

オーバレイネットワークにおける帯域消費が軽微な L2 ループ検出手法の提案

野呂 正明^{1,a)} 高野 陽介¹ 小口 直樹¹ 阿部 俊二²

概要：クラウドにおける IaaS や広域 LAN 接続など、オーバレイした L2 ネットワークを顧客に提供するサービスが増加している。こうしたサービスでは、顧客が構成した L2 ネットワークがループを起こした場合でも、サービス提供事業者は顧客所有の回線の遮断を行わず、顧客にループの発生を通知する機会が多い。従来のループ検出手法では、回線遮断を前提に、ブロードキャストストームの発生をループの検出手法としている。しかし、ループが発生した場合に顧客所有回線を遮断せず、顧客に問題解決を促す環境では、ブロードキャストストームを前提にした従来のループ検出手法は顧客の通信に大きな影響が出る。そこで本研究では、ループ発生時に顧客のネットワークを圧迫させることなく、ループを検出することが可能な手法を用い、顧客にループ解消を促す方法を提案する。本論文ではこれを評価し、ループ検出のために必要な帯域消費量、所要時間共に実用上問題ないことを確認した。

キーワード： extended Berkeley Packet Filter, eXpress Data Path, MAC 層ループ

Proposal of MAC layer loop detection method for overlay layer 2 network

Abstract: Today Layer 2 overlay networks such as wide area ethernet services and virtual networks sited in the IaaS of cloud services are increasing more and more. In such networks, service providers of overlay networks do not block customer networks even when they detect layer 2 network loops and only report to their customers the existence of loops.

However, as most of the conventional methods detect layer 2 network loop when broadcast storm is occurred, the quality of customer network is influenced by broadcast storm in this situation. Therefore, the authors propose layer 2 loop detection method that uses two kinds of unicast packet (probe packet and clear packet) and consume far less bandwidth.

In this paper, the authors also evaluate and verify the efficiency of the proposed method. We found that the necessary bandwidth of the proposed method was very small and the duration where test packets retain in the network was also very short.

Keywords: extended Berkeley Packet Filter, eXpress Data Path, MAC layer loop

1. 背景

クラウド技術の普及に伴い、複数組織で仮想化された MAC 層のネットワークを相互接続するケースが増加して

おり、ユーザが作成した Virtual Private Network (VPN) によりループが発生する可能性が増大しているが、エンドユーザに Spanning Tree Protocol (STP) [1] を正しく運用することを期待するのは難しい。また、LAN 機器に搭載されたループ検出機能は Virtual eXtensible Local Area Network (VXLAN) [2] や QinQ (IEEE 802.1ad) [3] といった VPN 技術の増加に追従できていないだけでなく、クラウド環境内の仮想ネットワークにはループ検出の機能を持つものがないため、ループ検出の仕組みを搭載した仮想計算機を接続して運用することが必要な状況にある。

¹ 富士通研究所 (Fujitsu Laboratories)
211-8588 神奈川県川崎市上小田中 4-1-1
4-1-1 Kamikodanaka, Kawasaki city, Kanagawa 211-8588, Japan

² 国立情報学研究所 (National Institute of Informatics)
101-8430 東京都千代田区一ツ橋 2-1-2
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

a) noro@fujitsu.com

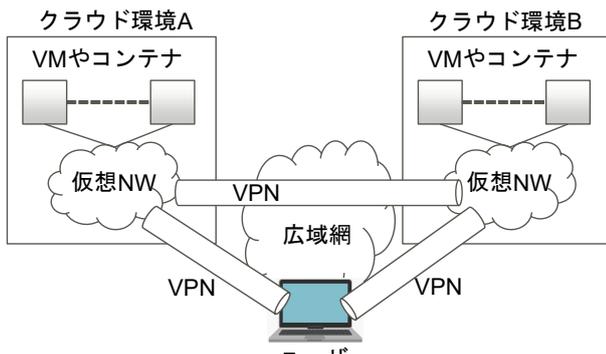


図1 マルチクラウド環境でのループ

Fig. 1 MAC layer loop between multi cloud environment.

