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の設計と評価

柏崎 礼生[1,a)] 西内 一馬[3,b)] 北口 善明[2,c)] 市川 昊平[4,d)] 近堂 徹[5,e)] 中川 郁夫[1,6,f)]
菊地 豊[7,g)]

概要：ICT 環境は生活にも業務にも不可欠になっており，ICT システムの堅牢性の確保が重要になっている．ICT システムを堅牢にするために，冗長性の増加や広域分散による単一障害点の回避手法がある．しかしながら，東南海大地震等の災害では同時多発的に障害が発生し，通常想定する範囲内に収まらない障害を想定しないとならない．本稿では，被災状況を模倣することで ICT システムの耐災害性・耐障害性を評価するプラットフォームを設計し，これを評価する．また本プラットフォームがより実践的な災害・障害対策に寄与するため，分散システムの罹災規模をトポロジとリスク評価の観点からの分類を提案する．

A design and evaluations of a platform that examine, evaluate, and reflect a disaster-resistance and a disaster tolerance on a distributed system.

Hiroki Kashiwazaki[1,a)] Kazuma Nishiuchi[3,b)] Yoshiaki Kitaguchi[2,c)] Kouhei Ichikawa[4,d)]
Tohru Kondou[5,e)] Ikuo Nakagawa[1,6,f)] Yutaka Kikuchi[7,g)]

Abstract: ICT environment is requisite for our life and business. It is important to ensure toughness and robustness of ICT systems. To ensure toughness and robustness, there are several solution such as redundant configurations, SPoF (single point of failure) less wide-are distribution. Nowaday, we must design our system to endure beyond the scope of assumption of disaster caused by southeaster sea (Tounankai) earthquake. In this paper, authors design and evaluate a platform to evaluate disaster tolerance of ICT system by immitating existing disaster records. To contribute to make countermeasures against disaster and disorder more practical, they propose classification from a point of view of a topology and risk evaluations.

Keywords: disaster recovery, disaster drill, distributed system

1. はじめに

分散システムとはネットワーク上に配置された計算機が互いにメッセージのやりとりによって通信し，連携するソフトウェアシステムである [1]．具体的な例として電話網や携帯電話網，インターネット，ワイヤレスセンサーネットワーク，WWW，P2P ネットワークなどが挙げられ，現

[1] 大阪大学
Osaka University
[2] 金沢大学
Kanazawa University
[3] 株式会社シティネット
City Net Inc.
[4] 奈良先端科学技術大学院大学
Nara Institute of Information Science and Technology
[5] 広島大学
Hiroshima University
[6] 株式会社インテック
Intec Inc.
[7] 高知工科大学
Kochi Insitute of Technology
a) reo@cmc.osaka-u.ac.jp
b) nishiuchi@city-net.jp
c) kitaguchi@imc.kanazawa-u.ac.jp
d) ichikawa@is.naist.jp
e) tkondo@hiroshima-u.ac.jp
f) ikuo@cmc.osaka-u.ac.jp
g) yu@kikuken.org

在の生活に必要不可欠な基盤を支えるシステムとなっている．このシステムの価値を，あるいはこのシステムの上で動作するサービスの価値を，利用者が高く評価すればするほどに，また利用者がこのシステムに依存すればするほどに，サービスの停止時間はより小さく運用されることが期待される．このような期待を寄せられたサービスは，サービスが獲得する利益を越えない範疇の設備投資で，その構成機器が冗長化され，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耐災害性・耐障害性は高まらない．

2011 年 3 月 11 日に起きた東北地方太平洋沖地震は福島第一原子力発電所の爆発事故を招いた．日本の原子力施設の緊急事態においては，緊急時迅速放射能影響予測ネットワークシステム (System for Prediction of Environmental Emergency Dose Information: SPEEDI) が放射性物質の拡散範囲を計算し，その影響予測を行う．しかしこの爆発事故においては，緊急時対策支援システム (Emergency Response Support System: ERSS) からの放出源情報が得られなかったため，SPEEDI 計算を行うことができなかった．ERSS が放出源情報を SPEEDI に提供できなかったのは，ERSS に原子炉内の情報等を提供する東京電力の緊急時対応情報表示システム (Safety Parameter Display System: SPDS) がデータを伝送できなかったことに起因する．

