Vol.2022-IOT-58 No.17
2022/7/13

# キャンパスネットワークにおける コアネットワークの費用対効果を考慮した冗長化について

大森 幹之[1,a)]

概要：新型コロナウイルス感染症 (COVID-19) の拡大に伴い，大学といった教育機関ではオンライン講義が急速に普及した．それに伴い，教育機関でのキャンパスネットワークの冗長化の重要性は増している．一方で，世界的な半導体不足によるネットワーク機器や保守費の高騰により，冗長化に投資できる費用も削減を求められている．
そこで，本稿では，費用対効果も鑑みつつ，キャンパスネットワークにおけるコアネットワークにおける冗長化を検討し，議論する．具体的には，鳥取大学における SINET 接続回線の冗長化と実際に発生した約 8 時間の通信断障害への効果について考察する．また，SINET6 への移行時における AWS との接続回線と BGP ピアリングの冗長化について，スタック構成で BFD (Bi-directional Forwarding Detection) を実装できないスイッチも念頭に検討する．
さらに，費用対効果のためにファイアウォールのバックアップ機を性能を抑えた機種にした場合の構成についても検討する．そして，OSPF と BFD を組み合わせることで，メンテナンスや障害発生の際にキャンパスネットワークの通信断を最小限に抑えることを試みる．

キーワード：コアネットワーク，冗長化，耐障害性，費用対効果

# On Redundancy of a Core Network in a Campus Network Considering Cost-Effectiveness

Motoyuki Ohmori[1,a)]

***Abstract:*** Because of the Coronavirus Disease 2019 (COVID-19) pandemic, online lectures have been rapidly deployed to educational institutions such as universities. As a result, redundancy of campus networks in educational institutions is becoming increasingly important. On the other hand, the cost of network equipment and maintenance is rising due to the global shortage of semiconductors, and there is a need to reduce the amount of money that can be invested in redundancy. In this paper, we examine and discuss redundancy in the core network of a campus network, taking cost-effectiveness into consideration. Specifically, we discuss the redundancy of SINET connection lines at Tottori University and its effect on the actual 8-hour communication breakdown that occurred. In addition, redundancy of the connection to AWS and BGP peering during the transition to SINET6 will also be discussed, taking into account switches that cannot implement BFD (Bi-directional Forwarding Detection) in a stacked configuration. This study is conducted with this in mind.
Furthermore, for cost-effectiveness, we will also examine the configuration when the backup firewall is a low-performance model. We will also try to minimize the communication breakdown of the campus network in the event of maintenance or failure by combining OSPF and BFD.

***Keywords:*** core network, redundancy, fault tolerance, cost-effectiveness

## 1. はじめに

新型コロナウイルス感染症 (COVID-19) の拡大に伴い，大学といった教育機関ではオンライン講義が急速に普及した．それに伴い，教育機関でのキャンパスネットワークの耐障害性の向上は重要度を増し，冗長化が望まれる．一方で，世界的な半導体不足によるネットワーク機器や保守費の高騰により，冗長化に投資できる費用も削減を求められている．

そこで，本稿では，費用対効果も鑑みつつ，キャンパスネットワークにおける対外接続やバックボーンを担うコアネットワークにおける冗長化を検討し，議論する．具体的には，鳥取大学における SINET (Science Information NETwork) 接続回線の冗長化と実際に発生した約 8 時間の通信断障害への効果について考察する．また，AWS (Amazon Web Services) との接続回線と BGP ピアリングの冗長化を検討する．検討にあたっては，スタック構成で BFD (Bi-directional Forwarding Detection) [1], [2] を実装できないスイッチも考慮する．検討した手法を，AWS 接続回線の学術情報ネットワーク SINET6 への切り替え時におけるダウンタイムなどで評価する．さらに，費用対効果のために次世代ファイアウォールのバックアップ機を性能を抑えた機種にした場合の構成についても検討する．そして，OSPF と BFD を組み合わせることで，メンテナンスや障害発生の際にキャンパスネットワークの通信断を最小限に抑えることを試みる．

