Vol.2015-IOT-28 No.12
2015/3/5

SDDE (Software Defined Disaster Emulation) プラットフォームを用いた経路冗長化環境に対する評価実験

北口 善明[1,a)]　西内 一馬[2]　近堂 徹[3]　市川 昊平[4]　柏崎 礼生[5]　中川 郁夫[5,7]　菊池 豊[6]

概要：インターネットに代表される広域分散システムにおいては，災害や障害に対する対策としてデータ分散や冗長化による堅牢性・可用性の確保が求められる．我々は，このような耐災害性・耐障害性が十分であることを評価するために，様々な災害や故障を模倣して発生させ，その状況を検証・評価する SDDE (Software Defined Disaster Emulation) プラットフォームを提案している．本稿では，提案する SDDE プラットフォームを利用し，BGP による経路冗長化環境に対する耐障害性の評価実験を行った結果を報告する．

キーワード：耐災害性, 耐障害性, SDN, BGP

An experimental evaluation of the routing system using Software Defined Disaster Emulation (SDDE) Platform

KITAGUCHI Yoshiaki[1,a)]　NISHIUCHI Kazuma[2]　KONDO Tohru[3]　ICHIKAWA Kohei[4]
KASHIWAZAKI Hiroki[5]　NAKAGAWA Ikuo[5,7]　KIKUCHI Yutaka[6]

Abstract: In the wide-area distributed systems such as the Internet, it is important to ensure the robustness and the availability through data distribution and redundancy for disaster and failure. We propose a platform which called SDDE (Software Defined Disaster Emulation) to evaluate disaster resistance and disaster tolerance by emulating some disaster and failure. In this paper, we report an evaluation experiment of the disaster resistance for BGP routing system using our proposed SDDE platform.

Keywords: Disaster resistance, Disaster tolerance, SDN, BGP

1. はじめに

インターネットに代表される広域分散システムは，今日の情報化社会における社会基盤のひとつとして人間の生活で必要不可欠となり，高いセキュリティ・安定性を確保しつつもユーザの利便性を損なわないサービス提供が必要とされるようになった．情報システムの信頼性確保に向けた動きが活発化し，突発的な災害やそれに伴う様々な障害に対しても高い可用性を確保することが求められている．

このようなサービスは，システム構成を多重化・冗長化することでその影響を局所化し，耐災害性・耐障害性を高めることができる．しかし機器やネットワーク構成，そして電力供給が冗長化されてもその運用が適切でなければ可用性への影響は避けられない．特に平成 23 年 3 月 11 日に発生した東北地方太平洋沖地震以降，BCP (Business continuity planning; 事業継続計画) や DR (Disaster Recovery; 災害復旧) への対応が現実的な課題となっている．

内閣官房情報セキュリティセンターの調査研究「耐災害性を強化した情報システムの在り方等に関する調査」[1] で

1 金沢大学 総合メディア基盤センター
Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
2 株式会社シティネット
3 広島大学 情報メディア教育研究センター
4 奈良先端科学技術大学院大学 情報科学研究科
5 大阪大学 サイバーメディアセンター
6 高知工科大学 地域連携機構
7 株式会社インテック
a) kitaguchi@imc.kanazawa-u.ac.jp

は，災害時・障害時への対応として，平時から利用するシステムの延長として災害時のシステムを構築することへの重要性や定期的な稼働確認・災害訓練の必要性が述べられている．システム導入時の机上訓練あるいはストレステスト等のスタートアップ時の訓練だけでなく，定期的かつ継続的な訓練こそが耐災害性を確保するための重要な要素であるとも考えることができる．

本研究ではこのようなニーズに対して，災害・障害を模倣することで情報システムの耐障害性・耐災害性を評価するための SDDE （Software-Defined Disaster Emulation）プラットフォームを構築することを目的とする．本プラットフォームは，特にインターネットを基盤とした分散システムに対して，複雑に発生する障害に対する可用性を客観的な根拠を持って構築・運用することを可能にするものである．

