情報処理学会論文誌 Vol. 50 No. 8 1810–1822 (Aug. 2009)

レイヤ2ネットワークにおけるループ障害のリモート診断方式

勝山恒男[†1] 安家武[†1] 野村祐士[†1]
若本雅晶[†1] 野島聡[†1]
木下和彦[†2] 村上孝三[†2]

IP スイッチの処理能力向上にともない，レイヤ 2 ネットワーク規模が拡大し，経済的に大規模なレイヤ 2 ネットワークの構築が可能となってきている．これにともない，障害の波及範囲も広域化し，障害復旧に時間を要し，可用性を低下させる要因となっている．レイヤ 2 ネットワークの典型的な大規模障害であるループ障害では，ループパケットは，消滅することなくネットワーク内を転送し続ける現象が起き，システム全体の稼働停止に至ることがよく知られている．この現象は，本来バス構造であった LAN セグメントをスタートポロジにしたが，ブロードキャスト型のプロトコルを変えていないために起きる本質的問題である．これを回避する従来技術として STP（Spanning Tree Protocol）が適用されているが，十分な効果が得られない場合も多い．そこで，本論文では，ループ障害を対象に，従来の障害事前防止型のアプローチではなく，その原因箇所をリモートホストから迅速に探索発見する診断型のアプローチを提案する．本方式は，第 1 ステップとして，診断探索を行う端末から送信するロングパケットによってノード負荷の低減を行い，また，架空 MAC アドレスからのブロードキャスト ARP 要求の送信によって MAC アドレス誤学習を訂正し，本来の通信機能の回復を行う．次いで，第 2 ステップとして，誤学習の成否と大量パケット受信ポートの分析によってループ箇所の特定を行う 2 ステッププロセスからなるリモート診断を特徴としている．

Remote Discovery of Loop Trouble in Layer2 Networks

Tsuneo Katsuyama,[†1] Takeshi Yasuie,[†1]
Yuji Nomura,[†1] Masaaki Wakamoto,[†1]
Satoshi Nojima,[†1] Kazuhiko Kinoshita[†2]
and Koso Murakami[†2]

Layer2 network expands widely for large networks as the layer2 switch improves the forwarding ability. Therefore, the impact of layer2 troubles is big, and it spends long time to find out the cause of the trouble because of geographical expanse. In the layer2 loop trouble which is a typical, large-scale trouble of layer2, the loop packets cause the phenomenon to keep forwarding in the network without disappearing, and it becomes the entire system failure. An essential issue is that the access protocol of the broadcast type has not been changed for long time. The existing technology to avoid loop failure such as STP is not enough. In this paper, it proposes not preventive approach to prevent a trouble but discovery approach to find out the cause part from a remote host system. After the node load is decreased with the long packet injection from the diagnosis terminal as the first step, and the recovery of the mis-learning of MAC address study, the layer2 network functions recover. In the next step, the loop point is identified by detecting the large amount traffic receiving port and reachability test by using the recovered network.

1. はじめに

イーサネットは，ローカルエリアネットワーク（LAN）で最も一般的な技術であるばかりでなく，広域ネットワークにも急速に適用され始めている．2000 年頃におけるイーサネットの適用範囲は，ビルやフロア内のたかだか数百台の端末からなるネットワークであり，あるセグメントから他のネットワークへの接続はルータ，あるいはレイヤ 3 スイッチを介して行われていた．しかし，近年では，VLAN によってブロードキャストセグメントと物理セグメントを分離できるようになり，イーサネットを広域ネットワークとして活用する事例が増えつつある．具体的には，複数のフロアにまたがった LAN の構築や，建物を越えたイントラネットへの適用，さらに，キャリアによる広域イーサネットサービスとしての利用などがあげられる[1)]．

しかし，本来のイーサネットは少数のセグメントからなるネットワークを対象としており，OSI のプロトコル階層におけるレイヤ 2 機能，すなわち，リンク制御が主たる機能である．各種ルーチングプロトコルや，障害などに対応する管理プロトコルは，レイヤ 3 で提供され，レイヤ 2 の管理プロトコルとしては，ループ経路を遮断するための STP（Spanning Tree Protocol）群[2)–4)] のみが規定されている．アドレス解決には LAN セグメントの接続形態がバス構造であった頃に設計されたブロードキャスト型の ARP（Address Resolution Protocol）が用いられている．

†1 株式会社富士通研究所
Fujitsu Laboratories Ltd.
†2 大阪大学
Osaka University

そのため，レイヤ 2 スイッチを多段に接続したイーサネットでは，何らかの障害，配線ミスなどのトラブルが発生した場合，ブロードキャストによる影響がネットワーク全域に及ぶ．また，管理機能の不足から，トラブル位置の検出と修復に多くの手間がかかる．最も重要な大規模ネットワーク障害と認識されているレイヤ 2 ネットワークのループ障害（以降，単にループ障害と呼ぶ）[5)–7)] は，現在でもイーサネットの大きな弱点であり，いったん，この障害が発生すると，ARP などのブロードキャストパケットがループ経路を巡回し続けるためにネットワークは高負荷になり，通信不能に陥る．この障害を回避する従来技術としては前述の STP があげられる．しかし，すべてのスイッチで STP を動作させなければ効果が得られないことなど，実システムへの適用性に課題があり，完全な防止手段はないのが現状である．

イーサネットの構成要素であるレイヤ 2 スイッチは，レイヤ 3 スイッチに比較して構成がシンプルであり，パケットのデータグラム転送という本質的な機能に特化したネットワークノードである．イーサネットをさらなる大規模ネットワークでも利用するためには，その特長を失うことなく，より安定で信頼度を高いものにする管理体系の構築が，重要かつ緊急の課題であるといえる．現実的には，すべてのループ障害を事前に防止することは不可能であることから，本論文では，事前防止型のアプローチではなく，原因箇所を迅速に探索発見する診断型のアプローチを提案する．

以下，2 章では，ループ障害によって通信が途絶する現象を通信帯域と通信経路形成の点から明らかにし，この現象を回避するための従来技術とその問題点を述べる．次に，3 章でループ障害の原因箇所特定方式を提案し，4 章で試作システムを用いた実験によって提案方式の有効性を明らかにする．最後に 5 章で今後の検討課題とその見通しを示す．

2. ループ現象解析と従来技術の問題点

2.1 ループ現象の解析

複数のレイヤ 2 スイッチ（以下，単にスイッチと呼ぶ）で構成された一般的なイーサネットの例を**図 1** に示す．このとき，1 つのスイッチの機器障害によって配下のすべての通信が途絶することを避けるために，図 1 の点線に示す冗長リンクを設け，たとえば SW-A の障害時には，この冗長リンクを利用することで，通信を継続できるような構成にすることが一般的である．ただし，この冗長リンクによってループ経路が生じないように，通常動作時は，STP によって，冗長リンクのポートを論理的に通信できない状態に設定している．ループが構成される原因としては，STP 機能に障害が発生し，通常動作時に冗長リンクでのパケット転送が行われてしまう場合や，STP による対処を行っていないスイッチの 2 つのポートを直接折り返し接続するケーブル誤接続などがある．その挙動は，原因によらず，また，複数のスイッチにわたって発生する場合も，単体のスイッチで発生する場合も同じである．以下，簡単のために，スイッチ単体でのループ障害発生を例にとって，その挙動を示す．

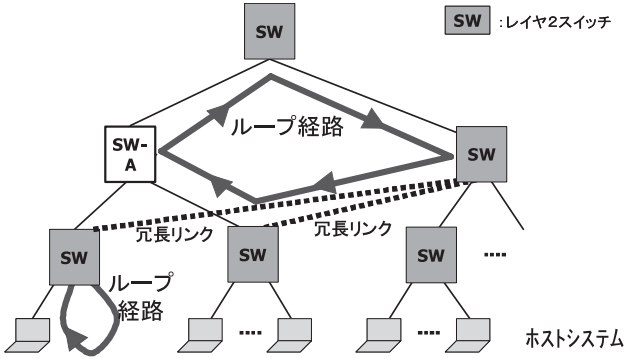


図 1　レイヤ 2 ループの発生例
Fig. 1　Typical example of layer2 loop failure.

