情報処理学会論文誌 Vol.57 No.3 967–975 (Mar. 2016)

一般組織において地理的に分散配置した複製サーバのフェイルオーバ・フェイルバックを可能にする冗長化構成

大隅 淑弘[1,a)] 山井 成良[2,b)] 岡山 聖彦[1,c)]

受付日 2015年6月22日, 採録日 2015年12月7日

概要：ICT システムの高可用性は重要な要件となっている．組織においてサービスの信頼性や可用性を向上させたり，BCP を整備したりするような場合においては，地理的に分散して設置されたサーバを冗長化する構成が有効である．しかし，クライアントによっては複数のサーバを指定できないものが存在する．また，複数のサーバを指定できるクライアントであっても，待機系サーバへフェイルオーバした後，主系サーバにフェイルバックしないものが存在する．そこで本論文では TCP を用いる一部のサービスに対して分散配置されたサーバ間で IP Anycast 技術を適用する方法を提案する．この方法では TCP 通信に対して pop switch が発生することを逆に利用し，フェイルオーバだけでなくフェイルバックも可能にする．また適用可能なサービスが満たすべき条件を明らかにする．提案方法を岡山大学の LDAP 認証サービスに適用した結果，提案方法の有効性を確認した．

キーワード：IP エニーキャスト，フェイルオーバ，フェイルバック，複製サーバ，冗長化

Redundant Configuration of Geographically Distributed Servers for Failover and Failback in a General Organization

Yoshihiro Ohsumi[1,a)] Nariyoshi Yamai[2,b)] Kiyohiko Okayama[1,c)]

Received: June 22, 2015, Accepted: December 7, 2015

Abstract: High Availability of ICT systems is an important requirement. In an organization, in order to improve the reliability and availability of services, and construct Business Continuity Planning, a configuration of redundant servers which are geographically distributed is valid. However, there exist some clients that cannot configure to use two or more servers. Among clients that can configure to use two or more servers, there exist some clients that can fail over to an alternative server but cannot fail back to the main server. In this paper, we propose a redundant configuration method that introduces IP Anycast technique to geographically distributed replication servers for some kinds of TCP based services. This method can not only fail over to an alternative server but also fail back to the main server by virtue of pop switch on TCP based services. We also show the conditions that should be satisfied by services applicable to the proposed method. According to our operation experience of the LDAP service in Okayama University, we confirmed the proposed method works effectively.

Keywords: IP anycast, failover, failback, replication server, redundancy

1. はじめに

組織における ICT（Information and Communication Technology）システムは，機器の高性能化と低価格化，ネットワークの高速化，クラウドコンピューティングの拡大

本論文は文献 [1] を発展させたものである．

[1] 岡山大学
Okayama University, Okayama 700–8530, Japan
[2] 東京農工大学
Tokyo University of Agriculture and Technology, Koganei, Tokyo 184–8588, Japan
[a)] oosumi@cc.okayama-u.ac.jp
[b)] nyamai@cc.tuat.ac.jp
[c)] okayama@cc.okayama-u.ac.jp

などにより，従来の集中設置型から，分散配置型への移行が進んでいる．さらに，2011 年 3 月の東北地方太平洋沖地震の教訓や，今後発生が予想されている東海・東南海・南海地震の被害の想定から，BCP（Business Continuity Planning：事業継続計画）への取り組みが強化され，情報資源の冗長化やディザスタリカバリの整備が進められている．このような情勢から，地理的に分散配置したサーバを冗長化し，サービスの信頼性と可用性を確保する情報基盤の整備が重要になっている．たとえば，岡山大学では多くのサービスで共通して用いられる認証サービス（LDAP（Lightweight Directory Access Protocol）[2]，RADIUS（Remote Authentication Dial-In User Service）[3]）を中心にいくつかのサービスで 2 つのキャンパスにサーバを分散配置している．

しかし，地理的な分散配置によりサーバを冗長化したとしても，クライアントによっては複数のサーバを指定できないものが存在する．また，複数のサーバを指定できるクライアントであっても，優先度の高いサーバ（主系サーバ）がダウンし優先度の低いサーバ（待機系サーバ）にアクセス先を切り替える動作（フェイルオーバ）を行った後，主系サーバが復旧しても主系サーバにアクセス先を戻す動作（フェイルバック）が行われないものが存在する．システムの構築後にこのようなクライアントの存在が明らかになった場合，自動的にフェイルオーバおよびフェイルバックを行うように修正することは困難である．

一方，分散配置されたサーバの冗長構成方法に IP Anycast [4] を用いる方法がある．IP Anycast は，ネットワーク的に分散配置された複数のホストに，1 つの IP アドレスを同時に割り当てることで冗長化を行う．ホストの障害時には経路が更新されることにより，他の場所にある別のホストに自動的に接続変更されてサービスを継続する．IP Anycast は，従来 DNS（Domain Name System）[5] のルートサーバや CDN（Contents Delivery Network）による WEB 配信サービスなど，インターネット全体にわたるような大規模なネットワーク上で展開されているサービスに用いられている．しかし，一般に IP Anycast は宛先 IP アドレスまでの最適な経路が変化すると宛先ホストが切り替わるため，特にステートフルなプロトコルである TCP で通信を行うと，宛先ホストの切替えにより TCP コネクションが突然強制終了される現象（pop switch）が発生する．このため，たとえば LDAP サーバの冗長化に IP Anycast が有効であるかどうかは自明ではない．文献 [6] では IP Anycast を用いれば TCP を用いたサービスについてもフェイルオーバが可能であることが述べられているが，フェイルバックに関しては言及されておらず，また pop switch 発生後の動作や適用可能なサービスの条件も示されていない．

