# 迅速な高パケットロスリンク検知のための アクティブ計測手法のP4技術を用いた設計

永倉 紘大[1,a)]　柴田 将拡[1,b)]　鶴 正人[1,c)]

**概要**：筆者らは，高パケットロス率の障害リンクを実時間で効率的に監視・特定するための計測システムを P4 技術を用いて設計した．先行研究では OpenFlow 技術を用いて計測ホストが一定数の計測パケットを送信し，コントローラがいくつかのスイッチに通過した計測パケットの統計情報を問い合わせて，障害リンクを特定する手法が開発された．しかし，特に大規模なネットワークでは検知時間と制御プレーン負荷の増大が課題であった．本報告では，P4 スイッチを使用してパケットロスを直接検出し，計測パケット自体で計測ホストに通知する計測システムを設計し，3 つのネットワークトポロジに対してエミュレーションを実施して設計を評価し，検知時間の短縮に有効であることを示す．

# A P4-based system design of active measurements to promptly locate high packet-loss links

***Abstract:*** The authors have designed a real-time measurement system for a P4 network that can efficiently detect and identify failed links with high packet loss rates. Previous research developed OpenFlow-based methods for locating faulty links in which a measurement host launches a certain number of probe packets and then a controller queries some switches the stats of probe packets passed there. However, those methods can often lead to longer detection times and increased control plane load, particularly in larger networks. This report suggests that the new measurement system, which uses P4 switches to detect packet loss directly and inform the measurement host by probe packets themselves, is effective in reducing latency. The system was evaluated by changing network topologies and positions of lossy links in Mininet emulation environment.

## 1. はじめに

デジタル化が進む昨今において，ネットワーク通信を前提に構成される製品やサービスが多くなっており，ネットワークはますます大規模となって複雑化している．一般的なネットワークではロードバランサやルータ等の機器を多段化して設置しているが，その各種機器の更新や構成変更を行う場合においては，影響のある全機器に対して設定を施し直す必要がある．特に柔軟なネットワーク構成変更や機器更新に実時間で対応するためには，既存のプロトコルの制約を超えた柔軟性・可用性が求められ，それを実現する有効な手段の 1 つとして，データセンターや通信キャリアにおいて SDN（Software Defined Networking）を導入する事例が増加している．SDN の代表的な実用化技術として，OpenFlow[7] と P4[8] がある．両技術は，スイッチやルータにおけるデータ転送機能（データプレーン）とルーティングのルールを決定する機能（制御プレーン）が分離されたアーキテクチャを持ち，中央集権的にネットワークを管理することができる．OpenFlow 技術ではデータプレーンを流れるパケットを読み取って，制御プレーンに関する制御に繋げることはできるものの，データプレーンを流れるパケットの改変はできなかった．一方で P4 技術は，データプレーンを流れるパケットの処理を制御でき，パケットの特定のフィールドに値を格納したり，パケット自体を複製して転送したりできるようになった．

ネットワークは各サービス運用の基幹となる位置付けであり，この品質が悪い場合にはサービスが正常に運用できず多くの損害を被る可能性が高いため，可及的速やかに原因を特定し，問題の解消に向けてネットワークの構成変更や各種制御を行う必要がある．ここで，高パケットロス率

[1] 九州工業大学 大学院 情報工学府 情報創成工学専攻
680-4 Kawazu, Iizuka, Fukuoka, Japan
[a)] nagakura@infonet.csn.kyutech.ac.jp
[b)] shibata@csn.kyutech.ac.jp
[c)] tsuru@csn.kyutech.ac.jp

のリンクを実時間で監視・特定するために，全リンクにおいて細粒度・実時間・高効率の品質推定を行う手段を考える．機器の状態情報の監視として広く浸透している SNMP は，実時間での監視には適さず，監視手法はパッシブ計測であるため，監視先ネットワークにパケットが流れていない場合には障害リンクの特定が行えない．そこで，全リンクに計測パケットを送信して，その状態を推定するアクティブ計測が重要となる．

先行研究では，OpenFlow ネットワークにおいて，計測用パケットが全二重リンクの上りと下りをそれぞれ 1 回ずつ通過するようなマルチキャスト経路木を生成し，多数の計測用パケットを経路木に沿って流した後で，各スイッチにおけるそれらの通過に関する統計情報を OpenFlow コントローラが必要な OpenFlow スイッチから適切な順番で集めることで，障害リンクを計測ホスト 1 台で効率的に特定する手法が提案された [1], [2]．パケットの遅延変動に着目する手法 [3] や，過去の計測結果を利用してロスが起きやすいリンクを計測経路の終端に置く手法 [4] が提案された．これらの手法は計測経路の作成においてダイクストラ法の最短経路アルゴリズムを用いており，パケットロス率の高い障害リンクの位置特定までの統計情報取得回数に影響を及ぼす結果となったため，より統計情報取得回数を少なくするべく，計測経路の作成にオイラー閉路分解を取り入れた計測経路作成手法が提案された [5]．

最も改善のあった手法においても，10000 個の計測パケットを送信した後に，スイッチ数 40 個・リンク数 122 個のトポロジにおいて，コントローラが最大 19.4 回・平均 11.6 回スイッチに通過数情報を問い合わせ，それを基に障害リンクを特定するため，スイッチやリンクの数が多い大規模ネットワークを考えると，検知までの時間と制御プレーン上のアクセス負荷が増大する課題があった．そこで，P4 技術を用いて，P4 スイッチ自身がリンクのパケットロスを検出して，計測パケットにその情報を格納することで，P4 コントローラを介さずにパケットロス発生リンク（障害リンクの候補）を計測ホストに通知する計測システムを検討する．実在するネットワークトポロジに対してエミュレーションを行い，計測システムによって推定されるパケットロス率について調べ，本検討の有効性を評価する．