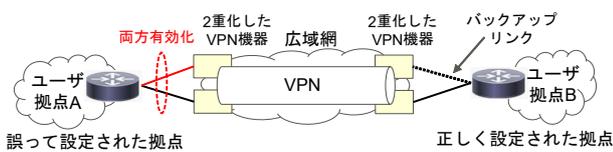


図2 マルチクラウド環境でのループ

Fig. 2 Virtual network loop over wide area network.

本研究の想定ユーザである、通信事業者や大きな組織における広域網の管理者は、エンドユーザがループを作成してしまった場合でもユーザ回線を勝手に切断することができないことが多く、ループが作られた場合に早期に検出して、ユーザに調査・対応をしてもらう時間的な余裕を作ることが重要である。

著者らは、ネットワークにループ検出のための計算機を外付けすることで、ループを検出・切断する手法を提案 [4][5] してきたが、本論文では、ループを検出しても切断できない運用ポリシーの元でも有効な手法を提案し、その基本的な性能を評価した。

2. 想定するネットワーク

本論文で想定するユーザは、クラウドサービスを提供する事業者の管理者、ユーザの複数拠点を Virtual LAN (VLAN) で接続する広域 LAN サービスを提供する通信事業者や大きな組織のネットワーク管理者である。

想定ユーザの顧客となるエンドユーザは、ネットワークの規格を始めとした技術的な知識が十分にない場合もある。そのため、クラウド環境では、複数のクラウドサービス上に作成したシステムと自らのネットワークを VPN で接続してループを発生させる事故 (図1) を発生させることが考えられる。また、広域 LAN サービスを提供する事業者は、ハード障害に備えて顧客の LAN 機器からの接続を複数のスイッチで多重化する場合もあるが、エンドユーザが図2のような接続をしてしまい、WAN 回線でのブロードキャストストームが発生し、回線の帯域を大量に消費する事故を起こす場合もある。

顧客にサービスを提供する事業者や、大きな組織で複数

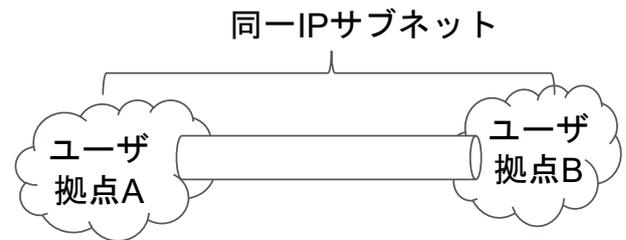


図3 広域網を経由した LAN の統合

Fig. 3 Integrated LAN over wide area network.

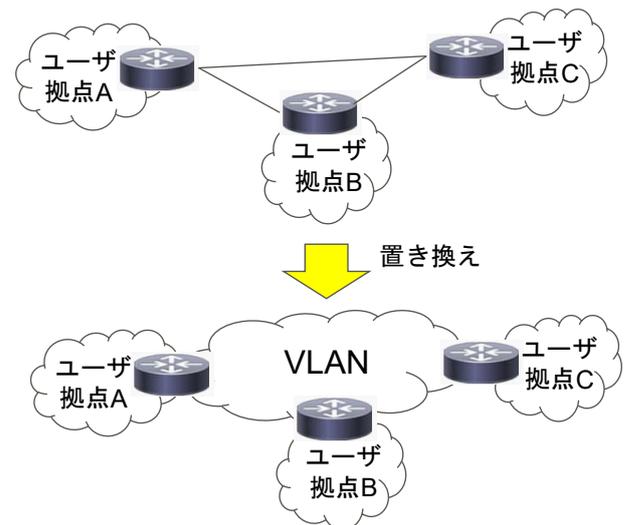


図4 LAN 機器間の接続の置き換え

Fig. 4 Replace link between network device by virtual network service.

