

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4615795号

(P4615795)

(45) 発行日 平成23年1月19日(2011.1.19)

(24) 登録日 平成22年10月29日(2010.10.29)

(51) Int.Cl.

F I

G 0 6 F 11/00 (2006.01)

G 0 6 F 9/06 6 3 0 B

請求項の数 6 (全 33 頁)

(21) 出願番号	特願2001-349299 (P2001-349299)	(73) 特許権者	500046438
(22) 出願日	平成13年11月14日(2001.11.14)		マイクロソフト コーポレーション
(65) 公開番号	特開2002-169702 (P2002-169702A)		アメリカ合衆国 ワシントン州 9805
(43) 公開日	平成14年6月14日(2002.6.14)		2-6399 レッドモンド ワン マイ
審査請求日	平成16年11月15日(2004.11.15)		クロソフト ウェイ
審査番号	不服2007-23318 (P2007-23318/J1)	(74) 代理人	100077481
審査請求日	平成19年8月23日(2007.8.23)		弁理士 谷 義一
(31) 優先権主張番号	09/713, 633	(74) 代理人	100088915
(32) 優先日	平成12年11月14日(2000.11.14)		弁理士 阿部 和夫
(33) 優先権主張国	米国 (US)	(74) 復代理人	100115624
			弁理士 濱中 淳宏
		(74) 復代理人	100115635
			弁理士 窪田 郁大

最終頁に続く

(54) 【発明の名称】 プログラムバイナリに対する最小デルタジェネレータ

(57) 【特許請求の範囲】

【請求項 1】

第1のプログラムバイナリと第2のプログラムバイナリとの間でデルタを生成するコンピュータが実施するデルタ生成方法であって、

コンピュータのプロセッサが、コンピュータ記憶媒体に記憶された第1のプログラムバイナリの第1の制御フローグラフ(CFG)表現を作成し、前記コンピュータ記憶媒体に記憶された第2のプログラムバイナリの第2のCFG表現を作成するステップであって、前記第1のCFG表現および前記第2のCFG表現は、それぞれ、前記第1のプログラムバイナリおよび前記第2のプログラムバイナリの各ブロックをノードで表し、各ノード間の制御フローをエッジで表した有向グラフであって、前記ブロックを表すノードには、当該ブロックの内容が同一であれば、第1のCFG表現及び第2のCFG表現に共通して同一のノード情報が付与され、当該ブロックの内容が同一でなければ、それぞれ異なるノード情報が付与された有向グラフを作成する前記ステップと、

前記コンピュータのプロセッサが、前記第1のCFG表現および前記第2のCFG表現の間でブロックを比較して、前記第1のCFG表現中および前記第2のCFG表現中でマッチするブロック(マッチするブロック)を識別し、それにより、前記第1のCFG表現中でマッチしない前記第2のCFG表現中のブロック(マッチしないブロック)を識別するステップであって、前記比較は、比較されるブロックを表すノードに対するノード情報と、比較されるブロックの周囲のブロックの拡張された局所近傍のブロックを表すノードに対するノード情報を前記第1のCFG表現および前記第2のCFG表現の間で比較する

10

20

ことにより行われ、各ブロックの局所近傍は、C F G 表現中の前記ブロックの近傍であるが前記 C F G 表現中の全てのブロックより少ないブロックからなり、前記ブロックの拡張された局所近傍は、前記ブロックの局所近傍と、前記ブロックの局所近傍より大きい近傍において該ブロックから固定長の様なランダムウォークを行う間に遭遇するブロックの集合とからなり、C F G 表現中の拡張された局所近傍は、前記 C F G 表現中の全てのブロックより少ないブロックからなる、識別するステップと、

前記コンピュータのプロセッサが、パッチを当てた前記第 1 のプログラムバイナリから作成される新たな第 1 の C F G 表現が前記第 2 の C F G 表現と同一になるように、パッチを当てた前記第 1 のプログラムバイナリが前記第 2 のプログラムバイナリと同一になるように、マッチしないブロックを前記第 1 のプログラムバイナリにパッチする編集操作を決定するステップと、

10

前記コンピュータのプロセッサが、前記マッチしないブロックと前記編集操作とを含むデルタを作成するステップと

を備えることを特徴とするデルタ生成方法。

【請求項 2】

前記プロセッサが、第 1 のプログラムバイナリを有するコンピュータに前記デルタを伝送するステップ

をさらに備えることを特徴とする請求項 1 に記載のデルタ生成方法。

【請求項 3】

前記プロセッサが、前記第 1 のプログラムバイナリのコピーが前記第 2 のプログラムバイナリと同一になるように前記コピーをパッチするステップであって、前記デルタはこのようなパッチングを導くステップ

20

をさらに備えることを特徴とする請求項 1 に記載のデルタ生成方法。

【請求項 4】

前記ブロックの局所近傍は、前記ブロックに隣接するブロックからなることを特徴とする請求項 1 に記載のデルタ生成方法。

【請求項 5】

請求項 1 に記載のデルタ生成方法をコンピュータに実行させるためのコンピュータ実行可能命令からなるプログラムを有することを特徴とするコンピュータ記憶媒体。

【請求項 6】

30

記憶媒体に記憶された第 1 のプログラムバイナリの第 1 の制御フローグラフ (C F G) 表現と、前記記憶媒体に記憶された第 2 のプログラムバイナリの第 2 の C F G 表現の間でブロックを比較して、前記第 1 の C F G 表現中と前記第 2 の C F G 表現中とでマッチするブロック (マッチするブロック) を識別し、それにより、前記第 1 の C F G 表現中でマッチしない前記第 2 の C F G 表現中のブロック (マッチしないブロック) を識別するように構成されたコンパレータであって、前記第 1 の C F G 表現および前記第 2 の C F G 表現は、それぞれ、前記第 1 のプログラムバイナリおよび前記第 2 のプログラムバイナリの各ブロックをノードで表し、各ノード間の制御フローをエッジで表した有向グラフであって、前記ブロックを表すノードには、当該ブロックの内容が同一であれば、第 1 の C F G 表現及び第 2 の C F G 表現に共通して同一のノード情報が付与され、当該ブロックの内容が同一でなければ、それぞれ異なるノード情報が付与された有向グラフであり、前記比較は、比較されるブロックを表すノードに対するノード情報と、比較されるブロックの周囲のブロックの拡張された局所近傍のブロックを表すノードに対するノード情報を前記第 1 の C F G 表現および前記第 2 の C F G 表現の間で比較することにより行われ、各ブロックの局所近傍は、C F G 表現中の前記ブロックの近傍であるが前記 C F G 表現中の全てのブロックより少ないブロックからなり、前記ブロックの拡張された局所近傍は、前記ブロックの局所近傍と、前記ブロックの局所近傍より大きい近傍において該ブロックから固定長の様なランダムウォークを行う間に遭遇するブロックの集合とからなり、C F G 表現中の拡張された局所近傍は、前記 C F G 表現中の全てのブロックより少ないブロックからなる、コンパレータと、

40

50

パッチを当てた前記第 1 のプログラムバイナリから作成される新たな第 1 の C F G 表現が前記第 2 の C F G 表現と同一になるように、パッチを当てた前記第 1 のプログラムバイナリが前記第 2 のプログラムバイナリと同一になるように、マッチしないブロックを前記第 1 のプログラムバイナリにパッチする編集操作を決定するように構成された編集操作決定機構と、

前記マッチしないブロックと前記編集操作とを含むデルタを作成するように構成された出力サブシステムと

を備えたことを特徴とするデルタジェネレータシステム。

【発明の詳細な説明】

【 0 0 0 1 】

10

【発明の属する技術分野】

本発明は一般に、プログラムバイナリ間で最小限のデルタを生成することに関する。

【 0 0 0 2 】

【従来の技術】

今日のソフトウェアシステムの重要な特徴の 1 つはアップグレードできることであり、これは「アップグレード性 (u p g r a d a b i l i t y) 」と呼ばれることもある。古いソフトウェアは、絶えずより新しいバージョンで置換されており、コードの再使用可能性およびモジュラ開発が、ソフトウェア設計の主要な特徴である。

【 0 0 0 3 】

正確さ

20

ソフトウェアを古いバージョンから新しいバージョンにアップグレードするときは、完全な正確さが極めて重要である。ターゲットコンピュータ中で新たにアップグレードしたソフトウェア中のあらゆるビットが、媒体ソースにある新しいソフトウェアと正確に合致 (マッチ) しなければならない。さもなければ、新しいソフトウェアは誤動作するか全く動作しない恐れがある。

【 0 0 0 4 】

完全な正確さを保証するために、従来の技術では、古いソフトウェアを新しいソフトウェアで完全に置換する。ソフトウェアプログラム (特に主要なアプリケーションスイートおよびオペレーティングシステム) のサイズおよび複雑さが増すにつれて、この十把ひとからげに置換して更新する手法は、このようなソフトウェアの顧客にとってより時間がかかり苛立たしいものになっている。

30

【 0 0 0 5 】

このような更新のソースを、ローカル、携帯型かつ高帯域幅のリムーバブルメディア (C D - R O M など) からリモート、集中型かつ比較的低い帯域幅のネットワークサーバ (インターネットウェブサーバなど) に移行する傾向により、事態は悪化している。 1 0 0 M B のソフトウェアをおそらくは C D - R O M から置換するのには数分かかることがあるが、同じ量のソフトウェアをダイヤルアップインターネット接続で置換するのには数時間かかることがある。

【 0 0 0 6 】

本明細書で、「完全な正確さ」および「ほぼ同一」は、元々製作された新しいソフトウェアとユーザのコンピュータ上に存在する新しいソフトウェアの間の小さく実質的でない差を考慮に入れている。

40

【 0 0 0 7 】

従来のデルタパッチング

通常、ソフトウェアのより新しいバージョンは、わずかな追加部分、ならびに、古い部分におけるいくつかの小さな変更を有する。したがって、古いバージョンを新しいバージョンで完全に置換する力任せの手法は行き過ぎである。ある代替手法は、古いバージョンから新しいバージョンを再構築できるように、これらの変更を「パッチ」に取り込むものである。古いバージョンと新しいバージョンの間に差があるので、この技法はしばしば「差分パッチング (d i f f - p a t c h i n g) 」と呼ばれる。本明細書では、古いバージ

50

ョンと新しいバージョンとの差を「デルタ」()と呼び、したがってこの差分パッチング技法を「デルタパッチング」(または「パッチング」)と呼ぶ。

【0008】

デルタパッチングに伴う問題は正確さである。何をパッチして何をパッチしないかを識別するのは難しい。このようなパッチの境界を正確に決定しないと、パッチしたバージョンは、ソフトウェアの所望の新バージョンとは異なるものになる。

【0009】

この結果、従来のデルタパッチングでは、正確さを達成する効率が損なわれる。一般に、サブモジュールファイル、データファイル、ライブラリファイル、およびこのようなファイルのグループは、どんな変更であれそれらの内に変更があるかどうかマークされる。このことは例えば、100KbのDLL(ダイナミックリンクライブラリ)ファイル内のソースコードの一行が変更される場合でもDLLファイル全体を置換することを意味する。既存のDLLファイル中のフラグメントを置換するのではなくこのようにするのは、一部には、置換が必要なフラグメントを選択してそのフラグメントだけを完全に正確に置換するのが難しいからである。しかし主に、従来の技術ではモジュール全体を置換の方が効率的なので、このことが行われる。ある小さなフラグメント中の小さな変更が、プログラム全体に及ぶ変更に見える場合もある。

【0010】

【発明が解決しようとする課題】

この従来の非効率的なデルタパッチングは、ソフトウェア全体の大規模な置換よりは効率的で速いものの、依然として可能な限り効率的ではない。古いソフトウェアバージョンと形の異なる、または古いソフトウェアバージョン中に存在しない、モジュールまたはサブモジュールのフラグメントだけをパッチする方が効率的であろう。フラグメントの例としては、サブルーチン、関数、オブジェクト、データ構造、インタフェース、メソッド、またはこれらのいずれかの一部が挙げられる。

【0011】

不変フラグメントの検出

フラグメントデルタを検出するための前提は、フラグメントの不変性を検出できることである。言い換えれば、プログラムモジュールをパッチできるようになる前に、2つのバージョン間でどのフラグメントが変更されていないかを決定する必要がある。このような不変性を検出してパッチを生み出すことは、各バージョンに対するソースコードを知っていればそれほど難しくはない。

