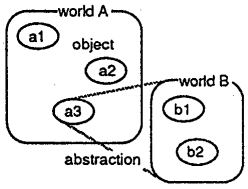
2W−1

# 広域分散システム環境の高信頼化に関する一手法

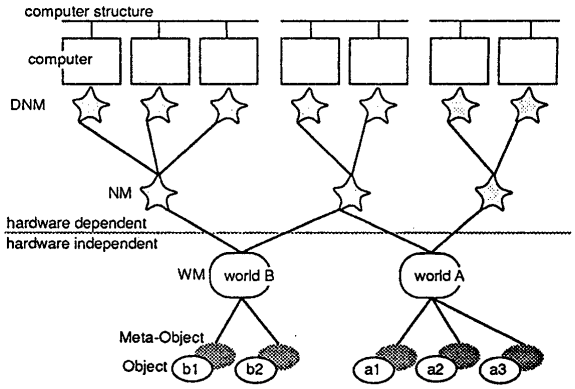
齊藤雅彦　村田悟　島田優　横山孝典
(日立製作所　日立研究所)

## 1．はじめに

極めて多数のオブジェクトを含有する広域分散システムに対して、我々は、関連あるオブジェクトを世界としてまとめて管理する「多重世界モデル」を提案した。更に、このモデルを実現するに当たって、処理と管理を別のレベルで行う「リフレクティブアーキテクチャ」を採用した。この構成においては、管理を実行するメタオブジェクトの高信頼化が重要である。本稿では、これらメタオブジェクトの高信頼化方式について述べる。



(a) 多重世界モデル



(b) ソフトウェアアーキテクチャ

図1　多重世界モデルとソフトウェアアーキテクチャ

## 2．広域分散システムモデルと信頼性の問題

多重世界モデルに基づくシステムは図1（a）に示すように、オブジェクトと世界から構成される。オブジェクトはデータとそれを管理する手続き群から成り、計算の基本単位である。世界はオブジェクトの存在する場であり、各オブジェクトはただ1つの世界に属する。オブジェクトは自己が属している世界内のオブジェクトのみ認識可能であり、通信可能な範囲も世界内に限られる。多重世界モデルでは抽象化により世界の階層化が可能である。世界の抽象化とは、下位世界に存在するオブジェクトの機能／データの一部または全てを上位世界のオブジェクトの機能／データとして利用可能とする方式である。世界間通信は世界を抽象化したオブジェクトと世界によって間接的に行われる。

多重世界モデルを実現するために、我々が開発中の広域分散システムのソフトウェアアーキテクチャは、図1（b）に示すような構成を有する。オブジェクト存在の場の提供を行うWM（世界メタオブジェクト）、ネットワーク固有のデータ管理と処理を行うNM（ネットワークメタオブジェクト）、計算機固有のデータ管理と処理を行うDNM（仮ネットワークメタオブジェクト）が存在し、これらがユーザプログラムであるオブジェクトを管理する。また、各オブジェクトにはそれぞれを管理するオブジェクト管理メタオブジェクトが存在する。

このような広域分散システムにおける高信頼化を実現するためには、故障の影響する範囲が大きいことを考慮しなければならない。広域分散システムにおいては、個々のオブジェクトのソフトウェア的故障と個々の計算機のハードウェア的故障・停止とを分離して考察するほうがよい。本稿では、後者の故障について考察する。すなわち、システムの最小故障単位を計算機と考え、1つ以上の計算機の完全停止を「故障」と定義する。したがって、仮にメタオブジェクト若しくはユーザプログラムの実行ログ、データの複製等が存在しても、それが同一計算機上であれば復旧に全く利用不可能な最悪な環境を考える。

