

ネットワーク接続された多数サーバの運用・構築容易化技術

高本 良史† 宇都宮 直樹†

舩生 真紀子†† 三林 弘和††

† (株) 日立製作所 中央研究所 †† (株) 日立超LSIシステムズ

インターネット／イントラネットを利用した電子商取引や企業情報システムが一般化するに従い、稼働するサービスも大規模化、多様化している。一方で、大規模化、多様化したシステムの運用や構築は複雑さを増しており、運用・構築コスト増大や構築ミスの大きな要因となっている。我々は、頻繁な構成変更が要求されるシステムの運用や構築を容易化することを目的とし、サーバとネットワークの結線管理を自動化する技術を開発した。特に、システムの初期構築時において、OS(Operating System)がインストールされていないサーバもネットワーク結線管理可能にするため、自律起動エージェントによりサーバからネットワーク結線情報を取得することが特長である。

Management technologies of large network systems

Yoshifumi TAKAMOTO†

Naoki UTSUNOMIYA†

Makiko FUNYU††

Hirokazu SAMBAYASHI††

†Hitachi. Ltd.. Central Research Laboratory ††Hitachi ULSI Svstems Co.. Ltd.

Recently, system needs of growing large-scale systems and management cost of a large-scale system increases complexly. We propose an technology to increase manageability of the system. We developed the technology that automated management of network topology. A feature is a network agent program that acquires network topology information on the server in which Operating System is not installed.

1. はじめに

インターネット／イントラネットの利用が一般化し、企業内センタあるいはアウトソーシングによるデータセンタのシステム規模が増加している。これに伴い、システムの運用管理が高コスト化、複雑化していることが大規模システムの運用において大きな問題となっている。インターネット／イントラネット環境は、サービスの種類や性能要件などが頻繁に変化するため、サービスを構成するサーバ環境も頻繁に再構築する必要がある。ここで再構築とは、新たなサーバの追加や、あるサービスで使用されているサーバを他のサービスに移行するといった作業を示す。インターネット／イントラネット環境で使用されるサーバはネットワークに接続されており、サーバの再構築に伴ってネットワークの結線も再接続するケースが多い。しかし、多数サーバ環境ではネットワークが複雑化しており、システムの再構築に伴うネットワークの再接続は容易ではない。そこで、我々は、ネットワーク接続された多数サーバ環境において、ネットワークのトポロジ（結線関係）管理を自動化することで計算機環境の運用・構築を容易化する技術を開発した。

2. 計算機環境の運用サイクル

ネットワーク接続された多数サーバ管理における新たな要件の抽出を目的として、運用サイクルの観点からサーバ管理を検討した。

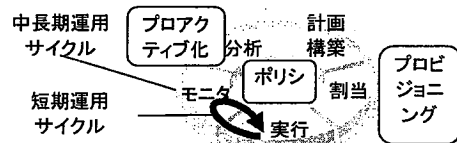


図1 サーバの運用サイクル

計算機リソースは図1に示すサイクルで運用されている[1]。短期運用サイクルはサーバの障害のように即時対応が必要な日々の運用に関するサイクルである。一方、中長期運用サイクルはシステム設計や再構築のように数ヶ月あるいは数年に渡る運用管理である。短期運用サイクルは、定型化業務が多く比較的自動化しやすい性質を持っており既に製品化が進んでいる。これに対し、中長期運用サイクルは人手を介する作業が発生するため、自動化が困難な性質を持っている。

中長期運用サイクルの自動化の中心となる技術として、ポリシー、プロビジョニング、プロアクティブ化がある。これらの技術は自律運用技術とも呼ばれている。ポリシーはルールに基づき自ら動作を決定するための技術である。プロビジョニングは業務に対して動的に計算機リソースを割り当てるための技術である。プロアクティブ化は、将来の障害発生や負荷の増減を予測し、予防する技術である。これらの技術を組み合わせることで、中長期運用サイクルの自動化を図っている。例えば、ある業務の負荷増大時にサーバを自動

的に追加するポリシーでは、プロアクティブ化技術により負荷増大を予測し、それに基づきプロビジョニング技術が業務に計算機リソースを割り当てといった一連の処理を自動化することができる。