【0012】

しかし、このようなフラグメントがバイナリとして表されたものを(そのソースコードを知らずに)扱うときは、フラグメントの不変性の検出はかなり難しくなる。主な困難は、プログラムモジュールの異なるバージョン中に、機能的には変更されていないが異なるように見えるコードが存在することある。コードは、その機能の変更を受けないこともあるが、種々の理由で2つのバージョン中で異なって見えることがある。このような理由の例としては、次のことが挙げられる。

・コードのある領域の変更により、別の(変更されない)領域が異なって見える可能性がある。

・小さな2つのバイナリコードシーケンスが、異なる機能を有するソースコードに対応する場合でも同一に見えることがある。

・2つの構造におけるレジスタ割付けの違い。

【0013】

変更が見かけ上の変更を引き起こす場合

コードのある部分の小さな変更は、しばしば近くで、また時として遠く離れたコード領域でも、連続的な変更を引き起こす。例えば、以下の2つのソースフラグメントを考えてみる。

Program P1**Program P2**

```

function f (int p)
int a = 3, b = 4;
if (b > p) {
    a = p;
return a;

```

```

function g (int p)
int b = 4, a = 3;
if (b > p) {
    a = p;
return a;

```

10

【 0 0 1 4 】

2つのプログラムP1およびP2中にある2つの関数fおよびgは、名前の違いを別にすれば実際は同じである。ソースコードを知っていれば、各フラグメント中の「if (b > p)」条件が同じでありパッチする必要がないことが確定するのは明らかである。しかし、これらに対応するバイナリを検査する場合、スタックのベースからのbのオフセットは、この2つのフラグメント中で異なることになる。これは、P1中でbの前にaが宣言されている形が、P2中でaの前にbが宣言されている形と異なるからである。したがって、他がすべて同じであっても、2つのフラグメントのバイナリは同一ではないことになる。当然、形の上でのこれらの違いは実質的には関係ないが、それでもやはり得られるバイナリは異なる。

20

【 0 0 1 5 】

次に、以下の抜粋を考えてみる。

Program P1

Program P2

x = f (10)

x = g (10)

【 0 0 1 6 】

この例では、前の例と同様に関数fおよびgが定義されていると仮定する。ここでもまた、呼び出される関数ならびに呼出し引数が同一なので、2つの呼出しは同一である。しかし、fとgの一致を知らない場合は、上の呼出しの一致もわからないことになる。これは、局所的な変更が、ともすれば遠くのコード領域を通過してどのようにカスケードする可能性があるかを示す一例である。

30

【 0 0 1 7 】

同一に見えるが同一ではない場合

時として2つのバイナリフラグメントが、対応するプログラムの構造中で異なる領域に対応するにもかかわらず同一に見える場合がある。以下の場合を考えてみる。

Program P1

Program P2

int a = atoi (argv [1]);

int b = atoi (argv [2]);

int b = atoi (argv [2])

if (b < 10) return;

if (a < 10) return;

...

if (b < 20) return;

...

40

【 0 0 1 8 】

P1中の条件「if (a < 10)」とP2中の条件「if (b < 10)」は、これらの機能が異なっても（これらのソースコードを検査すればはっきりわかるが）、両方とも同じバイナリコードに翻訳される場合がある。これは、P2中のスタック上のbのオフセットが、P1のスタック上のaのオフセットと同じである場合があるために起こる。2つの変数は、これらの上にあるコンテキストを見ればわかるように異なるプログラム引数によって定義されているので、明らかに異なる。しかし、上にあるソースコードコンテキストを

50

参照せずにこれらのバイナリに相当するものを比較すると、同一であるかのような錯覚が生じる可能性がある。このバイナリに相当するものの表現は、次のように見える場合がある。

```
mov eax, dword ptr [ebp+8h]
cmp eax, 0ah
jge L
ret
L: ...
```

【 0 0 1 9 】

レジスタ割付け

バイナリフラグメントの一致を検出する際のもう1つの問題は、レジスタ割付けによって生じる。コードの一部を変更すると、近くの領域が修正されていなくても、これらの後方の領域でレジスタ割付けの変更を引き起こす場合がある。したがって、バイナリを比較するときは、レジスタオペランドの変更のように見えるものが実際はレジスタの単純な名前変更によって識別が変えられたものだという可能性を考慮しなければならない。

【 0 0 2 0 】

【課題を解決するための手段】

本明細書では、少なくとも2つのプログラムバイナリ間で最小限のデルタを生成する技術について述べる。本明細書に述べる一実施形態では、バイナリフォーマットのソースプログラム（S）と、バイナリ形式のターゲットプログラム（T）が提供される。それぞれの制御フローグラフ（CFG）を構築する。SのCFGとTのCFGとの共通ブロックを突き合わせる（マッチング）。ブロックは、それらの内容およびそれらの局所近傍（例えばd近傍）に基づいてマッチングする。さらにブロックは、計算したハッシュ値に基づくラベルを使用してマッチングする。マッチングは複数のパスで行い、各パスは、マッチのための基準を緩和することによってマッチングを向上させる。さらに、ブロックを公正に比較できるように、レジスタ名の変更の問題を解決する。

【 0 0 2 1 】

上記の実施形態は中間出力を生み出すが、これはマッチしないブロックの内容である。このようなマッチしないブロックは、T中に見られるがS中には見られないブロックである。マッチしないブロックをSにマージするためのエッジ編集操作のセットを生み出す。マッチしないブロックと編集操作とを組み合わせたものがデルタである。SをパッチしてTの再構築コピーを生み出すには、デルタをSとマージする。

【 0 0 2 2 】

この概要自体は、本発明の範囲を限定するものではない。本発明をよりよく理解するために、以下の詳細な説明および頭記の特許請求の範囲を添付の図面と共に参照されたい。本発明の範囲は頭記の特許請求の範囲に示す。

【 0 0 2 3 】

図面全体にわたり、同じ要素および機構を参照するのに同じ番号を使用する。

【 0 0 2 4 】

【発明の実施の形態】

後続の記述では、頭記の特許請求の範囲に挙げる要素を組み入れた、プログラムバイナリに対する最小デルタジェネレータの具体的な実施形態を述べる。これらの実施形態は、法定の書面による記述、実施可能であること、および最良の形態の要件を満たすために具体的に述べるものである。ただし、この記述自体は本特許の範囲を限定するものではない。

【 0 0 2 5 】

本明細書では、プログラムバイナリに対する最小デルタジェネレータの、1つまたは複数の例示的な実施形態について述べる。発明者は、これらの例示的な実施形態を例として意図している。発明者は、これらの例示的な実施形態が、請求項に記載された本発明の範囲を限定するものであるとは意図しない。むしろ、請求項に記載された本発明が、現在または将来の他の技術を用いて他の形でも実施および実装できることを企図している。

【0026】

参照による組み込み

以下の同時係属の特許出願は、すべて1999年6月30日に出願され、Microsoft Corporationに譲渡されたものである。これらを参照により本明細書に組み込む。

「Translation and Transformation of Heterogeneous Programs」という名称の米国特許出願第09/343,805号

「Instrumentation and Optimization Tools for Heterogeneous Programs」という名称の米国特許出願第09/343,298号

「Shared Library Optimization for Heterogeneous Programs」という名称の米国特許出願第09/343,279号

「Application Program Interface for Transforming Heterogeneous Programs」という名称の米国特許出願第09/343,276号

「Cross Module Representation of Heterogeneous Programs」という名称の米国特許出願第09/343,287号

【0027】

簡単な概略

プログラムバイナリに対する最小デルタジェネレータの例示的な一実施形態は、例示的な「デルタジェネレータ」と呼ぶことができる。本明細書に述べる1つまたは複数の例示的な実施形態は、(全体的または部分的に)図1のデルタジェネレータシステム100および/または図11に示すコンピューティング環境によって実装することができる。

【0028】

効率および速度を促進するために、例示的なデルタジェネレータは、バイナリソースとターゲットプログラムとの最小限の差を見つけ、それを用いてソースプログラムを修正する。しかし、このような効率および速度のために正確さが損なわれることはない。

【0029】

これを達成するために、例示的なデルタジェネレータは、ソースプログラムとターゲットプログラムのバイナリフラグメントを比較する。これらの間で同じ(すなわち変更されていない)フラグメントを識別する。これらの間で異なる(すなわち変更された)フラグメントを識別することも必要である。変更されたフラグメントをデルタに含める。例示的なデルタジェネレータは、ソースをデルタでパッチしてターゲットプログラムに変形させる。

【0030】

デルタジェネレータシステムの高レベルな記述

図1に、デルタジェネレータシステム100を示す。この図には、バイナリソースプログラム112およびバイナリターゲットプログラム122が示されている。デルタジェネレータシステム100は、ソースプログラム112およびデルタ()142からターゲットプログラム122を再構築する。ターゲットプログラムとソースプログラムは、同じプログラムの異なるバージョンであると想定し、したがってこれらのコードの少なくともいくつかの部分は共通する。例えば、ターゲットプログラムはソースプログラムの新しいバージョンとすることができる。

【0031】

ソースCFG機構110が、ソースプログラム112の制御フローグラフ(CFG)を作成する。CFGについては、「用語」セクションでより詳細に述べる。同様に、ターゲットCFG機構120が、ターゲットプログラム122のCFGを作成する。当然、この2つのCFG機構110および120は共通のCFG機構としてもよい。

【0032】

図 1 には、ソース C F G およびターゲット C F G を入力として受け取ってこれらと比較するブロックマッチャ (m a t c h e r) 1 3 0 が示されている。これは、ソース C F G とターゲット C F G の両方に見られるバイナリフラグメントをマッチングする。

【 0 0 3 3 】

デルタジェネレータ 1 4 0 は、デルタ () フラグメントとしてマッチしなかった、ターゲット C F G 中の残りの (すなわちマッチしない) フラグメントを識別する。ブロックマッチャ 1 3 0 とデルタジェネレータ 1 4 0 がまとまって、ターゲット C F G 中のバイナリフラグメントを、ソース C F G 中でマッチするものとソース C F G 中でマッチしないものとに識別する。

【 0 0 3 4 】

さらに、決定機構 1 4 0 がまた、 をソース C F G 中にマージするためにどのようにソース C F G を編集するかを決定する。「編集」は 1 4 2 の一部であり、 1 4 2 は 決定機構 1 4 0 の結果である。

【 0 0 3 5 】

パッチャ 1 5 0 が、ソースプログラムをバイナリレベルでパッチする。ソースプログラム 1 1 2 を 1 4 2 と結合して、ターゲットプログラム 1 6 0 を再構築する。再構築したターゲットプログラム 1 6 0 は、ターゲットプログラム 1 2 2 と同一である。これは、意図される結果がターゲットプログラム 1 2 2 の同一コピーなので、同一である。このデルタジェネレータシステム 1 0 0 は、ソースプログラムを効率的にパッチすることによって誤りを導入することがないので、コピーは同一となる。

【 0 0 3 6 】

図 2 に、プログラムバイナリに対する最小デルタジェネレータの例示的な実施形態のためのサンプルの適用シナリオを概略的に示す。図 2 には、インターネットなどのネットワーク接続 2 1 5 を介してリンクされたサーバ側 2 1 0 とクライアント側 2 2 0 が示されている。サーバ側 2 1 0 は、バイナリソースプログラム (S) 1 1 2 のコピーおよびバイナリターゲットプログラム (T) 1 2 2 のコピーを有するサーバ 2 1 2 を含む。クライアント側 2 2 0 は、バイナリソースプログラム (S) 1 1 2 のコピーだけを有するクライアント 2 2 2 を含む。クライアントは最初、ターゲットプログラム (T) 1 2 2 のコピーまたはデルタ () 1 4 2 は有しない。

【 0 0 3 7 】

サーバ側 2 1 0 のサーバ 2 1 2 は、本明細書に述べる実施形態に従って 1 4 2 を作成する。サーバ 2 1 2 は、このような をクライアント側 2 2 0 のクライアント 2 2 2 に送信する。1 4 2 a および 1 4 2 b (ならびにこれらの矢印) の表現は、 がサーバからクライアントに送信されているのを示す。

【 0 0 3 8 】

クライアントは、新たに到着した 1 4 2 で S 1 1 2 をパッチして、T 1 2 2 を再構築する。したがって、例示的なデルタジェネレータの実施形態によれば、限られた帯域幅と思われるネットワーク 2 1 5 を介して最小サイズのデルタが送信され、それによりクライアントが S をパッチして T を正確に再構築する。

【 0 0 3 9 】

用語

プログラムバイナリ (またはバイナリプログラム) の一例は、プログラムモジュールがバイナリとして表されたもの (例えばソフトウェア) である。これには、プログラムモジュールに対するソースコードは含まれない。