本稿の構成は以下のとおりである．2 節では，鳥取大学を例に，キャンパスネットワークにおけるコアネットワークの冗長化の課題をより詳細に述べる．3 節では，課題の解決を目指した鳥取大学でのコアネットワークの冗長化について述べる．4 節では，提案手法などについて考察する．5 節では，関連研究に言及する．最後に，6 節で本論文をまとめる．

## 2. 課題

### 2.1 SINET 側部品の障害による約 8 時間に及ぶ通信断

本節では，鳥取大学において，2021 年に半年以内で 2 回発生し，共に復旧まで約 8 時間を要した通信断について紹介する．

1 回目として，2021 年 3 月 11 日 (木) 22:00 頃から 12 日 (金) 6:00 頃まで，約 8 時間の通信断が発生した．当時の鳥取大学では，対外接続は SINET のみで回線も 1 回線のみであったため，インターネットとの全ての通信が不達となった．平日ではあったものの，幸運にも深夜帯での発生であったため，鳥取大学内の利用者からの問い合わせは無かった．ネットワークスイッチのラインカード周辺の部品の障害と考えられ，交換により復旧した．

1 回目の部品の交換により，再発の可能性は極めて低いと考えていた．しかし，2 回目として，約 4 ヶ月後の 2021 年 7 月 11 日 (日) 5:00 頃から 12:30 頃まで，約 7 時間半に及ぶ通信断が発生した．1 回目の通信断と同様に，ラインカード周辺の障害が原因と推察された．発生日時が休日であったこと，また，1 回目の通信断発生後に鳥取大学では奇跡的に後述のように回線を冗長化していたこともあり，利用者からの問い合わせは無かった．

また，上記の 2 回の通信断の経験から，鳥取の様な地方では，障害発生から復旧まで約 8 時間の通信断は覚悟しなければならないと言える．特筆すべきは，SINET の様に 24 時間オンサイントの保守契約が締結されていても，約 8 時間を要している点である．もし，大学の機器の部品で平日 9 時から 17 時までの保守契約の場合は，より復旧までに時間を要することも予想される．

なお，他の同一の地方の国立大学においても，類似の障害がやはり数ヶ月を空けて 2 回発生していた．不運にも平日日中での発生であり，業務やオンライン講義などへの影響も大きかったと推察される．

以上のことから，1 回線の対外接続は，耐障害性に優れていると言えず，複数の回線によって接続することが望ましいと言える．

### 2.2 SINET6 移行などに伴う AWS 回線切り替えの通信断

鳥取大学では，SINET クラウド接続サービスにより，SINET L2VPN を用いて AWS と接続している．AWS 上では各種仮想サーバが動作しており，学内サービスを提供している．そして，SINET6 への移行といったメンテナンス作業時には，SINET L2VPN の切り替えが発生するため，通信断が発生する．このメンテナンス作業による通信断を回避するためには，SINET 大阪 DC と東京 DC それぞれと SINET L2VPN で大学を接続することで実現できる．実際に SINET6 への移行では，SINET 大阪 DC と東京 DC 経由の回線は同時には作業されず，時間差を設けて移行されたため，通信断を BGP で検知し自動的に検知して切り替えることが可能であった．

しかし，鳥取大学では，AWS との BGP ピアリングの切り替えに 180 秒以上を要していたことが判明した．この切り替え時の長時間の通信断は，BGP の標準的な生存確認のタイムアウト，つまり，ホールドタイマが 180 秒であることに依っていた．DNS や LDAP といった通信では，数十秒に及ぶ通信断が障害と検知され，認証が失敗していたこともあった．

BGP ピアリングの切り替えを高速化するためには，BFD

[1] 鳥取大学 情報基盤機構
Organization for Information and Communication Technology, Tottori University
[a)] ohmori@tottori-u.ac.jp

の併用が考えられる．実際，AWS は BFD にも対応しており，推奨もされている．しかし，鳥取大学では，BGP ピアリングのために採用していたアラクサラネットワークス社 (以下 AlaxalA という) の L3 スイッチ AX3660-48XT4QW (以下 AX3660 という) の機能の制限に起因して，BFD を併用できなかった．AX3660 では単体の 1 台だけであれば，BGP ピアリングの生存確認を BFD でも行える．しかし，スタックにより冗長構成を取ると BFD を併用できないのである．鳥取大学では耐障害性向上のために，AX3660 をスタック構成とし冗長化していた．皮肉なことに，それが仇となり，通信断が延びていた．