(1)　ブロードキャストストームによるノード負荷増大

ループが構成されたネットワークに送信された ARP，NetBIOS などのブロードキャストパケットは，サブネット内全域に転送され，ループを構成するスイッチ（以下では，L-SW と呼ぶ）では，同じパケットが巡回を続け，ブロードキャストパケットとして永久に転送され続ける．レイヤ 3 には，TTL（Time To Live）というホップ数カウンタがあり，レイヤ 3 中継装置でルーティング処理を行うごとにこのカウンタが減算され，0 になるとパケット廃棄される機構があるが，レイヤ 2 では同様の機構がないため，ブロードキャストストームと呼ばれる大量のトラヒックが発生する．この影響で，帯域が大幅に圧迫され，また，同一サブネット内のコンピュータシステム（以下では，サーバやクライアントシステムなどの IP レベルで通信可能な装置を総称して端末と呼ぶ）およびスイッチは高負荷となり，稼働停止や処理能力の著しい低下を引き起こす．

(2)　MAC アドレスの誤学習による通信途絶

加えて，同一サブネット内の L-SW が，ブロードキャストパケットを送信している端末であるかのように見える現象が起きる．たとえば，**図 2** に示す例では，発信端末-A からのブロードキャストパケットは，L-SW でループし，L-SW からさらに他のスイッチに発信さ

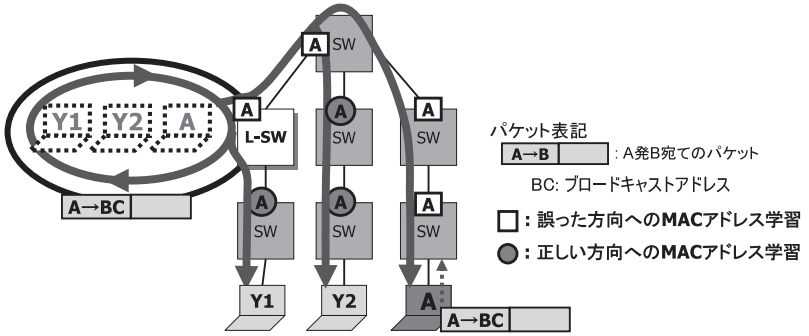


図 2　パケットループによる MAC アドレスの誤学習
Fig. 2　Mislearning of MAC address by broadcast packet loops.

れる．各スイッチでは，パケットを受信したポートの方向にパケットの発信端末が存在するとして MAC アドレスを学習するため，各スイッチでは L-SW に向かうポートに端末-A があると学習する．すなわち，ブロードキャストパケットの発信端末と L-SW の間にあるすべてのスイッチで，端末-A が L-SW に存在という MAC アドレス誤学習が行われる．その結果，ブロードキャストパケットの発信端末に宛てた通信は，すべて L-SW に向かい，発信端末には届かなくなる．

このように，ループ経路の形成は，ブロードキャストストームによる負荷増大と，MAC アドレスの誤学習という 2 点の問題を引き起し，正常な通信ができなくなるという重大な障害の原因となる．

2.2　従来技術とその問題点

従来技術としては，ループの発生を防止する技術と，発生したループの影響を軽減する技術がある[5),6)]．

防止技術としては，前述の STP が最もよく利用されている．本プロトコルでは，ループを構成する冗長リンクのポートからブロッキングポートを選定し，そのポートの通信機能を停止させる．ただし，STP には以下のような問題がある．

- CPU 障害などの場合，パケット転送は正常だが，STP が動作しないような装置故障が発生する．この場合ループ転送を防止できない．
- 冗長構成を形成する装置の中に，STP の制御パケットを中継しないスイッチが 1 台でもあると機能しない．

一方，ループの影響を軽減する技術としては，パケットを受信したときに，ループパケットを個別に検知してフィルタリングする技術[8)]がある．本技術は，通過パケットをリアルタイムに全数記憶しておき，受信時にパケットの内容あるいは送受信先アドレスなどを照合して同一であればループ発生と判断し，そのパケットを廃棄するものである．また，専用のループ検出用パケットを送信し，送信スイッチに検出用パケットが戻ってきたこと，あるいは，送信元アドレスが自スイッチであるパケットを受信したことでループ障害を検出し，以後，該当するポートに送られてくるパケットをすべてブロックする技術[9),10)]もある．本技術は，送信したスイッチのパケット情報のみを記憶するため，ループパケットを個別に検知してフィルタリングする技術と比較して必要な記憶容量を大きく削減できる．しかし，ループ障害を起こしたネットワークからのパケットをブロックすることで他のネットワークの機能維持を図ることが可能な場合もある一方，ブロックによってネットワーク全域の正常な通信にも悪影響を与えることもあり，本技術の効果は対象ネットワークのトポロジに強く依存するという問題がある．これらの技術の問題点を以下にまとめた．

- システム構成に依存するが，障害箇所を特定するためには，多数の本技術搭載スイッチを配備する必要があり，経済的でない．
- 通過パケットをリアルタイムに検査するために，高い処理能力が必要となる．
- ネットワークトポロジによっては，ポートのブロックでは対処できない場合がある．

なお，関連する技術として，ループパケットを検知するのではなく，各受信ポートにおいてブロードキャスト/マルチキャストアドレス宛てのパケット受信数に上限を設定しておき，これを超えたときに該当するポートの帯域制限あるいはブロックを行うストームコントロール技術[11)]もある．しかし，本技術では，正しい経路で送信されているパケットも制限または廃棄されるため，通常の ARP や NetBIOS などの送受信にも影響を及ぼし，正常な通信ができなくなるといった問題がある．

また，レイヤ 2 にも，レイヤ 3 と同様に，パケットの生存時間を規定する技術も考えられている．2004 年の IETF で議論が開始された[12)]が，以下の問題点がある．

- 現システムを構成するスイッチに新機能を導入する必要があり，既存網への適用は現実的でない．
- ブロードキャストの本質的な役割を維持する必要があることから，単なる転送回数規制として規定するだけで十分かなど，標準技術とするためには検討すべき点が多い．
- MAC アドレスの誤学習を回避することはできないため，この技術だけで通信の疎通性を確保することはできない．

このように，レイヤ 2 ネットワーク発展にともなって，対応技術が開発されつつあるが，完全な対策とはなっておらず，経済的で確実にループ障害に対応できる方式の開発が急務となっている．

3. ループ障害原因のリモート診断方式の提案

3.1 提案方式のアプローチと特長

ネットワーク利用形態の中には，一瞬も停止させられないミッションクリティカルなシステムもあるが，数分程度の停止を許容する代わりに経済的で簡易な管理が期待されるものも多い．このような場合には，ループ障害を予防する技術までは要求されず，発生した障害を迅速に復旧させる技術を経済的に実現することが重要となる．そのためには，ループ障害の原因箇所を正確に特定する技術が必要である．たとえば，ケーブル誤接続による障害の場合では，早期に誤接続箇所を特定できれば，原因である誤接続ケーブルを抜き取る作業そのものは非常に簡単で，ただちに障害を解消することができる．停止時間も長くならず，高価な付加機能を有するスイッチを導入する必要もないため，経済的である．しかし，従来は，ループ障害が発生するとネットワークの正常な機能はすべて失われるため，ネットワークを利用したリモート診断は不可能と考えられていた．