SPDS から ERSS へのデータ伝送は，福島第一原子力発電所からオフサイトセンターを経由して ERSS の計算機にデータを伝送する政府専用回線 (統合原子力防災ネットワーク) を用いて行われる．オフサイトセンターの免震重要棟に配置された SPDS サーバは，複数のネットワーク機器を経由し，研修棟の保安検査官室を経て統合原子力防災ネットワークへと接続されている．SPDS から統合原子力防災ネットワークに至る途中の研修棟保安検査官室には非常用発電機等が存在せず，その代わりに無停電電源装置 (Uninterruptible Power Supply: UPS) が中継機器とメディアコンバータに電力を供給していた．しかし接続されているはずの UPS とメディアコンバータは接続されておらず，地震発生から早い時点で機能を停止した [*1] (図 1)．

耐災害性・耐障害性を上げるための訓練は人間のコミュニティにおいて多く実践され，高い効果を上げている．岩手県釜石市では津波災害史研究家の山下文男が「津波てんでんこ」と呼ばれる防災標語を広めた．これは「津波が来たら，取る物も取り敢えず，肉親にも構わずに，各自てんでんばらばらに一人で高台へと逃げろ」の意味である [2]．この防災意識が浸透した結果，東北地方太平洋沖地震とこの地震に起因する津波に対して釜石市の小中学校 (児童数約 3,000 人) における生存率は 99.8%，釜石市の児童数の

[*1] 東京電力福島原子力発電所における事故調査・検証委員会最終報告 (本文編)IV 章 2(1) より

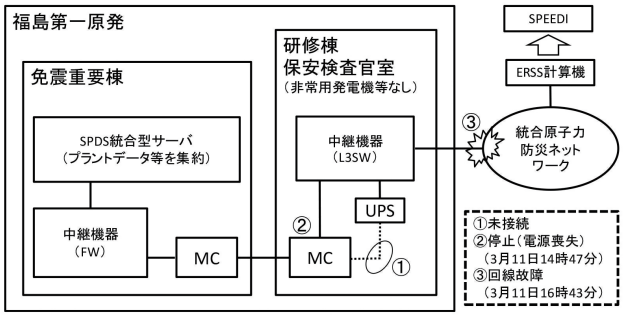

図 1　SPDS データの ERSS 回線への伝送状況
Fig. 1　A transmission diagram of SPDS data to ERSS line

二割が生活する鵜住居町における生存率は 100%であった．防災教育が極めて高い耐障害性を示したこの事例は「釜石の奇跡」として国内外のメディアで報じられた (図 2)．

図 2　NHK「釜石の “奇跡” いのちを守る　特別授業」の 1 シーン
Fig. 2　A screen capture from NHK “Kamaishi “Miracle” Save lives. Special lecture”

一方，訓練が適切に行われなかったことにより耐災害性・耐障害性が損なわれた事例もある．前述の釜石市鵜住居町では「釜石の悲劇」と呼ばれる事例が生じた．鵜住居地区では避難訓練への参加率を上げるため，本来の避難場所ではなく屋根のある釜石市鵜住居地区防災センターを避難訓練に利用した．この防災センターは平成 22 年に開所したばかりの標高 4.3m，最寄りの海岸線まで約 1.2km の距離にある鉄筋コンクリート造 2 階建の施設である．市指定津波避難場所は市街地から 1km ほどの距離があり，標高も 30m〜50m の地点にあった．この結果，市民には「避難場所」が防災センターであるとの認識が広がり，東日本大震災直前の 2011 年 3 月 9 日に起きた三陸沖地震の発生時においても防災センターに避難した市民がいた．このような事態を受けても市は住民に対して防災センターが一次避難場所ではないことの告知を行わなかった．この結果，東北地方太平洋沖地震とこの地震に起因する津波に対して防災センターで 69 人が遺体として収容された．また 200 人以

上が防災センターに避難していたという情報もある [*2].