そこで本論文では TCP を用いる一部のサービスに対して分散配置されたサーバ間で IP Anycast 技術を適用する方法を提案する．この方法では TCP 通信に対して pop switch が発生することを逆に利用し，フェイルオーバだけでなくフェイルバックも可能にする．また適用可能なサービスが満たすべき条件を明らかにする．

以下，2 章では，対象とする冗長構成とその問題点について述べる．次に 3 章では，提案するサーバの冗長化構成について述べ，4 章では評価システムの実装と評価について述べる．最後に 5 章ではまとめと今後の課題について述べる．

2. 対象とする冗長構成と問題点

本論文で対象とする冗長化の構成を図 1 に示す．この図に示すように，地理的に分散した複数の場所（図 1 では Location A，Location B の 2 カ所）に同一のサービスを提供するサーバが設置されているものとする．クライアントは TCP で優先度が高い主系サーバに接続し，何らかのサービスを受けるものとする．この図に示すように，主系サーバはクライアントの位置により異なっていても構わないものとする．クライアントは複数のサーバを指定することができ，主系サーバがダウンしている場合には優先度の低い待機系サーバにアクセスを試みるように動作する（以下，この動作をフェールオーバと記す）機能を有しているものとする．このような機能を有するクライアントとして Dovecot [7] がある．Dovecot において LDAP 認証を利用する場合，複数の LDAP サーバを設定しておくことによりフェイルオーバが可能になる．このようなクライアントとして，他にも RADIUS サーバの一実装である FreeRADIUS [8] や認証機能付きレイヤ 2 スイッチである Alaxala AX2400S シリーズがあることを確認しており，この他にも多くのクライアントが存在すると思われる．

ところが，優先度の高いサーバに障害が発生してフェイ

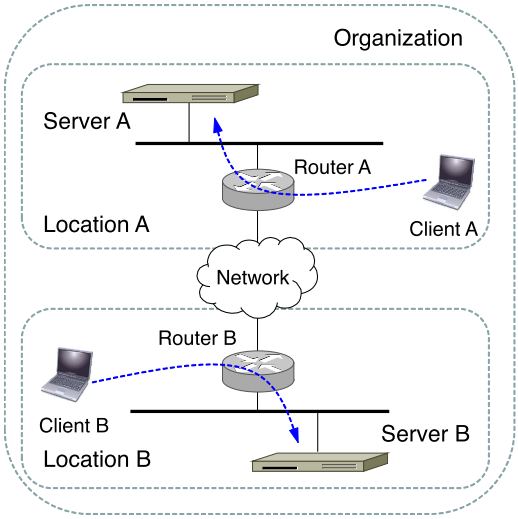

図 1 対象とする冗長構成
Fig. 1 Target redundant configuration.

ルオーバが起こり，その後サーバが障害から復旧した場合，本来であれば復旧した時点でアクセス先は優先度の高いサーバに戻る動作（以下，この動作をフェイルバックと記す）を行うべきであるが，少なくとも上記のクライアントのうち，Dovecot や FreeRADIUS では複数の LDAP サーバを指定した場合にフェイルバックしないことが判明している．このようなクライアントでは一度 LDAP サーバとの間で TCP コネクションが確立されるとこのコネクションを使用し続けるため，フェイルバックが発生しないと推測される．

このような状況では，複数のサーバ間で負荷分散が適切に行われないために待機系のサーバに負荷が集中したり，ネットワーク的に遠いサーバにアクセスするため無駄なトラフィックが発生したりする問題がある．岡山大学情報統括センターでは，POP3（Post Office Protocol）[9] や IMAP（Internet Message Access Protocol）[10] のサーバとして Dovecot を使用しており，地理的に離れて設置された 2 台の LDAP サーバを設定して参照しているが，実際にこのような問題が発生した．

この問題に対処するには，たとえば待機系サーバへのアクセスを維持しながら一定時間ごとにフェイルバックが可能であるかどうかを調べる機能の追加が考えられる．実際に，AX2400S シリーズでは複数の RADIUS サーバを指定した場合において主系サーバがダウンすると，タイムアウト（1〜30 秒，初期値 5 秒）と再送を何回か（0〜15 回，初期値 3 回）繰り返した後に次のサーバにアクセスし，指定時間（1〜1,440 分，初期値 10 分）経過後に主系サーバへのアクセスを試みる [11]．しかし，少なくとも上記の Dovecot や FreeRADIUS はこのような機能は有しておらず，また同機能の追加実装もたとえソースプログラムが公開されていたとしても一般的には容易ではない．