本稿ではまず，検証プラットフォームの設計概要について示し，実装すべき災害パターンの検討と評価について報告する．2 章では本研究で考える耐災害性・耐障害性検証プラットフォームの必要性について整理し，3 章ではそれらの内容をもとに構築する SDDE プラットフォームの設計について示す．4 章において，構築したテストベッドを用いた実証実験について報告し，最後に 5 章でまとめを記す．

2. 耐災害性・耐障害性検証の背景

ベストエフォート型であるインターネットを利用した通信は，災害やそれに起因する複数の障害が発生した場合，自律的な経路制御プロトコルにより，通信可能な経路があれば当該経路を自動的に利用する仕組みになっている．しかしながら，大規模災害などの複数箇所で同時多発障害が発生した場合に，全体として機能がどのように維持されるかを実際に稼働しているネットワークシステムにおいて確認することは非常に難しい．これは，既に運用されているものに対して実際の災害を模倣することが利用者の観点から困難であること，また，マルチステークホルダに対してこのような試験・訓練への合意を取ることが難しいことなどが課題として挙げられる．

一方で，平常時から障害を想定した訓練，システム強化，リスク・コミュニケーションを図ることで，災害時の情報通信システムの継続稼働に係る円滑な対応が可能になることも知られている [1]．例えば，高知県では県内の高等学術機関 5 組織が協働し，高知 IX，高知 PoP，南国 PoP が連携した冗長構成でのネットワーク接続が行われている．ここで，地域 IX に意図的な障害を発生させることで何が起こるのかを，DR や BCP の観点で観測するネットワーク防災訓練が行われており．一定の成果が得られている [2] [3]．このような防災訓練は，システムにおける欠陥や問題点を洗い出すだけでなく，組織的な対応を含め，広い範囲

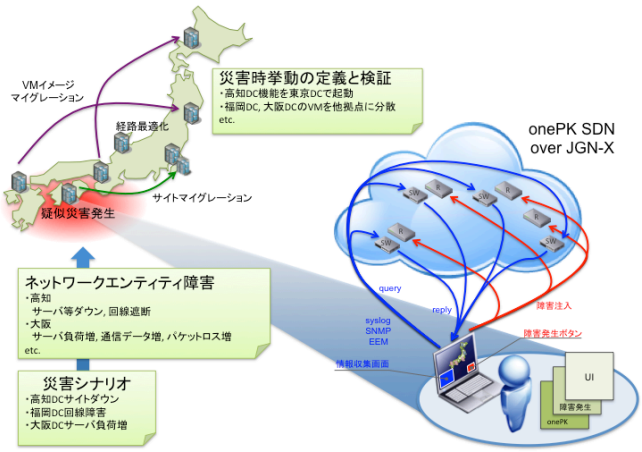

図 1 SDDE プラットフォームのイメージ図
Fig. 1 Image of the SSDE platform

での耐災害性・耐障害性を養うことができる点において，非常に有用な取り組みであると考えられる．

このように耐災害性・耐障害性検証への取り組みへの意識が高まりつつあるものの，実現に向けた技術的な課題は多い．文献 [2], [3] で示されたネットワーク防災訓練では，通信機器に手動で障害を発生させるため，障害の箇所が光のコネクタの抜差しや 1 行の CLI で表現できるような「手動かつ起こしやすい障害」に限定される点が問題となっていた．また，障害発生の前後での通信システム全体の状態把握については，ログ等の収集が機械的に迅速に行うことができず，十分な検証をすることが困難であったとの報告もされている．すなわち，ある災害シナリオに基づく再現可能な障害エミュレーションを自動的に実施する仕組みがなければ，客観的な評価・検証を行うことは難しいと考えられる．

本研究では，以上のような知見をもとに，災害から発生する多発的な障害シナリオを生成するフレームワークを備え，時間的な変化や空間的な違いを考慮した耐災害性・耐障害性のための検証プラットフォームを提案する．災害の現実に近い状況での評価が行えることで，運用システムが実用的に機能するか，組織対応として必要十分か等の検証を行うことができる．

3. SSDE プラットフォームの設計

3.1 概要

図 1 に，本研究で考える SDDE プラットフォームのイメージ図を示す．本プラットフォームで動作する検証アプリケーションは，災害シナリオの作成から，シナリオをもとにした実際のネットワーク装置（エンティティ）への障害エミュレートの投入および状態の収集までを一連の動きとして行うことを想定している．具体的には以下の機能を有し，時間的な変化あるいは空間的な変化をプログラマブルに再現することで，災害時における障害状況を模倣する．