### 2.3 オンプレミスな機器の障害に伴う全サービスの停止

鳥取大学の所有している対外接続 L3 スイッチや次世代ファイアウォール，SINET そのものに障害発生した場合，鳥取大学では全サービスが停止してしまっていた．鳥取大学内にのみ設置されたサーバのみで提供しているサービスであれば停止も止むを得ないと言える．しかし，クラウド化が進んだ現在，メールやオンライン会議などのサービスのサーバは学外のネットワークで提供されている．鳥取大学の所有している機器やネットワークに障害が発生していたとしても，これらのサービスの提供を継続できた方が良い．

### 2.4 次世代ファイアウォールの HA 構成と機器価格

運用性を保ちつつ，耐障害性を向上するためには，いわゆる，HA (High Availability) 構成とするのが望ましい．HA 構成では，同一機器を 2 台以上用意し，仮想的に 1 台として運用できる．そのため，同じ設定を複数台に投入する必要も無くなり，運用性も保てる．また，1 台に障害発生した際に，既に確立済みのコネクションを正常系に移し継続できる．しかし，以下の課題がある．

(1) 次世代ファイアウォールといった機器では購入費用が高額となる．
(2) 障害時に正常に切り替わるとは限らない．
(3) コネクションの維持といった高度な HA 構成は複雑になり，運用に機器特有の知識を必要とすることがある．

## 3. 費用対効果を考慮したコアネットワークの冗長化

本節では，2 節で述べた課題を解決するために，費用対効果を考慮したコアネットワークの冗長化を検討する．まず，鳥取大学におけるコアネットワークとその冗長化の例の概略を述べる．次に，冗長化のより詳細について説明する．具体的には，SINET 接続回線の冗長化，AWS との接続の冗長化，AWS 上の仮想マシンの学外ネットワークを経由しないインターネットとの通信，次世代ファイアウォールの IP での冗長化について述べる．

### 3.1 鳥取大学におけるコアネットワークの冗長化例の概観

図 **1** に鳥取大学におけるコアネットワークの冗長化例を示す．図 1 において，吹き出しを付した機器やリンクが本論文におけるコアネットワークを指す．図 1 から分かる様に，鳥取大学のキャンパスは大別すると，湖山キャンパスと米子キャンパスに別れている．各キャンパスは，鳥取情報ハイウェイを経由し，NTT 西日本の寺町 DC センターを経由して，SINET に接続している．鳥取大学では，対外接続としては SINET しか無いため，回線を冗長化している．また，吹き出しを付したコアネットワークの機器はおおよそ冗長化している．

なお，各キャンパスから鳥取情報ハイウェイへの接続は，規則の制約から 1 回線でしか接続できない．鳥取情報ハイウェイへの接続の冗長化については今後の課題とし，本稿では論じない．

### 3.2 SINET 接続のリンク集約による冗長化

2.1 節で述べた課題を解決するため，対外接続回線を冗長化した．冗長化にあたっては，SINET との接続回線の 10G-LR を 1 回線増やし，2 回線とした．接続にあたっては，SINET と鳥取大学各スイッチにて，2 回線の光ファイバを異なる通信モジュールで収容することとした．そして，10G-LR の回線を LAG (Link Aggregation: リンク集約) により，仮想的に論理的な一本の回線とすることとした．LAG の構成にあたっては，静的 LAG ではなく，動的 LAG である LACP (Link Aggregation Control Protocol) [3] を採用した．LACP では，LACPDU を双方向で送受信できていることを確認した上で，回線の疎通を確認する．そのため，物理的な光の受信による回線の疎通を確認する静的 LAG よりも，より精度が高いと考えられる．また，LACPPDU の送信間隔 (`periodic-timer`) は AX3660 のデフォルトの 30 秒 (`long`) とした．AX3660 では，1 秒 (`short`) も指定可能であったが，予期せぬ高負荷での誤ったリンクダウン検知を避けるため，30 秒のままとした．なお，LACP では SINET と鳥取大学のスイッチの相互運用性が懸念されたが，SINET と相談の上 LACP とすることとした．

