoundation (online), available from ⟨https://docs.perl6.org/⟩ (accessed 2018-11-21).

[10] ThePerlFoundation: Roast – Perl6 test suite, GitHub (online), available from ⟨https://github.com/perl6/roast⟩ (accessed 2018-11-21).

[11] TOKUMORI, K. and KONO, S.: Implementing Continuation based language in LLVM and Clang, *LOLA* (2015).

[12] Worthington, J.: Rakudo and NQP internals, EDUMENT (online), available from ⟨http://edumentab.github.io/rakudo-and-nqp-internals-course/⟩ (accessed 2018-11-21).

[13] Worthington, J.: Rakudo and NQP internals - day1, EDUMENT (online), available from ⟨http://edumentab.github.io/rakudo-and-nqp-internals-course/slides-day1.pdf⟩ (accessed 2018-11-21).

[14] 徳森海斗，河野真治：LLVM Clang 上の Continuation based C コンパイラの改良，琉球大学工学部情報工学科平成 27 年度学位論文 (修士) (2015).

[15] 光希宮城，優　桃原，真治河野：Gears OS のモジュール化と並列 API，技術報告 11，琉球大学大学院理工学研究科情報工学専攻, 琉球大学大学院理工学研究科情報工学専攻, 琉球大学工学部情報工学科 (2018).

[16] 笹田耕一, 松本行弘, 前田敦司, 並木美太郎：Ruby 用仮想マシン YARV の実装と評価，情報処理学会論文誌プログラミング（PRO）(2006).

[17] 大城信康，河野真治：Continuation based C の GCC 4.6 上の実装について，第 53 回プログラミング・シンポジウム (2012).

[18] 唐鳳：Pugs: A Perl 6 Implementation, Hackage (online), available from ⟨http://hackage.haskell.org/package/Pugs/⟩ (accessed 2018-11-21).

[19] 並列信頼研究室：CbC_gcc, 琉球大学 (online), available from ⟨http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_gcc/⟩ (accessed 2018-11-21).

[20] 並列信頼研究室：CbC_llvm, 琉球大学 (online), available from ⟨http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_llvm/⟩ (accessed 2018-11-21).

継続を基本としたOS Gears OS

清水 隆博[1,a)]　河野 真治[2,b)]

概要：継続を基本とする C と互換性のある言語、 Conitinuation Based C(CbC) を用いて OS の実装を考案した。状態遷移単位で OS の処理を実装することで、 処理の入出力が明確化され、 定理証明支援系に適した表現形式で処理が記述可能である。現在 CbC を用いて開発している OS、 GearsOS は Xv6 をベースに実機での動作を目指している。ここでは現在の GearsOS の状況、 今後の展望について考察する。

キーワード：システムプログラミング, CbC, 軽量継続, OS, CMake

1. 証明可能な OS

コンピュータ上で動作するあらゆるソフトウェアや資源を管理する OS は、 高い信頼性が保証されてほしい。信頼性の保証にはテストプログラムを用いた検証や、 形式手法を用いた証明を使う手法が存在する。頻繁に並列処理を行う OS では、スレッド間の共通資源の競合などの非決定的な実行を行う。このため、 OS の信頼性を保証する上で、 テストやデバッグを用いる手法では、 発生しうる状態を完全に保証するのは困難である。

テストを用いる方法ではなく、 形式手法的なアプローチを用いて OS の信頼性を保証したい。そのためには定理証明支援系などで証明が可能な形式と、 等価な形式で OS を記述する必要性がある。現在開発している GearsOS は、 継続を基本とする言語 Conitinuation Based C(CbC) で実装されている。CbC は状態遷移単位での実行であり、 他の状態に遷移する際に今までの環境を持たない。CbC で実装した処理は入出力が明確化され、 定理証明支援系で表現可能な形式にする事が可能である。

[1] 琉球大学大学院理工学研究科情報工学専攻
[2] 琉球大学工学部工学科知能情報コース
[a)] anatofuz@cr.ie.u-ryukyu.ac.jp
[b)] kono@ie.u-ryukyu.ac.jp

2. Continuation Based C

Continuation Based C(CbC) とは GearsOS の記述に利用しているプログラミング言語である。C 言語の下位言語として設計されており、C コンパイラである GCC、 LLVM/Clang 上に実装が存在する。CbC は通常の関数呼び出しとは異なり、軽量継続を基本としている。通常 C の関数呼び出しでは、 call 命令により、 スタックポインタを操作し、 ローカル変数や、 レジスタ情報をスタックに保存する。CbC の軽量継続は、 アセンブラレベルでは jmp で表現され、 スタックフレームを操作することなく次の状態に遷移する。CbC の状態は CodeGear と呼ばれる単位で記述される。