表 1　検討中の障害パターン
Table 1　Pattern of network failure under consideration

発生区分	障害要因	具体的な症状	実装する機能
制御・運用・ソフトウェア	通信規制制御	輻輳	遅延発生＋ N%パケットロス
			トラフィックシェーピング
	不正な経路伝播	ルーティングループ	RIB/FIB 強制書き換え
		ルートフラップ	
		宛先不達（経路障害）	
ネットワーク機器	装置故障（全体）	通信断（全体）	インタフェースダウン
	装置故障（部分）	通信断（部分）	
	リソース過負荷	パケットロス	N%パケットロス
		遅延	遅延発生
通信回線	拠点間通信ケーブル断	通信断	インタフェースダウン，100%パケットロス
	中継機・交換機故障		
	トラフィック集中	輻輳	遅延発生＋ N%パケットロス
			トラフィックシェーピング
設備環境	局舎損壊（浸水・火災等を含む）	通信断（全体）	インタフェースダウン，100%パケットロス
	電源喪失		
	空調故障		
		通信断（部分）	

(1) 災害時の災害シナリオからネットワーク障害を生成
(2) 実ネットワークにおける災害のエミュレートを実現
(3) 障害状態解除により正常状態への回復を容易に実施
(4) 障害発生前後におけるネットワーク状態の自動収集

次節以降では，上記のうち災害・障害パターンの検討と実装について述べる．

3.2　障害パターンの分類

災害・障害パターンには，自然災害に起因する障害から装置故障等に起因する障害まで，様々な要因が考えられる．また，影響範囲がどの程度か，あるいは空間的な変化を伴うか否か，時間的な推移をどのように表現するか，なども検討が必要となる．本節では，総務省「大規模災害等の緊急事態における通信 確保の在り方に関する検討会」[4] で示されている災害事象や「情報通信ネットワーク安全・信頼性基準」[5] 等の内容をもとに，災害時における通信設備等に対する障害に焦点を絞り，各事象に対してネットワーク装置に適用する制御について検討する．

現在検討中の分類表を表 1 に示す．通信障害は主に 2 つの原因に大別されると考える．ひとつはトラフィック集中による輻輳，もうひとつは回線や機器等のハードウェア・設備障害である．表 1 はこれらを細かな区分に分類し，それぞれの障害要因と具体的な症状の対応付けを行い，各々について本プラットフォームで実装する機能についてまとめたものである．このように，例えば通信断であっても，ネットワーク機器自体が故障する場合と中継機・交換機が故障する場合では，実際の通信でみられる症状が異なることが予想される．

3.3　SDN による制御

前節で述べたような災害・障害パターンを実ネットワーク上で実現するためには，様々な技術的要件が求められる．特に，障害を模倣したネットワークを確実に元の正常状態に戻すことができること，障害の模倣はソフトウェアで実装し自動化できること，という点が重要となる．そこで本研究では，これらの要件を満たすために SDN（Software-Defined Networking）技術を活用する．

本プラットフォームでは，3.1 節で示した機能をレイヤ構造で実現する．上位層から，システム管理者や災害訓練実施者による災害・障害シナリオの入力が行われ，それに基づきネットワークエンティティに対する障害パターンが構成される．シナリオの入力には API を用意し，形式的記述言語での表現を現在検討している．実際のネットワーク機器への注入や解除は，いくつかの API を用意することを検討しているが，今回は Cisco Systems 社が提供する SDN プラットフォームである onePK（One Platform Kit）を利用する．このキットは，同社のオペレーティングシステムである Cisco IOS 等に対して，C，Java，Python といった言語で操作できる API が提供され，ネットワーク管理者は CLI を想定したパーサーを用意することなく，API を使って柔軟にネットワーク機器を制御することが可能となる．

制御インタフェースとしては onePK のほか，従来より一般的に利用される SNMP[6] や NETCONF[7] などが存在するが，これらについての対応も検討しながら，OpenDaylight[*1] 等の SDN コントローラとの連携も視野に入れつつ実装を進めていく．