しかし、表1に示すようにポリシー技術、プロビジョニング技術、プロアクティブ化技術だけで中長期運用サイクルを全てカバーすることはできない。特に、計画・構築フェーズにおいては、現状では有効な自動化技術はない。これは、現状のポリシー技術、プロビジョニング技術、プロアクティブ化技術が Wire Once の考え方に基づいているためと考えられる。Wire Once は、ネットワークの物理結線は一度限りで、以降の運用は論理的な結線変更 (VLAN 変更等) で対応する考え方である。そのため、計画・構築の中でも特に構築に関するフェーズでは、ネットワークを未結線状態から結線状態へ移行する物理結線の作業を要する事から、Wire Once を前提とした現状の自律運用技術はサーバ管理の自動化には対応していない。

表1 運用サイクル自動化技術

運用サイクル	項目	自動化技術
中長期運用 サイクル	計画・構築	有効な技術なし
	計算機割り当て 実行	Wire Once に基づく 自律運用技術 ・ポリシー技術
	モニタ	・プロビジョニング技術 ・プロアクティブ化技術
	分析	
短期運用 サイクル		

3. 構築フェーズにおけるサーバ管理の要件

構築フェーズにおける運用管理の要件をより明確化する検討を行った結果を述べる。

構築フェーズにおいてはサーバ等の計算機リソース導入や、ネットワークの結線といった人手を要する作業が多く発生するため自動化技術があまり進んでいない。実際、新規にサーバを導入するケースでは、新たなサーバ追加作業はハードウェアの設定だけでなく、正確なネットワーク結線の把握に多くの時間を費やしている。もし、結線に誤りが生じてしまうと、現在稼働中の業務にも影響が及んでしまう可能性があるため慎重な調査が必要となる。この作業は、表2の結線管理台帳の例に示すように、現状のネットワークの結線が台帳により人手で管理されているために発生する。この結線管理を自動化できれば、構築を容易化することができる。

表2 結線管理台帳の例

SW ポート#	接続装置	MAC	接続先 SW ポート#	ラベル
1	SW1	MAC1	SW2 1	x1
2	SW2	MAC2	SW3 1	x2
3	サーバ1	MAC3	-	1
4	サーバ2	MAC4	-	2
5	サーバ3	MAC5	-	3
6	サーバ4	MAC6	-	4

構築フェーズにおけるサーバの状態を表3に示す。これは各運用サイクルにおける計算機リソースの状態を示している。自律運用技術が対応している計算機リソース割当～分析フェーズまでは、サーバ状態は OS がインストールされており、ネットワークは結線済みの状態となっている。本検討で対象とする構築フェーズでは、OS は未インストールであり、ネットワークの状態は未結線から結線済みに移行する。サーバに OS がインストール済みの場合、サーバを管理するためのプログラム (管理エージェント) を動作させることができる。

表3 サーバ状態

運用サイクル	項目	サーバ状態
中長期運用 サイクル	計画・構築	OS 未インストール 管理 Agt.動作不可
	計算機割り当て 実行	OS インストール済 管理 Agt.動作可
	モニタ	
	分析	
短期運用 サイクル		

一方、管理情報の取得の観点では、現在の運用管理ソフトウェアは、サーバ情報のほとんどを管理エージェントから得ている。管理エージェントは、OS 上で稼働するプログラムであり、そのため OS がインストールされていない構築フェーズにおいては運用管理ソフトウェアを自動化の目的で使用することはできない。また、ビジネス用途のサーバに管理インターフェース (SVP: Service Processor、BMC: Base-board Management Controller) が内蔵されている機種があるが、管理インターフェースを内蔵したサーバにおいても構築フェーズでは上記と同様の管理が必要となる。SVP を内蔵することで、OS 未インストール状態でもサーバの電源 ON/OFF、サーバの温度/ファンの回転数/障害情報といったハードウェアの基本情報を取得することはできる。しかし、サーバ管理に必要なネットワーク情報などは

管理エージェントから取得する必要がある。従って、構築フェーズからサーバの管理を行うためには、OS未インストール状態のサーバから情報を取得することができる技術が必要になる。