以下，2 節で計測システムについての検討を行う．そして，3 節では検討した計測システムを実現する手法について示し，4 節では計測システムの有効性を確かめるために，実ネットワークトポロジに対して，エミュレーションを実施して評価する．最後に，5 節でまとめを行う．

## 2. 計測システムの検討

従来のネットワーク上でのアクティブ計測は，トポロジの末端に計測ホストを設置し，ホスト間で計測パケットの送受信を行っていた．

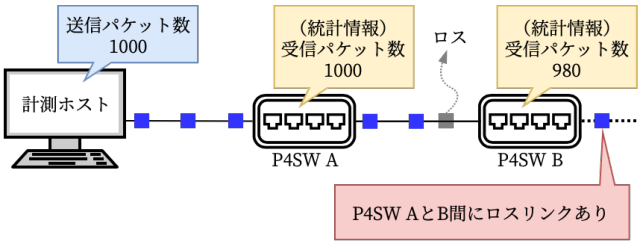


**図 1** 計測パケットの通過数の差を利用したロスリンクの検出

P4 技術では，スイッチを通過する計測パケットを活用したロスリンクの検出方法が提案できる．計測システムは，全二重リンクを想定し，パケットを転送する P4 スイッチ（P4SW）と，計測パケットを送受信する 1 台の計測ホスト（MH）から構成される．P4SW は接続されるポート数毎に計測パケット通過数を記録する内部カウンタを持っており，これを利用して MH が送信した計測パケットがリンクを通過し，計測パケットにそのリンク両端の P4SW の受信パケット数を格納することで，P4SW 単体によってリンクにパケットロスが発生しているのか判断できる．この方法でパケットロス率を算出するメリットの 1 つとして，複数の計測ホストを用いてパケットの送受信を行う必要がなく，1 台の計測ホストのみで計測が可能となる点がある．

統計情報を利用したパケットロス計測の模式図を図 1 に示す．P4SW A のパケットカウンタが 1000 であることを，計測パケットを介して P4SW B に伝え，P4SW B は自身のパケットカウンタが 980 であるため，P4SW A から P4SW B への方向に 2%のパケットロスが生じていると判定できる（図 1）．

本計測システムは大きく分けて 3 段階のプロセスに分かれており，各段階の動作について説明する．

(1) 計測経路の設定
MH はネットワークトポロジを事前把握し，それに基づき MH を出て，全ての全二重リンクを双方向に 1 回ずつ通過し MH に戻る一筆書き経路を決定する．経路を各 P4SW の送信ポート番号で表現し，計測パケットに通過する経路の順番通りに経路情報を格納する．

(2) 転送される計測パケットによるパケットロスの監視
MH は計測パケットを規定数送信する．計測パケットを受信した P4SW は，受信ポート毎にある通過数カウンタを 1 増やし，通過数カウンタの値を計測パケットに記録する．もし，計測パケットが計測開始要求通知を格納している場合は，P4SW の通過数カウンタをリセットする．P4SW の通過数カウンタと計測パケット内の直前の P4SW の通過数カウンタに差がある場合には，計測パケットにロス発生を記録する．続いて，計測パケットに格納されている経路情報を参照して次の P4SW へ計測パケットを転送する．MH は最後に送信する計測パケットに計測終了通知を格納する．

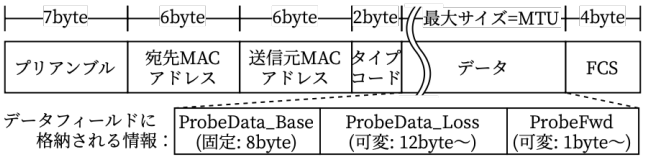


**図 2** Ethernet フレームのフォーマット

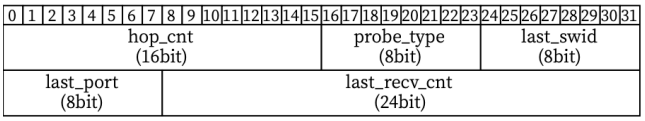


**図 3** ProbeData_Base のフィールド

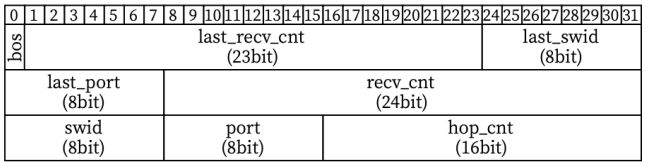


**図 4** ProbeData_Loss のフィールド（繰り返し）

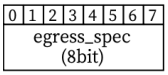


**図 5** ProbeFwd のフィールド

(3) 障害リンクの判定と表示

MH は受信した計測パケットに抱合されているロスリンクのデータを参照し，パケットロス率を計算する．算出したパケットロス率が，ネットワーク管理者が予め設定する閾値 $h$ 以上であったリンクである．パケットロス率が閾値未満であったリンクは正常リンクと判定する．障害リンクを検知する毎に，リンクとその両端にある P4SW の情報を表示し，計測パケットを規定数送信後，計測システムの動作が終了する．