本章では，従来の事前防止型のアプローチではなく，その原因箇所をネットワークに接続された端末から迅速に探索発見する診断型のアプローチをとる方式を提案する．**図 3** に手順を示すように，まず，過負荷およびアドレス誤学習によって通信機能を失っているネットワークに対して，障害箇所を探索する端末（探索端末-X と呼ぶ）をネットワークに接続し，高負荷の緩和とアドレス誤学習の訂正を行い，その機能を回復する．次に，通信機能が回復したネットワークを用いて，探索端末-X と診断の対象となる端末（対象端末-T と呼ぶ）との間で通信を行い，ループ箇所を診断する．このとき，ループトラヒック流をたどって L-SW を特定する診断方法と，誤学習の状況を分析して L-SW を特定する診断方法を組み合わせている．

3.2 通信機能の回復手順（第 1 ステップ）

リモート診断の第 1 ステップとなる通信機能の回復手法を以下に示す．

(1) 障害の検出

まず，障害の疑いのあるレイヤ 2 ネットワークに前述の探索端末を接続する．具体的には，以下に述べる診断手順を実行可能なパーソナルコンピュータなどを用いる．探索端末はパケットキャプチャにより，同一内容のブロードキャスト/マルチキャストパケットを一定

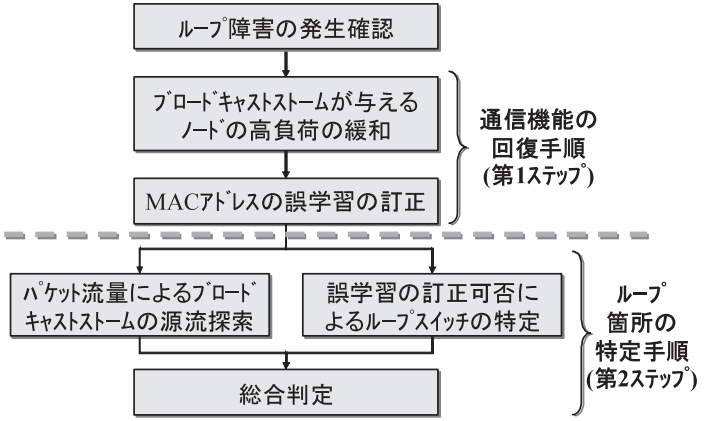


図 3　提案方式のリモート障害診断手順
Fig. 3　Proposed diagnosis method of packet loop points.

数以上受信したことで，ループ障害発生を検出する．パケット全数検査を行う必要はないため，探索端末に高い処理能力は求められない．

(2) ノード負荷の低減

一般に，平均パケット長が短くなるほど，単位時間あたりのノード到着パケット数が増大し，端末の処理負荷は大きくなる．ブロードキャストストームに含まれるパケットの多数を占めるのは，ブロードキャスト通信を行う ARP 要求パケットであり，これは通常 64 バイトのショートパケットである．スイッチ内バッファでの滞留時間はパケット長に比例するため，多数のロングパケットを送出することでショートパケットの廃棄率が高くなれば，ノードのスイッチング負荷が小さくなると期待される．そこで，探索端末からパケット長の長いブロードキャストパケットをネットワークに送出する仕組みを実施し，ショートパケットが時間とともに廃棄され，駆逐されていくことを狙う．

(3) MAC アドレス誤学習の訂正

2 章で述べたように，ループ障害発生時にはブロードキャストストームにより各スイッチで MAC アドレス誤学習が起きている．この状態では，探索端末からパケットを送出してもパケットは L-SW に向かうため，正しく通信できない．診断のためには，探索端末と診断の対象となる対象端末との間の双方向通信を正常化させる必要がある．そこで，探索端末から対象端末へ向かう送信経路の誤学習訂正を行い，次に，逆方向である探索端末への受信経

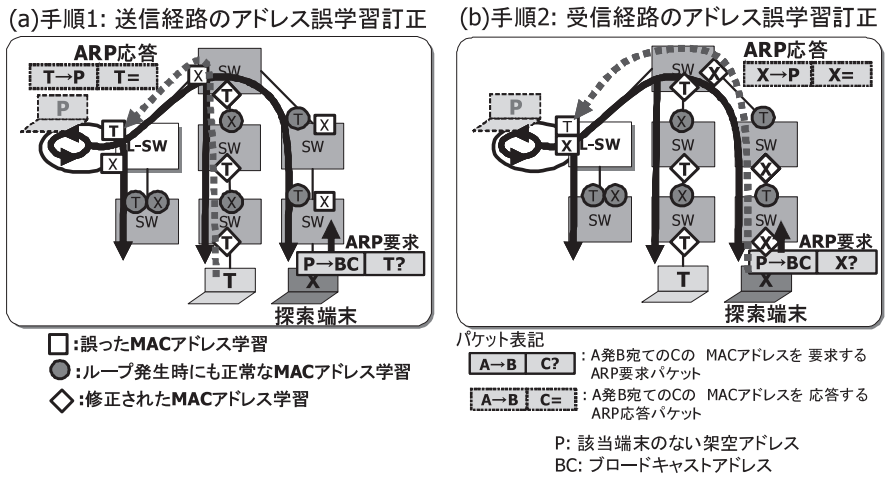


図 4　MAC アドレスの誤学習の訂正
Fig. 4　Correction of MAC address mislearning.

路の誤学習訂正という 2 つの手順に分けて訂正を行う．

以下に**図 4** (a)，(b) を用いて，探索端末の送信経路と受信経路のそれぞれの訂正手順例を示す．なお，誤学習訂正過程では，誤学習しても影響のない実在しない MAC アドレスを設定したブロードキャスト ARP 要求を発生させている．また，このブロードキャスト ARP 要求パケットは一般的なショートパケットではなく，ペイロード部にダミーのデータを付加したロングパケットとして送出することで，前項で示したノード負荷低減効果も図っている．

- 手順 1：探索端末の送信経路の誤学習訂正

探索端末-X から，架空アドレス（P）発で対象端末-T を要求する ARP 要求をロングパケットで送信し，ループパケットをこれで置換する．

L-SW から送信される本 ARP 要求に応える対象端末-T からの ARP 応答によって，図 4 (a) の破線経路上のスイッチの対象端末-T に関する誤学習が訂正される．なお，残りのスイッチの学習は，もともと誤っていないので，訂正の必要はない．

- 手順 2：探索端末への受信経路の誤学習訂正

探索端末-X から，架空アドレス（P）発で探索端末自身（X）を要求する ARP 要求を送信する．

L-SW から送信される本 ARP 要求に応える探索端末-X からの ARP 応答によって，図 4 (b) の破線経路上のスイッチにおける探索端末-X の誤学習が訂正される．なお，残りのスイッチの学習は，もともと誤っていないので，訂正の必要はない．

本手順の後，探索端末-X と対象端末-T 間の通信が回復し，MIB 取得や ping による疎通確認などが可能となる．本手順は，通信を回復させたい対象端末ごとに行う．

なお，以降の原因箇所の特定手順の前に対象端末-T の MAC アドレスを取得しておく必要がある．MAC アドレスがあらかじめ分からない場合には，MAC アドレス誤学習訂正手順に先立って，探索端末-X から対象端末-T のアドレスを解決する通常のブロードキャスト ARP 要求を行っておく．すると，T から X へのユニキャストの ARP 応答パケットがループ箇所で周回するようになる．この応答パケットは，誤学習が訂正されると同時に，ループから取り出され，X はそのユニキャスト ARP 応答を受信することができる．この手順であらかじめ必要なのは対象端末-T の IP アドレスだけであり，通信機能の回復手順の実行中に T の MAC アドレスを取得することができる．

3.3　ループ箇所の特定手順（第 2 ステップ）

通信機能が回復したネットワークを用いて，ループの原因箇所を特定する提案手法を以下に示す．原因箇所がブロードキャストストームの発生源となっていることから，その発生方向から問題箇所を絞り込むトラヒック流量検査方式と，L-SW だけは誤学習を訂正できないことを逆に活用した誤学習訂正の確認結果によって，ループ箇所を絞り込む疎通確認方式を併用することを特長としている．