4. 結線管理の自動化技術

4.1 結線管理の課題

本節では、管理者が手動でネットワークの結線管理台帳を作成する必要がある理由と、現状のネットワーク結線管理の課題を述べる。

サーバとネットワークを接続する場合、サーバ内にネットワークインターフェースカード(NIC)といった接続ボードを介して接続する。NICは、MACアドレスで示されるネットワーク固有の識別子を有している。MACアドレスを取得するには、接続ボードをソフトウェアから制御する必要がある。ハードウェアレベルで情報を取得する標準インターフェースが決まっておらず、接続ボードを構成する制御LSIのインターフェースがLSIメーカー独自になっている。そのため、MACアドレスはOSのドライバを介して取得する必要がある。しかし、構築フェーズにおいて、管理対象サーバはOS未インストール状態であるためエージェントなどのソフトウェアからのサーバ情報を得ることができない。サーバ管理において重要な情報は、サーバとその接続ボードが有するMACアドレスとの対応情報である。この対応情報が無ければ、あるMACアドレスを持つサーバがどれなのかを知ることができない。そこで現状は、管理者がMACアドレスをサーバの筐体に貼られているシールなどから取得し、サーバと接続ボードとを対応付ける結線管理台帳を作成している。

また、既存のネットワーク管理ソフトウェアやストレージ管理ソフトウェアを使用しても、構築時におけるネットワークの結線管理はできない。ネットワーク管理ソフトウェアは、MACアドレスを基本とした結線管理をしている。ネットワークを流れるパケットの宛先や送信元に格納されたMACアドレスを取得することで、前述の管理エージェントがなくても基本的な装置管理はできる。しかし、これではMACアドレスを持った装置が存在することはわかるが、どのサーバに対応する情報かを知ることはできない。さらに一般のネットワーク管理ソフトウェアは、いずれもOSをインストールしエージェントを起動する必要があり、本検討の要件である構築フェーズにおける結線管理はできない。

以上のことから、構築時における計算機リソース管理の自動化とは管理者が手動で作成していた結線管理台帳を自動的に生成することと考えられる。具体的には、OS未インストールのサーバのMACアドレスといったサーバのネットワーク結線情報を取得し、サーバと接続ボードとの対応付けを自動化する結線管理の実現が課題である。

4.2 自律起動エージェントによる結線管理の自動化

上記課題を解決するために、サーバと接続ボードとの対応付けを自動化する自律起動エージェント技術を開発した。自律起動エージェントは、管理対象サーバにネットワークを介して自動的に転送され、サーバと接続ボードとの対応情報を取得し、管理サーバに転送する機能を有するプログラムである。