### 3.3 BGP タイマの調整のみによる AWS 接続回線の冗長化と収束時間の短縮

2.2 節で前述したとおり，鳥取大学の対外接続 L3 スイッチの構成では BFD を有効化できない．そのため，BGP のホールドタイマが切れるまで，通信断を検知できず，切り替えられない．標準的な BGP の設定では生存確認メッセージの送信間隔は 60 秒である．そして，送信間隔の 3 倍がホールドタイマとなる．その結果，鳥取大学の環境では，通信断を検知するまで最大で 180 秒を要していた．

そこで, BGP の生存確認メッセージの送信間隔を AX8600

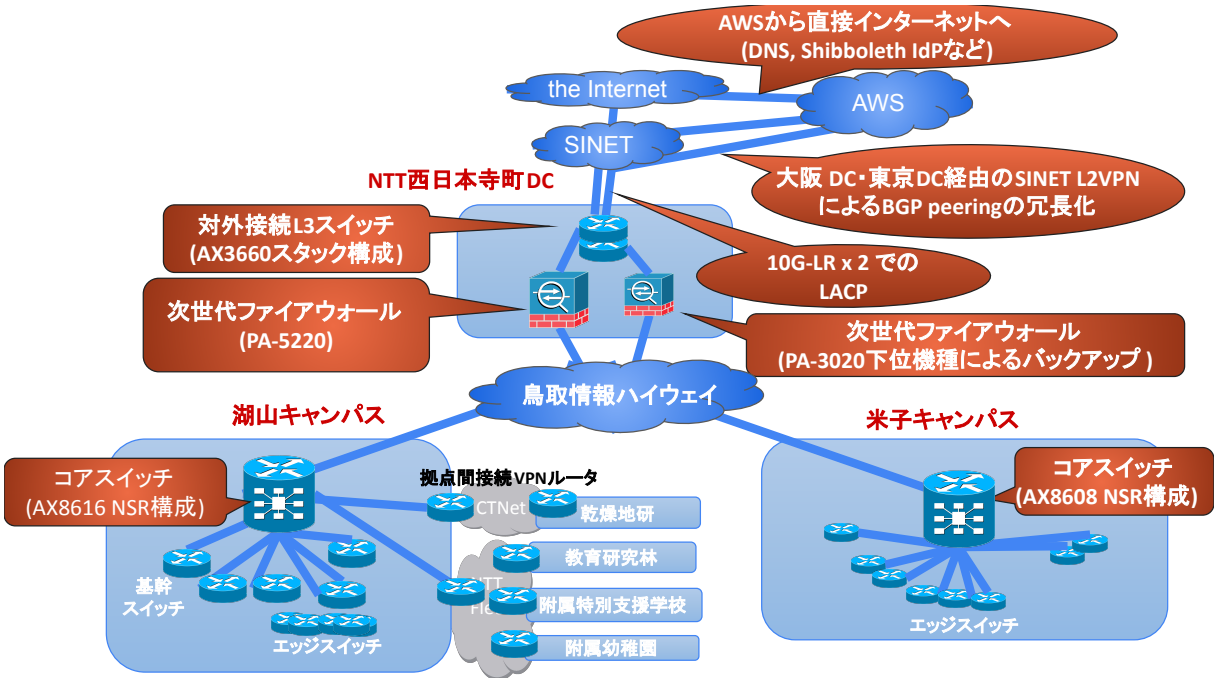


図 1　鳥取大学におけるコアネットワークの冗長化

で最小値として設定できる 1 秒とした．BGP においては，生存確認メッセージの送信間隔は相互に交換され，一致していることが求められるが，AWS の BGP の実装でも 1 秒の送信間隔が設定可能であった．これにより，BFD が設定できない状況でも，最大でも約 3 秒以内で通信断を検知し，切り替えることが可能となった．