部門の拠点間を接続するネットワークを管理する管理者は、エンドユーザが作成したループを発見しても、エンドユーザの回線を管理者の判断だけで切断することが許されていない場合がある。このような状況では、エンドユーザが作成したループを早期に発見し、関係者に連絡して、事故が大きくなる前に対処することが重要である。

ネットワークにループを作成した場合に大きな事故 (周回するブロードキャストやマルチキャストパケットで回線容量が埋まる) に至るまでの時間であるが、ループを作成した回線に流れるパケットのうち、ブロードキャストやマルチキャストパケットが占める割合が少なければ、ループが作成されてから事故が大きくなるまでの時間的な余裕がある。

たとえば、組織の分散した拠点を広域 LAN サービスで接続する場合は、多数の PC が接続しマルチキャストやブロードキャストパケットが交換される LAN を広域網で統合する例 (図3) よりも、拠点に設置された LAN 機器間の専用回線を広域 LAN サービスで置き換える例 (図4) の方が多く、後者の場合は LAN 機器間の接続であるため、ブロードキャストやマルチキャストのパケットが交換される比率は LAN の場合に比べて非常に低い。

同様に、クラウドに着目すると、クラウド上に展開して

いる VM はエンドユーザが所属する組織内のネットワークでのみ利用するものより、インターネットに公開しているものが多く、このようなサービスを動作させる OS は uPnP や Bonjour, CIFS のようなブロードキャストやマルチキャストを多様化するサービスはセキュリティ上の理由から停止させていることが多い。このようなネットワークでは、LAN より接続している装置の数が少ないだけでなく、ブロードキャストやマルチキャストを使うアプリを停止していることから、ネットワークで交換されるパケットのうち、ブロードキャストやマルチキャストパケットの占める割合が小さいと考えられる。

以上のような理由から、本研究のユーザ環境ではネットワークのループが発生しても、大事故に至るまでの時間は LAN 環境より長くなることが想定され、ループが発生した場合に、早期にそれを発見して管理者に通知することで大事故に至る前に修復することが期待できる。

3. 従来のループ検出手法

多くの LAN 機器に搭載されている従来手法は、あるポートで受信する単位時間あたりのブロードキャストやマルチキャストパケット数が、ユーザが設定した閾値を越えた場合にループありと判定する方法である。この方法は、ループが存在するネットワークにおいて、ブロードキャストやマルチキャストパケットが大量に周回し、回線が圧迫されるような状況にならないとループありと判定しないため、早めにループを発見して対処の時間を稼ぐという本研究の目的にはそぐわない。

これに対して、ブロードキャストもしくはマルチキャストパケットをスイッチから送信し (図 5)、それが自分に戻ってきたらループありと判定する方法 [6] と短時間に同じパケットが複数回受信された場合に、ループありと判定する方法 [7] を組み合わせると、ネットワークに接続した計算機から、ブロードキャストもしくはマルチキャストの検査パケット送信し、自分が送信した検査パケットを短時間に複数回受信した場合にループを検出することができる。

しかし、この 2 つの手法の組み合わせでは、検査パケットの宛先がブロードキャストもしくはマルチキャストアドレスであるため、ループを検出しても回線を切断できない環境では、検査パケットがループを周回し続け、ブロードキャストストームとなる (図 6)。そのため、ループの存在を確認する検査パケットを周期的に送信すると、周回するパケットが蓄積していき、最終的には回線容量が周回する検査パケットによって占拠される。

ループ発生の可能性があるネットワークに計算機 (もしくは仮想計算機) を接続し、その上でループを検出するプログラムを動作させて、早期にループを検出することを目的とする場合、検査パケット自身がループを周回し続けることを防止する必要がある。本研究の提案方式では、ルー

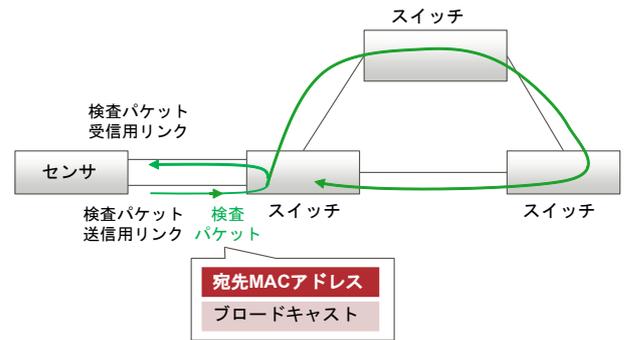


図 5 検査パケットの送信

Fig. 5 Sending loop detection test packet.