その他にも，複数の計算機によるクラスタ環境やクラウドサービスにおけるサーバ仮想化環境において HA（High Availability）クラスタを構成する方法なども考えられるが，すでに稼働しているサービスにおいて導入するには多くの条件に適合させたり，システムの導入経費が新たに必要になったりするため，本論文では考慮しない．

3. IP Anycast を用いた複製サーバ冗長化構成方法

前章で述べたように，既存のクライアントの中には，分散配置されたサーバの冗長化においてフェイルオーバが発生した場合，自動的にフェイルバックが行われず，複数のサーバ間で負荷分散が適切に行われなくなったり，無駄なトラフィックが発生したりする問題がある．この問題に対して，本論文では組織内ネットワークにおいて IP Anycast 技術を適用することにより解決する方法を提案する．

3.1 IP Anycast

IP Anycast は，IPv6 や IPv4 において 1 つの IP アドレスを複数のサーバに対して共通に割り当てるアドレッシング方法であり，分散配置された複数のサーバによる冗長化構成が可能である．IP Anycast では，クライアントからのパケットを，ルーティングによりそのクライアントから見てネットワークトポロジの観点で最も近いサーバに転送する．平均応答時間の短縮，DoS（Denial of Service）攻撃の局所化，DDoS（Distributed Denial of Service）攻撃の効果抑制の効果も有する．

従来，IP Anycast はインターネット全体にわたるような大規模なネットワークで展開されているサービスに用いられている．このような実装では，ルーティングプロトコルには BGP（Border Gateway Protocol）などの EGP（Exterior Gateway Protocol）を用いている．文献 [12] では，DNS のルートサーバを用いて IP Anyacst の実装を評価している．この他，IPv4 から IPv6 への移行に用いる 6to4 技法やマルチキャスト通信における PIM-SM（Protocol-Independent Multicast Sparse Mode）の RP（Rendezvous Point）の冗長化にも利用されている．

IP Anycast を用いた冗長化では，サーバとクライアント間の経路に変更があると，クライアントが送出したパケットが異なるサーバで受信されることがある．この場合，特に TCP ではコネクションが確立されていないサーバがパケットを受信すると RST フラグ付きパケットを返すため，pop switch が発生する．このため，従来 IP Anycast はステートレスな UDP のサービスについて適用されることが多い．ただし，文献 [6] のように TCP を用いたサービスに対して IP Anycast を適用した事例は存在する．

3.2 提案方法の概要

本論文では，図 1 に示した冗長構成において，組織内ネットワークに IP Anycast を適用することでサービスを冗長化し可用性の向上や負荷分散を図る方法を提案する．この方法では IP Anycast の特性により主系サーバに障害が発生した場合に経路切替えが発生するため，クライアントは待機系サーバにアクセスするように動作する．このとき pop switch が発生するが，その際クライアントがサーバへの再接続を試みる機能（リトライ機能）を持つなど一定の条件を満たしていれば，これを利用して自動的にフェイルオーバを行うことができる．また，pop switch は主系サーバが復旧した際にも発生するため，このリトライ機能を利用すれば自動的なフェイルバックも行うことができる．これにより 2 章で述べたような待機系サーバへの負荷の集中や無駄なトラフィックの発生がなく，またソースプログラムの修正も必要としないため，導入も容易である．ただし，IP Anycast に用いる IP アドレスの経路情報をネットワークに広告したり，後述するサービスの死活監視を運用

したりする作業など，サーバやネットワーク上でいくつかの作業を行う必要がある．

また，ルーティングプロトコルのメトリックを調整することで，冗長化したサーバを Active-Active や Active-Standby で動作させることができる．Active-Active は，複数の複製サーバを同時に稼働して処理を分散する．Active-Standby では，複数の複製サーバを動作させておき，Standby のサーバではサービス要求があればいつでもサービスを提供できるように，システムを動作状態にしておく．ルーティングプロトコルは IGP（Interior Gateway Protocol）を想定する．

3.3 冗長化構成の条件

提案方法では，クライアントは pop switch が発生した場合でも別のサーバと再接続を行い，正常に動作する必要がある．このためクライアントおよびサーバが満たすべき条件を以下に示す．なお，議論を分かりやすくするために，クライアント・サーバ間の通信は TCP で行われるものとする．

(1) クライアントは，サーバがダウンした場合などサーバとの通信に失敗した場合，少なくとも経路切替えが完了した後に再アクセスが発生する程度に十分長い間，継続して同一の IP アドレスを持つサーバへの再アクセスを試みることができる．
(2) すべてのクライアントは，サーバが停止してから経路切替えが完了するまでの間のサーバダウンに耐えられるものであること．
(3) クライアントがサーバへ再アクセスを行う場合，それまでに行われた通信内容を引き継がず，最初から通信し直すように動作する．
(4) すべてのサーバはクライアントからの通信に対して同じ動作を行う．
(5) すべてのサーバはそれぞれ独立して動作し，セッション管理や処理の継続，データ同期などは必要としない．