### 3.4　AWS 上の仮想マシンからインターネットへの直接通信

2.3 節で述べたとおり，鳥取大学が所有する機器に障害が発生したとしても，クラウド上のメールシステムや認証のために Shibboleth IdP，DNS などのサービスの提供は継続できた方が良い．そこで，図 1 に示すとおり，学外からの Shibboleth IdP や DNS へのアクセスは，AWS から学内ネットワークを介さずに，直接インターネット経由で実現することとした．実現にあたっては，後述する OSPF と BGP の経路情報の交換により実現した．AWS 内には，学内の IGP の経路を eBGP で広報し，学内からのアクセスは学内ネットワークを介して通信する様にした．

### 3.5　異なる次世代ファイアウォールの IP での冗長化

2.4 節で述べた様に，次世代ファイアウォールを HA 構成とするには費用がかかる．そこで，鳥取大学では，次世代ファイアウォールを HA 構成とはせず，プライマリ機 (`fw01`) よりも下位機種のモデルをバックアップ機 (`fw02`) とし，IP での経路制御による冗長化とした 図 **2**．パロアルトネットワークス社の次世代ファイアウォール (以降パロアルト) を採用した．バックボーンのネットワークに対しては，グローバル IP アドレスを付与した．一方，パロアルト配下の各キャンパスでは一部を除き，プライベート IP アドレスを付与し，パロアルトで NAT (Network Address Translation) や NAPT (Network Address Port Translation) を行った．SINET との経路交換には eBGP を採用し，SINET からはデフォルト経路のみ受信していた．一方，鳥取大学の IP アドレス空間を集約したもののみを広告していた．IGP としては OSPF を採用し，各リンクのメトリックを経路計算に反映するため，経路の再配布時のメトリックタイプは，一般的なデフォルトではないタイプ 1 とした．そして，バックアップ機のパロアルト以外のリンクに対してはメトリックを 10 とした．一方，バックアップ機では，対外接続 L3 スイッチ向けリンクを 200 として，各キャンパスから対外接続 L3 スイッチ向きのパケット転送ではプライマリ機が優先される様にした．そして，各キャンパス向けリンクを 1024 とし，対外接続 L3 スイッチから各キャンパス向け，つまり，インターネット側から各キャンパス向きのパケット転送でも，プライマリ機が優先される様にした．各キャンパス向けリンクを他のリンクよりも比較的大きなメトリックとしたのは，キャンパス内のネットワーク構成変更時に意図せずバックアップ機を経由するのを防止するためである．また，OSPF での経路収束を高速化するため，湖山キャンパス，米子キャンパスのコアスイッチでは，OSPF の SPF 計算の遅延時間と SPF 計算の間隔をそれぞれ設定できる最小値である 100msec，SPF 計算の間隔を 1sec とした．パロアルトでは，SPF 計算の遅延を 1 秒とした．