図2は、自律起動エージェントの動作概要を示している。自律起動エージェントは、自律起動エージェン

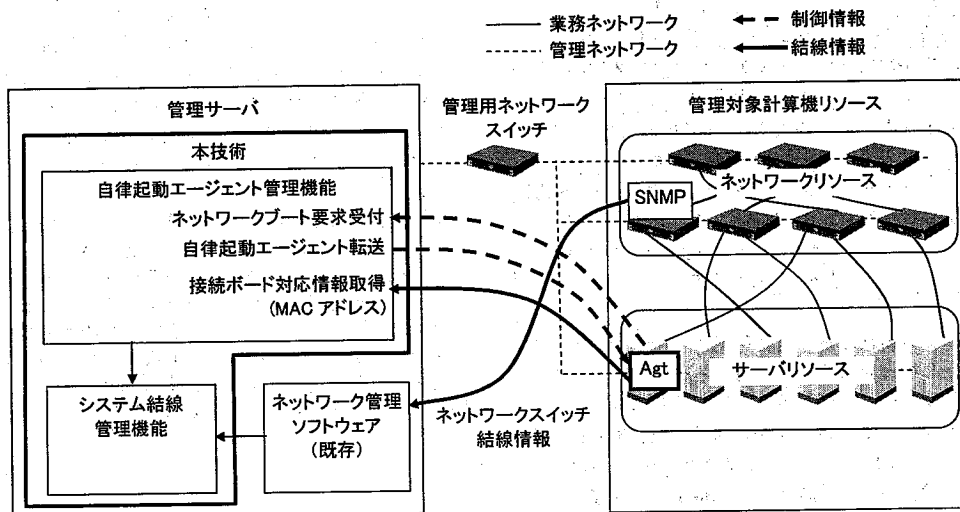


図2 自律起動エージェントの動作

ト管理機能から管理対象サーバに管理ネットワークを介して転送する。自律起動エージェントを管理対象サーバに転送するために、NIC の機能であるネットワークブート[2]の機能を使用する。ネットワークブート機能は最近のサーバでは一般的に使用可能な技術である。管理対象サーバの電源 ON に伴い、ネットワークブート要求が管理サーバに転送される。管理サーバ上の自律起動エージェント管理機能は、リクエストを受けると自律起動エージェントを管理対象サーバに転送する。自律起動エージェントには、OS 未インストールのサーバでも起動できるようにするための OS を内蔵している。転送された自律起動エージェントは管理対象サーバの接続ボード内に格納されている結線管理に必要な MAC アドレスを取得した後、管理対象サーバと MAC アドレスを対応付けした接続ボード対応情報を自律起動エージェント管理機能に転送する。一方、既存のネットワーク管理ソフトウェアは、ネットワーク装置が標準で有する SNMP(Simple Network Management Protocol)を利用し、ネットワークやストレージの結線情報を取得する。本技術のシステム結線管理機能は、自律起動エージェントが取得した接続ボード対応情報と、ネットワーク管理ソフトウェアが取得した結線情報との付き合わせを行う。これにより、サーバとネットワーク装置が関連づけられたシステム全体の結線管理が可能になる。

4.3 自律起動エージェントの機能

自律起動エージェントは、表 4 に示すように計算機リソースを管理するために必要な機能を有している。

表 4 自律起動エージェントの機能

機能	機能概要
結線情報取得機能	MAC アドレスを取得する機能
サーバ情報取得機能	サーバの CPU 情報、メモリ情報、PCI バス情報の取得機能
自律起動機能	OS を内蔵することで、他の OS を必要とせず起動する機能
ネットワークスイッチテーブル更新機能	NIC からネットワークパケットを強制的に送出することで、NIC に接続されたネットワークスイッチ内の制御テーブルを更新する機能

結線情報取得機能は、自律起動エージェントの基本機能であり管理対象サーバが有する MAC アドレスといった結線管理に必要な情報を取得する機能である。ここで取得した情報が、従来の結線管理台帳の代わりとなる。高信頼化のために複数の NIC が接続されている場合には、複数の接続情報を取得する。管理サーバ

に対して接続情報を通知する際には、サーバの識別子とペアで通知することでサーバと接続情報との対応付けが可能になる。例えば、サーバの UUID(Universally Unique Identifier)等がサーバの識別子となる。このようにサーバ特有の識別子とネットワークの識別子とを連携することで、サーバとネットワークの関係を正しく把握することができるようになる。サーバ情報取得機能は、管理サーバがサーバの管理に必要な情報を取得する機能であり、CPU 情報、メモリ情報、PCI バスに接続されたボードの情報等を取得し、管理サーバに通知する機能である。自律起動機能は、管理対象サーバに OS がインストールされていない状態でもサーバの情報を取得できるようにするため、自律起動エージェント自身が OS を内蔵している。ネットワークスイッチテーブル更新機能は、管理対象サーバが接続されているネットワークスイッチの内部にある制御テーブルを更新する機能である。この機能により、ネットワークスイッチの情報をサーバ起動と同時に更新することができるため、ネットワークの結線情報をリアルタイムに把握することができるようになる。自律起動エージェント起動時にネットワークスイッチにパケットを送出することで制御テーブルを更新する。

上記機能により、OS 未インストールのサーバでもリアルタイムな結線管理が可能になり、構築フェーズにおいて、従来は管理者が手動で行っていた結線管理の手間を削減することができるようになる。この技術は、業務やサービスの入れ替え時など、物理結線の変更を伴う計算機システムの再構築にも適用することができる。