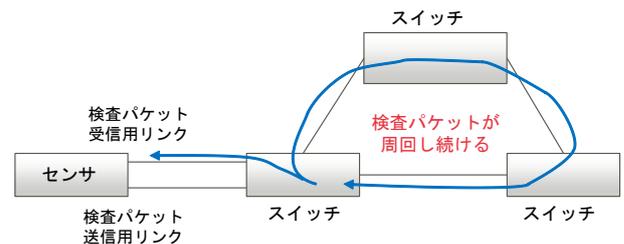


図 6 検査パケットの周回

Fig. 6 Loop of test packet.

プを検出するための検査パケットがループが存在するネットワークを周回し続けることを防止する機能を実現した。

4. 提案手法

本研究の提案手法では、3 章で説明した 2 つの従来手法 (文献 [6] と [7]) を組み合わせた場合の問題である、検査パケットの永続的な周回を解決するため、検査パケットにユニキャストパケットを用い、周回する検査パケットを打ち消すパケットを別途送信することで、検査パケットによるストームを解消する。センサで 2 つのネットワークインターフェイスを検査対象のネットワークに接続し、片方のインターフェイスは検査パケット送信専用とし、もう一方のインターフェイスは検査パケット受信と周回する検査パケットを打ち消すためのパケットを送信するために用いる。また、検査パケットの宛先にはユニキャスト (ネットワーク上に該当アドレスを持つ機器が存在しないアドレス) を使う。そのため、検査パケットの宛先アドレスはループを構成する経路上のスイッチ全てに学習されていないため、ブロードキャストされる。ネットワークがループしていた場合は、センサが送信したパケットは検査パケット受信用インターフェイスで非常に短時間の間に複数回受信される (図 7)。

センサは、これをもってループが存在すると判定し、ループを周回している検査パケットを消すため、発信元が検査パケットの宛先アドレスになっているパケットを送信する (図 8)。

MAC アドレス b を学習したスイッチは、周回する検査

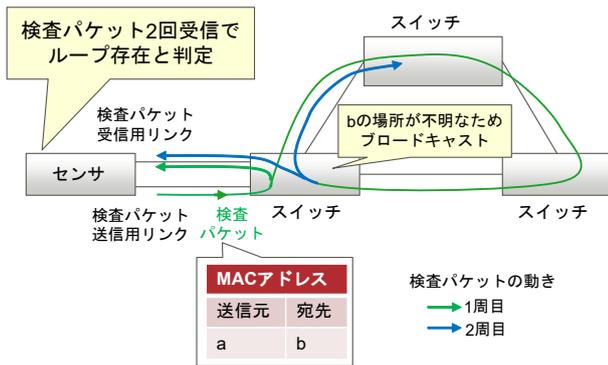


図 7 検査パケットの周回によるループの検出
 Fig. 7 Loop detection by rounding test packet.

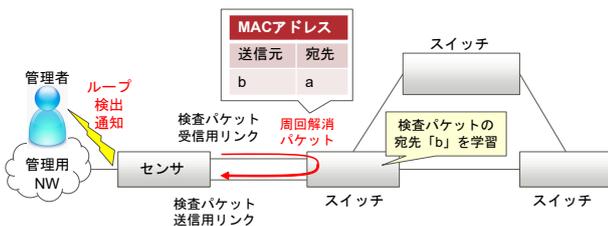


図 8 周回解消パケットの送信
 Fig. 8 Sending clear packet.

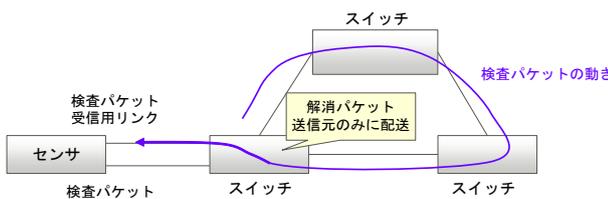


図 9 周回解消パケット送信後の検査パケットの流れ
 Fig. 9 Flow of test packet.