また，onePK には，Cisco IOS に搭載されているイベント

[*1] http://www.opendaylight.org/

図 2 検証プラットフォームの論理パス構成
Fig. 2 logical topology of the evaluation platform

マネージャ機能である EEM（Embedded Event Manager）への API も提供されており，検証プラットフォームにおけるネットワークエンティティの統合的な状態収集を実現する．EEM では SNMP や Syslog に代表される様々なイベント検出具術を扱う事ができるため，ネットワーク内に発生している状態を時系列で扱える環境を合わせて構築する予定である．

このように，SDDE プラットフォームでは，災害シナリオの入力からネットワーク機器への反映および情報収集までを自動的かつソフトウェアで行うことで，実環境でプログラマブルかつ再現可能な災害・障害エミュレーションを提供することが可能となる．

3.4 検証プラットフォームの構築

図 2 に，今回構築した検証プラットフォーム構成を示す．このプラットフォームは，JGN-X[*2] 上に配置された onePK 対応ルータを利用して構築した．現在，札幌，仙台，大手町，名古屋，大阪，岡山，広島，福岡の各アクセスポイント（AP）で利用可能である．配置されている onePK 対応ルータの構成は拠点毎で異なっており，大手町 AP および大阪 AP の 2 カ所が Cisco ASR9006，その他の AP では Cisco ASR1004 となっている．このような JGN-X の環境を利用し，全国 5 拠点のユーザセグメント（広島大，高知工科大，大阪大，奈良先端大，金沢大）と onePK 対応ルータを結ぶ論理パスをそれぞれ構築し，図 2 に示すように複数の論理パスで各ルータを結ぶネットワークとして構築した．金沢大のみ JGN-X への直接の接続性を持たないため，SINET4[*3] による L2VPN サービスを経由した接続形態としている．

今回のプラットフォームではユーザセグメント以外に管理セグメントも用意した．これは，onePK の制御用セグメントでもあり，オペレータが操作する制御用サーバから API 経由で各ルータと制御メッセージがやり取りされる．本セグメントはフラットな L2 で構成されており，サーバから各ルータに対して API 接続する際は TLS 認証が必須となる．なお，制御用サーバは高知工科大に設置しており，ここから検証プラットフォーム全体の制御を行う構成としている．

*2 http://www.jgn.nict.go.jp/

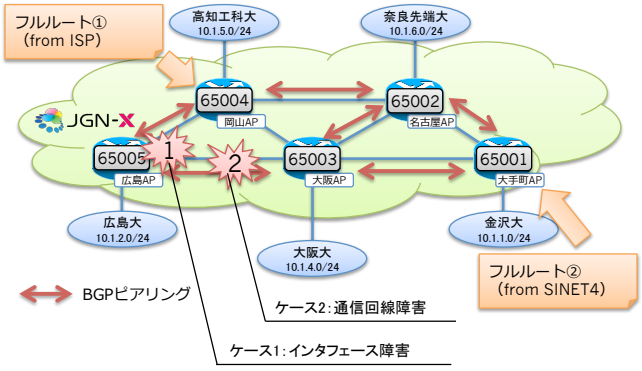

図 3 BGP 接続トポロジ
Fig. 3 BGP peering topology

表 2 BGP ピアリング設定のパラメータ
Table 2 Configuration parameter for BGP peering

設定項目	設定値
Keepalive パケット送信間隔（Hello Interval）	10 秒
待機時間（Hold Time）	30 秒
fast-external-fallover	有効
NSF/SSO/BFD	利用しない
MD5 認証	利用しない

4. 評価実験

本章では，構築した検証プラットフォームを用い，評価アプリケーションとして BGP の経路制御システムを用いた評価について述べる．

4.1 実験概要

図 3 に，評価に用いる BGP による経路冗長化環境を示す．構築した検証プラットフォーム上において，プライベート AS 番号を利用した eBGP ピアリング構成としている．また，BGP におけるフルルート情報の影響を評価するため，高知工科大学（ナインレイヤーズ [*4]）と金沢大学から外部のフルルート情報を注入可能な構成にしている．