## 3．再構築方式

本広域分散システム環境においては、「世界」を管理するWM、計算機ネットワークを管理するNMのような管理メタオブジェクトの信頼性保障が問題となる。最小故障単位を計算機とするため、他計算機上にデータの複製等を配置し、故障に備えなければならない。管理メタオブジェクトの高信頼化としては、プログラムの複製を動作させ、故障時に複製と本体を入れ替える方式も考案されているが、通常動作時に本体と複製との間でデータの一貫性を保障する通信が頻発し、本広域分散システムでは性能的に問題があると考える。本研究では、管理メタオブジェクトのデータの複製を各計算機のDNMに分散／共有配置させて、故障時にこれらのデータから管理メタオブジェクトを再構築する方式を採用した。

### 3．1　データ配置

データの複製等を配置する方式として、各計算機に一つ存在するDNMに計算機固有データの複製とシステム共通データの複製をそれぞれ、分散／共有配置する。故障時にはこれを用いて管理メタオブジェクトのデータを構築する。例えば、NMは計算機ネットワーク内で動作するオブジェクトのリスト、計算機ハードウェアのリスト、世界と計算機ネットワークとの対応表等を管理情報として有する。オブジェクト、ハードウェアリスト等の計算機固有データの複製は各DNMがそれぞれ所有し、世界とネットワークとの対応表等の共通データの複製は全てのDNMが共有する。NM再構築時には、全てのDNMから計算機固有のデータを、一つのDNMから共通データを収集する。WMも同様に、全てのNMからのネットワーク固有のデータと、一つのDNMからの共通データを入手して、再構築される。図

---

A Study of Fault-Tolerance under Widely Distributed Computing Environment
Masahiko SAITOH, Satoru MURATA, Masaru SHIMADA, Takanori YOKOYAMA
Hitachi Research Laboratory, Hitachi, Ltd.

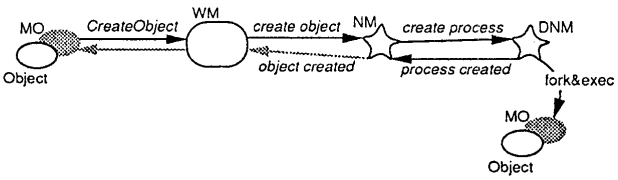



図２　オブジェクト生成に伴う処理依頼

１（ｂ）に示す上下関係を考慮すれば、ある管理メタオブジェクトはそれと直接／間接的に通信可能な上位の管理メタオブジェクト群から再構築されることになる。

本広域分散システムに限らず、全体の管理を行うプログラムに依頼された処理は何らかの形で個々の計算機を管理するプログラムに依頼を行うことが多い。例えば、本システムでは、オブジェクト生成はWMに対して要求するが、図２に示すように、WM、ＮＭでは計算機の負荷、性能等を計算してオブジェクト生成に最適な計算機を選択した後、その計算機を管理するＤＮＭに対して生成を依頼する。世界とネットワークとの対応が変化した場合等、特殊な状況でのみデータ変更をＤＮＭにブロードキャストしなければならないが、このような状況はシステム構成変更時を除き、通常は起こらない。このため、上述したデータの複製配置による通常動作時の性能への影響は無視できるほど小さい。

### ３．２　再構築動作

システム内の計算機が故障し、この計算機上に管理メタオブジェクトが存在した場合、各計算機上のＤＮＭが所有するデータの複製から新たな管理メタオブジェクトを他計算機上に生成する。以降、例としてＮＭを選択し、この再構築動作を示す（図３（ａ））。

（１）故障認識：故障は計算機ネットワーク内の各計算機のＤＮＭが認識する。故障を調査したいＤＮＭは定期的にネットワークに対する生存チェックのブロードキャストを行う。これに対して、その他の計算機に存在するＤＮＭが自己の計算機の生存を通知する。一定時間経過しても計算機から応答がなければ、その計算機が故障したと判定する（図中（１））。

（２）ＮＭ再構築：故障を認識したＤＮＭは計算機を一つ選択し、その計算機を管理するＤＮＭにＮＭの再構築を行わせる。ＮＭの再構築要求を受けたＤＮＭは新たなＮＭを起動する（図中（３））。