3. GersOS の基本単位

実行単位としては CbC で導入された CodeGear を用いる。CodeGear は関数よりも単位が小さく、かつアセンブラよりも単位が大きく処理を記述す

ることが可能である。そのため、 OS の必要な資源管理などのメタ計算を記述するのに適していると考えられる。

GearsOS では使われる情報を、 DataGear と呼ばれる単位で構成する。DataGear は C の構造体のように宣言するが、すべての DataGear は Context と呼ばれるデータ構造の中で、 共用体として管理されている。CodeGear では入出力を DataGear で管理している。CodeGear の入力で使用される DataGear を、 InputDataGear と呼び、 出力する DataGear を OutputDataGear と呼ぶ。この入出力の組を Task として定義し、 InputDataGear の依存関係が解決された Task から、 CodeGear が並列実行される。

4. GearsOS で記述された xv6

GearsOS の機能である Context などを用いて、実際に実機で動作する OS を作成したい。実機で動作する OS のベース実装として、システムコールなどのシンプルな UNIX の機能を持つ xv6 に着目した。xv6 は ARM プロセッサを持つ RaspberryPi 上で動作する、 xv6_rpi というバリエーションが存在する。GearsOS を実行で動作させるために、xv6_rpi のソースコードを GearsOS で一部再実装している。現在は xv6 のプロセスである proc 構造体に、GearsOS の context を導入し、GearsOS としても xv6 としても解釈可能な形で開発している。

5. GearsOS のクロスコンパイル

GearsOS は RaspberryPi 上での動作を目指している。RaspberryPi は ARM の CPU が搭載されている為、 動作には ARM のバイナリファイルが必要となる。しかし RaspberryPi を利用して GearsOS 自身のビルドを行うと、マシンパワーの問題でビルドに莫大な時間が掛かってしまう。著者らが使うことが多い、 資源が潤沢な x86 マシンから、 ARM にクロスコンパイルする必要がある。GCC 上に実装している CbC コンパイラは、ARM を出力するようにコンパイラを再構築する必要があった。他方 LLVM/clang 上に実装している CbC コンパイラは、 ARM のライブラリは必要であるものの、 本体を再度ビルドすることなくクロスコンパイラとして利用可能である。今回は RaspberryPi のデフォルト OS である Raspbian から、 ARM のライブラリを x86 マシン上に転送し、LLVM/clang 上に実装した CbC コンパイラを用いてビルドした。ビルドツールとしては CMake を導入している。CMake でクロスコンパイルを行う際に、クロスコンパイラなどを引数で指定する必要がある為、 引数の解決に一部 Perl スクリプトを利用している。

6. 今後の課題

現状は xv6 を GearsOS として書き直している段階であり、 システムコールで呼び出された後の kernel 部分の処理を順次 Interface として実装している。RaspberryPi 上で動作する様にクロスコンパイルをする環境は CMake を利用して構築出来たので、 実際に RaspberryPi 上で Interface を導入した GearsOS を動作させる必要がある。また xv6 は UEFI でのブートが組み込まれていので、 これを実装したい。UEFI でブートが可能になると、各種デバイスドライバを組み込むのが容易になる為、 USB3.0 の規格である xHCI などを xv6 上に実装することが可能となる。xHCI を実装する事によって xv6 を実機で動かした際に、 USB 接続をしたキーボードが使用可能となる。これらの実装には、CbC で実装された実装としても使用可能な仕様記述言語を用いる予定である。また、 実際に xv6 上での処理を定理証明支援系などで証明を行い、 証明しやすい実装と、 処理に適した実装に Interface を通して切り替える機構を実装することも課題である。

参考文献

[1] 宮城光希，桃原　優，河野真治：Gears OS のモジュール化と並列 API，技術報告 11，琉球大学大学院理工学研究科情報工学専攻, 琉球大学大学院理工学研究科情報工学専攻, 琉球大学工学部情報工学科 (2018).

[2] 並列信頼研究室：CbC_gcc, 琉球大学 (online), available from ⟨http://www.cr.ie.u-ryukyu.ac.jp/hg/CbC/CbC_gcc/⟩ (accessed 2018-11-21).

xv6の構成要素の継続の分析

清水 隆博[1,a)]　河野 真治[2,b)]

概要：アプリケーションやサービスの信頼性は、OS と結びついている。OS 自身が高い信頼性を持つ必要があり、その上で動作するソフトウェアの信頼性を OS が保証するような仕組みがあると良い。テストは本質的に有限なケースしか調べないので、テストだけで信頼性を保証するのには限界がある。アプリケーションと OS の状態を状態遷移を基本としたモデルに変換しモデル検査や Hoare Logic などの形式手法を用いて信頼性を高めたい。そのために状態遷移単位での記述に適した継続を基本とした言語である CbC(Continuation based C) を OS とアプリケーションの記述に用いる。最初の段階として小さな unix である xv6 kernel の CbC による書き換えを行っている。xv6 kernel の処理がどのような状態遷移を行うのかを分析し、CbC の継続ベースでのプログラミングに変換していく必要がある。本稿では xv6kernel の構成要素の一部に着目し、状態遷移系の分析と xv6 の書き換えを行う。