2. ネットワーク防災訓練

2012 年 8 月に内閣府が発表した「南海トラフの巨大地震による津波高・浸水域等 (第二次報告) 及び被害想定 (第一次報告)」によると，想定されるケース「『四国沖』に『大すべり域+超大すべり域』を設定」および「『四国沖〜九州沖』に『大すべり域+超大すべり域』を設定」において高知県幡多郡黒潮町および土佐清水市は最大津波高 (満潮位・地殻変動考慮) において国内最大の 34m と推定されている [*3]．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 を連携したさせた冗長構成でインターネットや相互接続を実現している (図 3)．既存の ICT システムに意図的に障害を起こすことにより，ICT システムの冗長性や障害対策の機能および ICT 関係者間での連絡体制等を確認・検討し，実際に機能する事業継続計画 (Business Continuity Plan: BCP) を策定することを目的として，TEReCo4[*4] プロジェクトは 2013 年度にネットワーク防災訓練を行った [*5]．この取り組みでは様々な障害要因をロジックモデル手法で可視化しており，3 つの障害パターンでの検証が行われた．その結果，本来不通になるはずの障害パターンにおいてもインターネットへの導通が確認されたことにより，運用者が把握していない冗長構成の存在が発覚するなど，防災訓練を行うことで，耐災害性・耐障害性を向上させるのみならず，本来の目的以外の効果が現れた事例の報告が行われている [3, 4].

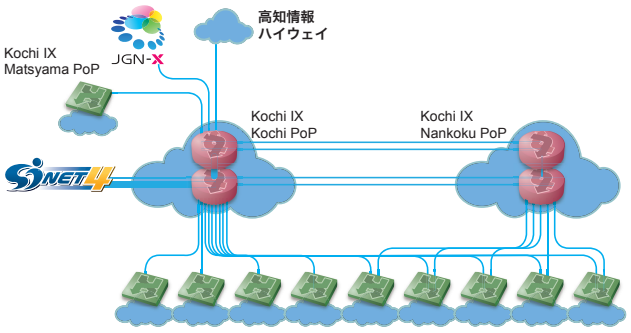

図 3　高知 IX を中心としたネットワークトポロジ
Fig. 3　A network topology centered around Kochi IX.

このような活動は，主に以下を確認し，課題がある場合にはそれを改善するための材料とすることを目的として行われている．

- ICT システムの障害に対する機能が正しく機能するか
  - 冗長性によって機能維持をする場合には，予備系が正しく機能するか
- ICT システムの障害発生時に，システムが障害を通知し，管理者が正しく障害を認識し，原因特定・復旧にいたる手段を実施できるか
  - 直ちに復旧できない場合に代替手段を講じるようになっている場合には，それを実施できるか

これまで ICT システム管理者を対象としてこのようなネットワーク防災訓練の実現が提案されてきたが [*6]，その反応は，活動の意義や効果は認めるものの，実施することに対する困難が大きいというものであった．また，ユーザが 365 日 24 時間利用しうるため，機能停止を伴う活動が困難であるという声も大きかった．ネットワーク防災訓練を実際に実施することで，ネットワーク防災訓練の有効性を実証することができ，また，ネットワーク防災訓練の方法論を確立して実施の手続きを整備し，小さな手間で実施することが可能となる．高圧以上の受電をしている組織が，電力設備の法定点検を義務付けられているのと同様に，一定以上規模の ICT システムを運用している組織に対する法令による点検義務が課せられるようになれば，特に地震の発生件数の多い環太平洋地域の国々において有益であると考えられる (図 4[*7]) .