*3 http://www.sinet.ad.jp/
*4 http://ix-layers.com/

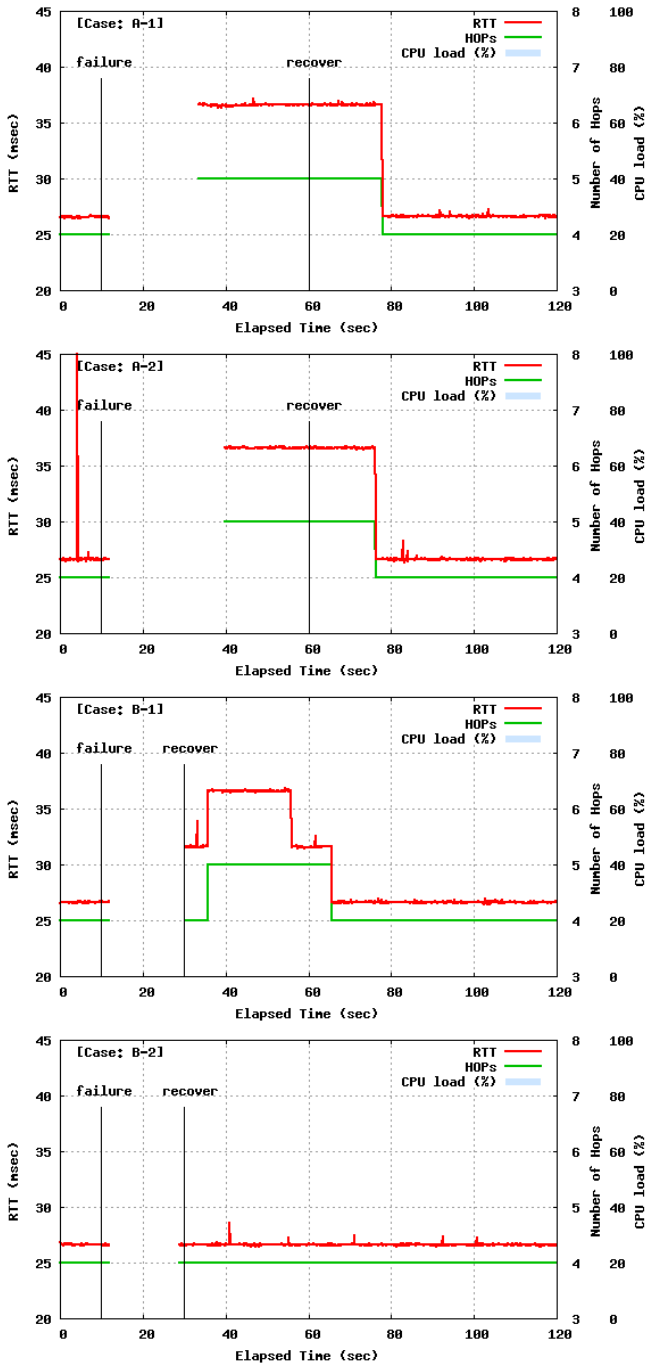

図 4 実験結果：フルルートなし
Fig. 4 Result of evaluation without full route

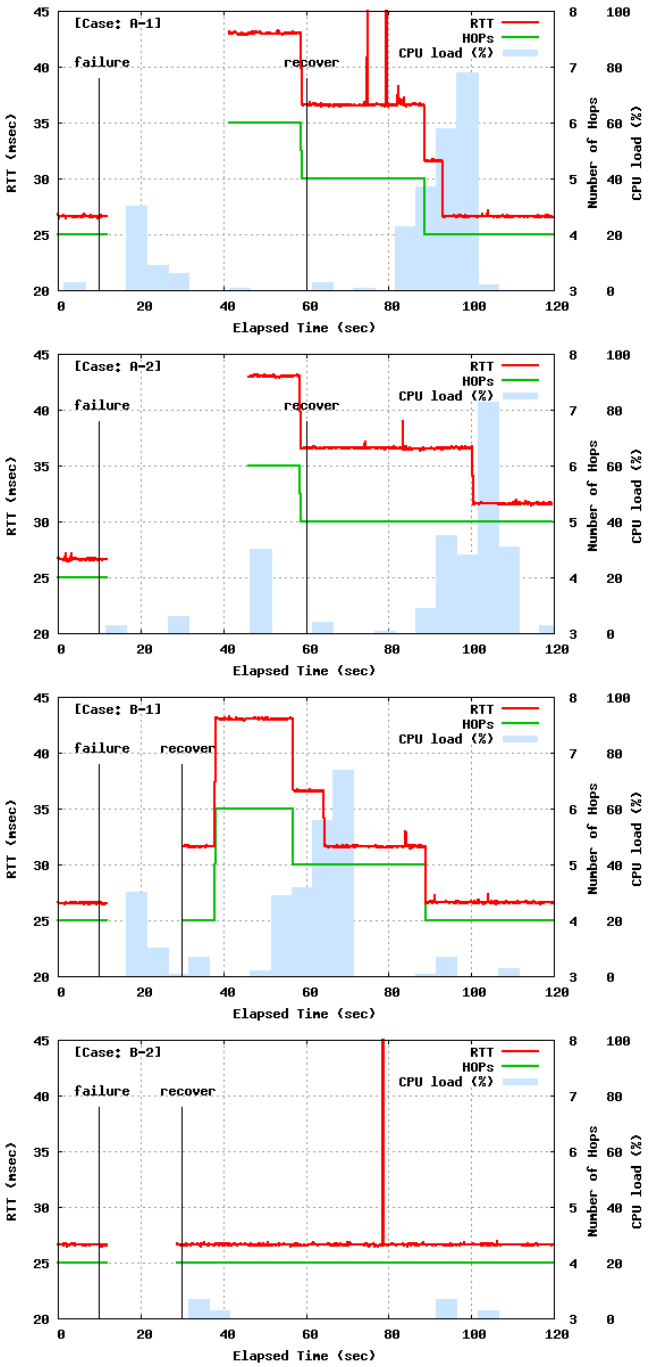

図 5 実験結果：フルルートあり
Fig. 5 Result of evaluation with full route

各 AS 間の基本的な BGP 設定を表 2 に示す．BGP 各種タイマの一般的な値としては，Hello interval: 30 sec, Hold Time: 90 sec が多いようであるが [*5]，検証時間の都合上，今回はそれぞれ 1/3 の値に設定して検証した．また，ネットワーク障害時におけるダウンタイムを小さくする技術である NSF（Non-Stop Forwarding）や SSO（Stateful Switch Over），BFD（Bidirectional Forwarding Detection）などの技術は利用しておらず，MD5 による認証も実施していない．

[*5] http://www.janog.gr.jp/doc/janog-comment/bcop-ebgp.txt

BGP の経路冗長化に対する障害時の挙動確認には，障害復旧のタイミングを設定する BGP タイマによる影響を見るために，次の二パターンで実施する．