1. アプリケーションの信頼性

アプリケーションの信頼性を向上させるためには、 土台となる OS 自体の信頼性が高く保証されていなければならない。OS そのものも巨大なプログラムであるため、テストコードを用いた方法で信頼性を確保する事が可能である。しかし並列並行処理などに起因するバグや、そもそも OS を構成する処理が巨大であることから、 テストで完全にバグを発見するのは困難である。テスト以外の方法で OS の信頼性を高めたい。

そこで数学的な背景に基づく形式手法を用いて OS の信頼性を向上させることを検討する。OS を構成する要素をモデル検査してデッドロックなどを検知する方法や、定理証明支援系 Agda[1] を利用した証明ベースでの信頼性の確保などの手法が考えられる。[2][3][4][5] 形式手法で信頼性を確保するには、 まず OS の処理を証明などがしやすい形に変換して実装し直す必要がある。[6] OS の内部処理の状態を明確にし、 状態遷移モデルに落とし込むことでモデル検査などを通して信頼性を向上させたい。しかし仕様記述言語や定理証明支援系では、 実際に動く OS と検証用の実装が別の物となってしまうために、 C 言語などでの実装の段階で発生するバグを取り除くことができない。実装のソースコードと検証用のソースコードは近いセマンティクスでプログラミングする必要がある。

1 琉球大学大学院理工学研究科情報工学専攻
2 琉球大学工学部工学科知能情報コース
a) anatofuz@cr.ie.u-ryukyu.ac.jp
b) kono@ie.u-ryukyu.ac.jp

OS 上のアプリケーションには本来行いたい処理の他に、メモリ管理やスレッド、 CPU などの資源管理がある。前者をノーマルレベルの計算と呼び、後者をメタ計算と呼ぶ。OS はメタ計算を担当していると言える。実装のソースコードはノーマルレベルであり検証用のソースコードはメタ計算だと考えると、OS そのものが検証を行ない、システム全体の信頼を高める機能を持つべきだと考える。ノーマルレベル上でのバグを例えばモデル検査のようなメタ計算によって発見し信頼性を向上させたい。

ノーマルレベルの計算とメタ計算の両方の実装に適した言語として Continuation Based C(CbC) がある。CbC は基本 `goto` で `CodeGaar` というコードの単位を遷移する言語である。通常の関数呼び出しと異なり、スタックあるいは環境と呼ばれる隠れた状態を持たない。このため、計算のための情報は `CodeGear` の入力にすべてそろっている。そのうちのいくつかはメタ計算、つまり、OS が管理する資源であり、その他はアプリケーションを実行するためのデータ (DataGear) である。メタ計算とノーマルレベルの区別は入力のどこを扱うかの差に帰着される。CbC は C と互換性のある C の下位言語であり、状態遷移をベースとした記述に適したプログラミング言語である。C との互換性のために、 CbC のプログラムをコンパイルすることで動作可能なバイナリに変換が可能である。CbC は `GCC`[7][8] あるいは `LLVM`[9][10] 上で実装されていて、通常の C のアプリケーションやシステムプログラムをそのまま包含できる。また CbC の基本文法は簡潔であるため、 Agda などの定理証明支援系 [11] との相互変換や、 CbC 自体でのモデル検査が可能であると考えられる。

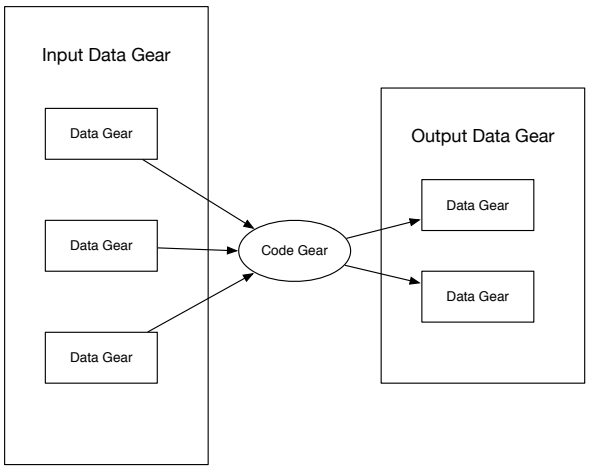

図 1: CodeGear と入出力の関係図

現在小さな unix である xv6 kernel を CbC を用いて書き換えている。書き換えの為には、 既存の xv6 kernel の処理の状態遷移を分析し、継続を用いたプログラムに変換していく必要がある。本論文ではこの書き換えに伴って得られた xv6 kernel の継続を分析し、 現在の CbC による書き換えについて述べる。