## 3. 計測システムの実装

### 3.1 独自プロトコルの計測パケット

計測に関わる機能を実現するために，ネットワーク通信で一般的に使用される DIX 仕様（EthernetⅡ 形式）の Ethernet フレームを使用する．

Ethernet フレームのデータ部には，既存の IPv4 等プロトコルに基づいたパケットフォーマットを使わず，本計測システムの動作に適した独自のパケットフォーマットを格納する．この独自に定義するパケットと，その他の IPv4 パケット等を見分ける識別子として，Ethernet フレームのタイプコードに独自プロトコルであることを示す番号を格納する．Ethernet フレームのフレームフォーマットを図 2 に示す．また，ProbeData_Base，ProbeData_Loss，ProbeFwd のパケットフォーマットをそれぞれ図 3，4，5 に示す．

1 つ目の ProbeData_Base ヘッダには，計測パケットのホップ数と P4SW の動作を制御する識別符号を示すフィールド，P4SW の統計情報の一時保存を行うフィールドがある．2 つ目の ProbeData_Loss には，ロスリンクのデータ格納場所となり，P4SW によってリンクにパケットロスが発生したと検知される度に，計測パケットに追加される．3 つ目の ProbeFwd には，計測パケット自身の経路情報が格納される．

### 3.2 計測パケットを受信した P4 スイッチの動作

2 節で検討した計測システムを機能させるには，3.1 節に示した計測パケットと，計測パケットの書き換え機能を有したデータプレーンプログラミングが必要となる．

P4 スイッチの機能は 3 種類に大別され，以下に挙げる．

**(a) フォワーディング**

各 P4SW において，どの受信ポートから受信したパケットも Ethernet フレームのヘッダを解析し，タイプコードフィールドによって動作を変化させる．

タイプコードフィールドが計測パケットを指し示す独自プロトコルである場合，ProbeData_Base ヘッダの中にある `hop_cnt` の値と ProbeFwd の `egress_spec` フィールドの対応する位置の送信ポートを取得し，その送信ポートに送信する．また，タイプコードフィールドが IPv4 パケットを示す場合，パケット内の宛先 IP アドレスを参照して，マッチアクションテーブルに照らし合わせて，テーブルにて指定される，宛先 MAC アドレスを Ethernet フレームに格納し，送信ポート番号を P4SW に設定する．次に，パケット内の TTL を参照して，取得した TTL から-1 した値を書き込んで送信する（フォワーディング以外の処理は未実装）．

**(b) 各 P4SW での計測パケットのロスの検知**

各 P4SW（$SW_A$ とする）においてある受信ポートから計測パケットを受信した時，それを送信した手前の P4SW（$SW_B$ とする）がその計測パケットの経路に沿って過去に受信した計測パケット数（当該計測パケットを含まない）が計測パケット内の Probedata_Base ヘッダの `last_recv_cnt` フィールドに示されている．よって，その値 $x$ と，$SW_A$ 内のレジスタが示すその受信ポートから過去に受信した計測パケット数 $y$（当該計測パケットを含まない）とを比較し，もしその差が 0 であれば，$SW_B$ から $SW_A$ へ向かうリンクでの計測パケットはこれまでロスしなかったことが判る．一方，その差が 1 以上であれば，(c) の方法で，そのリンク上での計測パケットのこれまでのロス率（$SW_B$ 内でのパケット廃棄を含む）を計算するための情報を MH に通知する．

**(c) リンクでのロス発生情報の MH への通知**

(b) において，リンク上での計測パケットのロスが以前に発生していた場合，「$SW_B$ がその計測パケットの経路に沿って過去に受信した計測パケット数 $x$」と

「$SW_A$ がその受信ポートから過去に受信した計測パケット数 $y$」を，$SW_A$ の位置（経路上の hop_cnt），$SW_A$・$SW_B$ の SW 番号とロスリンクに繋がっているポート番号と共に，計測パケットの可変長配列である ProbeData_Loss 内に格納し，(a) に従って次の P4SW へ転送する．ただし，多数の異なるリンクでロスが発生する場合，それらのロス発生情報を 1 個の計測パケットに格納できないことがあるため，$SW_A$ でのリンクのロス発生情報はすべての計測パケットではなく一定個数に 1 個の計測パケットに格納する．

計測システムの P4 スイッチの動作をステート毎に示す．トポロジ内の全ての P4SW で同じプログラムが動作する．

#### 3.2.1 Headers

Headers では独自プロトコルのヘッダ構造と，以降の各種処理にて必要なメタデータの定義を行う．

#### 3.2.2 Parser

Parser では，受信した計測パケットのプロトコルとホップ数によって動作を変化させる．

(1) Header で定義したヘッダの情報を基に，計測パケットを解析してメタデータとマッピングする．
(2) Ethernet フレームのヘッダを解析し，タイプコードフィールドを読み込んで，独自プロトコルの場合は次の (3)〜(6) を実行する．
(3) 計測パケットの ProbeData_Base ヘッダの hop_cnt フィールドを読み込んで，計測パケットが P4SW を通過した回数（ホップ数）を取得する．
(4) 計測パケットに格納されているホップ数が 0（最初のホップ）である場合，計測パケットのロスリンクのデータは空であるので，次の (5) に進む．計測パケットに格納されているホップ数が 1 以上である場合，ロスリンクのデータが 1 つ以上存在する可能性がある．パケットに記録されているロスリンクのデータを直近に保存されたものから順番（再帰的）に探索して，bos フィールドが 1 の場合は次の (5) に進む．
この処理では，ロスリンクのデータを格納している ProbeData_Loss は複数の収容可能（可変長）であり，複数の ProbeData_Loss が続いた後に ProbeFwd が記録されている計測パケットにおいて，ProbeData_Loss はまだ格納されているのかを確認し，bos フィールドのビットが 1 であった場合，ProbeData_Loss のデータ部分が終了し，次のビットから ProbeFwd の egress_spec フィールドが続くと判断できる．
(5) パケットに記録されている ProbeFwd を，ホップ数の回数だけ順番に探索して，次の P4SW に接続されているポート番号をメタデータフィールドに保存する．
この処理では，次に転送する P4SW への送信ポートを識別するために行う．現在の計測パケットのホップ数の回数分，再帰的に ProbeFwd を読み出すことで次に転送する P4SW の送信ポートがメタデータに保存される．このメタデータフィールドは，パケットを次の P4SW へ送信する際に用いられる．
(6) パケットを次のステート（Ingress）に転送する．