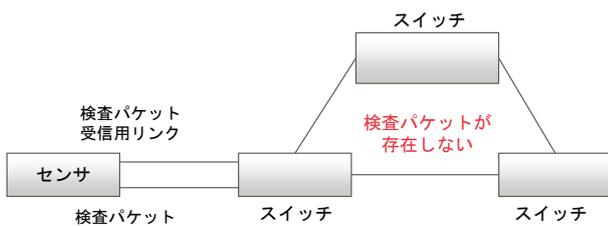


図 10 一定時間経過後のネットワーク
 Fig. 10 Loop network after sending clear packet.

パケットをセンサに送信する (図 9) ため、ループを周回する検査パケットはネットワーク上から無くなる (図 10)。

5. 評価

ループが存在するネットワークで提案手法を実施した場合、検査パケットがループを一定時間周回するため、検査パケットが周回する時間とその期間に消費する帯域を評価し、実用上問題ないことを確認する。図 11 は評価用に試作したセンサの実装であるが、ループ検出センサは python と extended Berkeley Packet Filter (eBPF) [8] で作成し、

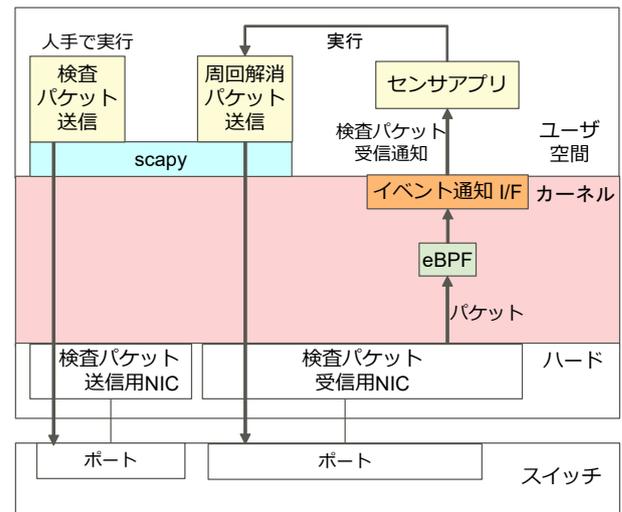


図 11 センサ用計算機のシステム構成
 Fig. 11 Architecture of loop detection sensor.

表 1 各 VM の仕様
 Table 1 Specification of virtual machines.

VM	仮想 CPU 数	メモリ容量	OS
センサ	4	8G	Ubuntu 20.04.1
ブリッジ用 Linux	2	8G	Ubuntu 20.04.1

検査パケット送信モジュールと周回解消パケットモジュールは Python のパケット操作ライブラリである scapy[9] を用いて開発した。検査パケット、周回解消パケット共に、arp を模擬したフォーマットの packets (MAC フレームのサイズ 60 バイト) を用いている。検査対象のネットワークに接続した 2 つのインターフェイスは、IPv4, v6 共にリンクローカルを含めてまったく設定していない状態である。

図 12 が評価環境であり、ループ検出センサを搭載した仮想計算機とスイッチの代わりとなるブリッジを構成する Linux VM のスペックは表 1 の通りである。また、図 12 の環境は、表 2 の環境で動作する Graphical Network Simulator-3 (GNS3) [10] の上に構築した。

本評価では、検査パケットがループのあるネットワークを周回開始から消えるまでの時間を測定するため、ループ上の 1 つの経路を Linux で構成したブリッジとし、このブリッジのインターフェイスの 1 つで eXpress Data Path (XDP) [11][12][13] を用いた観測用プログラムを動作させ、周回するパケットを監視する仕組みとなっている。なお、図 12 の環境では検査パケットが時計回りと反時計回りの両方向で周回するが、XDP は受信するパケットしか監視できないため、観測用プログラムは時計回りの周回のみを取り扱う。また、パケットがループを周回する時間を変化させるため、ブリッジを構成する Linux は tc コマンドを用いて、配送に遅延を追加している。

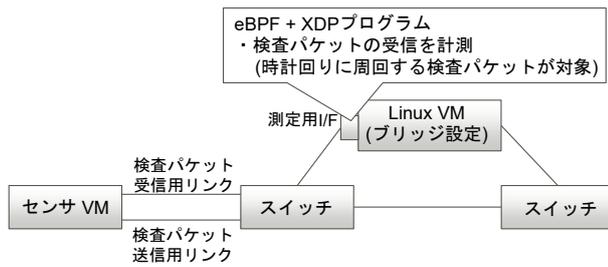


図 12 評価環境

Fig. 12 Performance evaluation environment.