【 0 0 4 0 】

「基本ブロック」 (または単に「ブロック」) は、単一の入口点を有し (外部からは最初の命令だけにしか到達できない) 、最後の命令に 1 つの出口点を有する命令シーケンスである。したがって、基本ブロック内部には一連の制御フローがある。基本ブロックはフラグメントの一例である。具体的には、コードフラグメントの一例である。基本ブロックは、「コードブロック」と呼ぶこともできる。

10

20

30

40

50

【 0 0 4 1 】

制御フローグラフ (C F G) は、プログラムを抽象化したものである。これは、プログラム中の基本ブロックをノードとし、ブロック間の可能なすべての制御フローをエッジ (視覚的にしばしば矢印で表される) で表した有向グラフである。図 3 ~ 図 5 は C F G を含んでいる。

【 0 0 4 2 】

プログラム中の隣接した静的データエリアを「データブロック」と呼び、これらもまた、C F G 中のノードを形成する。データブロックはフラグメントの一例である。具体的には、データフラグメントの一例である。

【 0 0 4 3 】

データブロックは、静的データ (「生」データと呼ばれる) を有し、プログラムの他の部分へのポインタを有する場合もある。これらのポインタは、例えばオブジェクトの仮想テーブルエントリとすることができる。このようなポインタは再配置可能データであり、したがって、データブロック中の他の静的データとは異なる。

【 0 0 4 4 】

C F G では、データブロック中のポインタを、データブロックからターゲットブロックへの有向エッジで表す。さらに、コードブロック中の命令中にあるアドレスオペランドを、コードブロックから対応アドレスにあるブロックへのエッジとして表す。

【 0 0 4 5 】

データブロックの内容にはその生データが含まれるが、ブロック中のポインタは除外する。同様に、コードブロックの内容にはその命令シーケンスが含まれるが、アドレスオペランドは除外する。C F G 中の基本ブロックの内容は、そのエッジと共に、プログラムを完全に指定する。C F G P 中に所与の任意のブロックがある場合、その親は、集合 { | が P 中のエッジ } であり、その子は、集合 { | が P 中のエッジ } である。

【 0 0 4 6 】

本明細書で述べる C F G は、プログラムのレイアウトに関する完全な情報を含む。C F G 中には、バイナリ中の関数に対応してプロシージャがあり、各プロシージャは、単一関数の基本ブロックによって誘導されるサブグラフである。関数のレイアウト (すなわちプログラムのアドレス空間における基本ブロックの構成) は、対応するプロシージャのすべてのブロックのリンクリストによって取り込まれ、したがって、リストの先頭から末尾まで走査すると、関数のレイアウト中の基本ブロックの順序が正確に記述される。

【 0 0 4 7 】

ツール V が、所与のプログラムバイナリに対し、本明細書で述べる C F G を生み出す。さらに V は、プログラムに多くの種類の修正を加えることを可能にする (C F G への) インタフェースを提供することもできる。これらの修正には、基本ブロックの追加および削除、それらの内容の修正、エッジの変更が含まれるが、これらに限定しない。

【 0 0 4 8 】

本明細書に述べるツール V をどのように実装して使用するかは、当業者なら理解する。さらに、参照により組み込む特許出願には、ツール V のようなツールを実装するのに使用できるコンポーネントが記載されている。さらに、以下の刊行物に、ツール V が使用できる C F G 方法に関する一般的な背景が提供されている。すなわち、Aho, Hopcroft, Ullman: 「Principles of Compiler Design」および 2) A. Aho, R. Sethi, J. Ullman, 「Compilers, Principles, Techniques, and Tools」(1986) である。

【 0 0 4 9 】

プログラムバイナリに対する最小デルタジェネレータの実施形態の 1 つまたは複数の例について述べる際、プログラムのソース (S) バイナリと、プログラムのターゲット (T) バイナリについて論じる。T と S の関係は、次のように数学的に記述することができる。

$Diff(S, T) =$

$Patch(S,) = T$

すなわち言い換えれば次のようになる。

SとTの差異がデルタであり、

SをデルタでパッチするとTになる。

【0050】

本明細書で述べるように、例示的な実施形態は を十分に最小化するが、SをパッチしてTを再構築する際に完全な正確さを達成する能力は維持する。本明細書では、パッチングに関して、「完全な正確さ」（および「ほぼ同一」、または同様の用語）は、元のTと再構築したTの間での実質的でない差を考慮に入れている。

10

【0051】

マッチング

バイナリプログラム（SやTなど）の間の変更を検出するための前提は、バイナリフラグメントの一致を検出できることである。言い換えれば、プログラムモジュールをパッチできる前に、2つのバイナリプログラム間でどのフラグメントが変更されていないかを決定する必要がある。例えば、プログラムモジュール中のあるフラグメントの一致は、別のプログラムモジュール（おそらく同じモジュールの先または後のバージョン）中で、変更されていない同じフラグメントを見つけることによって検出することができる。

【0052】

単純な例

20

図3～図5に、例示的なデルタジェネレータの方法論的な一実施形態の単純な一例を示す。この例では、最初の目標は、入力リストであるソース（S）310およびターゲット（T）320に対して 330を計算することである。

【0053】

図3にソース310を示すが、ソース310は6つのノード311～316を含み、これらのノードの内容はそれぞれH、O、O、V、E、Rである。ターゲット320は6つのノード321～326を含み、これらのノードの内容はそれぞれH、O、O、P、E、Rである。ソース310と比較して、ターゲット320は大きく変更されていない。違うのはノード314（「V」を含む）であり、これは、ターゲット中で欠けているが、ノード324（「P」を含む）で置換されている。

30

【0054】

例示的なデルタジェネレータは、変更されていないノードをマッチングする。

内容および相対的な配置に基づいて、ノード314および324を除く各ノードをマッチングする（すなわち識別する）。例えば、ノード311と321は、これらの内容（具体的には「H」）および相対位置が同一なのでマッチする。同様に、ノード312と322、ノード313と323、ノード315と325、ノード316と326がマッチする。

【0055】

図3に示すように、デルタ330は、ノード337の内容としてノード324の内容（「P」）を含む。デルタ330はまた、この新しいノードをどのようにソース中にパッチするかを指定する「編集」も含む。具体的には、デルタ330は、編集ボックス339内で「ADDEDGE(313, 337); ADDEDGE(337, 315)」を指定する。

40

【0056】

図4に、新しいノード337を挿入した後でソース310に追加すべき2つのエッジがあることを示す。したがってデルタは、ソース310中のノード313から新しいノード337へのエッジ339aを追加することになる編集操作ADDEDGE(313, 337)と、ノード337からソース中のノード315へのリンク339bを追加することを指定する編集操作ADDEDGE(337, 315)を含む。

【0057】

図5では、340aにソースとデルタを結合したものを示す。デルタ（新しいノード33

50

7 および新しい2本のエッジ (3 3 9 a および 3 3 9 b)) の追加により、ソースは、リストターゲット 3 2 0 の正確なコピーに変形する。ターゲットのこの正確なコピーは、再構築したターゲット 3 4 0 b と呼ぶことができる。

【 0 0 5 8 】

図 4 と図 5 を比較すると、ソースのノード 3 1 3 と 3 1 4 の間のエッジ 3 1 4 a と、ソースのノード 3 1 4 と 3 1 5 の間のエッジ 3 1 4 b が削除されていることに気付く。この削除は、エッジ 3 3 9 a および 3 3 9 b を追加することによって非明示である。

【 0 0 5 9 】

マッチするブロックの発見

よいブロックのマッチングは、編集操作を最小限にするマッチングである。内容の類似性だけに基づいて2つのブロックをマッチングすると、よいマッチングにならないことがある。これは、内容が同一であってそれぞれのグラフにおける場所が異なるだけのブロックが、通常、ソース中ならびにターゲット中に複数あるからである。例えば、図 6 に示すサブグラフを考えてみる。

【 0 0 6 0 】

図 6 には、ソース 4 1 0 およびターゲット 4 2 0 のサブグラフが示されている。この例では、内容からみて (4 1 2 、 4 1 6 、 4 2 2) 、 (4 1 3 、 4 1 7 、 4 2 3) 、 (4 1 4 、 4 1 8 、 4 2 4) 、 (4 1 1 、 4 2 1) 、 (4 1 5 、 4 2 5) の各グループ中のブロックはすべて同一であると仮定する。またこの例では、2つのブロックの内容が同一であればそれだけで、それらのブロックをマッチングすることができるとも仮定する。ブロック 4 2 2 をブロック 4 1 2 とマッチングする場合 (両方とも「 A 」を含む) 、他のブロックの最良のマッチングは次のようになる。

ブロック 4 1 1 (D_s) とブロック 4 2 1 (D_t)

ブロック 4 1 5 (E_s) とブロック 4 2 5 (E_t)

ブロック 4 1 3 とブロック 4 2 3 (両方とも「 B 」を含む)

ブロック 4 1 4 とブロック 4 2 4 (両方とも「 C 」を含む)

【 0 0 6 1 】

この場合、このマッチングに従ってローカルエッジ構造が保存されるので、編集操作は必要ない。一方、ブロック 4 2 2 をブロック 4 1 6 とマッチングする場合 (両方とも「 A 」を含む) 、他のブロックの最良のマッチングは次のようになる。

ブロック 4 1 1 (D_s) とブロック 4 2 1 (D_t)

ブロック 4 1 5 (E_s) とブロック 4 2 5 (E_t)

ブロック 4 1 7 とブロック 4 2 3 (両方とも「 B 」を含む)

ブロック 4 1 8 とブロック 4 2 4 (両方とも「 C 」を含む)

【 0 0 6 2 】

編集操作は、Delete (4 1 1 , 4 1 2) 、Add (4 2 2 , 4 1 6) 、Delete (4 1 6 , 4 1 9) 、Add (4 1 6 , 4 1 5) である。本明細書では、Delete (a , b) は、「ノード a からノード b までのエッジを削除する」ことを意味する。同様に、Add (a , b) は、「ノード a からノード b までのエッジを追加する」ことを意味する。

【 0 0 6 3 】

上の例は、内容だけに基づいている。これらの例は、各ブロックの周囲のブロックを考慮していない。これらの周囲ブロックは、そのブロックの近傍または局所近傍と呼ぶことができる。これらの例は、内容だけに基づく従来のマッチング手法がよいブロックのマッチングを生み出さないことを示す。これは誤ったマッチングさえ生み出す場合もある。

【 0 0 6 4 】

したがって、例示的なデルタジェネレータは、ブロックをマッチングしながらブロックの近傍を考慮する。

【 0 0 6 5 】

近傍の考慮

10

20

30

40

50

例示的なデルタジェネレータは、ブロックの内容および近傍に基づいてブロックをマッチングする。実際には、ソース C F G およびターゲット C F G のいくつかのパスを生み出し、各パスで近傍をサイズの大きい順に考慮する。先にサイズの大きい方の近傍に基づいてマッチングすることが望ましい。こうすることでより正確なマッチが得られる。近傍に基づくマッチを検査した後、最後のパスでブロックの内容だけにに基づいてブロックをマッチングする。必要条件ではないものの、最後に内容をマッチングするのが有利である。というのは、マッチしないブロックを記録するオーバーヘッドは普通、それを同一内容の任意のブロックとマッチングした後に必要となる編集操作のサイズよりも大きいからである。

【 0 0 6 6 】

同一内容のブロックを検出するために、コードブロックに対しては操作コード (o p c o d e) のシーケンスを得て、データブロックに対しては生の再配置不可能データを得ることから始める。前述のように、アドレスは普通、参照先のブロックが同じのままであっても変化するので、アドレスオペランドは内容の一部と見なさない。この問題は、アドレスオペランドを C F G 中のエッジと見なすことによって対処する。

【 0 0 6 7 】

レジスタオペランドは、扱いに注意を要することがある。いくつかのブロックを追加または削除すると、C F G の他のいくつかの近接領域にあるレジスタ割付けが、それらの領域に変更が加えられていなくても変化する場合がある。このようなオペランドは、各グラフのレジスタフロー分析を用いて、あり得るレジスタ名の変更を除外してモジュールをマッチングする。スタック変数のオフセットは普通、ローカル変数宣言の変化によって変化するので、これらは問題を呈する。この問題に対処するために、例示的なデルタジェネレータは、いくつかの即値オペランドが変化していてもブロックをマッチングする。上のそれぞれの場合では、オペランドタイプのシーケンスが同じのままであるときはコードブロックの内容は変化しないと見なす。