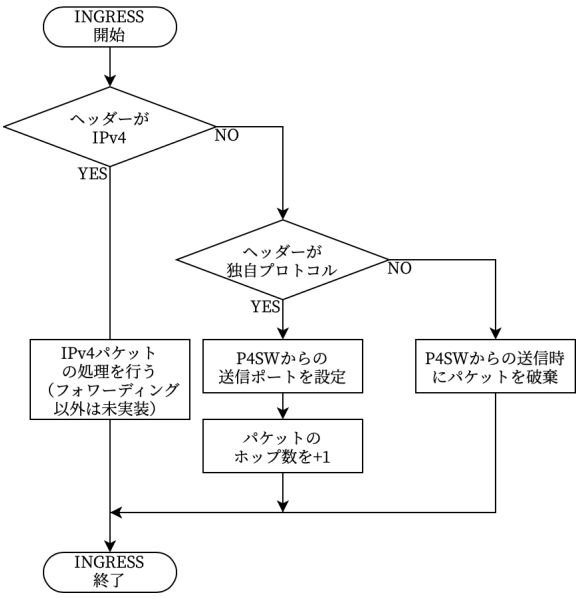


**図 6** Ingress 処理のフローチャート

#### 3.2.3 Ingress Pipeline

Ingress 処理のフローチャートを図 6 に示す．Ingress では，計測パケットの送信先ポートを再定義し，計測パケットに格納されるホップ数をインクリメントする機能を有する．

(1) パケットが独自プロトコルであり，ProbeData_Base ヘッダが存在する時に次の (2) に進む．
(2) 計測パケットの送信ポートを，Parser によって取得された送信ポートに設定する．
(3) ProbeData_Base ヘッダの hop_cnt フィールドに格納されているホップ数をインクリメントする．
(4) パケットを次のステート（Egress）に転送する．

#### 3.2.4 Egress Pipeline

Egress 処理のフローチャートを図 7 に示す．Egress では，計測パケットの ProbeData_Base ヘッダの probe_type フィールドを参照し，その値によって動作を変更する．

**probe_type が 255 であった場合**

(1) 計測パケットのカウンタを初期化する．
(2) ProbeData_Base ヘッダの hop_cnt フィールドが 1 であった場合は，計測パケットに ProbeData_Loss を新規作成し，bos フィールドに 1 を格納する．この処理で作成された ProbeData_Loss は，このパケットを次に受信する P4SW が，Parser ステートにおいて ProbeFwd の開始ビットを特定する際に使われる．
(3) パケットを次のステート（Deparser）に転送する．

**probe_type が 0〜249 であった場合**

計測パケットは受信した P4SW の計測パケットカウン

タを増やす機能とともに，ロスリンクの格納機能（ProbeData_Loss）を併せ持っている．計測パケットに対応付けされている収集番号は probe_type フィールドの値となり，収集番号 $n$ の時，ホップ数が $50n-1$ 以上かつ $50(n+1)$ 未満のリンクでパケットロスが発生した場合のみ，ProbeData_Loss にロスリンクのデータを書き込む．

(1) ProbeData_Base ヘッダの hop_cnt フィールドが 1 であった場合は，パケットに ProbeData_Loss を新規作成し，bos フィールドに 1 を格納する．
(2) 計測パケットの収集番号とホップ数の対応付けが合致した場合は次の (3) に進み，合致しない場合には，(4) に進む．
(3) P4SW の計測パケット数カウンタよりも，ProbeData_Base ヘッダの last_recv_cnt が大きい場合には，前回送信された P4SW とのリンクにて計測パケットがロスしているので，パケットに ProbeData_Loss を新規作成し，ロスリンクがあるホップ数，直近の P4SW が送信した時に記録したスイッチ ID，送信ポート番号，パケット数カウンタと，現在の P4SW のスイッチ ID，送信ポート番号，パケット数カウンタを格納して，次の (4) に進む．P4SW の計測パケット数カウンタと，ProbeData_Base ヘッダの last_recv_cnt が同じ場合には，リンクはパケットロスをしていないので，次の (4) に進む．
(4) 計測パケットのカウンタをインクリメントする．
(5) 現在の P4SW のスイッチ ID を ProbeData_Base ヘッダの last_swid，送信ポート番号を last_port，パケット数カウンタを last_recv_cnt に格納する．
(6) パケットを次のステート（Deparser）に転送する．

#### 3.2.5 Deparser

Parser ステートにおいて読み取ることができたフィールドを参照して，順番通りに再度編成しなおして，パケットを送信する．

## 4. エミュレーションの実施と評価

### 4.1 各種パラメータ

計測システムの有効性を確かめるために，The Internet Topology Zoo[6] がデータセットとして提供する実ネットワークトポロジの中から Geant，Uunet，Uninett を使用して，mininet[9] と P4SW の bmv2[10] を用いてエミュレーションを行った．使用した実ネットワークトポロジを図 8，図 9，図 10 に示す．