（３）管理情報の構築：新たに生成されたＮＭはシステム構成に即した管理情報を構築しなければならない。そこで、自己の属する計算機ネットワークに対してブロードキャストを行い、各計算機上のＤＮＭに情報提供を依頼する（図中（４））。これに対してＤＮＭは自己の現在の情報を通知する（図中（５））。この情報には、計算機上で動作するオブジェクトのリスト、自己の計算機のハードウェア情報、世界とネットワークとの対応表等がある。世界とネットワークとの対応表は全ての計算機上のＤＮＭが共有して所有しているので、一つのＤＮＭから得られた情報をそのままＮＭの情報とする。また、オブジェクトリスト、ハードウェア情報は各計算機のＤＮＭがそれぞれ自己に関する情報を分散して所有しているので、各ＤＮＭからの情報を収集してリストを構築する（図３（ｂ））。

ＷＭの動作している計算機が故障した場合にも、同様にして再構築される。

本方式は、管理メタオブジェクトの複製を動作させる方式に比べて、故障からの復旧に時間を要する。しかし、故障自体が頻繁に発生することはないこと、および、通常動作への影響の小さいことを考慮して、本方式を採用した。

### ５．おわりに

我々が開発中の広域分散システムにおける管理メタオブジェクトの高信頼化手法について述べた。現在、多重世界モデルに基づく分散処理環境のプロトタイプが完成し、本高信頼化手法を導入したシステムを作成中である。

### ［参考文献］

１）横山他：リフレクティブアーキテクチャに基づく分散処理環境（１）−分散処理モデル−，情報処理学会第４３回全国大会

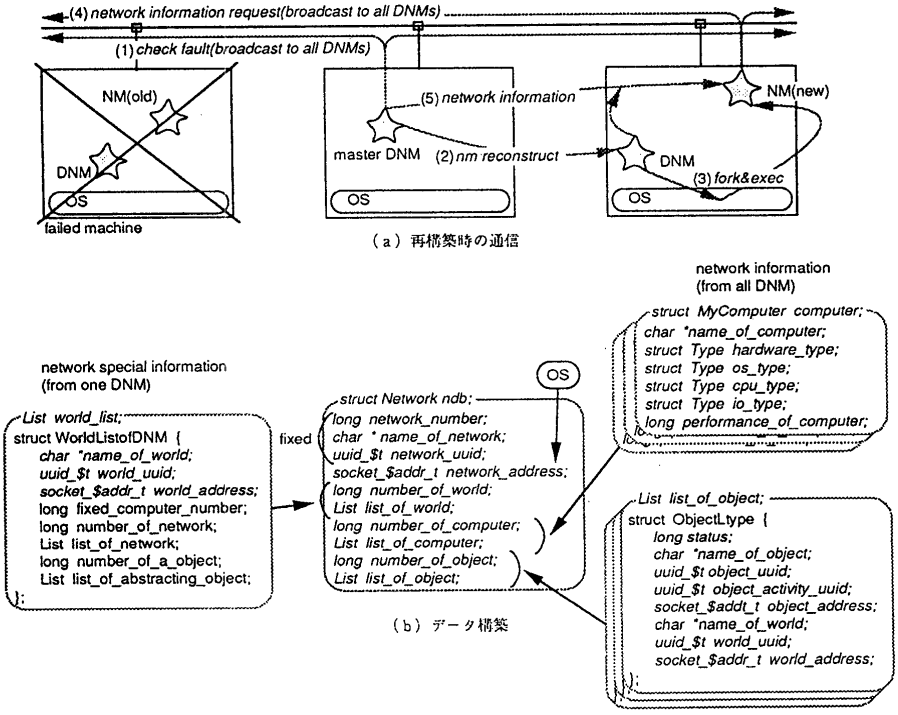



図３　再構築動作