【 0 0 6 8 】

C F G P 中のノード の i 近傍 (すなわち局所近傍) は、P に対応する無向グラフ中から i の距離にあるブロックの集合である。したがって、例えば 0 近傍は集合 { V } であり、1 近傍は およびその親と子である。例示的なデルタジェネレータは、ブロックに対するマッチを見つけようとする間に、異なる値の d についてその d 近傍を計算する。ブロック に対し、 からその子のうちの 1 つへのエッジを子エッジとし、 の親から へのエッジを親エッジとする。

【 0 0 6 9 】

例示的なデルタジェネレータは、ノード から開始する C F G の幅優先走査を行いながら、 の i 近傍を計算する。 からの i よりも遠くなるまで、ブロックの子エッジならびに親エッジを (逆方向に) 走査する。別法として、例示的なデルタジェネレータは、すべての親エッジを無視し、子エッジだけを走査して近傍を計算することもできる。

【 0 0 7 0 】

例えば、異なる 3 つのブロック b_1 、 b_2 、 b_3 から呼び出される S 中のプロシージャ f にある最初のブロック を考えてみる。これらすべてのブロックならびにプロシージャ f は T 中で変更されないが、ブロック b_4 からの呼出しが余分に f に追加されると仮定する。 の i 近傍 ($i = 1$ の場合) はこの追加の呼出しによって変化し、したがって、 はこれによりマッチングされずに残ることがある。

【 0 0 7 1 】

しかしこの例で、 f 自体は何の変更も受けておらず、したがって の近傍を計算する間に の親を無視する方が有利な場合があることを考える。時として近傍は、異なるプロシージャ中にあるブロックはどれも除外することができる。

【 0 0 7 2 】

一般に、マッチングの候補のあいまいさを解決する点で、考慮する近傍サイズが大きいほどマッチングは正確である。しかしこれはマッチがより少なくなり、望ましくない。したがって、例示的なデルタジェネレータは、小さい近傍サイズ (通常 3 つまたは 2 つ以下)

10

20

30

40

50

を、後で述べる「ランダムウォーク」を行うことによって得られるより長いランダム成分と共に用いる。この「ランダムウォーク」ヒューリスティックがよく機能することが、実験から観測された。

【 0 0 7 3 】

「ランダムウォーク」の一例を次に示す。ある人が、いくつかのバス停 A、B、...、N（この順序で）のある、まっすぐな道でランダムウォークを行わなければならないと仮定する。この人は、A（道の一端）から出発する。この人は右にしか行けないので、右すなわち B に行かなければならない。次に B では、2 つの選択肢がある。左（A）に行くか右（C）に行くかである。統一されたランダムウォークモデルでは、公平なコインを投げて、表が出たら一方に行き、裏が出たら他方に行く決めてる。到達したあらゆるバス停でこれを続けると、ランダムウォークを行ったことになる。

10

【 0 0 7 4 】

同じ概念を、直線に代えて一般的なグラフに容易に拡張することができる。以下は、前述の「ランダムウォーク」の例示に基づいたヒューリスティックである。無向グラフを R とするが、この無向グラフは、C F G のノードを頂点とし、C F G 中の対応する頂点の対がいずれかの方向の（有向）制御フローエッジを有する場合に限って頂点の対の上にエッジがある。明らかに、R は最大次数 3 のグラフである。元のノード w から開始して、いずれかのノード x で、d x エッジのうちの 1 つを一樣な確率でとる（d x は x の次数である）。プロシージャ境界（別のプロシージャの呼出しまたはそれへの分岐）に遭遇したときに中止するか、あるいはパスの長さが所定の限度を超えたときに中止する。

20

【 0 0 7 5 】

他のどんなブロックとも内容がマッチしないブロック b を考えてみる。このようなブロックがいずれかのブロックの近傍に含まれている場合、これはのマッチングを妨げる場合がある。というのは、を何らかのブロックとマッチングするには、その全近傍をマッチングしなければならないからである。このブロック b を、「不良」ノードまたはアウトライア（outlier）と呼ぶ。このようなアウトライアは、マッチングを行う前にフィルタリングして、これらの近隣すべての近傍から除去することができる。

【 0 0 7 6 】

例示的なデルタジェネレータは、漸進的に緩和される基準に基づいて、C F G のいくつかのパスでマッチを検出する。例示的なデルタジェネレータは事前に、マッチするブロックの数の事前推定値に基づいて S と T のプロシージャをマッチングする。プロシージャの例には、バイナリプログラムのルーチンおよびサブルーチンがある。

30

【 0 0 7 7 】

この事前マッチングは、大域レベルで行う（すなわち、ブロックが位置するプロシージャに関係なくブロックをマッチングする）。その後、マッチするプロシージャにブロックが属する場合に限り、これらのブロックをマッチングする。このようなマッチングはまた、局所マッチングと呼ぶこともできる。プロシージャマッチ情報を用いて呼出し命令のターゲットを検査して、異なる 2 つの呼出しが同じ関数（名前が変更されているか最初のブロックが異なる場合もある）をターゲットにしているかどうかを検出することもできる。

【 0 0 7 8 】

マッチングパスの間、現在のマッチ基準を使用して、各ノードの短いハッシュ値を計算する。ブロックのハッシュ値は次のようにして計算する。

40

【 0 0 7 9 】

その d 近傍（何らかの適した値 d に対する）中の各ブロックをその内容に基づいてハッシュして、各ブロック自体に対するラベルを作成する。

【 0 0 8 0 】

これらのラベルすべてを連結し、得られるストリングを再度ハッシュして、に対する単一のラベルを作成する。これを、このブロックの「d ラベル」と呼ぶ。

【 0 0 8 1 】

各グラフ中のノードをこれらのラベルによってソートし、同一のラベルおよびマッチする

50

プロシージャを有するブロックをマッチングする。

【0082】

マッチングを向上させるために、他のいくつかのヒューリスティックも使用する。本明細書では、これらのヒューリスティックについては特に後述の「追加の実施形態の詳細」セクションで述べる。例えば2つのブロック c_1 および c_2 が、何らかの i に対して前述の順序付けに従ってそれぞれブロック b_1 および b_2 の i 番目の子である場合、これらのブロックを b_1 および b_2 の対応子と呼ぶ。マッチングをより精密するのに使用するヒューリスティックの一例は次のとおりである。すなわち、マッチするブロックが、内容ラベルにおいてマッチする（ただしおそらく d ラベルにおいてはマッチしない）対応子を有する場合、これらの子を実行する。

10

【0083】

この段階の終わりには、例示的なデルタジェネレータは、 S と T のブロックの部分的マッチングを計算し終えている。

【0084】

編集の計算

マッチングを完了した後、次の段階は の計算である。図3～図5の小さな例で の計算を簡単に示した。図7に、 を生み出す際に、上で計算したマッチングをどのようにグラフ（すなわち CFG ）の場合に使用できるかを示す一例を示す。

【0085】

図7には、ソース510およびターゲット520のサブグラフが示されている。マッチング段階で、511と521（「A」）、512と522（「B」）、513と523（「C」）、516と526（「E」）のブロックの対がマッチしたと仮定する。ターゲットサブグラフ520中のマッチしなかったブロックは、ブロック525（ F_t ）およびブロック524（ D_t ）である。ブロック525（ F_t ）およびブロック524（ D_t ）の内容と、編集操作 $Add(512, 525)$ および $Add(512, 524)$ とを が含んでいれば、ソースサブグラフ510からターゲットサブグラフ520を再構築することができる。

20

【0086】

編集操作 $Delete(512, 514)$ および $Delete(512, 515)$ は、に明示的に含める必要はない。これらは非明示的に含まれる。本発明の趣旨および範囲を逸脱することなく非明示的な編集を特に に含めることもできることは、当業者なら理解する。しかし、非明示的な編集は のサイズを不必要に増大させるので、本明細書に述べる例示的な実施形態の には含めない。

30

【0087】

したがって例示的なデルタジェネレータは、ターゲット中のマッチしなかったブロックの内容および追加する必要があるエッジを（ある時点で）出力する。これが を構成する。

【0088】

例示的なデルタジェネレータの方法論的な実施形態

図8および9に、デルタジェネレータシステム100（またはその一部）によって行う、例示的なデルタジェネレータの方法論的な実施形態を示す。これらの方法論的な実施形態は、ソフトウェア、ハードウェア、またはこれらの組合せにおいて行うことができる。

40

【0089】

図8には主に、例示的なデルタジェネレータの方法論的な実施形態の「マッチング」段階を示す。図9には主に、例示的なデルタジェネレータの方法論的な実施形態の「編集」段階を示す。「編集」段階は、「マッチング」段階の後に続く。「編集」段階の結果がデルタである。

【0090】

図8では、620で、例示的なデルタジェネレータが、コードブロックに対しては操作コードのシーケンスを得て、データブロックに対しては生の再配置不可能データを得るのが示されている。このようなブロックは、ソースプログラムバイナリ（ S ）612からのブ

50

ロシージャのブロックと、ターゲットプログラムバイナリ (T) 6 1 4 からのプロシージャのブロックである。

【 0 0 9 1 】

6 2 4 で、例示的なデルタジェネレータは、ブロックに対するマッチを見つけようとする間に、異なる d の値についてその d 近傍を計算する。さらに、幅優先走査によって d 近傍を計算する。

【 0 0 9 2 】

6 2 8 で、各ノードについて短いハッシュ値を計算する。さらに、例示的なデルタジェネレータは、ノードの d 近傍 (すなわち局所近傍) でハッシュ値に基づいてノードに対するラベルを決定する。6 2 8 ではまた、いくつかの即値オペランドが変更されている場合でも、例示的なデルタジェネレータはブロックをマッチングする。6 3 0 で、各グラフ中のノードをこれらのラベルによってソートし、同一ラベルおよびマッチするプロシージャを有するブロックをマッチングする。この段階の終わりの 6 3 2 で、例示的なデルタジェネレータは、S と T のブロックの部分的なマッチングを計算し終えている。この逆の、マッチしないブロックもまた計算する。

【 0 0 9 3 】

デルタジェネレータのこの方法論的な実施形態の「マッチング」段階に関するこれ以上の詳細は、後述の「追加の詳細」セクションに提供する。

【 0 0 9 4 】

図 9 では、7 2 0 で、例示的なデルタジェネレータが P_t (すなわちターゲットプログラムバイナリからのプロシージャ) 中のマッチしないブロックの内容 7 1 0 を受け取るのが示されているが、この内容 7 1 0 は、図 8 の方法論的な実施形態からの出力の一部である。すべてのエッジ情報はブロックの内容から除外する。したがって、内容が記録されているマッチしないブロック がアドレスオペランド (コードブロックの場合) またはポインタ (データブロックの場合) を有する場合は、エッジ編集情報を用いてこのようなオペランドまたはポインタを訂正しなければならないことを示す「ダミー」アドレスでこれらのオペランドまたはポインタを置換する。

【 0 0 9 5 】

P_s および P_t は、マッチング段階でマッチした (S および T それぞれの中の) すべてのプロシージャの対を表す。ある対につき、 P_s はソースからのプロシージャであり、 P_t は T からのマッチするプロシージャである。例示的なデルタジェネレータは、ソースプログラムバイナリ (S) 中のプロシージャ P_s 7 1 2 と、ターゲットプログラムバイナリ (T) 中のプロシージャ P_t 7 1 7 との対を受け取る。

【 0 0 9 6 】

7 3 0 で、 P_s のリンクリストから P_t のリンクリストを再構築するのに必要なすべての情報を、計算した の一部として記録する。プログラムのアドレス空間における各プロシージャのレイアウト (すなわちブロックの配置) をブロックのリンクリストによって取り込むが、これにより、このリストの先頭から末尾まで順次走査すると、プロシージャのレイアウトを正確に記述することができる。

P_s から P_t を再構築する際の最初のタスクは、 P_s のリンクリストから P_t のリンクリストを再構築することになる。したがって、一方のリンクリストから他方のリンクリストを再構築するのに必要なすべての情報を、計算した の一部として記録する。

【 0 0 9 7 】

7 3 2 で、エッジ編集操作を計算する。前述のように、エッジ編集操作はエッジの明示的な追加である (非明示的な削除ではない) 。各エッジは、そのソースおよびターゲットによって完全に指定することができる。編集のコンテキスト内で、用語「ソース」は、このエッジが表す正確なオペランド (コードブロックの場合) または再配置可能ワード (データブロックの場合) を指す。さらに、編集のコンテキスト内で、用語「ターゲット」は一方、このオペランドまたは再配置可能ワードの参照先であるアドレスのブロックを表す。例示的なデルタジェネレータは、ソースおよびターゲットに対する固有の識別子を別々の