計測パケットが走査する経路の区間を 3 分割し，区間 A・区間 B・区間 C とする．3 区間のうち 2 区間内の複数リンクに対して以下の混雑度を設定する．区間内でリンクのパケットロスを設定する場所は，ホップ数 $i$ と $i+1$ が連続して同じリンクの上りと下りを通るリンクとして，パケットロス率は混雑度に応じて設定する．

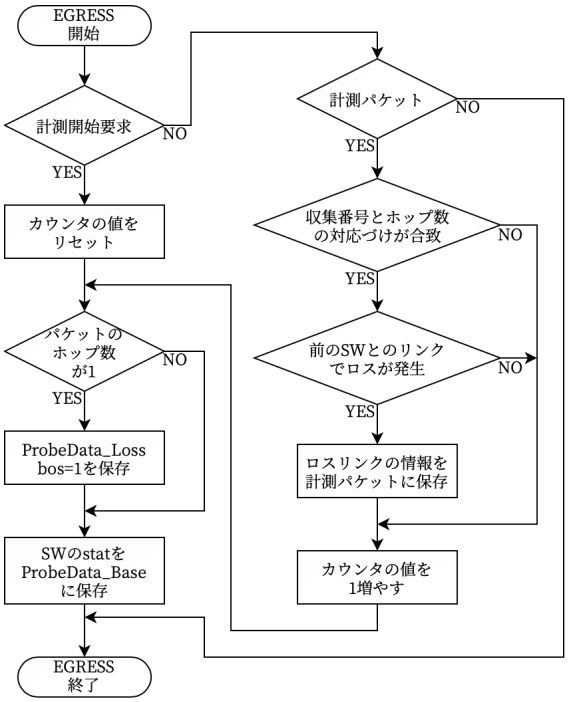


**図 7** Egress 処理のフローチャート

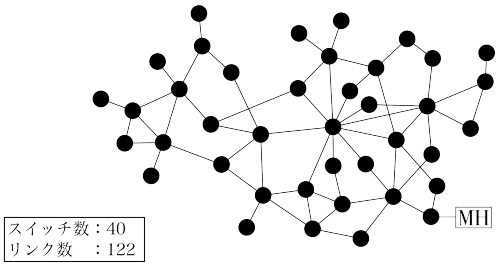


**図 8** Geant トポロジ

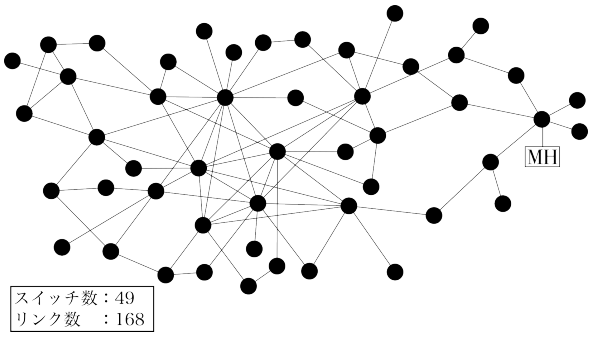


**図 9** Uunet トポロジ

**大混雑度** 設定リンク数 2 本（計測経路上で連続する 2 リンク），設定パケットロス率 10.5 %
**中混雑度** 設定リンク数 4 本（計測経路上で連続する 2 リンクを 2 箇所），設定パケットロス率 1.5 %
**全リンク** 設定リンク：上述 6 本を除いた全リンク，設定パケットロス率 0.1 %
**閾値** $h$ 障害リンクと判定するパケットロス率 閾値 $h \geq 8\%$

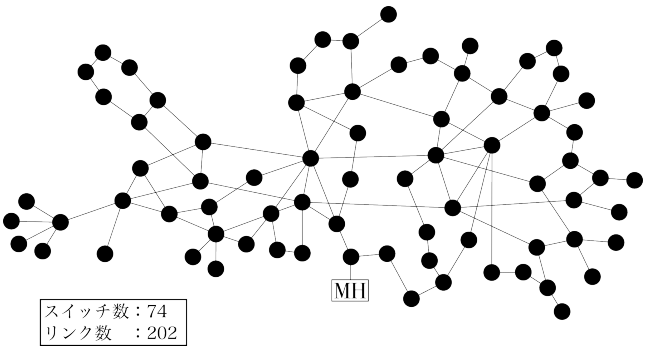


**図 10**　Uninett トポロジ

### 4.2　エミュレーション手順

MH が P4SW に「計測開始要求」を送信すると処理が開始される．計測手順を以下に示す（図 11）．

(1) 計測対象のトポロジにパケットロス率を設定
　トポロジ内でパケットロスを発生させるリンクを指定し，パケットロス率を設定する．

(2) 計測パケットを送信
　2 節にて示した，計測経路の設定と計測パケットによるパケットロスの監視を実施する．

(3) 計測パケットの受信と障害リンク表示
　MH は収集番号ごとに計測パケット内の収集データを一時保存しており，計測パケットが届くたびに MH の収集データを上書きしている．計測終了通知を示す識別符号が格納されている計測パケットを受信すると，MH に一時保存していた収集データを参照し，パケットロス率を計算する．算出したパケットロス率が，閾値 $h$ 以上であったリンクを障害リンクとして扱い，障害リンクの情報を標準出力する．