これらの条件のうち，1 と 2 は提案方法はフェイルオーバを IP Anycast における経路切替えにより実施しているために必要な条件である．経路切替えに要する時間はルーティングプロトコルに依存するが，標準的には RIP の場合には 180 秒，OSPF の場合には 40 秒であり，ある程度は調整可能である．サーバダウンから経路切替えが行われるまでには，最悪の場合で後述する死活監視の間隔時間がさらに必要となる場合がある．再アクセスの発生間隔は短いほど経路切替え後に実際にフェイルオーバが起こるまでの無駄な待ち時間を短くできる．また，pop switch 発生時など，TCP コネクションが強制切断された場合，ただちに再アクセスする機能があれば，フェイルバックの際に素早いサーバ切替えが可能になる．

そのほかの条件はクライアントから見てサーバはいつでも切り替わる可能性があるため，TCP コネクションの情報を除いてクライアント・サーバ間で HTTP 通信における Cookie のような共有情報をいっさい持たないことを意味する．

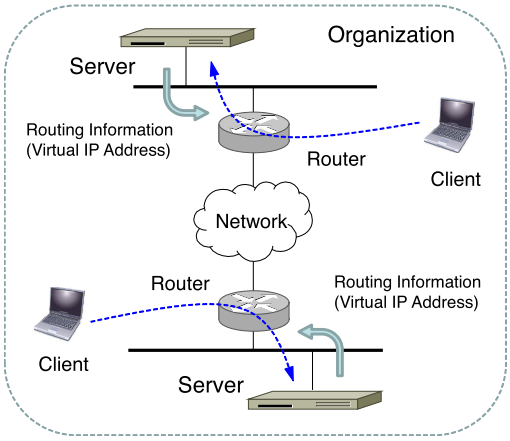

図 2 IP Anycast による冗長構成の例
Fig. 2 Example of redundant configuration by IP Anycast.

このような条件は 2 章で述べた Dovecot や FreeRADIUS と LDAP サーバとの通信が当てはまる．これ以外にも上記の条件を満たしていれば提案方法は適用可能である．

3.4 システムの構成

まず，各複製サーバにおいて IP Anycast アドレスとしてネットマスクが 32 ビットの仮想 IP アドレスを設定し，その経路情報をネットワーク上に広告するためのルーティングデーモンを動作させる．経路情報の広告方法によって，Active-Active や Active-Standby の構成が可能である．Active-Active の場合は，クライアントから見て最寄りのサーバのメトリックが，もう一方のサーバのものよりも小さくなるように設定する．Active-Standby の場合は，クライアントから見て Active のサーバのメトリックが，Standby のサーバのそれよりも小さくなるように設定する．32 ビットのネットマスクは，経路情報の最長一致のためである．

この構成だけでも動作するが，目的のサービスだけが停止した場合，すなわち，サーバは動作しており経路情報は広告しているが，サービスが停止している場合には，ブラックホールサーバとなり，サービス障害が発生する．この対策としてサービスの死活監視が必要である．この機能については，3.6 節で述べる．2 台のサーバで冗長構成する例を図 2 に示す．

なお，一般組織におけるルーティングプロトコルは RIP（Routing Information Protocol）[13] や OSPF（Open Shortest Path First）[14] が利用されるため，本論文では，RIP や OSPF の利用を想定して説明をする．

3.5 提案方法の動作手順

本提案方法により，フェイルオーバ，フェイルバックす

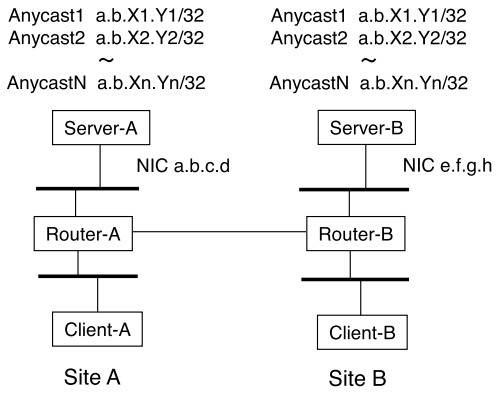

図 3 複数の IP Anycast アドレスの運用
Fig. 3 Operation of multiple IP Anycast address.

る動作手順を説明する．Active-Active，Active-Standby とも同様である．

(1) 障害の発生によりサーバへの経路が変更される．
クライアントの接続先は他のサーバに変更される．
(2) クライアントでは pop switch が発生するが，サーバへの接続をリトライすることでサーバの切替を完了する（フェイルオーバ）．
(3) 障害のサーバが復旧し経路が以前の状態に戻る．
クライアントの接続先は以前のサーバに戻る．
(4) クライアントでは pop switch が発生するが，サーバへの接続をリトライすることでサーバの切替を完了する（フェイルバック）．

3.5.1 複数のサービスへの IP Anyacst 適用

1 台のサーバが IP Anycast によって複数のサービスを提供する場合には，各サービスは独立して冗長化を行う必要がある．IP Anycast では一般的には，ループバックインタフェースに IP alias で仮想 IP アドレスを設定して動作させるが，サービスが複数ある場合には複数の仮想 IP アドレスの経路情報を独立して制御できない場合がある．たとえば，サーバのルーティングデーモンとして，Quagga Routing Suite [15] が一般によく用いられるが，Quagga では，IP alias の場合にはそれぞれの経路情報を個別に制御することができない．