3.3.1　トラヒック流量検査方式

L-SW がトラヒックの大量発生源となることから，L-SW 以外のスイッチでは，L-SW がある方向の 1 ポートからパケットを大量受信するが，L-SW では，ループを構成する 2 ポートから大量受信することとなる．この特性は，**図 5** (a) の単一スイッチでループが構成される場合だけでなく，図 5 (b) の複数スイッチでループが構成される場合においても同様である．そこで，各スイッチの各ポートのパケット受信量（パケット数）を取得し，以下のように判定する．なお，ループ障害では，ループ障害に起因する受信パケットは他のパケットと比べて圧倒的に多くなるので，判定は容易である．

- パケット大量受信ポートが 1 つ：本スイッチは L-SW でなく，そのポート方向に L-SW が存在する．
- パケット大量受信ポートが 2 つ：本スイッチが L-SW であり，そのポート対でループしている．

図 5 (a) の単一スイッチループの場合は，SW4 から大量受信ポートが 2 つ抽出されること

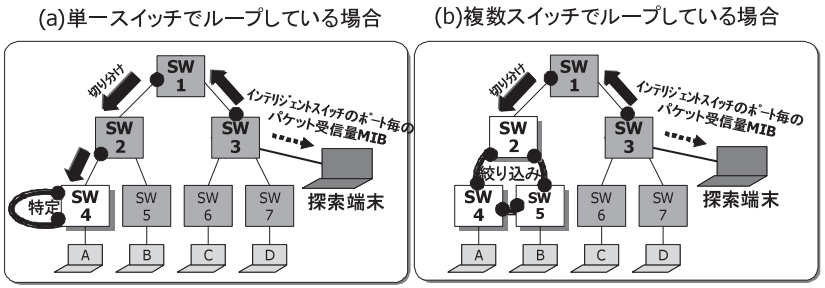


図 5　ポート通過量によるループ箇所の特定
Fig. 5　Loop point exploration by port traffic analysis.

になり，このスイッチが L-SW であり，ループポートも特定できる．図 5 (b) の複数スイッチループの場合は，SW2，SW4，SW5 で大量受信ポートがそれぞれ 2 つ抽出されることになり，この 3 台のスイッチでループしていると判断できる．

なお，本方式は，各スイッチのポートごとのパケット受信量の取得が前提である．図 5 (b) の例で，もし，SW5 のパケット受信量が取得できない場合には，SW2 と SW4 の情報を参考にして形成ループを人手などで絞り込んでいくことになる．1 台でもポートごとのパケット受信量が取得できるスイッチがあれば，その大量受信ポートを遡ることで，L-SW がある方向を，絞り込むことができる可能性がある．本方式は，以下の特長を持つ．

- スイッチのポート単位にループ箇所を特定/切り分けが可能である．
- スイッチ内に，多くの場合具備されている一般的なポートごとのパケット受信数カウンタを利用しているため，経済的な実現が可能である．

3.3.2　疎通確認方式

前節の MAC アドレス誤学習訂正処理を行った後も，探索対象の接続関係によっては通信を回復できない端末が存在する．すなわち，誤学習訂正の手順では，L-SW を周回し続けるパケットによって，再度，誤学習されてしまうために，L-SW は誤学習されたままになる．その結果，探索端末から見て，L-SW より先に位置する端末との間の通信は回復できない．

図 6 を用いて，詳しく説明する．同図は，MAC アドレス誤学習訂正処理における ARP ユニキャスト応答パケットの振舞いを示したものである．探索端末-X は，架空アドレス（P）発で Y1 および Y2 の MAC アドレスを解決する ARP 要求を送信すると（図 6 (1)），ARP 要求はループし L-SW から大量に送出され，端末-Y1，Y2 は ARP 応答（図 6 (2)）を要求

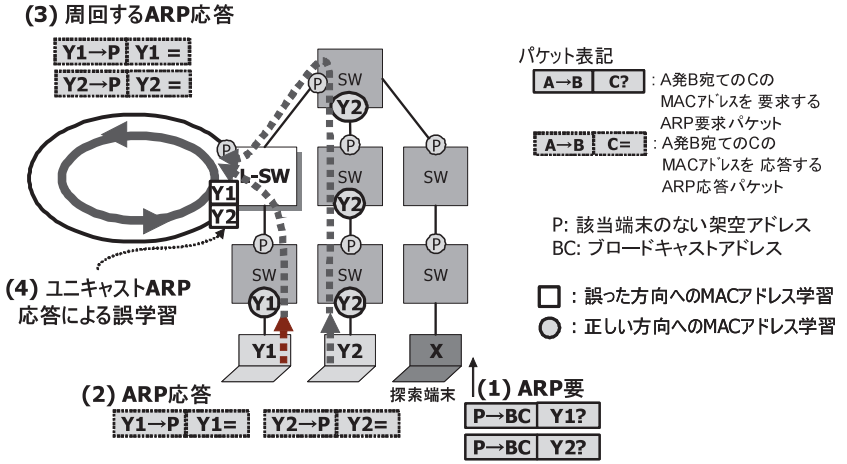


図 6　MAC アドレス誤学習訂正時のユニキャストパケットの振舞い
Fig. 6　Mislearning of MAC address by unicast packet loops.

元である P 宛にユニキャストで返信する．ARP 要求によってアドレス（P）は全スイッチで L-SW の方向に学習されているため，これらの ARP 応答パケットは L-SW に引き込まれるように転送され，各スイッチの Y1，Y2 に対する MAC アドレス誤学習が訂正される．しかし，L-SW において，このユニキャスト ARP 応答は周回し続けるため（図 6 (3)），L-SW でのみ再び誤学習状態となる（図 6 (4)）．つまり，L-SW での誤学習だけは訂正されない．

この結果，探索端末-X から見て L-SW を経由しない位置に接続されている端末-Y2 への通信は回復するが，L-SW を経由する位置に接続されている端末-Y1 への通信は回復しないことになる．この特性は，**図 7** (a) の単一スイッチでループが構成される場合だけでなく，図 7 (b) の複数スイッチでループが構成される場合においても同様である．

この現象を活用すると，探索端末から，各端末に対して，ping などの一般的な手段により疎通確認を行うことで，以下のように判定できる．

- 疎通可：探索端末と対象端末との経路上に L-SW はない．
- 疎通不可：探索端末と対象端末との経路上に L-SW がある．

図 7 (a) の単一スイッチループの場合は，端末-A のみ疎通不可となることから，SW4 が L-SW であると特定できる．図 7 (b) の複数スイッチループの場合，端末-A，B が疎通不可となることから，その共通する経路上のスイッチである SW2 が L-SW であり，SW2 の単

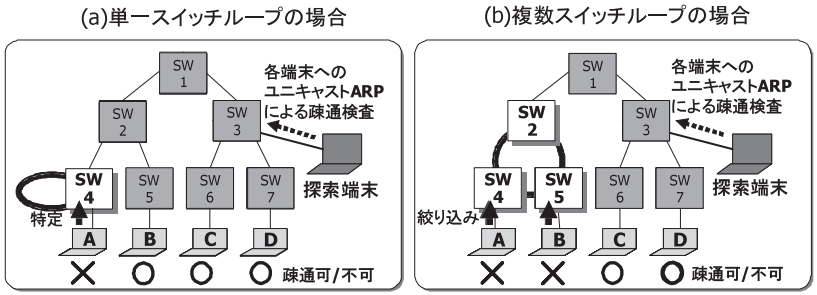


図 7　疎通確認によるループ箇所の特定
Fig. 7　Loop point exploration by reachability analysis.

表 1　診断の前提条件と方式特徴
Table 1　The premise of discovery and the proposed methods.