- ケース A：60 秒後に復旧（Hold Time より大きな値
- ケース B：20 秒後に復旧（Hello Interval と Hold Time の間の値）

また，発生させる障害としては，現時点までで開発が完了している以下の障害発生インタフェースを利用して，BGP に与える影響の違いを観測する．

- ケース 1：インタフェース障害（リンクダウン）
- ケース 2：通信回線障害（100%パケットロス）

さらには経路収束の影響を見るために，フルルートありとなしの場合で評価する．

評価は，広島大のユーザセグメントから金沢大のユーザセグメントまでの遅延時間とホップ数の変化をそれぞれ計測し，経路変更が行われるまでの時間やネットワークのダウンタイムを観測した．経路収束の影響として，広島 AP に設置してある ASR の CPU load 値を観測し，コンバージェンス時間の評価に活用する．CPU load 値の収集は，最終的には EEM 等による自動収集を目指しているが，今回は CLI による取得を行っている．

4.2 実験結果

図 4 にフルルートがない状況での結果を，図 5 に二系統のフルルートを注入した場合の結果をそれぞれ示す．各グラフは上から『ケース A-1，ケース A-2，ケース B-1，ケース B-2』の順で掲示してあり，通信遅延の変化を赤線，ホップ数の変化を緑線，広島 AP の ASR における CPU load を青棒でそれぞれ描画している．すべての障害ケースにおいて，計測開始 10 秒後に障害を発生させ，指定の時間後に復旧させている．