表 2 評価用マシンスペック

Table 2 Machine spec of evaluation environment.

CPU 数	1
コア数	6
仮想 CPU 数	12
メモリ容量	32G
OS	Windows 10 pro (20H2)

5.1 周回する検査パケットが消費する帯域

本提案手法では、一定時間検査パケットがループが存在するネットワークを周回するため、検査パケットが消費する帯域を評価する。

フレームのサイズが $P(\text{bit})$ のパケットが $\tau(\text{ミリ秒})$ でループを周回する場合、1秒で周回する回数は $1000/\tau$ となる。すると、このパケットの周回で消費される帯域 B は、以下の式 (1) で計算できる。

$$B = P \times \frac{1000}{\tau} (\text{bps}) \quad (1)$$

図 12 の環境において、ブリッジとなっている Linux マシンの遅延を 0 に設定し、検査パケットを止めずに周回させた場合、1秒あたり 2千回以上周回し、その平均値 τ は 0.38ms 以上であった。

検査パケットは 60 バイト (480bit) であるため、周回する検査パケットが消費する帯域 B と周回時間 τ の関係を式 (1) をグラフにすると、図 13 となる。 $\tau=0.3\text{ms}$ の場合で 1.6Mbps、 $\tau=500\text{ms}$ では 1kbps 未満となり、検査パケットが周回した場合に消費される帯域は非常に少なく、同じネットワークに同居する利用者に問題となる帯域ではない。

今回の評価では、ループを周回するパケットの周回時間は最大で 500ms を想定しており、これは有線回線でパケットが地球一周するより長い時間であるため、通常のネットワーク構成ではループの周回時間としては十分な時間である。

5.2 周回する検査パケットが無くなるまでの所要時間

図 14 は、図 12 における観測用のネットワークインターフェイスを通過した検査パケットの通過時刻から計算した周回時間の平均値 (ミリ秒単位) を横軸に、最初の検査パケットが同インターフェイスで受信された時刻と最後に検

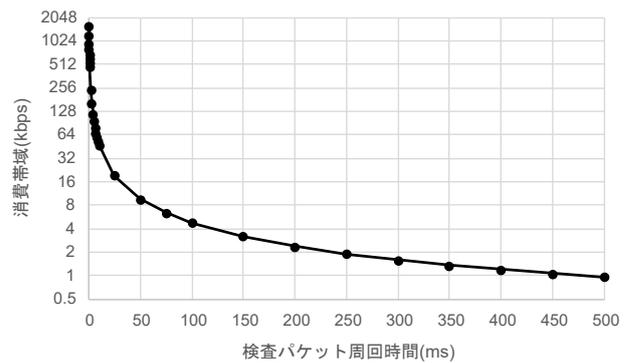


図 13 検査パケットの消費帯域

Fig. 13 Bandwidth consumed by rounding test packet.

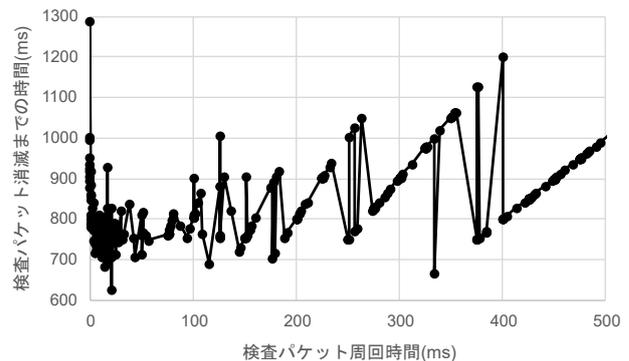


図 14 検査パケットが消えるまでの時間

Fig. 14 Duration until to clear test packet loop.