このような場合には，ルーティングプロトコルに OSPF を使用し，IP トンネルの生成や VLAN インタフェースの生成によって，仮想 IP アドレスを複数作成する方法を用いることができる．OSPF ではインタフェースごとにコストの変更が可能なため，各仮想 IP アドレスに対応したサービスが独立して IP Anycast による経路制御を行うことができる．

なお，RIP などのようにインタフェースごとにメトリックを変更できない場合には，サーバのルーティングデーモンとルータの間に経路情報を制御する機構を組み込むなどの方法が考えられる．複数のサービスへ IP Anycast を適用するモデルを図 **3** に示す．

3.6 サービスの死活監視

本提案による冗長化構成では，サーバあるいはルーティングデーモンが停止したときには，経路情報が更新されてバックアップとなるサーバに自動的にフェイルオーバするが，サービスのプロセスだけが停止した場合には，経路情報が更新されないため自動的にフェイルオーバしない．このため，障害を検知してただちに経路情報を更新し，フェイルオーバさせるための機能が必要である．なお，サーバのルーティングデーモンとネットワークが正常に動作している状態であれば，経路情報の変更によるフェイルオーバの時間は，ルーティングプロトコルのコンバージェンス時間に基づく．サーバが突然ダウンしたような場合には，RIP では Flash タイマ，OSPF では Dead 間隔に基づいた時間でフェイルオーバする．サービスの死活監視にはいくつかの方法がある．なお，ネットワーク通信障害の死活監視については，提案方法に限らず経路の冗長化など既存技術を適用することが可能であるので，本論文では議論しない．

3.6.1 死活監視用に別システムを運用する方法

別に運用する監視用のシステムでヘルスチェックを行い，障害が発生したときには，ルータやサーバに指示を出してフェイルオーバさせる（以下，別システム監視と記す）．この方法では，ネットワークやサーバの状態を総合的に管理できるため，柔軟な運用が可能になる利点があるが，単一障害点になる可能性があり，監視用システムの冗長化なども検討する必要がある．

3.6.2 冗長化するサーバ自身が死活監視を行う方法

別の監視用システムを利用することなく，サーバ自身がヘルスチェックを行うことができる．別システム監視と同様にヘルスチェックで障害を検出した場合には，ルータやサーバに指示を出してフェイルオーバさせる．この場合には，ヘルスチェックとフェイルオーバでそれぞれ 2 つの方法がある．

ヘルスチェックでは，自分自身の動作チェックをする方法と，他のサーバの動作チェックをする方法がある．他のサーバの動作をチェックする方法では，他のサーバやネットワークの状況に応じて動作することができる．

フェイルオーバでは，自分への経路の優先度を上げる方法と下げる方法がある．自分の優先度を上げる方法では，特に冗長の多重度が大きい場合には負荷が集中する場合のあること，優先度を上げたサーバの動作に異常がある場合には，ブラックホールとなってサービス障害を起こす場合がある．自分自身のヘルスチェックを行い，障害時には自分への経路の優先度を下げる方法をセルフ監視と記す．また，他のサーバのヘルスチェックを行い，障害時には自分への経路の優先度を上げる方法を相互監視と記す．セルフ監視では，監視機能がそのサーバ内だけで動作するためシンプルな構成が可能であるが，他のサーバやネットワークと協調した動作はしない．相互監視では，他のサーバや

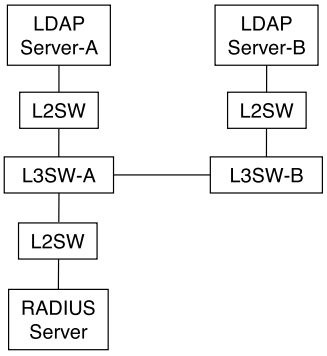

図 4 評価システムの接続構成
Fig. 4 Evaluation system.

ネットワークの状況に応じて動作することができるが，前述のとおり負荷集中やブラックホールとなる場合があるため注意が必要である．

これらの方法は，別システム監視のような総合的な管理機能が必要でない場合に有効であり，単一障害点や通信経路のボトルネックの問題を回避することができる．ライトウェイトな冗長化システムとして運用することが可能である．冗長化の要件に応じて，これらの構成方法を選択することができる．本論文ではシンプルな冗長化システムとして，主にセルフ監視を想定する．

4. 評価システムの実装と評価

TCP に IP Anycast を適用した冗長化においてサーバの切替りを検証するため，岡山大学のキャンパス情報ネットワークに評価システムを構成した．また，同システムにおいてフェイルオーバ，フェイルバックの動作を検証し，切替え時間の評価を行った．