項目＼方式	提案方式 (トラヒック流量検査)	提案方式 (疎通確認)	STP
狙い	ループ発生箇所の特定 (原因箇所診断)		ループ発生の防止
方式の概要	スイッチポートから、ブロードキャストストームの方向を検知して診断	検査対象端末と探索端末間のルート上のループ構成スイッチの存在を検知して診断	ループを形成するひとつのポートを論理的に通信不可に設定
対象とするネットワーク構成の条件	各スイッチのポート毎のパケット受信量(パケット数)の取得	制約なし	全てのスイッチでのSTP対応が必要
使用するプロトコル	SNMP, Telnet	Ping, ARP	STP
探索のための条件/情報	SNMPのコミュニティ名、スイッチのIPアドレス (ただし、探索中にアドレス情報は取得可能)	端末のMACアドレス (ただし、探索中にアドレスは取得可能)	

一スイッチループ，あるいは，SW2 と，SW4 または SW5 を含む複数スイッチループのいずれかであると絞り込むことができる．本手法の特長を以下にまとめる．

- ループが発生したサブネットに，端末が接続されているだけでよいので汎用性が高い．

3.3.3　両方式の組合せ

トラヒック流量検査方式では，ノードから流量方向を検査するのに対し，疎通確認方式では，端末までのルート，つまり，方向からノードを絞り込む診断を行っている．ノードからルート方向の探索をするトラヒック流量方式に対して，ルート方向からノードを探索する疎通確認方式との位置づけである．そのため，どちらか 1 つの方式では L-SW，および，ループポートを完全には特定できない場合でも，両方式を併用することで，被疑範囲を狭められる場合がある．たとえば，図 7 (b) の場合，SW2 単独あるいは，3 スイッチでのループ構成の可能性を残していたが，トラヒック流量検査方式を併用して，SW2 と SW4 でそれぞれ 2 つの大量受信ポートを検出していれば，複数スイッチループであることが特定できる．

表 1 に，2 種の提案方式の特質や前提条件をまとめて示す．前提条件としては，トラヒック流量方式では，検査対象のスイッチのポートごとのパケット受信量（パケット数）が取得可能なことである．一般的には，SNMP に対応したインテリジェントスイッチであれば，本条件を満たすことが多い．また，疎通確認方式では，ping などが利用可能なことが条件となるが，これも特殊な条件ではない．ping のほかには，ユニキャスト ARP も同じ用途に適用できるので，ネットワークの構成条件に合わせて選択しうる．なお，表中には，参考として，STP の条件もあわせて示した．

なお，本方式では，トポロジ情報の取得を前提とはしていない．トポロジ情報がなくても，トラヒック流量検査方式では，L-SW のポートまで特定でき，また，疎通確認方式では，疎通不可となったルート上に L-SW が存在することまでが分かる．一方，トポロジ情報がある場合，トラヒック流量検査方式では，ループを形成する一部のスイッチの流量が不明でもトポロジ情報を活用してループ形状を絞り込み，すべての L-SW を特定できる可能性がある．また，疎通確認方式では，トポロジ情報を活用して複数の疎通確認結果を照合できるようになり L-SW の位置を絞り込むことができる．

4.　提案方式の評価

4.1　試作システムと診断結果

提案方式の有効性を検証するため，試作システムを開発した．試作システムは，**図 8** に示すように，フリーソフトであるキャプチャエンジン WinPcap とソケット API を用い ARP や SNMP 送受信を行うプロービングモジュール，および，リモート探索・診断ロジックを実行する探索モジュールから構成され，WindowsXP あるいは Windows2000 の.NET Framework 上で動作する．診断は，本システムを探索端末として障害の発生しているサブネットの任意のスイッチに接続し，GUI を通じて実行する．診断結果として，トラヒック流量検査方式では，大量パケット受信している同一スイッチのポート対あるいは L-SW が

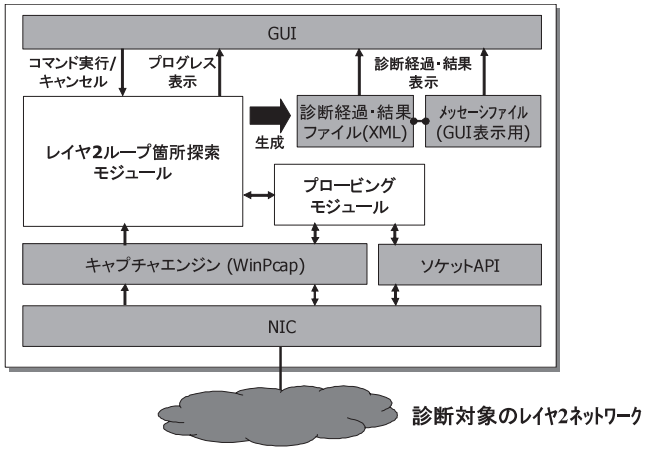
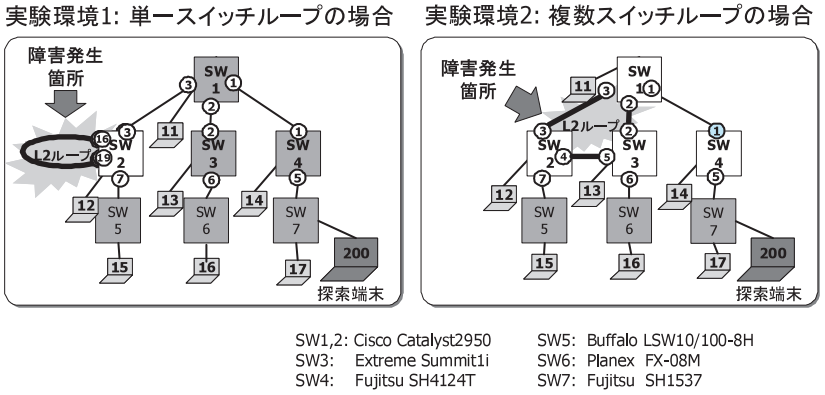


図 8　提案方式の試作システム
Fig. 8　Experimental system applying proposed technology.



SW1,2: Cisco Catalyst2950　　SW5: Buffalo LSW10/100-8H
SW3: Extreme Summit1i　　SW6: Planex FX-08M
SW4: Fujitsu SH4124T　　SW7: Fujitsu SH1537

図 9　実験ネットワークの構成
Fig. 9　Experimental network topology.

存在する方向のポート番号を表示し，疎通確認方式では，探索対象端末までの経路に L-SW が存在するか否かを表示する．

図 9 に示す 2 つの実験環境でループ障害を発生させ，診断結果の評価を行った．実験環境 1 は単一スイッチ（SW2）による折り返しループの場合であり，実験環境 2 は複数スイッチ（SW1，SW2，SW3）でのループの場合である．トラヒック流量検査方式では SW1〜4 の 4 スイッチを探索対象に，疎通確認方式では端末 11〜17 の 7 端末を探索対象とした場合，期待される診断結果は以下のようになる．

(1)　実験環境 1 の場合

- トラヒック流量検査方式
  SW2 のポート 16 と 19 の対でループしている．
  SW1 のポート 3 の方向に L-SW が存在する．
  SW3 のポート 2 の方向に L-SW が存在する．
  SW4 のポート 1 の方向に L-SW が存在する．
- 疎通確認方式
  端末 12，15 との経路上に L-SW が存在する．
  端末 11，13，14，16，17 との経路上に L-SW は存在しない．

(2)　実験環境 2 の場合

- トラヒック流量検査方式
  SW1 のポート 2 と 3 の対でループしている．
  SW2 のポート 3 と 4 の対でループしている．
  SW3 のポート 2 と 5 の対でループしている．
  SW4 のポート 1 の方向に L-SW が存在する．
- 疎通確認方式
  端末 11，12，13，15，16 との経路上に L-SW が存在する．
  端末 14，17 との経路上に L-SW は存在しない．

パケット受信量取得や疎通確認中にも，探索対象の端末やスイッチからブロードキャストパケットが送信されるなどにより，通信状態は動的に変化するため，期待される結果を得られない場合もある．そこで，実験環境 1，2 において，各方式ともに 300 回の診断実験を行い，診断結果を評価した．その結果を**表 2** に示す．ループを特定または絞り込める診断成功の割合は，実験環境 1，2 のいずれの場合も，トラヒック流量検査方式では 91%以上，疎通確認方式でも 78%以上であった．また，いずれの方式でも診断誤りは認められなかった．このことから，提案方式が有効であるといえる．

なお，ポートやスイッチの特定まで至らず，絞り込みとなる場合は，トラヒック流量検査方式では，大量受信ポート判定で期待される結果の一部しか取得できない，あるいは，アド

表 2　診断結果
Table 2　Results of experiments.