5. 自律起動エージェントの実装

自律起動エージェントに関してプロトタイプを開発し動作検証を行った。プロトタイプでは、管理対象サーバの OS を使用せず、自律起動エージェントによる結線管理自動化技術により、サーバが接続されたネットワークスイッチのポート番号まで正確に把握することができる。以降では、プロトタイプの実装について述べる。

図 3 は、プロトタイプの全体構成を示している。ネットワークは、業務ネットワークと管理ネットワークから構成されている。業務ネットワークは、アプリケーションが占有できるネットワークであり、管理ネットワークは計算機リソースを管理するために必要なデータを転送するために設けた。管理対象サーバおよび業務ネットワークスイッチは、業務ネットワークと管理ネットワークの両方に属するよう接続する。

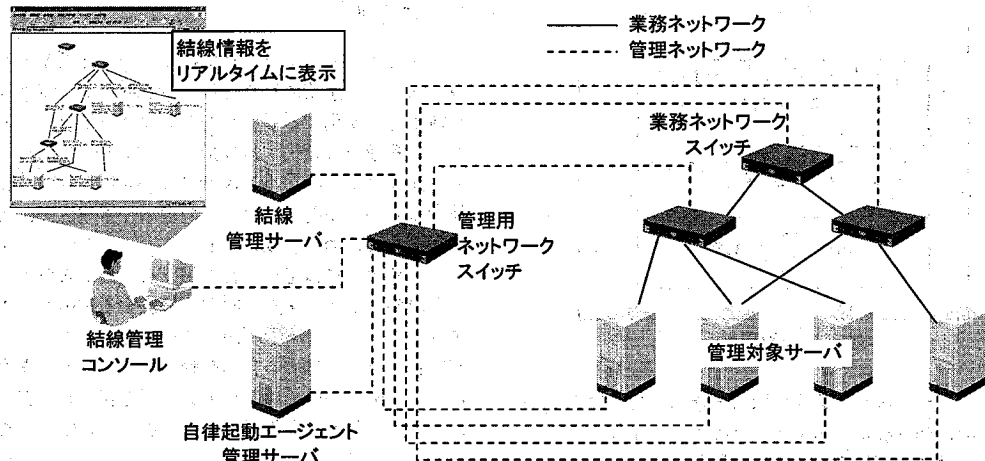


図3 プロトタイプ全体構成図

図4は自律起動エージェント管理サーバで稼動するソフトウェアを示している。自律起動エージェント管理サーバは、管理対象サーバへの自律起動エージェント転送および管理サーバ情報の管理を行う。

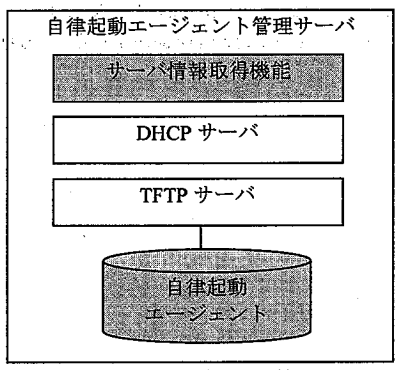


図4 自律起動エージェント管理サーバ

自律起動エージェント管理サーバは管理対象サーバに自律起動エージェントを転送するためのDHCP(Dynamic Host Configuration Protocol)サーバおよびTFTP(Trivial File Transfer Protocol)サーバと、自律起動エージェントから転送される結線などの情報を受信するサーバ情報取得機能を有する。DHCPサーバは、管理対象サーバがネットワークブートで使用する際の一時的なIPアドレスを割り当てるために用いた。開発した自律起動エージェントは、TFTPサーバを介して管理対象サーバに転送される。自律起動エージェントは、管理対象サーバのディスクは一切使用せず、サーバのメモリ上で動作する。自律起動エージェントは起

動すると、結線情報と共にサーバのハードウェア情報を取得する。取得した情報は、サーバ情報取得機能に転送される。これらの一連の動作により、図5に示すように、自律起動エージェント管理サーバ内に、従来管理者が手動で作成していた結線管理台帳に相当する接続ボード対応情報を自動的に生成することができる。接続ボード対応情報は、サーバ識別子とサーバが有するNICのMACアドレスをペアで管理しており、これにより管理者の結線管理台帳のメンテナンスの手間を削減することができる。