パロアルトのプライマリ機のダウンとアップの検知を

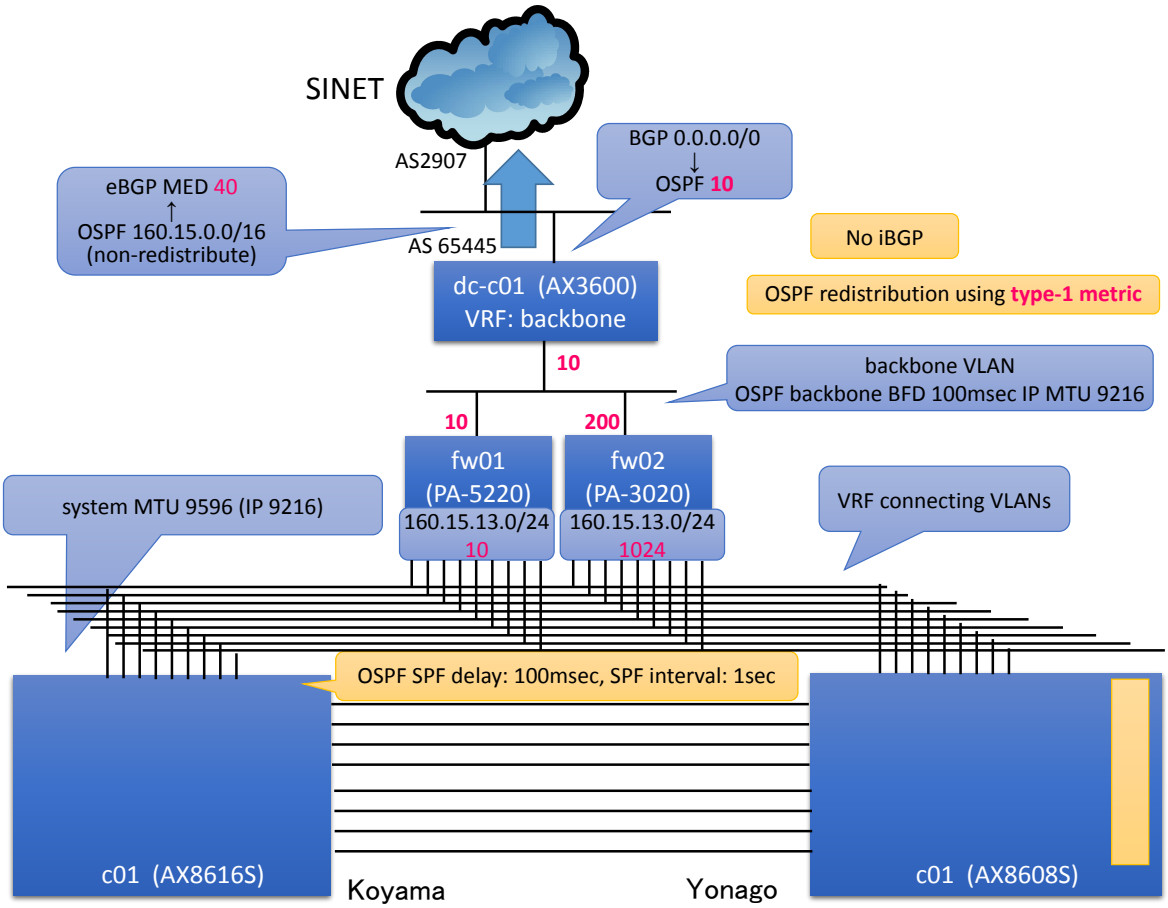


図 2　鳥取大学におけるコアネットワークの IP での冗長化

表 1　各機器の BFD の最小送信間隔，最小受信間隔

| 機種 | 最小送信/受信間隔 (msec) |
|---|---|
| AX8600 | 10 |
| PA-5220 (fw01) | 50 |
| PA-3020 (fw02) | 100 |
| AWS | 300 |

速めるため，OSPF に対して BFD を併用した．OSPF と BFD を有効にした機器で設定できる最小送信間隔，または，最小受信時間隔を表 1 に示す．表 1 から，最小送信間隔/最小受信間隔を 100msec とし，障害検出乗数を 3 とした．これにより，300msec 以上の間 BFD のメッセージが受信されなかった機器をダウンしたものと見なし，OSPF において 1 秒未満の経路収束を実現できた．

なお，キャンパスネットワークは単純な構成であることが多いため，IGP としては技術的には RIP でも十分であったが，導入ベンダでの都合により OSPF の採用となった．

## 4. 考察

### 4.1 SINET6 移行時の AWS BGP ピアリング切り替え

SINET6 移行時の様なメンテナンス作業の場合は，SINET 大阪 DC と東京 DC それぞれの回線切り替え作業に合わせて，大学側で手動により BGP ピアリングを shutdown すれば，BFD がなくとも，通信断を回避できたと考えられる．手動作業による回避も技術的には可能であったが，以下の理由から実施しなかった．

(1) 大阪 DC と東京 DC が同日に切り替えられ，前日の平日に手動で切り替えができない．
(2) 休日出勤が必要になる．
(3) 休日出勤に見合う効果が見込まれない．
(4) 手動による誤操作に起因するより長時間の通信断の可能性を拭えない．

回線切り替えを SINET DC 毎に別日に設定されていれば，切り替え前日に手動作業により BGP ピアリングを shutdown し通信断を回避できたかもしれない．しかし，それは鳥取大学での作業負担を SINET や AWS に押し付けているとも考えられ，全体としては好ましい対応とは言えない．やはり，BFD は優先度の高い，実装されるべき機能と考えられる．