実験環境1: 単一スイッチでループの場合

方式	診断成功				診断不能	診断誤り
	ポート特定	ポート絞り込み	スイッチ特定	スイッチ絞り込み		
トラヒック流量検査方式	250 (83%)	27 (9%)	0 (0%)	21 (7%)	2 (1%)	0 (0%)
疎通性確認方式	-	-	187 (62%)	59 (20%)	54 (18%)	0 (0%)

実験環境2: 複数スイッチでループの場合

方式	診断成功				診断不能	診断誤り
	ポート特定	ポート絞り込み	スイッチ特定	スイッチ絞り込み		
トラヒック流量検査方式	101 (34%)	164 (55%)	0 (0%)	7 (2%)	28 (9%)	0 (0%)
疎通性確認方式	-	-	163 (54%)	73 (24%)	64 (21%)	0 (0%)

ポート特定:　LSWのループポートを全て特定できた場合
ポート絞り込み:　LSWのループポートの一部(片側)まで特定できた場合
スイッチ特定:　LSW(の一部)を特定できた場合
スイッチ絞り込み:　LSWを複数スイッチのいずれかまで絞り込めた場合
診断不能:　診断失敗、または、結果に矛盾があった場合
診断誤り:　誤ったループポートやスイッチが特定された場合

レス誤学習訂正が持続しないために通信疎通が安定せず，あるスイッチの MIB がまったく取得できないことによるものであった．また，疎通確認方式では，対象端末のブロードキャストパケットが再度ループし，その対象端末に対する診断ができない場合であった．たとえば，実験環境 1 において，トラヒック流量検査方式で SW2 からポート 16 の片側ループポートしか特定できない場合，あるいは，疎通確認方式で端末 12 に対してアドレス誤学習訂正を複数回行っても，対象端末発のブロードキャストパケットがない状況を形成できず，診断ができなかった場合（この場合，SW2 と SW5 が被疑スイッチと診断される）などがそれにあたる．

また，診断不能となる場合は，トラヒック流量検査方式では，ループに関係するスイッチすべての流量検査ができない場合であり，疎通確認方式では，疎通可能となるべき端末に対して疎通不可となるなど，診断結果に不整合が生じた場合である．たとえば，実験環境 2 で端末 14 とは疎通できているが，その途中に位置していて疎通されるはずの端末 17 との疎

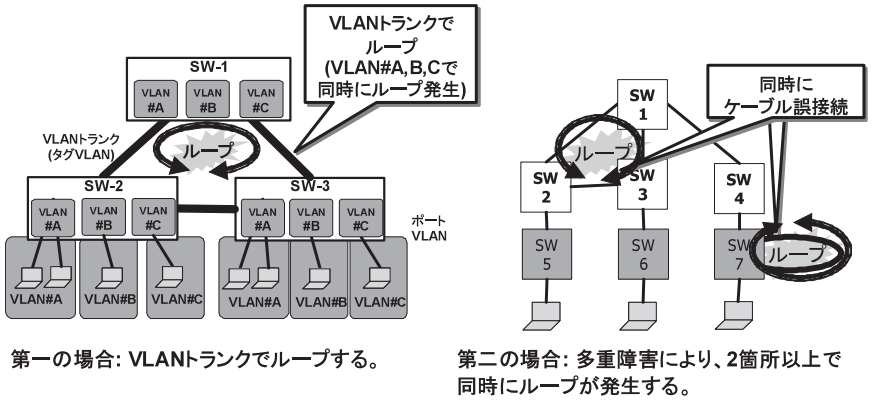


図 10　複数のループが同時に発生する例
Fig. 10　Example of occurring two or more loops at the same time.

通を確認できない場合などがそれにあたる．

いずれの場合も，診断時間中に通常の ARP 要求などの端末動作が発生したことで，訂正されたアドレス学習が再度誤学習状態に戻るという確率的な現象が起きたことに起因している．これは診断処理とは独立な事象であるので，再度，誤学習訂正を行い，試行を繰り返すことにより，診断成功の確率を高めることが可能である．

なお，探索端末が L-SW に直接接続された場合にも，トラヒック流量検査方式では，L-SW のループポート対を抽出できるため，診断可能である．また，疎通確認方式では，すべての探索対象端末に対して疎通不可となることから，探索端末が接続されたスイッチが L-SW，あるいは，複数 L-SW のうちの 1 つであると判断できる．

また，試作システムにおける探索時間は，以下のとおりであった．

- 流量検査方式で，スイッチあたり平均 20 秒
- 疎通確認方式では，端末あたり平均 15 秒

探索処理内容は，スイッチあるいは端末ごとへのパケット受信量取得や疎通確認であるため，台数に比例して増加する．たとえば，一般的なレイヤ 2 ネットワークとして，スイッチ数を 20，平均的なポート数を 10，端末数を 200 程度とした場合の総診断時間は，およそ 7〜8 分間となる．従来のループ障害によるネットワークダウン時間が半日以上に及ぶ場合が多いことを考えると，診断時間の点からも十分に実用的であるものと評価できる．

なお，複数のループが同時に発生する場合について，以下で説明する．**図 10** 左に示すよ

うに，第 1 の場合として，単一障害が複数の VLAN トランクに影響して複数のループが構成される場合がある．このほかに図 10 右に示す第 2 の場合として，多重障害により 2 カ所以上で同時にループを構成する場合もある．現実的に起こりやすいのは第 1 の場合であり，この場合，VLAN ごとに個別に考えると，同一 VLAN 内には，複数のループが同時に発生することはなく，これまでに説明した手順どおり，各 VLAN で独立して診断を行うことが可能である．しかも，どの VLAN 環境で診断を行った場合でも，同じ被疑箇所（ループ構成にある VLAN トランク）を示すことができるため，よりループ箇所を特定しやすくなる．第 2 の場合も，同一サブネット内で複数のループが発生するケースでなければ，第 1 の場合と同様に，各サブネットで独立して診断可能である．一方，図 10 右のように，同一サブネット内で複数のループが発生する場合，単一ループの場合と異なり，ループを周回するパケットが，他のループによって増幅される現象が発生し，第 1 ステップの通信機能の回復手段が有効に機能しなくなり，診断できない場合がでてくる．しかしながら，第 2 の場合は，一般的には 7〜8 分間と考えられる診断時間の中で，複数障害が発生する場合であって，単一のループ障害や上述の第 1 の場合と比べてきわめて発生頻度が小さいと考えられる．したがって，第 2 の場合に対応できないとしても，提案方式は，十分に実用範囲にあるものと考えられる．

4.2　通信機能回復の効果と持続性評価

ループ障害が発生している中での探索であるだけに，3.2 節に示した負荷低減や疎通確保をいったん行っても，端末が発信する ARP 要求などの新たなブロードキャストパケットの送信で通信機能の回復が持続しない可能性がある．そこで，この持続性を評価し，本方式が実用上十分であることを示す．