10

20

30

40

50

リストに記録することにより、編集したエッジ（追加または削除したエッジ）を記録する。

【0098】

734で、ソース中の所与のプロシージャ（ P_s ）に対するデルタを出力する。このようなデルタは、 P_t 中のマッチしないブロックと732で計算したエッジ編集操作を含む。

【0099】

例示的なデルタジェネレータの「編集」段階に関する前述の方法論的な実施形態は、レジスタ名、即値、および操作コードにおけるいくつかの小さい変更まで、ターゲットプログラムバイナリを再構築するのに十分な情報を記録する。変更されたこれらのレジスタ名、即値、および操作コードは、別々のリストに記録する。

10

【0100】

追加の実施形態の詳細

例示的な実施形態のマッチング段階ではCFGのいくつかのパスを生み出すが、以下に、各パスについてより詳細に述べる。

【0101】

マッチング段階における複数パス

マッチング段階における各パスは、SとT中のブロックを比較するのに異なる基準を使用するが、これは各ブロックについてハッシュ値またはラベルを計算することによって行う。例示的な実施形態では、例示的な2つのサブプロセスを採用し、これらをいくつかのパスで呼び出す。本明細書では、これらをComputeLabelおよびComputeDLLabelと呼ぶ。

20

【0102】

ComputeLabel：ブロックの内容のハッシュ値を計算する。これはブロックの内容ラベルとも呼ぶ。標準的なハッシュを使用してハッシュ値を計算する。このような標準的なハッシュの一例はMD5であり、これは、所与のバイトストリングのハッシュとして16バイトストリングを生成する。

【0103】

このComputeLabelと呼ぶ例示的なサブプロセスは、異なるマッチ基準を可能にするようにパラメータ化されており、ブールパラメータは、Immediate、RegisterChain、OpcodeString、およびProcIdである。Immediateは、コードブロック中の即値オペランドをハッシングで使用するべきかどうかを示す。同様に、RegisterChainは、レジスタ名変更を使用すべきかどうかを指定し、OpcodeStringは、操作コードがその正確な名前で表されるか、あるいはそのグループ識別子（「グループID」）で表されるかを示す。

30

【0104】

類似する操作コードにグループIDを使用すると、機能に影響しない、命令の小さな変更が捉えられる。例えば、命令「jge eax, 10」が「jle eax, 10」に変わったのが唯一の変更であるSの一部を考えてみる。これは例えば、不等式の向きが正しくないバグが発見された場合に起こり得る。このような場合、「jge」と「jle」の両方をこれらの操作コードグループ「branch」で表すことができ、この結果、対応するブロックをマッチングすることができる。

40

【0105】

第4のパラメータ、ProcIdは、プロシージャマッチングがすでに行われたかどうかを示し、すでに行われた場合は、プロシージャはマッチするプロシージャに合う固有の識別子を有する。このパラメータが設定されている場合、関数呼出しのターゲットは、それらのプロシージャ識別子で表される。サブプロセスComputeLabelに対する擬似コードの一例は、以下のとおりである。

ComputeLabel (Block, Immediate, RegisterChain, OpcodeString, ProcId)

If (Blockがデータブロックである)

 ブロック中の再配置可能な各ワードを反復し、クリアしてゼロにする。

 得られるデータバッファをハッシュし、ハッシュ値を返す。

Endif

If (Blockがコードブロックである)

 ハッシュすべきデータを保持できるバッファを生み出す。 10

 最後の命令を除く各命令ごとに:

 If (OpcodeStringがtrueである)

 バッファ中の操作コードストリングを記録する

 Else

 バッファ中の操作コードグループを記録する

 End

 命令中の各オペランドごとに: 20

 Case OperandType

 Pointer:定数ストリング「Block」を記録する。

 Register:If (RegisterChainがtrueである)

 レジスタフローIDを記録し、レジスタフローID

を定義する (先に計算する)。

 Else

 レジスタ名を記録する。 30

 Endif

 Immediate:If (Immediateがtrueである)

 即値定数を記録する

 Endif

 End

 End

End 40

 最後の命令に対して:

```

    If (OpcodeStringがtrueである)
        バッファ中の操作コードストリングを記録する
    Else
        バッファ中の操作コードグループを記録する
    End
    If (ProcIdがtrueである)
        命令のターゲットブロックを計算する。
        Ifターゲットが異なるプロシージャ中にある
            ターゲットプロシージャのプロシージャIDを記録する
        Endif
    Endif
    End
    バッファをハッシュしてハッシュ値を返す。
Endif
End ComputeLabel

【0106】
ComputeDLablel：内容ならびに近傍に基づいてハッシュ値を計算する。これは、すでに計算した全近傍（ルートノードを含む）を検査し、これらの内容ラベルを連結する。次いで、連結したこのリストのハッシュ値を返す。

【0107】
この実施形態は、標準的なハッシュを使用してハッシュ値を計算する。このような標準的なハッシュの一例はMD5であり、これは、所与のバイトストリングのハッシュとして16バイトストリングを生成する。

【0108】
例示的なデルタジェネレータのマッチング段階の方法論的实施形態に関する追加の詳細図10に、例示的なデルタジェネレータのマッチング段階に関する広範な方法論的实施形態を示す。この広範な方法論的实施形態は、デルタジェネレータシステム100（またはその一部）によって実施することができる。この広範な方法論的实施形態は、ソフトウェア、ハードウェア、またはこれらの組合せにおいて実施することができる。

【0109】
以下に、この広範な方法論的实施形態の追加の詳細について述べる。
タスク1（図10の810）：基本ブロックの事前マッチング
このタスク（図10の810）は、基本ブロックを大域レベルでマッチングする。
・例示的なデルタジェネレータは、SならびにT中の各ブロックごとにComputeLabelを呼び出し、それに内容ラベルを割り当てる。この段階では、ComputeLabelに対する4つのパラメータはそれぞれ偽である。
・例示的なデルタジェネレータは、計算したラベルに基づいてSのブロックをTのブロックとマッチングする。S中の重複ブロック（すなわち同じハッシュ値を有するブロック）は、同じラベルを有するT中のブロックのいずれかとマッチングする。2つのブロックをマッチングするときは常に、後で参照するために各ブロックに固有のMatchIdを割り当てる。このタスクの後でマッチしないブロックは、アウトライアとして指定する。
・例示的なデルタジェネレータは、各ブロックについて、このブロックから開始する幅優先走査を行うことによってd近傍（d=2）を計算する。走査では、ブロックに子がない

```

10

20

30

40

50

場合はその親に行く。

- ・例示的なデルタジェネレータは、上で計算した不良ノードをフィルタリングして、すべての d 近傍から除去する。

- ・例示的なデルタジェネレータは、各ブロックごとに `ComputeDLabe1` を呼び出し、それに d ラベルを割り当てる。 d ラベルは呼出しから返される値である。

- ・例示的なデルタジェネレータは、同一の d ラベルを有する S と T のブロックをマッチングする。

- ・いずれかの2つのマッチするブロック b_1 および b_2 は、それから出て行くエッジ (`out-edge`) を同数有さなければならない。これらの出て行くエッジを、ブロック中のそれらの位置によって順序付ける。2つのブロック c_1 および c_2 が、何らかの i に対して前述の順序付けに従ってそれぞれ b_1 および b_2 の i 番目の子である場合、この2つのブロックを b_1 および b_2 の対応子と呼ぶ。例示的なデルタジェネレータは、以下の規則を用いてマッチングをより精密にする。すなわち、マッチするブロックが、内容ラベルにおいてマッチする (ただしおそらく d ラベルにおいてはマッチしない) 対応子を有する場合、これらの子をマッチングする。

【0110】

タスク2 (図10の812) : プロシージャのマッチング

このタスク (図10の812) は、上で計算した事前マッチングを用いてプロシージャをマッチングする。

- ・例示的なデルタジェネレータは、 T_s 中の各プロシージャ P_t を反復し、 S 中の各プロシージャ P_s について、 P_s 中のブロックにマッチする P_t 中のブロックの数 $m_{s,t}$ を計算する。

- ・例示的なデルタジェネレータは、プロシージャの対をそれらの $m_{s,t}$ の値に基づいてマッチングする。 $m_{s,t}$ の値は、プロシージャマッチ基準と呼ぶこともできる。 $m_{s,t}$ の値を用いてプロシージャをマッチングするこの技法については、後で考察する。

- ・例示的なデルタジェネレータは、マッチするプロシージャに同じ固有の識別子を割り当てる。例示的なデルタジェネレータは、他のすべてのプロシージャにも固有の識別子を割り当てる。

【0111】

プロシージャをマッチングするための例示的なデルタジェネレータは、次の条件を満たす。すなわち、任意の P_s と任意の2つ P_{t1} および P_{t2} があるとして、 P_s を P_{t1} にマッチングする場合は $m_{s,t1} = m_{s,t2}$ であり、あるいは P_{t2} を P_s にマッチングするので $m_{s,t2} = m_{s,t1}$ である。以下の技法がこの条件を満たす。

- ・ $(P_t, P_s, m_{s,t})$ トリプレット s, t を生み出し、これらを $m_{s,t}$ に関してソートする。

- ・ソートしたトリプレットリストを降順で反復する。いずれかの反復で P_t と P_s が両方ともマッチしない場合は、それらをマッチングする。

【0112】

タスク3 (図10の814) : 基本ブロックの局所マッチング

このタスク (図10の814) は、前のタスクで得られたプロシージャマッチング情報を用いて、ブロックの局所マッチングを計算する (すなわち、マッチするプロシージャ間だけでブロックをマッチングする)。

【0113】

マッチするブロックがない状態から始める。例示的なデルタジェネレータは、 S ならびに T 中の各ブロックごとに `ComputeLabel` を呼び出し、それに内容ラベルを割り当てる。`ComputeLabel` を呼び出すとき、`Immediate`、`OpcodeString`、および `ProcId` のパラメータは真である。

【0114】

例示的なデルタジェネレータは、計算した内容ラベルに基づいて S のブロックを T のブロックとマッチングする。 S 中の重複ブロック (すなわち同じハッシュ値を有するブロック

10

20

30

40

50

）は、同じラベルを有するT中のブロックのいずれかとマッチングする。2つのブロックをマッチングするときは常に、後で参照するために各ブロックに固有のMatchIdを割り当てる。ブロックが前に割り当てたMatchIdをすでに有する場合は、それを再度割り当てる。このタスクの後でマッチしないブロックは、アウトライアとして指定する。

【0115】

d = 3、2、1として、例示的なデルタジェネレータは以下のことを行う。各ブロックについて、このブロックから開始する幅優先走査を行うことによってd近傍を計算する。d近傍はランダム成分も有する。ランダム成分は、元のノードから固定長の一様なランダムウォークを行う間に遭遇するブロックの集合である。ランダムウォークは、実際の制御フローエッジだけを見て、一様な確率でそれらを選択する。プロシージャ境界（例えば別のプロシージャの呼出しまたはそれへの分岐）に遭遇したときに停止する。

上で計算した不良ノードをフィルタリングして、すべてのd近傍から除去する。

各ブロックごとにComputeDLabe1を呼び出し、それにdラベルを割り当てる。

同一のdラベルを有するSとTのブロックをマッチングする。T中の2つのブロックがS中の同じブロックにマッチする場合は、それらのうち的一方だけを実際にマッチングする。タイは任意に断つことができる。

以下のガイドラインに従ってマッチングを向上させる。すなわち、マッチするブロックが、内容ラベルにおいてマッチする（ただしおそらくdラベルにおいてはマッチしない）対応子を有する場合、これらの子をマッチングする。

【0116】

いずれの段階でも、前にマッチしたブロックが再度マッチすること、またはマッチしないことはない。

【0117】

例示的なデルタジェネレータは、マッチするブロックの対にそれぞれ固有のMatchIdを割り当てる。以後これを、この2つのブロックに対する内容ラベルとして使用する。

【0118】

タスク4（図10の816）：レジスタ名変更の問題の解決

このタスク（図10の816）は、レジスタ名変更の問題を解決する（すなわち、レジスタ名変更の可能性があってもマッチするブロックを検出する）。例示的なデルタジェネレータは、SおよびT中のすべてのレジスタフロー連鎖を計算し、それぞれに識別子を割り当てる。マッチするブロックの場合は、対応するレジスタフローに同じ識別子を割り当てる。