### 4.2 BGP の短いホールドタイマ

BGP の生存確認メッセージの送信間隔を 1 秒にすることで，AX3660 の様に BFD を有効化できない構成の場合でも，最大約 3 秒くらいで障害を検知し，経路収束可能とすることができた．数秒であれば，利用者も気付かない程

度であると予想され，実情上は問題無いかもしれない．しかし，実時間のオンラインによるコミュニケーションでは数秒間の通信断も無視できない場合があるかもしれない．

表 1 で示した様に，AWS では BFD の送信間隔，または，受信間隔を 300msec とできる．そして，最小障害検出乗数は 3 である．このことから，BFD を併用することで，約 900msec での障害の検知を期待できる．これは，本稿で示した BGP の短いホールドタイマによる検知遅延時間の 1/3 以下である．そのため，BFD が併用できる場面では，BFD を利用した方が良いと言える．

また，BGP は TCP によりメッセージ交換するため，パケット損失が極めて短時間で発生した場合，再送により生存確認メッセージの受信が遅れてしまい，誤って障害を検知してしまう可能性もある．一方，BFD は UDP によってメッセージ交換するため，同様のパケット損失が発生しても，誤って障害を検知する可能性は BGP よりも低い．このことからも，可能であれば，BGP のホールドタイマを短くするよりも，BFD を用いた方が好ましいと言える．

## 5. 関連研究

Parker らは，リンク状態型経路制御プロトコルである IS-IS において，ハローメッセージの送信間隔を短縮する手法を提案した [4]．しかし，IS-IS のハローメッセージには双方向通信を確認する以外の大きな役割も担っているため，送受信の負荷が大きく，障害を速く検知できないという疑念を拭いきれない．

Katz らは，より軽量なハローメッセージを持つ BFD を提案した [1]．BFD のハローメッセージは双方向通信の疎通を確認する役割を担うため，ラインカード内で実装することも可能で，経路制御を担う経路制御エンジンへの負荷が軽いという利点を持つ．BFD は BGP や OSPF といった経路制御と併せて用いられ，BGP や OSPF の隣接ルータの発見と消失をより高速に検知できる．

Francois らは，IP ネットワークにおいて，IGP で達成可能な経路収束時間を明らかにした [5]．この研究は，実際のネットワークトポロジを用いてシミュレーションによって，理論的には 1 秒以下の経路収束を達成可能であると結論付けている．

## 6. おわりに

本稿では，キャンパスネットワークにおけるコアネットワークの費用対効果を考慮した冗長化手法について論じた．鳥取大学での障害事例から，鳥取の様な地方では，24 時間のオンサイト保守契約が締結されていても，通信復旧に約 8 時間を要することが明らかとなった．そのため，対外接続回線して SINET としか接続していない教育基幹においては，回線を冗長化することが必要であると考えられる．また，SINET L2VPN を利用した AWS との接続では，SINET 大阪 DC と東京 DC の 2 系統で BGP ピアリングすることが有効であった．BGP ピアリングでは，BGP と BFD を併用した障害検知が通信断の時間を短縮するのに有効であり，可能な限り BFD を併用するのが好ましいことが明らかとなった．機器の機能制限により BFD を併用できない場合であっても，BGP の生存確認メッセージの送信間隔を 1 秒にし，ホールドタイマを 3 秒とすることで，数秒以内で通信断を検知し，切り替えられることが明らかとなった．

参考文献

[1] Katz, D. and Ward, D.: Bidirectional Forwarding Detection (BFD), RFC 5880 (Proposed Standard) (2010).
[2] Katz, D. and Ward, D.: Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop), RFC 5881 (Proposed Standard) (2010).
[3] IEEE Std. 802.3ad-2000: *Local and Metropolitan Area Networks: Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications-Aggregation of Multiple Link Segments*, The IEEE Standards Association (2000).
[4] J. Parker, D. M. and Alaettinoglu, C.: *Short Adjacency Hold Times in IS-IS*, Internet draft (2001).
[5] Francois, P., Filsfils, C., Evans, J. and Bonaventure, O.: Achieving Sub-Second IGP Convergence in Large IP Networks, *SIGCOMM Comput. Commun. Rev.*, Vol. 35, No. 3, p. 35–44 (online), DOI: 10.1145/1070873.1070877 (2005).