2. Continuation Based C

Continuation Based C(CbC) とは C 言語の下位言語であり、 関数呼び出しではなく継続を導入したプログラミング言語である。CbC では通常の関数呼び出しの他に、関数呼び出し時のスタックの操作を行わず、次のコードブロックに `jmp` 命令で移動する継続が導入されている。この継続は Scheme などの環境を持つ継続とは異なり、 スタックを持たず環境を保存しない継続である為に軽量である事から軽量継続と呼べる。また CbC ではこの軽量継続を用いて `for` 文などのループ文を実装する。これは関数型プログラミングでの Tail call スタイルでプログラミングすることに相当する。実際、Agda よる関数型の CbC の記述も用意されている。実際の OS やアプリケーションを記述する場合には GCC 及び LLVM/clang 上の CbC 実装を用いる。

CbC では関数の代わりに CodeGear という単位でプログラミングを行う。CodeGear は通常の C の関数宣言の返り値の型の代わりに`__code` で宣言を行う。各 CodeGear は DataGear と呼ばれるデータの単位で入力を受け取り、その結果を別の DataGear に書き込む。入力の DataGear を InputDataGear と呼び、 出力の DataGear を OutputDataGear と呼ぶ。CodeGear がアクセスできる DataGear は、 InputDataGear と OutputDataGear に限定される。これらの関係図を図 1 に示す。

CbC を利用したシステムコールのディスパッチ部分を Code 1 に示す。この例題では特定のシステムコールの場合、CbC で実装された処理に `goto` 文をつかって継続する。

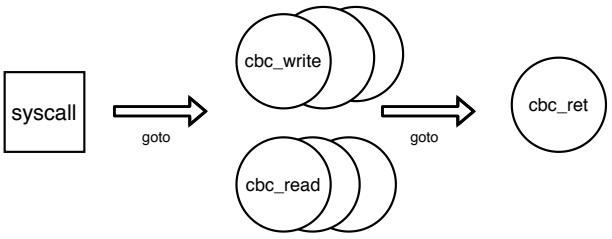

図 2: CbC を利用したシステムコールディスパッチの状態遷移

例題では CodeGear へのアドレスが配列 `cbccodes` に格納されている。引数として渡している `cbc_ret` は、 システムコールの返り値の数値をレジスタに代入する CodeGear である。実際に `cbc_ret` に継続が行われるのは、 `read` などのシステムコールの一連の処理の継続が終わったタイミングである。

```
void syscall(void)
{
    int num;
    int ret;

    if((num >= NELEM(syscalls)) && (num <= NELEM(
         cbccodes)) && cbccodes[num]) {
        proc->cbc_arg.cbc_console_arg.num = num;
        goto (cbccodes[num])(cbc_ret);
    }
```

Code 1: CbC を利用したシステムコールのディスパッチ

Code1 の状態遷移図を図 2 に示す。図中の `cbc_read` などは、 `read` システムコールを実装している CodeGear の集合である。

CodeGear は関数呼び出し時のスタックを持たない為、一度ある CodeGear に遷移してしまうと元の処理に戻ってくることができない。しかし CodeGear を呼び出す直前のスタックは保存されるため、 部分的に CbC を適用する場合は CodeGear を呼び出す `void` 型などの関数を経由することで呼び出しが可能となる。

この他に CbC から C へ復帰する為の API として、 環境付き goto という機能がある。これは GCC では内部コードを生成、 LLVM/clang では `setjmp` と `longjmp` を使うことで CodeGear の次の継続対象として呼び出し元の関数を設定することが可能となる。したがってプログラマから見ると、通常の C の関数呼び出しの返り値を CodeGear から取得する事が可能となる。

3. CbC を用いた OS の実装

軽量継続を持つ CbC を利用して、 証明可能な OS を実装したい。その為には証明に使用される定理証明支援系や、 モデル検査機での表現に適した状態遷移単位での記述が求められる。CbC で使用する CodeGear は、 状態遷移モデルにおける状態そのものとして捉えることが可能で

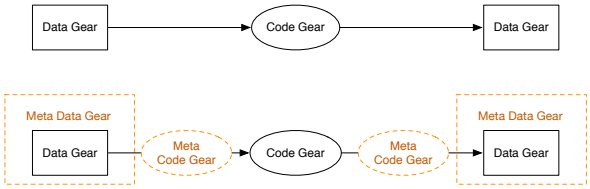

図 3: CodeGear と MetaCodeGear

ある。CodeGear を元にプログラミングをするにつれて、CodeGear の入出力の Data も重要であることが解ってきた。CodeGear とその入出力である DataGear を基本とした OS として、 GearsOS の設計を行っている。[12] 現在の GearsOS は並列フレームワークとして実装されており、実用的な OS のプロトタイプ実装として既存の OS 上への実装を目指している。