(4) 計測パケット送信数の変更
　計測パケット送信数を 100 から 200 までは 25 間隔で増やし，200 から 2500 まで 100 間隔で増やして観察する．計測パケットによって算出された閾値 $h$ 以上のロスリンクのパケットロス率を観察する．

(5) パケットロスリンク・トポロジの変更
　手順 1 のパケットロスが発生するリンクの場所を 4 回変えて，更に 6 つのパターンを変化させ，(2)〜(4) を再実施する．また，3 つのトポロジを変えて，合計 72 回の測定を行う．

### 4.3　評価結果

#### 4.3.1　計測パケット送信数に対する精度

Geant における，計測パケット送信数に対する障害リンクの誤検知率と推定パケットロス率を図 12 に示す．計測パケットを 150 回送信すると，推定されるパケットロス率の最大値は 15.0%で，最小値は 4.90%，このときの第一四分位数は 8.46%，中央値を示す第二四分位数は 10.3%，第三四分位数は 12.4%であり，50%の確率でパケットロス率を

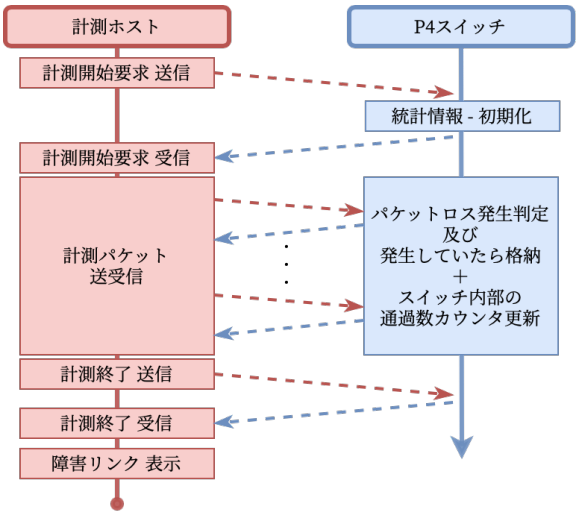


**図 11**　計測システムの動作手順の概略

8.46%〜12.4%と推定でき，その全ては閾値以上であるため障害リンクと判定される．最小値〜第一四分位数の区間にパケットロス率を推定した回数は全 54 回の試行において 11 回であり，うち 6 回は閾値を下回って正常リンクと誤検知された．障害リンクと正しく判定される計測パケット送信回数を探すと，600 回となり，推定されるパケットロス率の最大値は 12.8%，中央値は 10.4%であり，設定パケットロス率 10.5%に対して，±2.5%の範囲まで絞り込めた．

Uunet における同様の結果を図 13 に示す．障害リンクと正しく判定される計測パケット送信回数を探すと，500 回となり，計測パケットを 500 回送信すると，推定されるパケットロス率の最大値は 13.8%，中央値は 10.1%であり，設定パケットロス率 10.5%に対して，±3.3%の範囲まで絞り込めた．

Uninett における同様の結果を図 14 に示す．障害リンクと正しく判定される計測パケット送信回数を探すと，1200 回となる．計測パケットを 1200 回送信すると，推定されるパケットロス率の最大値は 13.2%，中央値 10.6%であり，設定パケットロス率 10.5%に対して，±2.7%の範囲まで絞り込めた．

障害リンクを正常リンクと取り違える誤検知率は，計測パケット送信回数を増やすほど 0%に近づくが，計測パケット送信数が 175 回から 200 回に増加すると，誤検知率が増加する現象を観測した．計測パケット送信数 200 回の時に正常リンクと誤検知した大混雑度リンクは，175 回送信時点で 8.33%と推定されていたが，200 回送信時点では 7.45%と推定された．このリンクの両端に接続される P4SW のパケットカウンタは，175 回送信時点で送信側 144 回・受信側 132 回で差は 12，200 回送信時点で送信側 161 回・受信側 149 回で差は 12 を記録しており，送信回数を重ねるものの，当該のリンクでパケットロスが発生しなかった結果，推定されるパケットロス率が閾値を下回って正常リンクと誤検知した．Uunet においても同様の原因で障害リン