【0119】

例示的なデルタジェネレータは、RegisterChainを真としてComputeLabe1を呼び出し、各ブロックに内容ラベルを割り当てる。例示的なデルタジェネレータは、これらのラベルおよびd近傍（d = 2）に基づいて、マッチしていないブロックをマッチングする。例示的なデルタジェネレータは、マッチがそれ以上見つからなくなるまでこのタスクを繰り返す。こうするのは、新しいマッチするブロックによってさらにいくつかのレジスタ連鎖に同じ識別子が割り当てられ、さらにマッチが見つかることがあるからである。

【0120】

これについては後で、「レジスタ名変更の問題の解決」という題のセクションでより詳細に論じる。

【0121】

タスク5（図10の818）：最終パス

最後のマッチングタスク（図10の818）で、同一の内容ラベル（前のタスクで計算したもの）を有するブロックの対があればそれらをマッチングする。

【0122】

レジスタ名変更の問題の解決

例示的なデルタジェネレータは、各プロシージャについてその `use - def` 連鎖、`def - use` 連鎖、およびドミネータ情報を計算する。`use - def` 連鎖は、レジスタ `use` から開始して、この `use` に到達するそのレジスタのすべての定義 (`def`) を通るリストである。同様に、`def - use` 連鎖は、レジスタ `def` から開始して、この `def` により到達する同じレジスタの各 `use` を通るリストである。ブロック `d` を、別のブロック `b` のドミネータと呼ぶ。`b` は `d` から到達可能であり、プロシージャの入口点から `b` へのどんなパスも最初に `d` を通過しなければならない。通常、ドミネータ情報の計算は、プロシージャ中のあらゆるブロックのすべてのドミネータを計算することを意味する。

【0123】

10

例示的なデルタジェネレータは、各プロシージャ中の各基本ブロックごとに、ブロック中のすべてのレジスタ `def` を見て、`def id` と呼ぶ各ブロック固有の ID をそれぞれに割り当てる。例示的なデルタジェネレータは、そのブロック中のすべてのレジスタ `use` を見て、いずれかの `use` が同じブロック中で唯一の `def` を有する場合に、その `use` に `def` 識別子を割り当てる。したがってこの段階では、すべての `def` およびいくつかの `use` に ID が関連付けられている。未定義の ID を有するレジスタ `use` がいくつかある場合もある。`ComputeLabel` (上で定義した) が `RegisterChain = true` を伴って呼び出されたとき、ID が定義されていない場合はそれらの名前を使用する。

【0124】

20

例示的なデルタジェネレータは、`S` および `T` 中のマッチしない各ブロックごとに `ComputeLabel` (`RegisterChain = true` を伴う) を呼び出し、返されるラベルをその内容ラベルに割り当てる。例示的なデルタジェネレータは、マッチしない各ブロックについて `d` 近傍 (`d = 2`) を計算する。

例示的なデルタジェネレータは、`ComputeDLLabel` を呼び出してそれに `d` ラベルを割り当て、新たに計算した `d` ラベルに基づいて `S` と `T` をマッチングする。

【0125】

コードブロックの現在マッチングしているある対 (`bs`、`bt`) に対して、例示的なデルタジェネレータは以下のことを行う。

【0126】

30

2つのブロック中の対応する (同じ場所にある) レジスタ `def` を反復する。

【0127】

プロシージャ全体で固有の固有識別子 (正) を、現在検査中の2つの `def` それぞれに割り当てる。

【0128】

これらの識別子は、その `def` が有することのある、前に割り当てたどんな `def` 識別子にも置き換わる。2つの `def` がマッチすることを記録し、これらの両方に単一の識別子を関連付ける。

【0129】

40

次に、例示的なデルタジェネレータは、新たに割り当てた `def` 識別子に基づいてレジスタ `use` をマッチングする。例示的なデルタジェネレータは、`S` と `T` の両方に対して各基本ブロック中のすべてのレジスタ `use` を反復する。このような各 `use` について、それに到達しそれに支配されないすべての `def` を検査する。例示的なデルタジェネレータは、それらの `def` 識別子を合計し、合計を現在の `use` の識別子に割り当てる。したがって、各レジスタオペランド、`def` または `use` に識別子が関連付けられる。マッチするブロックは、対応するレジスタオペランドに関連するマッチング識別子を有する。より具体的には、正確にマッチする `def` を有する2つのレジスタオペランドは、レジスタ名が異なる場合があっても、マッチング識別子も有することができる。

【0130】

例示的なデルタジェネレータが `RegisterChain = true` を伴って `Comp`

50

uteLabelを呼び出すこと（前述）を一番最近に実行した際に新しいマッチが発見された場合、プロセスはそこにループバックする。

【0131】

例示的なデルタジェネレータの編集段階の方法論的な実施形態に関する追加の詳細

例示的なデルタジェネレータは、binと呼ぶ新しいダミーCFGを生み出すが、このbinは、Tのすべての追加ブロックのためのコンテナとして働くことのできる単一プロシージャを有する。これらのブロックは、例えばリンクリストの形で維持する。

【0132】

例示的なデルタジェネレータは、T中の各ブロックを順次走査し、マッチしない各ブロックに連続した識別子（0で始まる）を割り当てる。例示的なデルタジェネレータはまた、マッチしないブロックをbinにダンプして、コードブロック中のアドレスオペランドおよびデータブロック中のポインタが、これらのポインタを含むブロックを参照する「ダミー」アドレスに確実に修正されるようにする。

10

【0133】

例示的なデルタジェネレータは、S中の各ブロックを順次走査し、マッチする各ブロックに連続した識別子を割り当てる。同じ識別子を、T中のマッチするブロックに割り当てる。このときおよび前のタスクで割り当てる識別子をNew_Node_Idと呼ぶ。

【0134】

例示的なデルタジェネレータは、ソースグラフ中のマッチするブロックを検査する。S中のブロックxがT中のブロックyとマッチする場合、それぞれから出るエッジをチェックする。各エッジx z₁につき、対応するy z₂がなければならない。

20

ケースi z₂がz₁とマッチする：何もしない。

ケースii z₂がz₃ z₂とマッチする：リンクx z₃を記録する。

ケースiii z₂がマッチしない：リンクx z₂を記録する。

【0135】

エッジは次のように記録する。TargetId、OperandIndex、およびSourceProcFirstBlockという名前の3つのリストを生み出す。エッジx yを記録する場合、yのNew_Node_IdをTargetIdに挿入する。このエッジが現プロシージャのi番目のエッジであった場合、iをリストOperandIndexにプッシュする。最後に、S中のいずれかのプロシージャP_sのエッジが少なくとも1つ記録されている場合、その最初のブロックのNew_Node_IdをSourceProcFirstBlockにプッシュする。

30

【0136】

例示的なデルタジェネレータは、ターゲットグラフ中のマッチしないブロックすべてを検査する。例示的なデルタジェネレータは、それぞれから出るエッジx yを見る。yもまたマッチしないブロックである場合は、xからyに行くエッジをbin中に生み出す。yがソースグラフのzとマッチする場合は、zのNew_Node_IdをTargetOperandTargetsと呼ぶリストにプッシュすることにより、リンクx zを記録する。

【0137】

40

例示的なデルタジェネレータは、T中のマッチするプロシージャP_tを有するS中の各プロシージャP_sごとに、P_tのブロックのリンクリストを走査する。

走査のいずれかの時点で、ptr₁およびptr₂がT中の現ノードおよび次のノードを指す場合、mptr₁およびmptr₂をこれらの「マッチング」ポインタとする。ptr₁がT中のブロックxを指す場合、そのマッチングポインタmptr₁は、S中にマッチするブロックがあればそれを指すか、あるいはbin中にダンプしたxのコピーを指す。走査のいずれかの段階で、mptr₁およびmptr₂がリスト中で（この順序で）隣接しないノードを指す場合、エッジmptr₁ mptr₂を記録する。これは、これらに対応するブロックのNew_Node_IdをLinkedListsLeftおよびLinkedListsRightと呼ぶ別々のリストにプッシュすることによって行う

50

。さらに、 P_t の最初のブロックのNew__Node__Idを、後の部分のサイズ（ブロック数で表す）および P_t 中の記録済みリンクの数と共に、ProcedureInfoと呼ぶ別個のリストに記録する。

【0138】

例示的なデルタジェネレータは、上で計算した編集情報をSと共に用いてTを再構築する。

【0139】

別法として、例示的なデルタジェネレータは、元バージョンのTと再構築バージョンのTとで異なるレジスタ名、即値、および操作コードを検出し、これらの差異をMinorDifferencesと呼ぶ別個のリストに出力することもできる。

10

【0140】

小さな差異

デルタジェネレータの例示的な一実施形態では、これらのリストすべてがbinと共に、計算された を構成する。

【0141】

デルタ圧縮

以上のセクションは、Sおよび がある場合に例示的なデルタジェネレータを使用してどのようにTのコピーを再構築するかを十分に述べている。図2に示すように、生成した はサーバ212からクライアント222に送信することができる。

【0142】

20

この は、ネットワーク215上で利用可能な帯域幅をより効率的に使用するために圧縮することができる。したがって例示的なデルタジェネレータは、パッチをよく圧縮できる方式でフォーマットすることができる。パッチはいくつかの部分に分割するが、それぞれの部分は異なる種類の情報を保持し、したがって各部分は別々に圧縮することができる。この手法は、異なるタイプのデータ（例えばコードに対して整数リスト）は異なる圧縮アルゴリズム（すなわちエンジン）で最もよく圧縮できるという観測結果に基づく。

【0143】

この手法は、先に「追加の実施形態の詳細」セクションで述べたステップと組み合わせることができる。例示的なデルタジェネレータにおいて、本明細書でいう「いくつかの部分」とは、前述の編集段階で言及した異なるいくつかのリスト、およびそこで言及したダミーマー「bin」である。一実施形態では、これらの各リスト（およびbin）を別々のファイルにストアし、したがって別々に圧縮する。使用できる圧縮アルゴリズムの一例は、LZWである。

30

【0144】

これは、大きな実行可能ファイル（約数百キロバイト）の場合によりよく機能すると思われる、我々の最初の実験では、このような大きな実行可能ファイルの場合に、例示的なデルタジェネレータは平均して、我々の知る利用可能な他のどんなパッチングツールよりも小さなパッチを作成することが示されている。

【0145】

例示的な実施形態に対する適用例

40

例示的なデルタジェネレータの実施形態に対して考えられる適用例には、帯域幅の制約のある媒体を介してソフトウェアのアップグレードまたはパッチを送信するためのツールとしての適用例がある。インターネット上でソフトウェアを入手できることが日々ますます一般的になりつつある中で、このようなパッチングツールはソフトウェアの使用にとって有益であろう。

【0146】

例示的なデルタジェネレータの実施形態に対して考えられる別の適用例は、ソフトウェアの海賊行為防止の分野にある。例示的なデルタジェネレータの一実施形態を用いれば、海賊行為防止ツールがプログラムの類似性を検出することができる。考えられる別の適用例は、ソフトウェアの異なるバージョン間の変更の性質を分析することにある。

50

【 0 1 4 7 】

例示的なコンピューティング環境

図 1 1 に、例示的なデルタジェネレータを例えば実装できる、好適なコンピューティング環境 9 2 0 の一例を示す。

【 0 1 4 8 】

例示的なコンピューティング環境 9 2 0 は、好適なコンピューティング環境の一例でしかなく、例示的なデルタジェネレータの使用または機能の範囲に関するいかなる制限も示すものではない。コンピューティング環境 9 2 0 は、例示的なコンピューティング環境 9 2 0 中に示すコンポーネントのいずれか 1 つまたは組合せに関して、いかなる依存または要件も有すると解釈すべきではない。

10

【 0 1 4 9 】

例示的なデルタジェネレータは、他の多くの汎用または専用コンピューティングシステム環境または構成で動作可能である。例えば例示的なデルタジェネレータと共に使用するのに適した、周知のコンピューティングシステム、環境、および/または構成の例としては、パーソナルコンピュータ、サーバコンピュータ、シン(t h i n)クライアント、シック(t h i c k)クライアント、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能な消費者電子機器、ネットワーク P C、ミニコンピュータ、メインフレームコンピュータ、前述のシステムまたはデバイスのいずれかを含む分散コンピューティング環境などが挙げられるが、これらに限定されない。

20

【 0 1 5 0 】

例示的なデルタジェネレータは、例えば、コンピュータによって実行されるプログラムモジュールなどのコンピュータ実行可能命令の一般的なコンテキストで述べる。一般にプログラムモジュールは、特定のタスクを実行するか特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。例示的なデルタジェネレータはまた、通信ネットワークを介してリンクされたリモート処理装置によってタスクが実行される分散コンピューティング環境で実施することもできる。分散コンピューティング環境では、プログラムモジュールは例えば、メモリ記憶装置を含めたローカルとリモートの両方の記憶媒体中にある。