GearsOS では、 CodeGear と DataGear を元にプログラミングを行う。遷移する各 CodeGear の実行に必要なデータの整合性の確認などのメタ計算は、 MetaCodeGear と呼ばれる各 CodeGear ごと実装された CodeGear で計算を行う。この MetaCodeGear の中で参照される DataGear を MetaDataGear と呼ぶ。また、 対象の CodeGear の直前で実行される MetaCodeGear を StubCodeGear と呼ぶ。MetaCodeGear や MetaDataGear は、プログラマが直接実装することはなく、 現在は Perl スクリプトによって GearsOS のビルド時に生成される。CodeGear から別の CodeGear に遷移する際の DataGear などの関係性を、図 3 に示す。

通常のコード中では入力の DataGear を受け取り CodeGear を実行、 結果を DataGear に書き込んだ上で別の CodeGear に継続する様に見える。この流れを図 3 の上段に示す。しかし実際は CodeGear の実行の前後に実行される MetaCodeGear や入出力の DataGear を MetaDataGear から取り出すなどのメタ計算が加わる。これは図 3 の下段に対応する。

遷移先の CodeGear と MetaCodeGear の紐付けや、計算に必要な DataGear を保存や管理を行う MetaDataGear として `context` がある。context は処理に必要な CodeGear の番号と MetaCodeGear の対応表や、DataGear の格納場所を持つ。計算に必要なデータ構造と処理を持つデータ構造であることから、 context は従来の OS のプロセスに相当するものと言える。cotnext と各データ構造の関わりを図 4 に示す。

コード上では別の CodeGear に直接遷移している様に見えるが、 実際は context 内の遷移先の CodeGear に対応するスロットから、対応する MetaCodeGear に遷移する。MetaCodeGear 中で、次に実行する CodeGear で必要な DataGear を context から取り出し、 実際の計算が行われる。

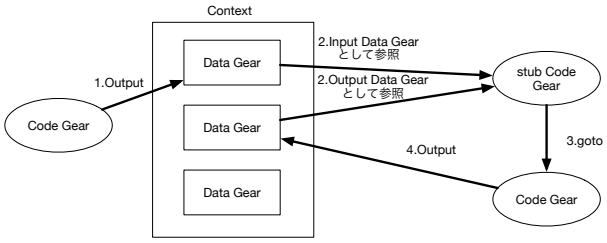

図 4: Context と各データの関係図

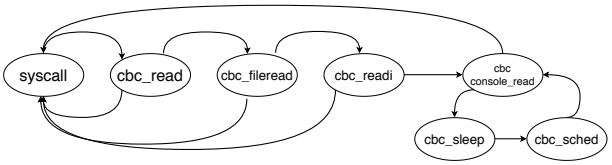

図 5: read システムコールの状態遷移

4. xv6 kernel

xv6 とはマサチューセッツ工科大学で v6 OS[13] を元に開発された教育用の UNIX OS である。[14] xv6 は ANSI C で実装されており、 x86 アーキテクチャ上で動作する。Raspberry Pi[15] 上での動作を目的とした ARM アーキテクチャのバージョンも存在する。[16] 本論文では最終的に Raspberry Pi 上での動作を目指しているために、 ARM アーキテクチャ上で動作する xv6 を扱う。

xv6 は小規模な OS だがファイルシステム、 プロセス、システムコールなどの UNIX の基本的な機能を持つ。またユーザー空間とカーネル空間が分離されており、 シェルや ls などのユーザーコマンドも存在する。

本論文では xv6 のファイルシステム関連の内部処理と、システムコール実行時に実行される処理について分析を行う。xv6 kernel のファイルシステムは階層構造で表現されており、 最も低レベルなものにディスク階層、 抽象度が最も高いレベルのものにファイル記述子がある。

本論文では xv6 の継続の分析をシステムコール部分とファイルシステム、 仮想メモリなどの OS の根幹部分でそれぞれ行った。

5. xv6 のシステムコールの継続の分析と書き換え

xv6 の処理を継続を中心とした記述で書き換えを行う。この際に、 xv6 のどの処理に着目するかによって継続の実装が異なっていくことが実装につれてわかった。

まず xv6 の `read` システムコールに着目し、 システムコール内部でどのような状態を遷移するかを分析した。[17] 分析結果を CbC の CodeGear に変換し、 状態遷移図におこしたものを図 5 に示す。

CbC で書き換えた `read` システムコールは、 xv6 の `read` システムコールのディスパッチ部分から、

cbc_readCodeGear に goto 文で軽量継続される。継続後は read する対象によって cbc_readi や、 cbc_consoleread などに状態が変化していく。各 CodeGear の遷移時には DataGear がやり取りされる。DataGear は xv6 のプロセス構造体に埋め込まれた context を経由して CodeGear に渡される。

この実装の利点として、 CodeGear の命名と状態が対応しており、 状態遷移図などに落としても自然言語で説明が可能となる点が挙げられる。しかし実際には cbc_readi の状態はさらに複数の CodeGear に分離しており、 実際に read システムコールを実装する CodeGear の数は図の状態より多い。この事から、 複数の CodeGear を 1 つにまとめた上で見た状態と、 各 CodeGear それぞれの状態の 2 種類の状態があるといえる。