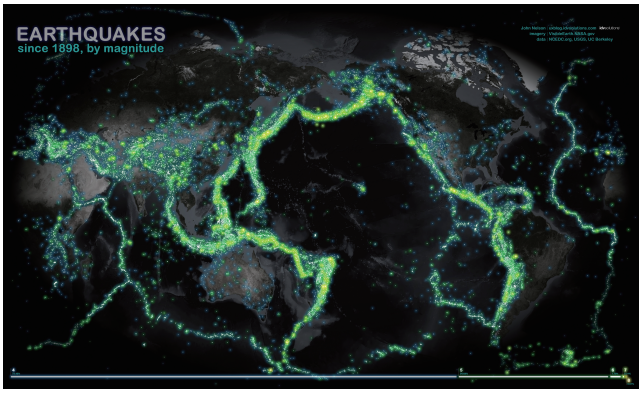

図 4　1898 年以来発生した地震の分布 (マグニチュード順)
Fig. 4　A distribution of earthquakes since 1989, by magnitude

情報システムの中でも特に分散システムは防災訓練を行うために複数のステイクホルダーの了承を得ることが求められるために困難であったが，それと同時に高い耐災害性・耐障害性が求められるシステムでもある．そこで我々は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を開発する．

[*2] 釜石市鵜住居地区防災センターにおける東日本大震災津波被災調査報告書 http://www.city.kamaishi.iwate.jp/index.cfm/6,28416,c,html/28416/20140312-130741.pdf より
[*3] 南海トラフの巨大地震による津波高・浸水域等 (第二次報告) 及び被害想定 (第一次報告) 資料 1-2 都府県別市町村別最大津波高一覧表＜満潮位＞ http://www.bousai.go.jp/jishin/nankai/taisaku/pdf/1_2.pdf
[*4] Traffic Engineering for Regional Communities, version 4
[*5] 福本昌弘ら「災害時に事業継続性を発揮する情報通信インフラのための運用計画改善手法および冗長化技術の研究開発」総務省地域 ICT 振興型研究開発 (平成 25 年度〜26 年度, 四国総合通信局)http://www.soumu.go.jp/main_content/000284013.pdf
[*6] 菊池豊「防災訓練! 本当に切るとしたら何処を切りたい何を知りたい?」裏ジャノ 2013@ ミクシィ，など.
[*7] IDV soltions “Earthquakes since 1898” http://uxblog.idvsolutions.com/2012/06/earthquakes-since-1898.html”

3. SDN による実装

既に述べた高知県での取り組みにおいてはネットワーク障害の発生は人為的に，かつ実際に人間の手による手動操作で行われたものである．しかし例えば組織内においてネットワークトラフィックの少ない時間帯である深夜から朝にかけての時間帯において障害を発生させようとすると，必然的に人力で訓練のための障害を発生させるコストを要する．前述の取り組みでも障害がもたらす影響を加味して，1 月 5 日の午前 5 時から訓練を開始しているため，定期的に行うことは困難であることが考えられる．我々は，この訓練が定常的に高い頻度で，しかも多様な障害で行われることが必要であると考えている．人力で行うためには多様な障害シナリオを記述するコストが必要であり，なおかつ高い頻度で障害を発生させ，その評価を行い，検証するコストを算出すると，人力で実現することは現実的ではない．また先の検証においては複数箇所で同時に発生する障害を人為的に作り出すことや，障害後のネットワーク情報を収集すること，そして障害発生後に元の状態へ戻すことの困難さを明らかにしている．そこで本稿で扱う検証・評価・反映を行う機構は，障害を形式的に記述し，かつ障害時にネットワーク情報を収集することが可能であり，評価後にネットワーク状態を元に戻すことが可能であることが要件となる．また自動化により運用コストを低減させるため，これらがソフトウェアで実装されることが重要である．

自動化の実現にはネットワーク機器の制御をソフトウェアで行う必要がある．Open Network Foundation (ONF) による White Paper“Software-Defined Networking: The New Norm for Networks”[*8] では Software Defined Network (SDN) を “低レイヤーの機能を抽象化することでネットワーク管理者にネットワークサービスの管理を委託するコンピュータネットワーキング” として定義している．我々の開発が必要としているものはこれに合致する．現在，SDN といえば OpenFlow といった風潮があるが [5] トラフィックの制御をすることのみが SDN でないことは ONF の White Paper にも記述されている．我々が開発しようとしているソフトウェアは OpenFlow のようなトラフィックの制御のみでは実現することが困難である．