【 0 1 5 1 】

図 1 1 に示すように、コンピューティング環境 9 2 0 は、コンピュータ 9 3 0 の形をとる汎用コンピューティングデバイスを含む。コンピュータ 9 3 0 のコンポーネントには、1 つまたは複数のプロセッサまたはプロセッシングユニット 9 3 2 と、システムメモリ 9 3 4 と、システムメモリ 9 3 4 を含めた種々のシステムコンポーネントをプロセッサ 9 3 2 に結合するバス 9 3 6 を含めることができるが、これらに限定しない。

30

【 0 1 5 2 】

バス 9 3 6 は、メモリバスまたはメモリコントローラ、周辺バス、A G P (a c c e l e r a t e d g r a p h i c s p o r t)、および種々のバスアーキテクチャのいずれかを使用するプロセッサバスまたはローカルバスを含めた、いくつかのタイプのバス構造のうちのいずれか 1 つまたは複数を表す。限定ではなく例として、このようなアーキテクチャには、I S A (I n d u s t r y S t a n d a r d A r c h i t e c t u r e) バス、M C A (M i c r o C h a n n e l A r c h i t e c t u r e) バス、E I S A (E n h a n c e d I S A) バス、V E S A (V i d e o E l e c t r o n i c s S t a n d a r d s A s s o c i a t i o n) ローカルバス、およびメザニンバスとも呼ばれる P C I (P e r i p h e r a l C o m p o n e n t I n t e r c o n n e c t s) バスが含まれる。

40

【 0 1 5 3 】

コンピュータ 9 3 0 は通常、種々のコンピュータ可読媒体を備える。このような媒体は、例えばコンピュータ 9 3 0 からアクセス可能な任意の利用可能媒体であり、揮発性媒体と不揮発性媒体、取外し可能媒体と取外し不可能媒体の両方がこれに含まれる。

50

【0154】

図11では、システムメモリは、ランダムアクセスメモリ(RAM)940などの揮発性メモリ、および/または、読出し専用メモリ(ROM)938などの不揮発性メモリの形をとるコンピュータ可読媒体を含む。ROM938には、起動中などにコンピュータ930内の要素間で情報を転送するのを補助する基本ルーチンを含むBIOS(basic input/output system)942がストアされる。RAM940は通常、プロセッサ932からすぐにアクセス可能な、かつ/またはプロセッサ932が現在作用している、データおよび/またはプログラムモジュールを含む。

【0155】

コンピュータ930はさらに、その他の取外し可能/取外し不可能、かつ揮発性/不揮発性のコンピュータ記憶媒体を備えることもできる。例として示すだけだが、図11には、取外し不可能かつ不揮発性の磁気媒体(図示していないが通常「ハードドライブ」と呼ばれる)に対して読み書きするためのハードディスクドライブ944と、取外し可能かつ不揮発性の磁気ディスク948(例えば「フロッピー(登録商標)ディスク」)に対して読み書きするための磁気ディスクドライブ946と、CD-ROM、DVD-ROM、またはその他の光学媒体など取外し可能かつ不揮発性の光ディスク952に対して読み書きするための光ディスクドライブ950とが示されている。ハードディスクドライブ944、磁気ディスクドライブ946、および光ディスクドライブ950はそれぞれ、1つまたは複数のインタフェース954によってバス936に接続される。

【0156】

ドライブおよびそれらに関連するコンピュータ可読媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、およびコンピュータ930に対するその他のデータの、不揮発性記憶域を提供する。本明細書に述べる例示的な環境は、ハードディスク、取外し可能磁気ディスク948、および取外し可能光ディスク952を採用するが、磁気カセット、フラッシュメモリカード、デジタルビデオディスク、ランダムアクセスメモリ(RAM)、読出し専用メモリ(ROM)など、コンピュータからアクセス可能なデータをストアすることのできる他のタイプのコンピュータ可読媒体をこの例示的な動作環境で使用することもできることを、当業者は理解されたい。

【0157】

ハードディスク、磁気ディスク948、光ディスク952、ROM938、またはRAM940には、いくつかのプログラムモジュールが例えばストアされるが、これらのプログラムモジュールには、限定ではなく例として、オペレーティングシステム958、1つまたは複数のアプリケーションプログラム960、その他のプログラムモジュール962、およびプログラムデータ964が含まれる。

【0158】

このようなオペレーティングシステム958、1つまたは複数のアプリケーションプログラム960、その他のプログラムモジュール962、およびプログラムデータ964(またはこれらの何らかの組合せ)はそれぞれ、例示的なデルタジェネレータの一実施形態を含むことができる。より具体的には、これらはそれぞれ、デルタジェネレータシステム、コンパレータ、編集操作決定機構、および出力サブシステムの一実施形態を含むことができる。

【0159】

ユーザは、キーボード966やポインティングデバイス(「マウス」など)968などの入力デバイスを介してコンピュータ930にコマンドおよび情報を入力することができる。その他の入力デバイス(図示せず)には、マイクロホン、ジョイスティック、ゲームパッド、衛星受信アンテナ、シリアルポート、スキャナなどを含めることができる。これらおよび他の入力デバイスは、バス936に結合されたユーザ入力インタフェース970を介してプロセッシングユニット932に接続されるが、例えばパラレルポート、ゲームポート、ユニバーサルシリアルバス(USB)など、他のインタフェースおよびバス構造でも接続される。

【 0 1 6 0 】

モニタ 9 7 2 または他のタイプの表示装置もまた、ビデオアダプタ 9 7 4 などのインタフェースを介してバス 9 3 6 に接続される。モニタに加え、パーソナルコンピュータは通常、スピーカやプリンタなど他の周辺出力デバイス（図示せず）も備え、これらは出力周辺インタフェース 9 7 5 を介して接続される。

【 0 1 6 1 】

コンピュータ 9 3 0 は、リモートコンピュータ 9 8 2 など 1 つまたは複数のリモートコンピュータへの論理接続を用いて、ネットワーク化された環境で動作することができる。リモートコンピュータ 9 8 2 は、コンピュータ 9 3 0 に関して本明細書に述べた要素および機構の多くまたはすべてを備えることができる。

10

【 0 1 6 2 】

図 1 1 に示す論理接続は、ローカルエリアネットワーク（LAN）9 7 7 および一般的なワイドエリアネットワーク（WAN）9 7 9 である。このようなネットワーキング環境は、オフィス、企業全体のコンピュータネットワーク、イントラネット、およびインターネットによくあるものである。

【 0 1 6 3 】

LAN ネットワーキング環境でコンピュータ 9 3 0 を使用するときは、ネットワークインタフェースまたはアダプタ 9 8 6 を介してコンピュータ 9 3 0 を LAN 9 7 7 に接続する。WAN ネットワーキング環境で使用するときは、コンピュータは通常、モデム 9 7 8 を備えるか、あるいは WAN 9 7 9 を介して通信を確立するための他の手段を備える。モデム 9 7 8 は、例えば内蔵または外付けであり、例えばユーザ入力インタフェース 9 7 0 または適した他の機構を介してシステムバス 9 3 6 に接続される。

20

【 0 1 6 4 】

図 1 1 には、インターネットを介した WAN の具体的な実施形態が示されている。コンピュータ 9 3 0 は通常、モデム 9 7 8 を備えるか、あるいはインターネット 9 8 0 を介して接続を確立するための他の手段を備える。モデム 9 7 8 は、例えば内蔵または外付けであり、インタフェース 9 7 0 を介してバス 9 3 6 に接続される。

【 0 1 6 5 】

ネットワーク化された環境では、パーソナルコンピュータ 9 3 0 に関して示したプログラムモジュールまたはこれらの一部が、リモートメモリ記憶装置に例えばストアされる。図 1 1 には、限定ではなく例として、リモートコンピュータ 9 8 2 のメモリ装置上にあるリモートアプリケーションプログラム 9 8 9 が示されている。図示および記述したネットワーク接続は例示的なものであり、コンピュータ間に通信リンクを確立する他の手段も例えば使用されることを理解されたい。

30

【 0 1 6 6 】

例示的な動作環境

図 1 1 には、例示的なデルタジェネレータを実装できる、適した動作環境 9 2 0 の一例が示されている。具体的には、図 1 1 中のいずれかのプログラムモジュール 9 6 0 ~ 9 6 2 および/またはオペレーティングシステム 9 5 8、あるいはこれらの一部により、本明細書に述べた例示的なデルタジェネレータが（全体的または部分的に）実装される。

40

【 0 1 6 7 】

この動作環境は、適した動作環境の一例でしかなく、本明細書に述べる例示的なデルタジェネレータの機能の範囲または使用に関するいかなる制限も示すものではない。例示的なデルタジェネレータとともに使用するのに適した他の周知のコンピューティングシステム、環境、および/または構成には、パーソナルコンピュータ（PC）、サーバコンピュータ、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、プログラム可能な消費者電子機器、無線電話機および無線機器、汎用および専用アプライアンス、特定用途向け IC（ASIC）、ネットワーク PC、ミニコンピュータ、メインフレームコンピュータ、以上のシステムまたはデバイスのいずれかを含む分散コンピューティング環境などが挙げられるが、これらに限定しない。

50

【0168】

コンピュータ実行可能命令

例示的なデルタジェネレータの一実施形態は、1つまたは複数のコンピュータまたはその他のデバイスによって実行される、プログラムモジュールなどのコンピュータ実行可能命令の一般的なコンテキストで述べるができる。一般にプログラムモジュールは、特定のタスクを実行するか特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。通常、プログラムモジュールの機能は、所望の種々の実施形態で組み合わせることも分散させることもできる。

【0169】

コンピュータ可読媒体

例示的なデルタジェネレータの一実施形態は、何らかの形のコンピュータ可読媒体上にストアすることもでき、それを介して伝送することもできる。コンピュータ可読媒体は、コンピュータからアクセス可能な任意の利用可能媒体とすることができる。限定ではなく例として、コンピュータ可読媒体には、「コンピュータ記憶媒体」および「通信媒体」を含めることができる。

【0170】

「コンピュータ記憶媒体」には、コンピュータ可読命令、データ構造、プログラムモジュール、またはその他のデータなどの情報をストアするために任意の方法または技術で実装される、揮発性および不揮発性、取外し可能および取外し不可能媒体が含まれる。コンピュータ記憶媒体には、RAM、ROM、EEPROM、フラッシュメモリ、またはその他のメモリ技術や、CD-ROM、DVD(digital versatile disk)、またはその他の光学記憶装置や、磁気カセット、磁気テープ、磁気ディスク記憶装置、またはその他の磁気記憶装置や、所望の情報をストアするのに使用できコンピュータからアクセスできるその他の任意の媒体が含まれるが、これらに限定しない。

【0171】

「通信媒体」は通常、コンピュータ可読命令、データ構造、プログラムモジュール、またはその他のデータを、搬送波やその他の搬送機構など変調されたデータ信号に具現化する。通信媒体にはまた、任意の情報送達媒体も含まれる。

【0172】

用語「変調されたデータ信号」は、情報が信号に符号化される形でその1つまたは複数の特徴が設定または変更された信号を意味する。限定ではなく例として、通信媒体には、有線のネットワークや直接有線接続などの有線媒体と、音波、RF、赤外線などの無線媒体と、その他の無線媒体が含まれる。これらの任意の組合せもまた、コンピュータ可読媒体の範囲に含まれる。

【0173】

結び

プログラムバイナリに対する最小デルタジェネレータについて、構造上の特徴および/または方法論的ステップに特定した言葉で述べたが、頭記の特許請求の範囲に定義するプログラムバイナリに対する最小デルタジェネレータは、上述した特定の特徴またはステップに必ずしも限定しない。むしろ、これらの特定の特徴およびステップは、特許請求する本発明の実施の好ましい形として開示するものである。