複数の CodeGear をまとめた状態は、 抽象化した API の操作時におけるアルゴリズム上の問題が無いかの確認として使用出来る。対して各 CodeGear それぞれはモデル検査や、 特定の関数の中の処理が適しているかどうかの検査として見ることが出来ると考えられる。

この事から GearsOS では、 各 CodeGear のモジュール化の仕組みである Interface 機能を導入している。Interface の導入によって CodeGear を定義することで状態数を増やしても、 抽象化された API を利用することで細部の状態まで意識する必要が無くなった。xv6 の処理を CbC で書き換える際には、 対象の継続の API をまず決定しモジュール化を図る必要がある。

6. xv6 のシステムコール以外の継続の分析

xv6 はシステムコール以外に、 ファイルシステムの操作やページテーブルの管理などの処理も存在している。これらは OS の立ち上げ時やシステムコールの中で、ファイルシステムの操作に対応した関数や構造体などの API を通して操作される。システムコールの一連の流れに着目するのではなく、 特定の対象の API に着目して継続の分析を検討した。

xv6 のファイルシステムに関する関数などの API は主に fs.c 中に記述されている。API の内部を CodeGear に分割をすると、 API を呼び出す時点で API 細部の継続を考慮する必要がある。細部の継続を隠蔽するために、 抽象的に複数の CodeGear をまとめる機能である Interface を導入したい。Code2 に示す様に、 fs.c 中に定義されている API を抜き出し、 CbC の Interface として定義した。__code から始まる CodeGear の名前が、 それぞれ抽象化された CodeGear の集合の最初の継続となる。

```
typedef struct fs<Type,Impl> {
    __code readsb(Impl* fs, uint dev, struct
        superblock* sb, __code next(...));
    __code iinit(Impl* fs, __code next(...));
    __code ialloc(Impl* fs, uint dev, short type,
        __code next(...));
    __code iupdate(Impl* fs, struct inode* ip, __code
        next(...));
    __code idup(Impl* fs, struct inode* ip, __code
        next(...));
    __code ilock(Impl* fs, struct inode* ip, __code
        next(...));
    __code iunlock(Impl* fs, struct inode* ip, __code
        next(...));
    __code iput(Impl* fs, struct inode* ip, __code
        next(...));
....
} fs;
```

Code 2: ファイルシステム操作の API の一部

Code2 内の readsb などは fs.c 内で定義されている C の関数名と対応している。この C の関数を更に継続ごと分割するために、 関数内の if 文などの分岐を持たない基本単位である Basic Block に着目した。

CbC の CodeGear の粒度は C の関数とアセンブラの中間であるといえるので、 BasicBlock を CodeGear に置き換える事が可能である。したがって特定の関数内の処理の BasicBlock を分析し、 BasicBlock に対応した CodeGear へ変換することが可能となる。実際に BasicBlock 単位で切り分ける前の処理と、切り分けたあとの処理の一部を示す。例として inode のアロケーションを行う API でる ialloc の元のコードを Code3 に示す。

```
struct inode* ialloc (uint dev, short type)
{
    readsb(dev, &sb);
    for (inum = 1; inum < sb.ninodes; inum++) {
        bp = bread(dev, IBLOCK(inum));
        dip = (struct dinode*) bp->data + inum % IPB;

        if (dip->type == 0) { // a free inode
            memset(dip, 0, sizeof(*dip));
            ...
            return iget(dev, inum);
        }
        brelse(bp);
    }
    panic("ialloc:␣no␣inodes");
}
```

Code 3: ialloc の元のソースコード

ialloc はループ条件である inum < sb.ninodes が成立しなかった場合は panic へと状態が遷移する。この for 文での状態遷移を CodeGear に変換したものを Code4 に示す。

```
__code allocinode_loopcheck(struct fs_impl* fs_impl,
    uint inum, uint dev, struct superblock* sb,
    struct buf* bp, struct dinode* dip, __code next
    (...)){
    if( inum < sb->ninodes){
        goto allocinode_loop(fs_impl, inum, dev, type,
```

```
            sb, bp, dip, next(...));
    }
    char* msg = "failed␣allocinode...";
    struct Err* err = createKernelError(&proc->
        cbc_context);
    goto err->panic(msg);
}
```