4.1 評価システムの実装

3.3 節の 4 つの条件をすべて満たす代表例として LDAP サービスの冗長化において提案方法が正しくフェイルオーバ，フェイルバックを行えるかどうかを評価するシステムを実装した．このシステムでは**図 4** に示すように 1 台の RADIUS サーバが LDAP クライアントとして LDAP サーバを参照する構成になっている．機器構成を**表 1** に示す．L3SW-A と L3SW-B の間はバックボーンであり，ルーティングプロトコルは OSPF を運用した．また，L3SW-A と L3SW-B の支線では RIP を運用し，RIP のエージングタイマーは 180 秒に設定した．各 LDAP サーバでは RIP を運用し，IP Anycast の仮想 IP アドレスの経路情報をネットワークに広告した．RADIUS サーバは，認証情報として LDAP サーバを参照するため，参照先として LDAP サーバの IP Anycast アドレスを指定した．RADIUS サーバは FreeRADIUS を，LDAP サーバは OpenLDAP を使用した．

表 1 評価システムの機器構成
Table 1 Specifications of the evaluation system.

	機器構成
L3SW-A,B	ALAXALA Networks Corp. AX6708S
LDAP Server-A	CPU Xeon 3065（2.33 GHz） メモリ 2 GB OS CentOS 5.10 64 bit
LDAP Server-B	CPU Pentium D（3 GHz） メモリ 1.5 GB OS CentOS 5.10 64 bit
RADIUS Server	CPU Pentium D（3 GHz） メモリ 1.5 GB OS CentOS 6.5 64 bit

4.2 評価システムによる動作検証および評価

評価システムにおいて，RADIUS サーバが現在参照している LDAP Server-A の LAN ケーブルを抜き差しすることで，疑似的に障害と復旧を発生させ，RADIUS サーバの参照先の切替りを確認した．なお，RADIUS サーバの参照先が切替る契機は，サーバの停止，ルーティングデーモンの停止，サービスのプロセスの停止である．LAN ケーブルが抜かれることは，LDAP サーバからの経路情報が届かなくなることであり，サーバでのルーティングデーモンが停止することと同様である．また，サービスのプロセスが停止した場合は，3.6 節のサービスの死活監視によってサーバへの経路を変更するため，サーバでのルーティングデーモンが停止することやサーバが停止したことと同様である．LDAP Server-A，LDAP Server-B では tcpdump を実行し，サーバの通信状態を確認した．また，RADIUS サーバでは，localhost に対して radtest を 1 秒間隔で実行し，認証の成功と失敗を確認した．検証実験の手順を以下に示す．

- 障害の発生によるフェイルオーバ
  (1) Server-A の LAN ケーブルを抜く
  (2) RADIUS サーバの参照先が LDAP Server-B に移る
- 障害の復旧によるフェイルバック
  (3) Server-A の LAN ケーブルを差す
  (4) LDAP Server-A のネットワークインタフェースの通信が復活する
  (5) RADIUS サーバの参照先が LDAP Server-A に戻る

この検証を 5 回行い，切替わり時間を測定した結果を**表 2** に示す．フェイルオーバの時間が平均 171 秒であり，RIP のエージングタイマが 180 秒であることから想定どおりの時間といえる．また，フェイルバックの 11 秒についても，RIP の Update が 30 秒であることから想定どおりの時間といえる．pop switch は手順 (2) と (5) で発生するが，サーバの切替わりとほぼ同時にサービスが再開されることを確認した．特に，フェイルバックの際に発生す

表 2 切替わり時間
Table 2 Switching time.

動作	最小 (秒)	最大 (秒)	平均 (秒)	検証手順の対応
フェイルオーバ	163	180	171	(1)–(2) 間
LAN ポートのネゴシエーション	30	31	30	(3)–(4) 間
フェイルバック	3	19	11	(4)–(5) 間

る pop switch に関しては，手順 (2) から手順 (5) の間はクライアントは LDAP Server-B を参照し続けることから，サービスの停止は事実上なかった．なお，LAN ポートのネゴシエーションは本論文とは無関係であるが，検証手順の明確化のため記載した．

学内ネットワークの設定を変更することは運用に支障が生じる可能性があるため実施できなかったが，もしすべてのルーティングを OSPF で運用したとすると，経路の切替わり時間が短縮される．フェイルオーバは，OSPF の Dead 間隔である 40 秒程度に，フェイルバックは OSPF のコンバージェンス時間に短縮されると考えられる．

4.3 岡山大学 LDAP サーバへの適用

次に提案方法の実システムへの適用例として，2012 年 12 月に行った岡山大学統合認証システムへの適用事例を説明する．この事例では特に参照する LDAP サーバを 1 台しか指定できない，既存のクライアントに対する可用性の向上を目的とした．

当時の岡山大学では統合認証システムのサービスを行っており，約 5 km 離れた津島キャンパスと鹿田キャンパスに各 1 台の LDAP サーバを運用していた．2 台の LDAP サーバはレプリケーションによって同じ情報を有する複製サーバとなっていた．これらの LDAP サーバは本学の教員，職員，学生，来訪者の一時アカウントなどすべてのユーザのアカウント情報を有しており，ロケーションフリーネットワークシステムの 2 台の RADIUS サーバ，教務システムの 2 台のサーバおよび 16 台の PC 端末から認証サーバとして利用されていた．LDAP クライアントであるこれらのサーバや PC 端末では，運用しているシステム構成により，2 台の RADIUS サーバを除いて LDAP サーバとして設定できるホストが 1 台に限られる問題があった．通常では各 LDAP クライアントはどちらかの LDAP サーバを参照しているが，LDAP サーバで障害が発生すると，各 LDAP クライアントの設定を手作業で変更し，もう一方の LDAP サーバを参照させなければならなかった．このため，長時間のサービス停止と多大な作業工数が発生する問題があり，障害が発生した場合でも，数分以内[*1]に自動的にサービスを復旧させる必要があった．