査パケットを受信した時刻の差分 (検査パケット消滅までの経過時間) をミリ秒単位で計算したものを縦軸にしたものとなっている。図 14 のグラフからわかるように、本評価で用いた実装では、1.3 秒未満の時間で周回する検査パケットをネットワークから消すことができる。

ネットワークに輻輳が発生した場合や、パケット廃棄率が大きくなった場合でも、エンドユーザが大きな影響を受けるのは、コネクションが切断される場合、ルータ間のルーティング情報の交換に失敗するような場合である。TCP のコネクションタイムアウトやルーティング情報交換の間隔は 30 秒であることが多いため、1.3 秒であればそれほど大きな問題にはならない。

6. まとめ

本論文では、クラウド環境における仮想ネットワークや広域ネットワークに計算機もしくは仮想計算機を取り付け、ネットワークに発生したループを検出するのに適した仕組みを提案し、その性能を評価した。本提案方式では、ループが存在するネットワークでは一定時間検査パケットがネットワークを周回してしまうが、想定するネットワーク環境では 1.3 秒未満で検査パケットは消え、周回する検査パケットが消費する帯域も最大でも 1.6Mbps であり、実用上問題がない範囲であることを確認することができた。

なお、本研究では、早期にループを検出した後、ループ

がどこに存在するかの調査とその切断をエンドユーザが行うことを想定している。このような環境では、ループが作成された場合にどの程度の時間で、大きな問題となるかを知る必要があるため、この評価が今後の課題である。

参考文献

- [1] IEEE: IEEE Standard for Local and metropolitan area networks: Media Access Control (MAC) Bridges, *IEEE Std 802.1D-2004 (Revision of IEEE Std 802.1D-1998)*, pp. 1–281 (online), DOI: 10.1109/IEEESTD.2004.94569 (2004).
- [2] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M. and Wright, C.: Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks, RFC 7348 (2014).
- [3] IEEE: IEEE Standard for Local and Metropolitan Area Networks—Virtual Bridged Local Area Networks—Amendment 4: Provider Bridges, *IEEE Std 802.1ad-2005 (Amendment to IEEE Std 802.1Q-2005)*, pp. 1–74 (online), DOI: 10.1109/IEEESTD.2006.6044678 (2006).
- [4] 野呂正明, 高野陽介, 小口直樹, 阿部俊二: eBPF による MAC 層ループ対策, 情報処理学会研究報告. マルチメディア通信と分散処理研究会報告, Vol. 2020, No. 62, pp. 1–7 (2020).
- [5] 野呂正明, 高野陽介, 小口直樹, 阿部俊二: 広域ネットワークで複数拠点を接続する環境での MAC 層ループ対策の評価, 第 28 回マルチメディア通信と分散処理ワークショップ論文集, pp. 209–214 (2020).
- [6] Tzeng, S.: LOOP DETECTION FOR A NETWORK DEVICE, US Patent 0285.499A1 (2006).
- [7] 武藤亮一, 杉谷樹一: ループフレーム検知装置およびループフレーム検知方法, 特開 2006-33275 (2006).
- [8] Gregg, B.: Performance Superpowers with Enhanced BPF, *USENIX*, USENIX Association (2017).
- [9] Philippe Biondi et al.: Scapy Packet crafting for Python2 and Python3. <https://scapy.net/> (2021 年 01 月 20 日参照).
- [10] Galaxy Technologies: Graphical Network Simulator-3. <https://www.gns3.com/> (2021 年 01 月 20 日参照).
- [11] Høiland-Jørgensen, T., Brouer, J. D., Borkmann, D., Fastabend, J., Herbert, T., Ahern, D. and Miller, D.: The EXpress Data Path: Fast Programmable Packet Processing in the Operating System Kernel, *Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies*, Vol. CoNEXT, No. 18, New York, NY, USA, ACM, pp. 54–66 (online), DOI: 10.1145/3281411.3281443 (2018).
- [12] Choudhury, D. G.: XDP-Programmable Data Path in the Linux Kernel., ; *login.*, Vol. 43, No. 1 (2018).
- [13] IO-Visor: XDP eXpress Data Path. <https://www.iovisor.org/technology/xdp> (参照 2021 年 01 月 20 日).