OpenFlow 以外の SDN 実装で我々の開発の要求に合致するものとして SNMP や NETCONF[*9] が挙げられる [6,7]．SNMP は RFC3411，NETCONF は RFC6241 としてインターネット標準になっており，ネットワーク機器に対して設定を投入し，操作，削除をする機構を提供するプロトコルである．SNMP は管理情報ベース (Management Information Base: MIB) をもとに機器の制御を行うため運用が複雑になり，現在は専らネットワーク機器の情報を収集するプロトコルとして利用されている．一方で NETCONF には取り立てて欠点も見当たらないが，使われている事例というものがあまり見当たらない理由として，ツールの決定的な不足を挙げることができる．2000 年代前半に提案されたこともあり，XML 幻想に囚われているのかもしれないが，XML は人間が書くようなものではないので何らかのツールが必要となる．しかしそのツールがなければツールを作るほかない．そのツールを作るコスト，作る余裕はネットワーク管理者にはないのであろう．

とはいえ，大量のネットワーク機器を遠隔で制御し，設定情報を一括管理し，なおかつセキュアであるツールのニーズは消え失せたわけではなく，当時と変わらず今もなお確実に存在している．そこで Cisco Systems 社は 2012 年に “one Platform Kit” (onePK) をリリースした．このキットは C，Java，Python で操作できる API として提供されており，ネットワーク管理者はそのプロトコルに対応するためのパーサーを書くことに煩わされず，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である．本開発では Cisco Systems 社の協力のもと，onePK を用いた実装を進め，今後の余裕次第では NETCONF や，他の onePK, NETCONF 非対応ネットワーク機器についても CLI によるコンフィギュレーションを行う API を整備することで，本来あるべき広義の SDN を用いた検証・評価・反映を行うプラットフォームを実現する．

3.1 Cisco onePK

前述のように，Cisco Systems 社が開発した onePK は開発，自動化，迅速なサービス構築などを可能にするツールキットとして登場した．現在，onePK に対応した Cisco 製品はまだ限定的だが (ASR シリーズほか)，今後は全ての Cisco 製品に対応するロードマップとなっている．onePK はネットワークの機能制御を行うだけでなく，SNMP と同様にネットワーク機器の情報の取得を行うことも可能である．また onePK は経路制御テーブルを操作することが可能であるため，OpenFlow と同様にフローの制御を実現することができる．ネットワークに限らずあらゆる管理業務において，時間経過とともに自動化可能なタスクと自動化不可能なタスクとに分類され，自動化可能なタスクはソフトウェアで実装され，管理者の自由な時間が増大する．一方，自動化不可能なタスクが減少するので管理者はより自由な時間を獲得できるように思われるが，実際には顧客からさらなる無理難題を要求されることにより，高度な制御を求められることになる．そのため，単純な自動化で対処することが可能なタスクは枯渇し，より複雑な自動化を求められる．そのため計算機言語を用いた制御によりネットワーク制御を行うことが今後より強く要請されることにな

[*8] https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf
[*9] http://datatracker.ietf.org/wg/netconf/charter/

ることが予想される．onePK はこのような，ネットワーク管理者にとって切ない時代の要請に応じたツールキットであると言える．

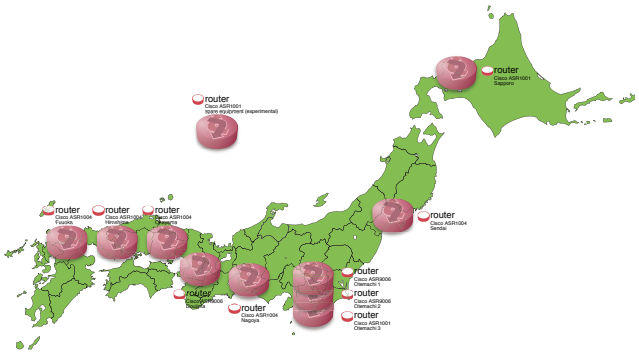

図 5　JGN-X 上の onePK リソース
Fig. 5　onePK resouces on JGN-X

onePK Software Developer Kit (SDK) は 2014 年 8 月に Version 1.3 がリリースされた [*10]．国内においては，onePK 対応ルータは新世代通信網テストベッド JGN-X[*11] 上に配置されており，現在札幌，仙台，東京，名古屋，大阪，岡山，広島，福岡の各アクセスポイントで利用することが可能である．東京と大阪に設置された ASR9006 は 10GbE，東京に配置された 2 台のうち 1 台は 40GbE/100GbE 対応である．札幌以外の拠点に配置された ASR1004 は 10GbE に対応しており，札幌に配置された ASR1001 は GbE 対応となっている (図 5)．既にこの機器を利用して，さっぽろ雪祭りの映像伝送実験が 2013 年に行われている (図 6)．