そこで，IP Anycast によって 2 台の LDAP サーバを構成し，Active-Active としてシステムを実装した．これは，どちらかの LDAP サーバで障害が発生すると，もう一方にフェイルオーバしてサービスを継続し，またサーバが復旧し経路が戻るとフェイルバックするように動作させるためであった．サービスの死活監視には，セルフ監視を採用した．2 台の LDAP サーバは運用状態のため，本論文のために切替え時間などをテストすることはできなかったが，システムの構築作業において障害を想定した動作試験を行い，次のとおり確認した．

(1) 津島キャンパスの LDAP サーバのネットワークインタフェースを down した．
(2) その結果，津島キャンパスの LDAP クライアントの接続先が鹿田の LDAP サーバにフェイルオーバした．
(3) その後，津島キャンパスの LDAP サーバのネットワークインタフェースを up した．
(4) その結果，津島キャンパスの LDAP クライアントの接続先が津島の LDAP サーバにフェイルバックした．

津島キャンパスの LDAP クライアントでは，それぞれ 3 分以内で自動的にサービスが復旧した．鹿田キャンパスの LDAP サーバ，LDAP クライアントでも同様の試験結果を得た．なお，LDAP サーバは，NEC 社の Enterprise Directory Server（以下，EDS と記す）であり，OS は RedHat Enterprise Linux 5 であった．津島キャンパスと鹿田キャンパスには，それぞれコアスイッチとなるレイヤ 3 スイッチが設置され，両キャンパス間は岡山情報ハイウェイ [16] の光ファイバによって 2 回線の 10GbE で接続されている．IP Anycast による LDAP サーバの冗長構成を**図 5** に示す．

この事例では各 LDAP サーバのループバックインタフェースに IP alias で共通の仮想 IP アドレス 150.46.X.Y を割り当て，150.46.X.Y/32 の経路情報を RIP で広告するようにしてから LDAP クライアントの参照先 LDAP サーバを 150.46.X.Y に変更した．このため，構築作業中でも各 LDAP サーバの固有 IP アドレスである 150.46.c.d や 150.46.e.f を使用し続けることが可能で，LDAP サーバを 1 台しか設定できないクライアントであっても参照先 LDAP サーバの変更時および動作試験時を除いてサービスの停止はなかった．なお，教務システムの 2 台のサーバおよび 16 台の PC 端末は 2014 年 7 月の事務システムの更新まで利用したが，運用期間中にはフェイルオーバ，フェイルバックの発生は確認されなかった．

[*1] 利用者からの苦情がそれほど多くない程度のサービス停止時間以内．

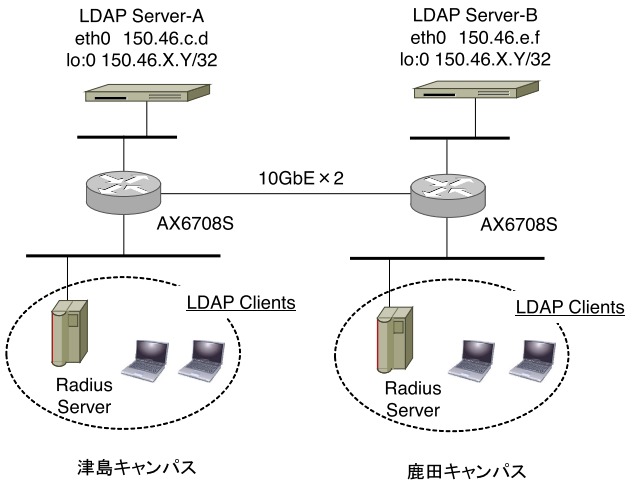

図 5 LDAP サーバの冗長構成

Fig. 5 Redundant configuration of LDAP servers.

5. まとめ

本論文では地理的な分散配置によりサーバを冗長化したシステムにおいて，複数のサーバを指定できないクライアントや，複数のサーバを指定できるクライアントのうちフェイルバックが自動的に行われないものの一部に対して，IP Anycast 技術を利用してフェイルオーバおよびフェイルバックを自動的に行う方法を提案し，またこの方法を適用できるサービス，クライアントおよびサーバが満たすべき条件を示した．また，この条件を満たすサービスの代表例として LDAP による認証サービスをあげ，実際にフェイルバックおよびフェイルオーバが経路の切替わりに要する時間程度で行われることを確認した．さらに岡山大学において実際に運用し，フェイルオーバ，フェイルバックの発生事例は本論文執筆時までに確認されていないものの，動作確認では問題がないことを確認した．これらの結果から，提案方法は図 5 と同様の構成および少なくとも LDAP による認証サービスに対して有効であるといえる．