4.2.1　ネットワーク負荷低減の効果と持続性

本項では，ロングパケットの注入によって，ループしているショートパケットが置換されてネットワーク負荷の低減効果が得られたか否かと，その持続性について評価する．**図 11** は，処理負荷軽減のためにロングパケットを注入し続けた状態でのネットワーク流量を計測したものである．なお，評価したネットワークは，1 台のループ状態にあるスイッチに，探索端末に相当するパーソナルコンピュータ 1 台をパケットフィルタ装置を介して縦列に接続する方法をとったネットワークである．接続機器として，スイッチには Cisco 社製 Catalyst 2950 を，探索端末には Mobile Pentium4（3.2 GHz）搭載機を用いている．同図に見られるように，最初にループしていた 64 バイトのショートパケットが，送出した 1,514 バイトのロングパケットに 10 秒未満ですべて置換されており，3.2 節で述べた負荷低減の仕組みの

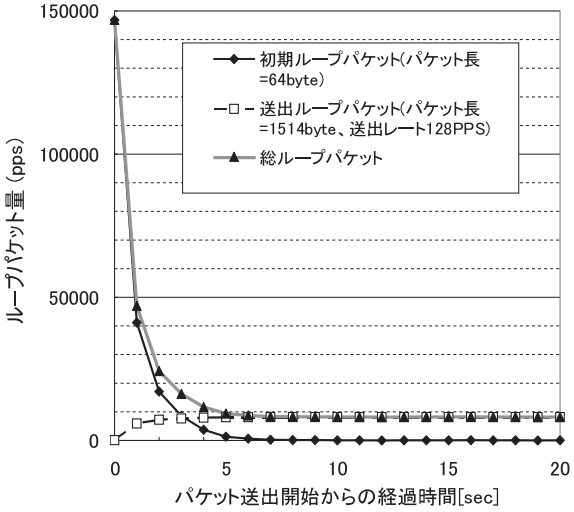


図 11　ロングパケット注入による負荷低減効果
Fig. 11　Node load reduction effect of long packet injection.

効果が得られていることを確認できた．また，同図では，ロングパケットの注入中に，毎秒 1 つの新たな ARP 要求を発生させているが，パケット流量の増加は見られておらず，ループパケット数削減の効果が確認できる．なお，送出するパケットのサイズは 1,514 バイトで，送出レートは毎秒 128 パケットであるため，必要な通信速度は約 1.5 Mbps となり，一般的な処理能力を持つ探索端末から十分送出可能である．

パケット数削減をノードに与える負荷量で評価するため，ループパケット長とノード負荷率の関係を測定したグラフを**図 12** に示す．パケット長が 64 バイトでは，Mobile Pentium4（3.2 GHz）クラスの高性能端末でも，端末負荷は 80%に至るが，パケット長が 1,514 バイトでは，端末負荷は 5%程度と低く安定する．このことから，ロングパケット注入により，端末の負荷は 1/20 に軽減されているといえ，非常に効果的であると評価できる．また，代表的なインテリジェントスイッチである Cisco 社製 Catalyst 2950 の場合は，パケット送受信処理がハードウェアで実現されるアーキテクチャのため，ループの影響を受けにくく，CPU 負荷はつねに 20%〜30%程度で安定している．つまり，スイッチについては，ロングパケット注入による CPU 負荷の軽減効果は認められないが，つねに通信には十分な能力を持つことを示している．以上より，ノード負荷に関しては，端末負荷の低減が重要であり，

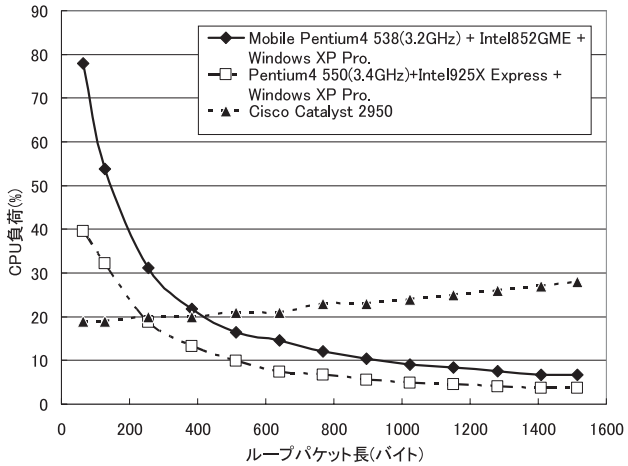


図 12　パケット長とノード負荷の関係
Fig. 12　Relationship between packet length and node load.

ロングパケット注入はこれに対して，大きな効果があるといえる．

一方，ネットワーク帯域の点からは，ブロードキャストストーム発生中は伝送路帯域がほとんどすべてブロードキャストパケットで占有されている．ロングパケットに置換しても，基本的にこの状況は変わらない．この状態で，SNMP や ARP などの診断用パケットが利用可能であるかを評価した．MAC アドレスの誤学習が発生せず，ロングパケットのみがループしている環境において，2 台の端末間で 50 ミリ秒間隔で 64 バイトおよび 1,514 バイトのユニキャスト ARP 要求パケットをそれぞれ 10 万回送信したときの受信 ARP 応答数からロス率を計測した．計測は，インテリジェントスイッチである Cisco 社製 Catalyst 2950 と，安価な非インテリジェントスイッチである Planex 社製 FX-08M とについて行った．この結果，どちらの場合もパケットロスはまったく発生しなかった．ブロードキャストストームが発生している環境でも，第 1 ステップを実行することで正常に通信ができることが確認された．なお，本評価環境でショートパケットがループしているときの ARP 要求のロス率を評価したところ，約 65%であった．

以上のことから，本方式によれば，診断中に，ロングパケットを注入し続けることで，新たなループパケットの増大も抑制し，必要な通信を可能とするものと評価できる．

4.2.2　アドレス誤学習訂正の持続性

次に，通信疎通のために行った誤学習の訂正の持続性を評価した．4.1 節の診断結果から，対象端末あたり約 20 秒間の探索時間を要しているが，この間に対象端末からの新たな ARP 要求が送信されると，訂正したアドレス学習が誤った状態に戻ってしまう．提案方式では，対象端末の ARP 送信を監視しつつ探索診断を行っており，ARP 送信が検出されると，誤学習訂正の動作をやり直す．しかし，疎通確認探索動作中に頻繁に ARP 送信がなされると，誤学習訂正を繰り返すことになり，実質的には診断ができなくなる．そこで，ARP 送信頻度を調査するため，端末あたりの ARP 送信間隔を実際に利用されているネットワークで実測した．なお，評価に利用したネットワークは，企業内の研究開発部門において一般的な日常業務に使用しており，OS として Windows が稼動する端末が接続されている．端末の電源は必要に応じて投入され，主なトラヒックはメールおよび WWW アクセスによるものである．利用者の行動パターンが異なる可能性を考慮して，以下の実測 1 は業務日の午前，実測 2 は業務日の午後に行った．

- 実測 1：稼働端末数 71，1 時間 10 分計測の場合，平均送信間隔 191 秒（標準偏差 125 秒）
- 実測 2：稼働端末数 73，1 時間 55 分計測の場合，平均送信間隔 251 秒（標準偏差 115 秒）