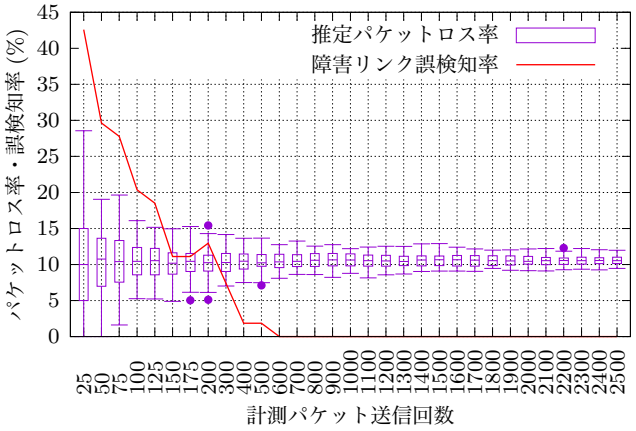


**図 12** Geant における大混雑度設定リンクの推定パケットロス率

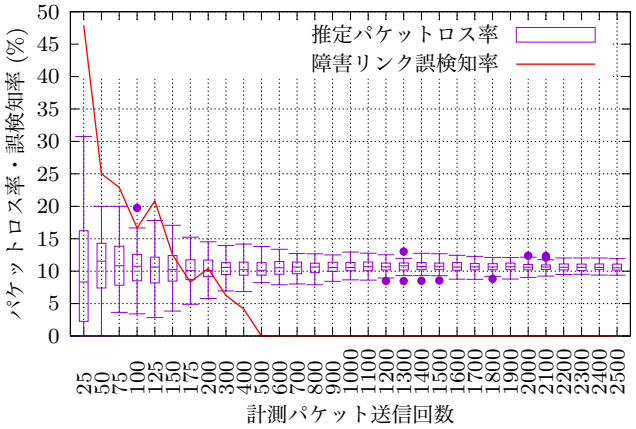


**図 13** Uunet における大混雑度設定リンクの推定パケットロス率

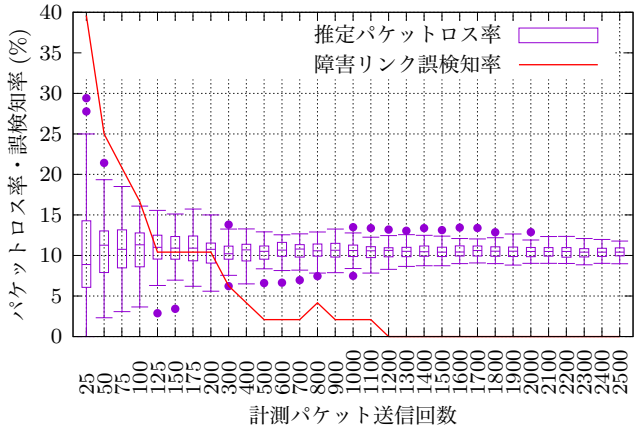


**図 14** Uninett における大混雑度設定リンクの推定パケットロス率

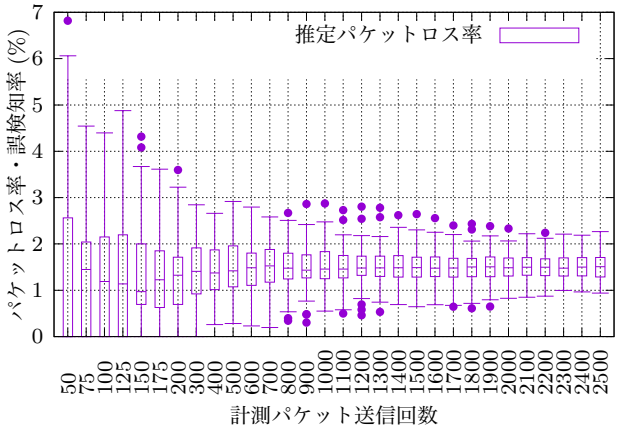


**図 15** Geant における中混雑度設定リンクの推定パケットロス率

クの誤検知が発生している．この誤検知を無くすには計測パケットの送信回数を増やす対応が必要となる．

大混雑度程度の大きなパケットロスリンクがあれば，計測パケットを 1200 回の送信することで，Uninett のような比較的複雑で大きいネットワークにおいても，障害リンクを特定できることが判る．計測パケット送信回数をこれ以上増やすと，中央値は設定したパケットロス率に収束している．

続いて，中混雑度程度のパケットロスリンクを特定する目的で，閾値 $h$ が低い場合を仮定して，中混雑度リンクの推定パケットロス率を評価する．

Geant における，計測パケット送信数に対する中混雑度リンクの推定パケットロス率を図 15 に示す．図 15 より，推定パケットロス率の外れ値が見られなくなるのは，計測パケット送信回数が 2300 回の時である．推定されるパケットロス率の最大値は 2.21%で，最小値は 1.00%，このときの第一四分位数は 1.30%，中央値を示す第二四分位数は 1.48%，第三四分位数は 1.70%であり，50%の確率でパケットロス率を 1.30%〜1.70%と推定できる．

大混雑度リンクで得られた結果と比べ，送信回数に対して外れ値を観測することが多い．つまり，中混雑度程度のパケットロス率を特定するには，大混雑度程度の大きなパケットロス率の測定に比べて，計測パケット送信回数を大きく増やす必要があることが判る．

### 4.3.2 MH へ到達した計測パケット数

図 16 は，3 つのトポロジにおいて，計測パケットを 2500 回送信し，MH への計測パケット到達数と到達率を示したグラフである．

トポロジ毎の計測パケット到達数とその割合は，Geant 1676 個 > Uunet 1596 個 > Uninett 1549 個，Geant 67.4% > Uunet 64.0% > Uninett 61.6%である．トポロジ内のリンク数を考えると，Geant 122 個 < Uunet 168 個 < Uninett 202 個である．

Geant において，1549 個の計測パケットが MH に到達した際の計測パケット送信回数は 2200〜2400 回の間であった．トポロジ内の全リンクの設定パケットロス率を合計すると，Geant は 59.8%，Uninet は 75.8%であった．

パケットロス設定リンクがトポロジ内のどの区間にあっても，最終的に MH へ到達する計測パケット数は変化しないことは自明である．しかし，リンク数が変化すると，

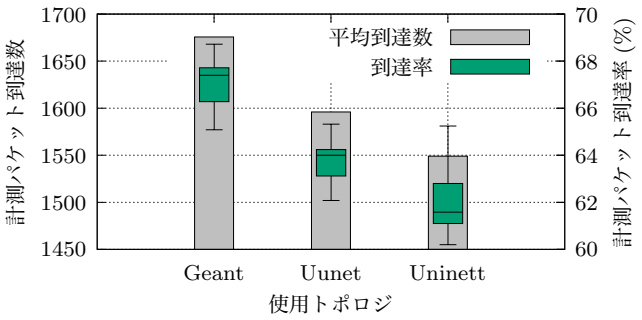


**図 16** MH まで到達した計測パケットの数と割合

大・中混雑度のパケットロス設定リンク数に変わりは無いが，大・中混雑度以外の全リンクはパケットロスが 0.1%に設定されるため，トポロジのリンク数が増えるほど，トポロジでロスするパケットが増える．よって，トポロジ内のリンクのパケットロス率の合計が変わる場合には，計測パケット送信数を調整する必要があることが判る．