【図面の簡単な説明】

【図1】プログラムバイナリに対する最小デルタジェネレータの一実施形態を示す概略ブロック図である。

【図2】プログラムバイナリに対する最小デルタジェネレータの別の実施形態を示す概略ブロック図である。

【図3】プログラムバイナリに対する最小デルタジェネレータの一実施形態の単純化した適用例を示す図である。

【図4】プログラムバイナリに対する最小デルタジェネレータの一実施形態の単純化した適用例を示す図である。

10

20

30

40

50

【図 5】プログラムバイナリに対する最小デルタジェネレータの一実施形態の単純化した適用例を示す図である。

【図 6】プログラムバイナリに対する最小デルタジェネレータの実施形態の適用例を示す図である。

【図 7】プログラムバイナリに対する最小デルタジェネレータの実施形態の適用例を示す図である。

【図 8】プログラムバイナリに対する最小デルタジェネレータの方法論的な一実施形態を示す流れ図である。

【図 9】プログラムバイナリに対する最小デルタジェネレータの別の方法論的な実施形態を示す流れ図である。

【図 10】プログラムバイナリに対する最小デルタジェネレータの別の方法論的な実施形態を示す流れ図である。

【図 11】プログラムバイナリに対する最小デルタジェネレータの一実施形態を実装できるコンピューティング動作環境の一例を示す図である。

【符号の説明】

- 1 0 0 ジェネレータシステム
- 1 1 0 ソース C F G 機構
- 1 1 2 ソースプログラム
- 1 2 0 ターゲット C F G 機構
- 1 2 2 ターゲットプログラム
- 1 3 0 ブロックマッチャ
- 1 4 0 デルタ決定機構
- 1 4 2 デルタ
- 1 5 0 デルタパッチャ
- 1 6 0 再構築したターゲットプログラム
- 2 1 0 サーバ側
- 2 1 2 サーバ
- 2 1 5 インターネット
- 2 2 0 クライアント側
- 2 2 2 クライアント
- 9 2 0 コンピューティング環境
- 9 3 0 コンピュータ
- 9 3 2 プロセッサまたはプロセッシングユニット
- 9 3 4 システムメモリ
- 9 3 6 バス
- 9 3 8 R O M
- 9 4 0 R A M
- 9 4 2 B I O S
- 9 4 4 ハードディスクドライブ
- 9 4 6 磁気ディスクドライブ
- 9 4 8 磁気ディスク
- 9 5 0 光ディスクドライブ
- 9 5 2 光ディスク
- 9 5 4 データ媒体インタフェース
- 9 5 8 オペレーティングシステム
- 9 6 0 アプリケーションプログラム
- 9 6 4 プログラムデータ
- 9 6 6 キーボード
- 9 6 8 ポインティングデバイス
- 9 7 0 ユーザ入力インタフェース

10

20

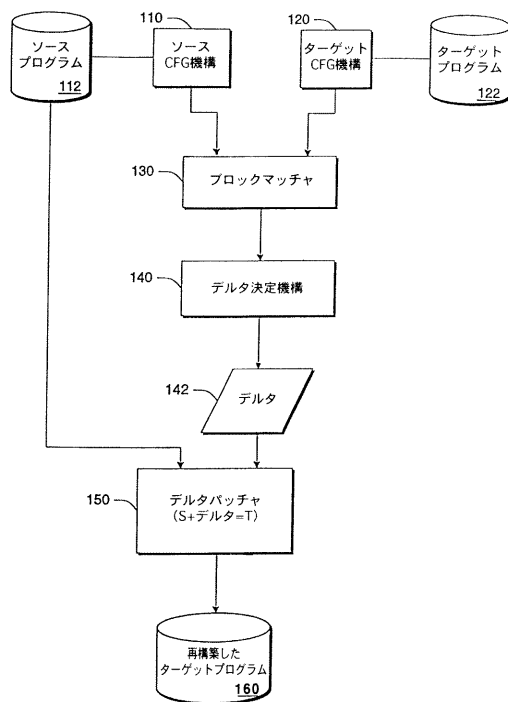
30

40

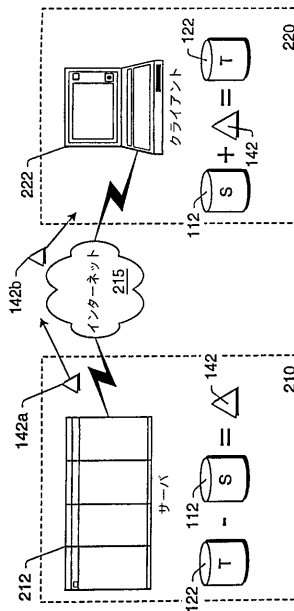
50

- 9 7 2 モニタ
- 9 7 4 ビデオアダプタ
- 9 7 5 出力周辺インタフェース
- 9 7 7 L A N
- 9 7 8 モデム
- 9 7 9 W A N
- 9 8 0 インターネット
- 9 8 2 リモートコンピュータ
- 9 8 6 ネットワーク
- 9 8 9 リモートアプリケーション

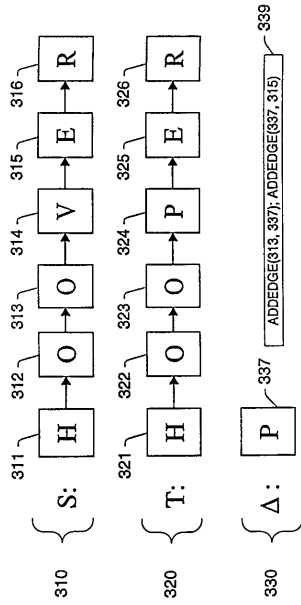
【図 1】



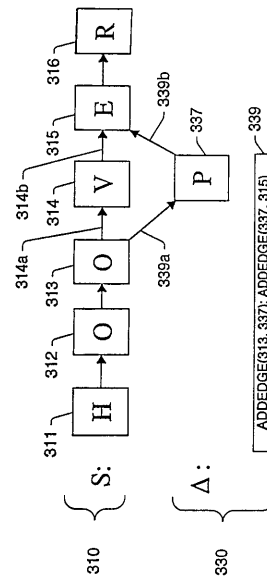
【図 2】



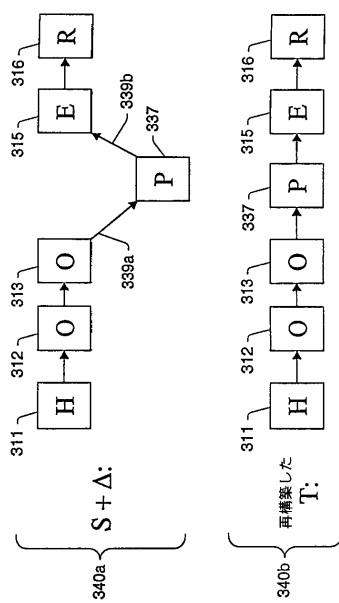
【図 3】



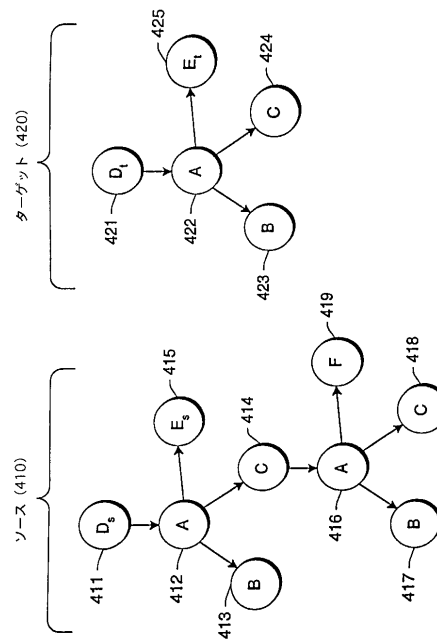
【図 4】



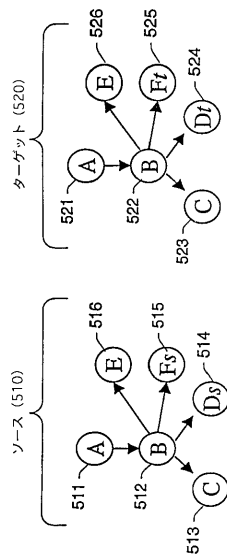
【図 5】



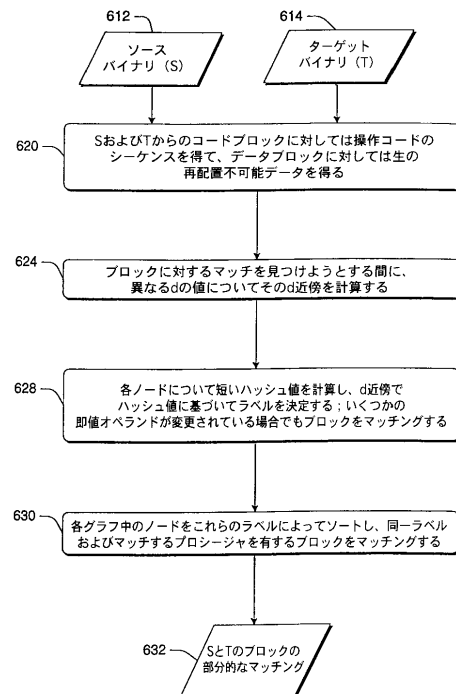
【図 6】



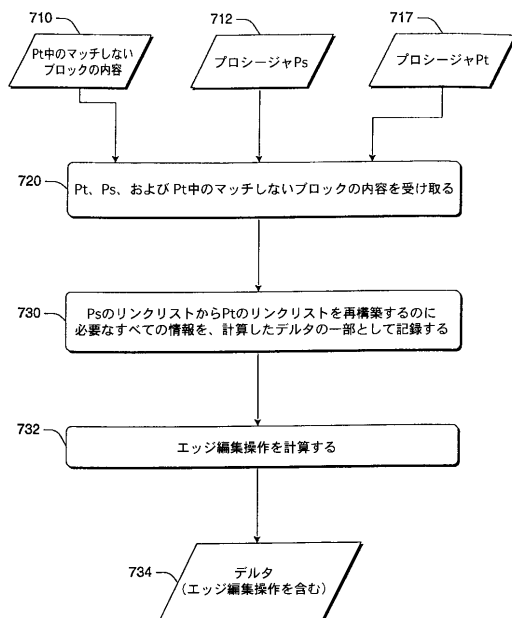
【図 7】



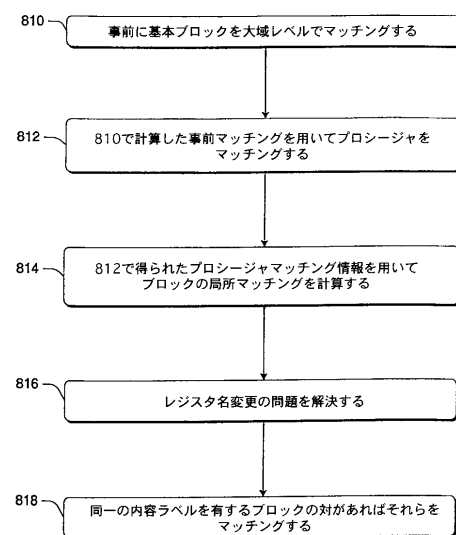
【図 8】



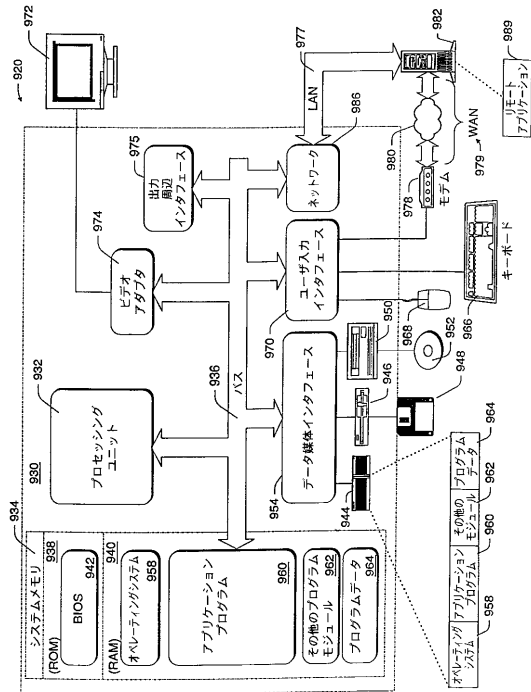
【図 9】



【図 10】



【図 11】



フロントページの続き

(72)発明者 ラマラスナム ベンカテサン

アメリカ合衆国 9 8 0 5 2 ワシントン州 レッドモンド ノースイースト 2 2 コート 1
7 2 0 8

(72)発明者 サウラブ シンハ

アメリカ合衆国 9 8 1 0 3 ワシントン州 シアトル オーロラ アベニュー ノース 4 7 1
0 ナンバー 2 0 2

合議体

審判長 赤川 誠一

審判官 宮司 卓佳

審判官 石井 茂和

(56)参考文献 国際公開第 9 9 / 1 1 5 4 9 号パンフレット (W O , A 1)

特開平 9 - 1 6 3 8 9 号公報 (J P , A)

(58)調査した分野 (Int . Cl . , D B 名)

G06F 9/06 630B