Code 4: ループ条件を確認する CodeGear

Code4 ではループ条件の成立を if 文で確認し、ループ処理に移行する場合は allocinode_loop へ遷移する。goto 文の中の引数の 1 つ next(...) は、 API として呼び出した ialloc の次の継続の CodeGear に対して、 context などの環境を渡す構文である。ループ条件が満たされなかった場合は、 コンテキストから panic に関する CodeGear の集合を取り出し、 集合中の panic CodeGear へと遷移する。オリジナルの処理では、 ループ中に dip->type == 0 が満たされた場合は関数から return 文により関数から復帰する。CodeGear では Code5 内で、 状態が分けられる。この先の継続は、 復帰用の CodeGear かループの先頭である allocinode_loopcheck に再帰的に遷移するかになる。

```
__code allocinode_loop(struct fs_impl* fs_impl, uint
     inum, uint dev, short type, struct superblock* sb
     , struct buf* bp, struct dinode* dip, __code next
     (...)){
    bp = bread(dev, IBLOCK(inum));
    dip = (struct dinode*) bp->data + inum % IPB;
    if(dip->type = 0){
        goto allocinode_loopexit(fs_impl, inum, dev,
            sb, bp, dip, next(...));
    }

    brelse(bp);
    inum++;
    goto allocinode_loopcheck(fs_impl, inum, dev, type
        , sb, bp, dip, next(...));
}
```

Code 5: ループ中に復帰するかどうかの確認をする CodeGear

この継続の分析方法の利点として、 既存のコードの Basic Block 単位で CodeGear に変換可能であるため機械的に CodeGear への変更が可能となる。既存の関数上のアルゴリズムや処理に殆ど変更がなく変更可能であるために、 CodeGear で細分化して表現することは容易となる。

現在は従来の xv6 の関数呼び出しを元にした API の中で CodeGear に分割している。このために既存の API 内の処理の細分化は可能とはなったが、 API そのものを CodeGear を用いた継続に適した形には表現できていない。API の内部の CodeGear はあくまで Basic Block 単位に基づいているために、 状態遷移図で表現した際に自然言語で表現できない CodeGear も存在してしまう。

さらに、 for ループを CodeGear に分割することを考えるとループ中にループの index を利用している場合は、その index も次の継続に渡さなければならない。このため index を使用していない CodeGear でも継続の引数として index を受け取り、 実際に index を利用する CodeGear に伝搬させる必要がある。これらの問題を解決する為には、 API を分割した CodeGear それぞれの DataGear に型を与え、 どの継続で DataGear の意味が変わるかを追求する必要がある。API を分割して作成した CodeGear の DataGear は、現在各 API に対応した 1 つの巨大な構造体に隠蔽されている。巨大な構造体で管理するのではなく、構造体で次の CodeGear の状態に影響を与える要素を適宜組み合わせた DataGear を作る必要がある。

7. CbC を用いた部分的な xv6 の書き換え

CbC では CodeGear、 DataGear からなる単位を基本とし、 それぞれにメタな Gear が付随する。また実行に必要な CodeGear と DataGear をまとめた context という MetaDataGear が存在する。この機能を元に xv6 の書き換えを検討した。

xv6 内で CbC の軽量継続に突入する際は、 元の処理関数に通常の方法では戻ってくることができず、部分的に書き換えていくのが困難である。今回は呼び出し関数に戻れるスタックフレームを操作したい為に、 ダミーの void 関数を用意した。この関数内で CodeGear に goto 文を用いて遷移することで、 CbC から帯域脱出した際に void 関数の呼び出し元から処理を継続し、部分的に CbC に書き換えることが可能となった。Code6 では、 userinit 関数へ戻るために、 cbc_init_vmm_dumy を経由している。

```
void cbc_init_vmm_dummy(struct Context* cbc_context,
     struct proc* p, pde_t* pgdir, char* init, uint sz
     )
{
    struct vm* vm = createvm_impl(cbc_context);
    goto vm->init_vmm(vm, pgdir, init, sz , vm->
        void_ret);
}

void userinit(void)
{
...

    if((p->pgdir = kpt_alloc()) == NULL) {
        panic("userinit:␣out␣of␣memory?");
    }

    cbc_init_vmm_dummy(&p->cbc_context, p, p->pgdir,
        _binary_initcode_start, (int)
        _binary_initcode_size);

    p->sz = PTE_SZ;
    memset(p->tf, 0, sizeof(*p->tf));
...
```

```
}
```

Code 6: 部分的に CbC を適応する例

Code6 中で、 CodeGear への遷移が行われる `goto vm->init_vmm()` の `vm->void_ret` は `init_vmm` の次の継続の CodeGear 名である。この `vm->void_ret` は `return` するのみの CodeGear であり、 void 型関数と組み合わせることで呼び出し元へと復帰する事が可能となる。

8. xv6 の今後の書き換え

xv6 ではカーネルパニックの発生時や、 inode のキャッシュなどをグローバル変数として利用している。グローバル変数を使用してしまうと、 CodeGear で定義した状態が DataGear 以外のグローバル変数によって変更されてしまう。グローバル変数を極力使わず継続を中心とした実装を行いたい。

context は現在プロセス構造体に埋め込まれており、kernel そのものの状態を制御するためには各 context を管理する機能が必要であると考えられる。