ARP 送信の発生には規則性はないが，ループ状態では ARP 送信は成功しないため，発生した ARP 送信は必ずリトライされる．たとえば，Windows の場合は，3 秒間隔で 3 回再送信される．ただし，この再送信の時間を経た後で，前述の ARP 送信を検知した場合の再診断を行うこととすれば，計 4 回の ARP 送信を 1 つの事象と見なすことができる．つまり，診断時間 20 秒間に発生する事象は，ARP 要求が送信されないか，4 回送信されるかの 2 種類のみとなる．この考えで，診断失敗となる確率を以下のように算出した．まず，2 つの実測結果から，平均送信間隔を 210 秒，標準偏差を 120 秒とした．次に，ARP 送信の発生がポアソン過程に従うと仮定すると，20 秒間に 1 つ以上の ARP パケットが送信される確率は，$\lambda = 20/210$ とした場合のポアソン分布の一般式 $f(x) = e^{-\lambda}\lambda^x/x!$ を用いて $1 - f(0)$ で求められ，約 9.1%となる．つまり，疎通確認に失敗する確率は 9.1%となる．診断時に診断失敗を検知した場合に，誤学習訂正ならびに疎通確認方式を n 回まで再試行するとすれば，最初の診断失敗と合わせて合計 $(n+1)$ 回の診断がすべて失敗する確率は 9.1%の $(n+1)$ 乗となる．3 回の再試行，すなわち $n = 3$ 程度の再試行でも失敗確率を 0.0069%と大きく低減できることから，提案方式には十分な実用性があると評価できる．

4.3　提案方式の有効性と今後の課題

試作システムを用いた実環境による評価の結果，提案方式の有効性が確認できた．提案方

式の最大の特長は，ループ発生によって失われているネットワークの通信機能を，ロングパケット注入による負荷低減およびアドレス再学習によって回復させることである．診断には，ブロードキャスト ARP 要求のような一般的なパケットを用いるため，スイッチに特別な機能を必要とせず汎用性も高い．また，診断の全過程を探索端末からリモートで行うことができ，実用性が高い．

提案方式でさらに検討すべき点として，スイッチで QoS 制御が設定されている場合，パケットの種類によって処理優先度が異なるため，注入するロングパケットで，スイッチや端末の負荷を軽減できない場合がありうることである．また，スイッチ内のバッファ障害が発生している場合には，ロングパケット注入が有効に働かない可能性も検討すべきである．これらの解決策は今後の課題であるが，探索端末の接続箇所を変えるなど，運用面である程度対応できるものと考えられる．

今後の技術課題としては，ネットワークトポロジの事前取得との連携が重要である．本提案では，対象端末の IP アドレスは必要であるが，ネットワークトポロジなどの構成情報の取得は前提とはしていない．しかしながら，もし，トポロジ情報が既知であれば，流量検査方式では流量取得できないスイッチがあっても L-SW を絞り込めること，疎通確認方式では複数の疎通結果を照合した診断ができることが期待される．近年，トポロジの自動取得を行う技術の研究開発が進んでおり，これらの機能との連携による適用範囲の拡大や診断精度向上について検討する予定である．

5. おわりに

IP ネットワークの社会普及にともない，大規模化し，重要性を増す広域イーサなどのレイヤ 2 ネットワークにおける最も重大な障害であるループ障害を解析し，その原因箇所をリモート診断する技術を提案した．ループ障害の発生を防止する技術が従来の中心であるが，防止技術自身の障害や全スイッチに適用しないと効果が確実でないなど，完全な回避には問題があった．そこで，本提案では，回避よりも，発生原因を迅速に特定診断する技術を考案した．従来は，ループ障害が発生すると，ネットワーク機能はすべて失われると考えられてきたが，通信機能を回復する手段が存在することを示した．これを利用して，機能が回復したネットワークを用いたリモートからのループ障害箇所を特定する方式を提案した．次いで，本方式を試作し，診断結果ならびに診断時間を評価し，その有効性を確認した．

今後の残された検討課題として，実ネットワークでの実際の多数の障害事例への適用を通した評価がある．また，ネットワークトポロジの自動探索技術との併用による診断精度の向上や適用範囲の拡大も今後の技術として重要である．

参 考 文 献

1) 岩田　淳ほか：広域イーサネット標準化動向と NEC における広域イーサネットへの取り組み（その 1），ITU ジャーナル，Vol.33, No.10, pp.4–11 (2003).
2) 802.1D-2004 IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges.
3) 802.1W-2001 IEEE Standard for Local and metropolitan area networks - Common Specifications - Part 3: Media Access Control (MAC) Bridges: Rapid Configuration.
4) 802.1S-2002 IEEE Standard for Local and Metropolitan Area Networks - Amendment 3 to 802.1Q Virtual Bridged Local Area Networks: Multiple Spanning Trees.
5) 安留多伎明良：広域イーサネット技術概論，電子情報通信学会 (2005).
6) 岩田　淳：広域イーサネット技術概論，電子情報通信学会東京支部講演会資料 (2008).
7) 安藤雅人：広域イーサネットの運用管理技術，信学技報，IA2007-12 (July 2007).
8) フレーム中継装置，特開 2001-197114 (2001).
9) 堀川健史ほか：GAVES におけるユーザ網ループ検出，NTT 技術ジャーナル，Vol.18, No.4 (2006).
10) 長嶋　潤ほか：次世代広域イーサネット網向けスイッチングハブの開発，日立電線，No.24, pp.13–16 (2005).
11) シスコシステムズ株式会社 LAN スイッチワーキンググループ：Cisco Catalyst LAN スイッチ教科書，p.142, インプレス (2004).
12) IETF IP Virtual Link eXtension BOF (2004). http://www.ietf.org/ietf/04aug/ipvlx.txt

(平成 20 年 9 月 2 日受付)
(平成 21 年 5 月 13 日採録)

勝山　恒男（正会員）

1974 年慶應義塾大学工学部計測工学科卒業．1976 年同大学大学院修士課程修了．同年富士通（株）入社．富士通研究所において，ネットワークアーキテクチャならびにネットワークサービス，サービスプラットフォーム技術および ICT システムの自律型運用管理技術に関する研究開発に従事．博士（情報科学）．電子情報通信学会会員．

安家　　武

1996 年大阪大学工学部通信工学科卒業．1998 年同大学大学院博士前期課程修了．同年（株）富士通研究所入社．ネットワーク監視技術に関する研究に従事．電子情報通信学会会員．

野村　祐士

1992 年北海道大学工学部情報工学科卒業．1994 年同大学大学院工学研究科情報工学専攻修士課程修了．同年富士通研究所入社．以来，ネットワークアーキテクチャ，ネットワークにおける障害検出に関する研究開発に従事．2001〜2002 年コロンビア大学客員研究員．電子情報通信学会会員．

若本　雅晶（正会員）

1982 年横浜国立大学大学院工学研究科修士課程修了．同年富士通（株）入社．富士通研究所において，交換ソフトウェア，インテリジェントネットワーク，IP サービス制御アーキテクチャ，IP ネットワーク計測・運用技術の研究開発に従事．電子情報通信学会会員．

野島　　聡

1978 年早稲田大学大学院理工学研究科電子通信専攻修士課程修了．同年株式会社富士通研究所入社，現在に至る．主にパケットネットワーク，スイッチング技術の研究に従事．電子情報通信学会会員．

木下　和彦

1973 年生．1996 年大阪大学工学部情報システム工学科卒業．1997 年同大学大学院工学研究科情報システム工学専攻博士前期課程修了．1998 年 3 月同博士後期課程退学後，同年 4 月より大阪大学大学院工学研究科情報システム工学専攻助手．2002 年大阪大学大学院情報科学研究科情報ネットワーク学専攻助手，2007 年同助教を経て，2008 年同准教授．現在，ネットワークアーキテクチャ，モバイルネットワーク，エージェント通信システムに関する研究に従事．博士（工学）．電子情報通信学会，IEEE 各会員．

村上　孝三（正会員）

1971 年大阪大学工学部電子工学科卒業．1973 年同大学大学院修士課程修了．同年富士通（株）入社．富士通研究所通信研究部門，同所マルチメディアシステム研究所を経て，1995 年大阪大学大型計算機センター教授．1998 年同大学大学院工学研究科情報システム専攻教授．2002 年同大学大学院情報科学研究科情報ネットワーク学専攻教授．マルチメディア情報通信システム，フォトニックネットワーク，インテリジェントネットワーキングの研究に従事．IEEE フェロー，電子情報通信学会フェロー．工学博士．