障害ケース A では，いずれも大きなダウンタイムが発生していることが分かる．インタフェース障害と通信回線障害において通信再開までの差が数秒程度であるが見られる．障害ケース B では，復旧と同時に通信が回復しているが，インタフェースダウン（障害ケース 1）では経路の再計算が障害発生直後に開始されていて，経路が不安定になっている様子が観測された．

また，ネットワーク負荷として導入したフルルートの影響は，若干のダウンタイムが延びることよりも，経路再構成の伴うネットワークの不安定さに大きく現れることが見て取れた．

4.3 考察

今回の評価により，通信断という同じ障害においても，障害要因によりネットワークシステムに与える影響が異なることを明確に示すことができたと言える．Hold Time を小さくすることで，ネットワークのダウンタイムを短くすることが可能であるが，短い期間の障害においては経路再構成により通信が不安定になることが示された．ネットワーク障害が発生した際には，ダウンタイムを小さくすることが求められ，Hello interval や Hold Time の調整だけでは難しいと考えられる．提案する検証プラットフォームを利用する事で，前述した BFD のような機能による効果も評価できると考えており，実際に適用した評価を今後実施したいと考えている．

5. まとめ

本稿では，広域分散システムに対する災害などによる障害の影響を評価するプラットフォームの設計について述べ，BGP による経路冗長化環境に対する評価実験の結果を報告した．発生させるネットワーク障害ケースにより，システムの挙動がことなる点を確認することができ，本プラットフォームによる災害の模倣による評価は有用であると考える．

現在開発中の検証プラットフォームでは，今回利用した 2 つのネットワーク障害のほかにも表 1 に挙げたものを順次実装中である．今後は，災害シナリオに基づいたネットワーク障害の適用について検討し，複数の障害が同時多発的に発生する場合や，段階を追って影響範囲が拡大するような場合での評価検証が可能であることを示す予定である．

謝辞 本研究は，総務省戦略的情報通信研究開発推進事業（SCOPE）先進的通信アプリケーション開発推進型研究開発「分散システムの耐災害性・耐障害性の検証・評価・反映を行うプラットフォームとビジネスモデルの開発（140201003）」の助成を受けています．また，日本学術振興会産学協力研究委員会インターネット技術第 163 委員会（ITRC）地域間インタークラウド分科会（RICC），コンピュータリソースのご提供をいただいた各大学，SINET4 の回線をご提供頂いた国立情報学研究所，および JGN-X の回線をご提供頂いた情報通信研究機構の関係者各位に感謝致します．

参考文献

[1] 内閣官房セキュリティセンター. 耐災害性を強化した情報システムの在り方等に関する調査. http://www.nisc.go.jp/inquiry/pdf/taisaigaisei_gaiyou.pdf(参照 2015-01-30).
[2] 岡村健志, 菊池豊, 福本昌弘, 豊永昌彦, 佐々木正人, 今井一雅, 山田覚, 風間裕, 一色健司, 名和真一, 高畑貴志. 地域 ix における人為的障害による耐災害性の検証. マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム, pp. 485–489, July 2014.
[3] 佐々木正人, 豊永昌彦. 回線切断を含むネットワーク防災訓練に関する報告. 大学 ICT 推進協議会 2014 年度年次大会論文集, November 2014.
[4] 総務省. 大規模災害等緊急事態における通信確保の在り方に関する検討会. http://www.soumu.go.jp/main_sosiki/kenkyu/saigai/(参照 2015-01-30).
[5] 総務省. 情報通信ネットワーク安全・信頼性基準. http://www.soumu.go.jp/manu_seisaku/ictseisaku/net_anzen/anshin/(参照 2015-01-30).
[6] D. Harrington, R. Presuhn, and B. Wijnen. An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. RFC 3411 (INTERNET STANDARD), December 2002. Updated by RFCs 5343, 5590.
[7] R. Enns, M. Bjorklund, J. Schoenwaelder, and A. Bierman. Network Configuration Protocol (NETCONF). RFC 6241 (Proposed Standard), June 2011.