## 5. まとめ

本報告では，全二重リンクを持つ P4 ネットワーク上で，障害リンク検知までの時間短縮と，障害リンクの検知・位置特定が制御プレーンに与える影響を低減することを目的に，P4 スイッチ自身がリンクのパケットロスを検出して，計測パケットに統計情報を格納することで，P4 コントローラを介さずにパケットロス発生リンクを計測ホストに通知する計測システムの設計について検討した．

本設計を評価するため，3 種類の実ネットワークトポロジに対して，計測パケットの送信回数を 25 から 2500 まで変動させながら，大・中混雑度リンクの場所を変えてエミュレーションを行い，障害リンクを誤検知なく特定するのに要する計測パケット送信数を調べた．

得られた評価として，大混雑度程度の大きなパケットロス率のリンクがあれば，計測パケットを 900 回〜1200 回送信することで障害リンクを誤検知なく特定でき，中混雑度程度のパケットロス率のリンクを特定するには，計測パケットを 2000 回程度送信することで，ロスリンクの存在が検知できる．また，パケットロス設定リンクがどの区間にあっても，設定リンクの数が変動しない限り，最終的に MH へ到達する計測パケット数は変化しないが，トポロジ内のリンクのパケットロス率の合計が変わる場合には，計測パケット送信数を調整しなければいけないことが判った．

従来の OpenFlow 技術における計測システムでは，規定数の計測パケットを送信した後に，コントローラが各スイッチに通過数情報を問い合わせ，それを基に障害リンクを特定していたが，本検討における P4 技術を用いた計測システムでは，P4 スイッチから計測パケット内に情報を格納することによって，制御プレーンを介さずに全リンクのアクティブ計測を行いながらパケットロス発生リンクを検知することが実現し，より早く障害リンクを特定できた．ネットワーク状態の急変によってリンクのパケットロスが急増しても，これを基に迅速なネットワーク制御に繋げられる可能性がある．

今後の課題として，以下の検討・改善すべき点がある．

- P4SW の内部カウンタに関する問題がある．計測パケット通過数を保存する内部カウンタをポート毎に用意するため，計測経路のホップ数に紐づけたが，計測経路を変えた場合，計測パケットの経路上のホップ数と P4SW の内部カウンタが合致しなくなる．
- 計測パケットのロスリンク検知の際に過剰な情報を格納している問題がある．計測ホストはパケットロスリンクのホップ数と，その両端に接続される P4SW の通過数カウンタの情報だけで障害リンクの検知が行える．
- 全ての計測パケットに経路情報を格納している問題がある．計測ホストから送信される計測パケットは例外なく同じ経路を走査するため，P4SW に宛先ポート番号を記憶させる実装を行うことで，経路情報を全ての計測パケットに格納する必要がなくなる．

**謝辞** NICT の委託研究 JPJ012368C05501 および JSPS の科研費研究 20K11770 により，本研究成果は得られた．ここに謝意を表す．

### 参考文献

[1] 月岡祐太, 鶴正人, “OpenFlow ネットワークでの全リンクパケットロス率計測の効率化”, 電子情報通信学会技術研究報告, IN2015-84, pp. 77-82, 2015.
[2] 藤村悠樹, 月岡祐太, 鶴正人, “OpenFlow における全リンクパケットロス率計測のための統計情報取得順序の最適化”, 電子情報通信学会技術研究報告, CQ2016-45, pp. 73-78, 2016.
[3] 永田隼也, 鶴正人, “OpenFlow におけるリンク毎パケット遅延変動の監視と劣化リンク特定の効率化”, 電子情報通信学会技術研究報告, ICM2018-19, pp. 47-52, 2018.
[4] S.Goto, M.Shibata and M.Tsuru, “Dynamic optimization of multicast active probing path to locate lossy links for OpenFlow network”, Proc. the 34th ICOIN, pp. 628-633, 2020.
[5] 佐野由一, 柴田将拡, 鶴正人, “障害リンク検知のためのオイラー閉路分解を用いたパケットロス計測経路の設計”, 電子情報通信学会 2023 年総合大会, 電子情報通信学会大会講演論文集, B-14-9, 2023.
[6] “The Internet Topology Zoo”, [Online]. Available: http://www.topology-zoo.org/ (参照 2024-01-24).
[7] N.McKeown, et al., “OpenFlow : Enabling innovation in campus networks”, SIGCOMM Computer Communication Review, Vol. 38, pp. 69-74, 2008.
[8] P.Bosshart, et al., “P4: Programming protocol-independent packet processors”, SIGCOMM Computer Communication Review, Vol. 44, pp. 87-95, 2014.
[9] “Mininet: An Instant Virtual Network on your Laptop (or other PC)”, [Online]. Available: http://mininet.org/ (参照 2024-01-21).
[10] “BEHAVIORAL MODEL(bmv2)”, [Online]. Available: https://github.com/p4lang/behavioral-model (参照 2024-01-21).