今後の課題としては，3.3 節で述べた条件を満たす，LDAP による認証サービス以外のサービスでの検証があげられる．また，ルーティングプロトコルとして OSPF を用いた場合のフェイルオーバ時間の評価や RIP，OSPF におけるコンバージェンス時間に関係するパラメータの最適化も行いたい．さらに，適用範囲を拡大するためにサーバ間でのセッション管理，処理の継続，データ同期などが可能なサービスでの検証やマルチホーミング環境への対応方法の検討も今後の課題としてあげられる．

参考文献

[1] 大隅淑弘，山井成良，岡山聖彦，河野圭太，藤原崇起：地理的に分散したサーバ間のフェイルオーバ・フェイルバックを可能にする複製サーバ冗長化構成，第 12 回情報科学技術フォーラム講演論文集，Vol.4, pp.23–27 (2013).

[2] Wahl, M., Howes, T., Kille, S.: Lightweight Directory Access Protocol (v3), RFC2251, IETF (1997).

[3] Rigney, C., Willens, S., Rubens, A. and Simpson, W.: Remote Authentication Dial In User Service (RADIUS), RFC 2865, IETF (2000).

[4] Partridge, C., Mendez, T. and Milliken, W.: Host Anycasting Service, RFC1546, IETF (1993).

[5] Mockapetris, P.: DOMAIN NAMES - CONCEPTS AND FACILITIES, RFC1034, IETF (1987).

[6] Weiden, F. and Frost, P.: Anycast as a Load Balancing feature, *LISA '10 Proc. 24th International Conference on Large Installation System Administration*, pp.1–6 (2010).

[7] Dovecot (online), available from ⟨http://dovecot.org/index.html⟩ (accessed 2015-09-30).

[8] The FreeRADIUS Server Project and Contributors: FreeRADIUS: The world's most popular RADIUS Server (online), available from ⟨http://freeradius.org/⟩ (accessed 2015-9-30).

[9] Myers, J. and Rose, M.: Post Office Protocol - Version 3, RFC1939, IETF (1996).

[10] Crispin, M.: INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1, RFC3501, IETF (2003).

[11] ALAXALA Networks Corporation: AX2400S ソフトウェアマニュアルコンフィグレーションガイド Vol.2 (online), 入手先 ⟨https://www.alaxala.com/jp/techinfo/archive/manual/AX2400S/PDF/11_7/CFGUIDE2/CFGUIDE2.PDF⟩ (参照 2015-09-30).

[12] Ballani, H., Francis, P. and Ratnasamy, S.: A measurement-based deployment proposal for IP anycast, *IMC '06 Proc. 6th ACM SIGCOMM Conference on Internet Measurement*, pp.231–244 (2006).

[13] Malkin, G.: RIP Version 2, RFC2453, IETF (1998).

[14] Moy, J.: OSPF Version 2, RFC2328, IETF (1998).

[15] Quagga Routing Suite (online), available from ⟨http://www.nongnu.org/quagga/⟩ (accessed 2015-05-03).

[16] 岡山県県民生活部情報政策課，岡山県：OKIX (online), 入手先 ⟨http://www.pref.okayama.jp/page/detail-8208.html⟩ (参照 2015-01-09).

大隅 淑弘 （正会員）

昭和 58 年近畿大学理工学部電気工学科卒業．昭和 63 年静岡大学電子工学研究所技官．平成 4 年岡山大学総合情報処理センター（現，情報統括センター）技官を経て，現在，同技術専門職員．平成 23 年岡山大学大学院自然科学研究科（産業創成工学専攻）博士後期課程入学，平成 26 年単位取得後退学．ネットワーク運用管理，情報セキュリティにかかわる業務に従事．博士（工学）．

山井 成良 （正会員）

昭和 59 年大阪大学工学部電子工学科卒業．昭和 61 年同大学大学院博士前期課程修了．昭和 63 年同大学大学院基礎工学研究科（物理系専攻情報工学分野）博士後期課程退学．同年奈良工業高等専門学校情報工学科助手．同講師，大阪大学情報処理教育センター助手，同大学大型計算機センター講師，岡山大学総合情報処理センター（現，情報統括センター）助教授を経て，平成 18 年同教授．平成 26 年より東京農工大学大学院工学研究院教授．分散システム，ネットワーク運用管理，ネットワークセキュリティの研究に従事．IEEE，電子情報通信学会各会員．博士（工学）．

岡山 聖彦 （正会員）

平成 2 年大阪大学基礎工学部情報工学科卒業．平成 4 年同大学大学院基礎工学研究科博士前期課程修了．同年同大学院基礎工学研究科博士後期課程を退学し，同大学工学部助手．奈良先端科学技術大学院大学情報科学研究科助手，岡山大学工学部助手，同大学総合情報基盤センター助教を経て，平成 22 年同大学情報統括センター助教．平成 23 年同准教授．博士（工学）．インタネットアーキテクチャ，ネットワーク管理，ネットワークセキュリティの研究に従事．電子情報通信学会会員．