現状は xv6 の全ての機能をまだ CbC を用いて書き換えていない。ファイルシステムや仮想メモリにまつわる処理などは API 単位では書き換えているが、 API を呼び出す箇所は C の関数上で部分的に呼び出している。そのために OS そのものを状態遷移単位で完全に書き直す必要が存在し、 そのためには全ての処理に対して状態を定義していく必要がある。

また OS 上で DataGear と CodeGear の位置づけを明確に定義する必要も存在する。DataGear の依存関係や CodeGear の並列実行など、 プロセスベースで実装していた処理を CodeGear などで意味がある形式にする必要がある。

9. まとめ

本稿では xv6 を継続を用いた単位での書き換えを検討し、実際にいくつかの処理を書き換えた。書き換えはシステムコールに着目し CodeGear へ分割する手法と、BasicBlock ごとに CodeGear へ分割する手法で行った。現状はまだ xv6 の実装を利用した証明や、 xv6 にモデル検査機能の実装を行いたい。また Agda などの定理証明支援系で証明された処理から、 CbC の CodeGear に変換する処理系の実装なども検討する。

参考文献

[1] Norell, U.: Dependently Typed Programming in Agda, *Proceedings of the 4th International Workshop on Types in Language Design and Implementation*, TLDI '09, New York, NY, USA, ACM, pp. 1–2 (online), DOI: 10.1145/1481861.1481862 (2009).

[2] Yang, J. and Hawblitzel, C.: Safe to the Last Instruction: Automated Verification of a Type-safe Operating System, *Proceedings of the 31st ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '10, New York, NY, USA, ACM, pp. 99–110 (online), DOI: 10.1145/1806596.1806610 (2010).

[3] Klein, G., Elphinstone, K., Heiser, G., Andronick, J., Cock, D., Derrin, P., Elkaduwe, D., Engelhardt, K., Kolanski, R., Norrish, M., Sewell, T., Tuch, H. and Winwood, S.: seL4: Formal Verification of an OS Kernel, *Proceedings of the ACM SIGOPS 22Nd Symposium on Operating Systems Principles*, SOSP '09, New York, NY, USA, ACM, pp. 207–220 (online), DOI: 10.1145/1629575.1629596 (2009).

[4] Sigurbjarnarson, H., Bornholt, J., Torlak, E. and Wang, X.: Push-button Verification of File Systems via Crash Refinement, *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, Berkeley, CA, USA, USENIX Association, pp. 1–16 (online), available from ⟨http://dl.acm.org/citation.cfm?id=3026877.3026879⟩ (2016).

[5] Chen, H., Ziegler, D., Chajed, T., Chlipala, A., Kaashoek, M. F. and Zeldovich, N.: Using Crash Hoare Logic for Certifying the FSCQ File System, *Proceedings of the 25th Symposium on Operating Systems Principles*, SOSP '15, New York, NY, USA, ACM, pp. 18–37 (online), DOI: 10.1145/2815400.2815402 (2015).

[6] Nelson, L., Sigurbjarnarson, H., Zhang, K., Johnson, D., Bornholt, J., Torlak, E. and Wang, X.: Hyperkernel: Push-Button Verification of an OS Kernel, *Proceedings of the 26th Symposium on Operating Systems Principles* (2017).

[7] GNU Compiler Collection (GCC) Internals: *http://gcc.gnu.org/onlinedocs/gccint/*.

[8] 大城信康，河野真治：ContinuationbasedC の GCC4.6 上の実装について，第 53 回プログラミング・シンポジウム予稿集，Vol. 2012, pp. 69–78 (2012).

[9] Lattner, C. and Adve, V.: LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation, *Proceedings of the 2004 International Symposium on Code Generation and Optimization (CGO'04)*, Palo Alto, California (2004).

[10] TOKKMORI, K. and KONO, S.: Implementing Continuation based language in LLVM and Clang, *LOLA 2015* (2015).

[11] MASATAKA, H. and KONO, S.: GearsOS の Hoare Logic をベースにした検証手法，ソフトウェアサイエンス研究会 (2019).

[12] 河野真治，伊波立樹，東恩納琢偉：Code Gear、Data Gear に基づく OS のプロトタイプ，情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS) (2016).

[13] Lions, J.: *Lions' Commentary on UNIX 6th Edition with Source Code*, Computer classics revisited, Peer-to-Peer Communications (1996).

[14] Russ Cox, Frans Kaashoek, Robert Morris: xv6 a simple, Unix-like teaching operating system, https://pdos.csail.mit.edu/6.828/2018/xv6/book-rev11.pdf.

[15] : Raspberry Pi, https://www.raspberrypi.org.

[16] Wang, Z.: xv6-rpi, https://code.google.com/archive/p/xv6-rpi/ (2013).

[17] 坂本昂弘，桃原　優，河野真治：継続を用いた x.v6 kernel の書き換え，情報処理学会システムソフトウェアとオペレーティング・システム研究会 (OS)，No. 4 (2019).