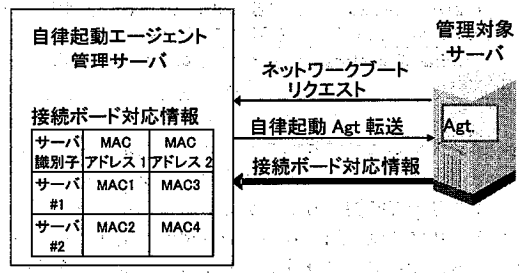


図5 自律起動エージェント管理サーバ動作

図6は、結線管理サーバの構成図を示す。結線管理サーバは、業務ネットワークスイッチから結線情報を取得するネットワークスイッチ情報取得機能と、自律起動エージェントが取得した管理対象サーバ情報およびネットワークスイッチ情報とを合わせてシステム全体の結線情報管理を行うネットワーク結線管理機能、結線図生成機能を有する。

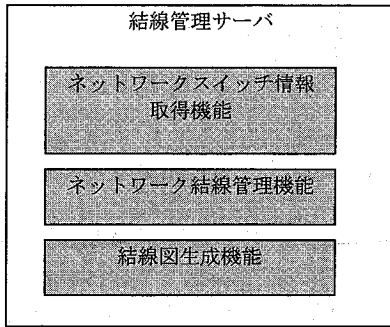


図 6 結線管理サーバ

ネットワークスイッチ情報取得機能は、SNMP(Simple Network Management Protocol)を使用し、ネットワークスイッチから表 5 に示すような MIB(Management Information Base)情報を取得する。ネットワークスイッチ情報には、ポート毎の MAC アドレス、接続先機器の MAC アドレスなどが含まれる。ネットワーク結線管理機能は、図 5 の接続ボード対応情報と表 4 のネットワークスイッチ情報とをマージすることでシステム全体の結線情報を生成する。

表 5 取得するネットワークスイッチ情報例

使用 MIB 名	意味
ifPhysAddress	ポートの MAC アドレス
ifName	ポート番号
vtpVlanName	VLAN 名
vmMembershipSummaryMemberPorts	該当 VLAN にポートが所属するか否かのフラグ列
dot1dBasePortIfIndex	該当ポート番号に対応するインデックス

上記開発した自律起動エージェント技術により、システム全体のネットワーク結線をスイッチのポート番号まで含めて把握することができるようになった。プロトタイプでは、約 10 秒で結線情報を更新する事ができる。従って、ほぼリアルタイムにシステム全体の結線を正確に把握することができることを確認できた。また、結線情報を自動的に取得できることから、結線管理台帳の更新を簡略化でき、また構築ミスを低減できると考える。さらに、結線の把握以外の効果として OS のデプロイが容易になる効果もある。OS デプロイは、一般に MAC アドレスをキーにサーバを特定し、OS イメージをサーバに転送する方式が一般である。しかし、MAC アドレスだけでは正しいサーバに対するデプロイかどうか不明であり、本技術により OS インストール前にサーバと対応付けて管理することで OS デプロイ管理も容易になる。

6. まとめ

本稿では、ネットワークに接続された多数サーバ環境において、頻繁に変更される計算機システムの運用や構築を容易化する事を目的とし、ネットワークのトポロジ管理を自動化する技術について述べた。

中長期運用サイクルの構築フェーズにおける自動化技術の要件を検討し、サーバの追加や構成変更に対応するために、OS 未インストールのサーバからネットワーク結線管理を自動化する技術が必要であることがわかった。また、課題を解決する技術として、ネットワークブートを利用しサーバとネットワークの結線管理を自動化する自律起動エージェントを提案した。本技術についてはプロトタイプを開発し、実機による結線管理の自動化が実現できることを確認した。

参考文献：

- [1] 日立:ブループリント - サービスプラットフォーム編「Harmonious Computing が描くサービスプラットフォームの将来像」
http://www.hitachi.co.jp/Prod/it/harmonious/download/hcbp_pfm.pdf
- [2] Preboot Execution Environment Version 2.1
<http://www.intel.com/labs/manage/wfm/wfmspecs.htm>