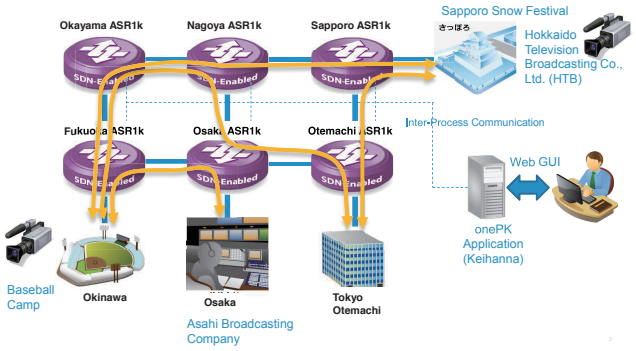

図 6　NICT/JGN-X テストベッドを用いたさっぽろ雪祭り伝送実験の模式図 (シスコシステムズ合同会社提供)
Fig. 6　A diagram of NICT/JGN-X Testbed Experimental demonstration in Sapporo Snow Festival (according to Cisco Systems Inc.)

現在，onePK は C，Java，Python で API が提供されているが，全て同じ機能を使えるわけではない．Version 1.3 ではかなりの改善が行われ，Data path サービスセット以外ではほぼ全ての言語で全ての API がサポートされている．表 1 に Data path サービスセットにおける言語による API サポートの差異を示す．

表 1　onePK における言語によるサポートの差異
Table 1　differences of support by language on onePK

API 名	Python	C	Java
Dpass Aplly Qos Cir Cbs	×	○	×
Dpss Assign Egress	×	○	×
Dpss Assign Next Hop	×	○	×
Dpss Bypass Flow	×	○	×
Dpss Bypass Flow Until	×	○	×
Dpss Bypass Flow	×	○	×
Dpss Drop Packet	×	○	×

表に示すまでもなく C でのサポートが最優先され，その後に Java，Python のサポートが行われる．C 以外の言語で記述するときはラッパーを書くか，バージョンアップで対応されることを待つのが賢明である．

4.　検証環境

本開発は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であるため，実際に分散システムに対して防災訓練を行う必要がある．国内で研究用途で利用することが可能な分散システムとして前述の JGN-X があるが，JGN-X は日本国内の広域に分散して配置されているため，定常的な障害発生を繰り返す事を何の調整を行う事は困難である．そこで JGN-X および学術情報ネットワーク SINET4 上に展開されるオーバーレイフレームワークである広域分散仮想化基盤 “distcloud” を対象として検証を行うこととした．

4.1　distcloud

distcloud は，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) の地域間インタークラウド分科会 (RICC) を中心として研究開発が行われている広域分散仮想化基盤である [8]．災害回復の手法としてのライブマイグレーションに着目し，広域ライブマイグレーションを行ってもその前後で IO パフォーマンスの劣化の少ないストレージ技術を実現している．2014 年 9 月現在，広島大学，金沢大学，国立情報学研究所，および京都大学からなる 4 拠点で構成されており，2014 年度中に奈良先端科学技術大学院大学，大阪大学，北海道大学，高知工科大学が新たに接続される予定となっており，そのほか北陸先端科学技術大学院大学，九州大学，九州工業大学などが接続を予定している (図 8)．これまでは SINET4 のみからなるネットワーク構成であったが，2013 年 11 月に米コロラド州デンバーで開催された国際会議 SuperComputing2013[*12] で

[*10] SDK がダウンロード可能になったのは 9 月．
[*11] http://www.jgn.nict.go.jp/
[*12] http://sc13.supercomputing.org/

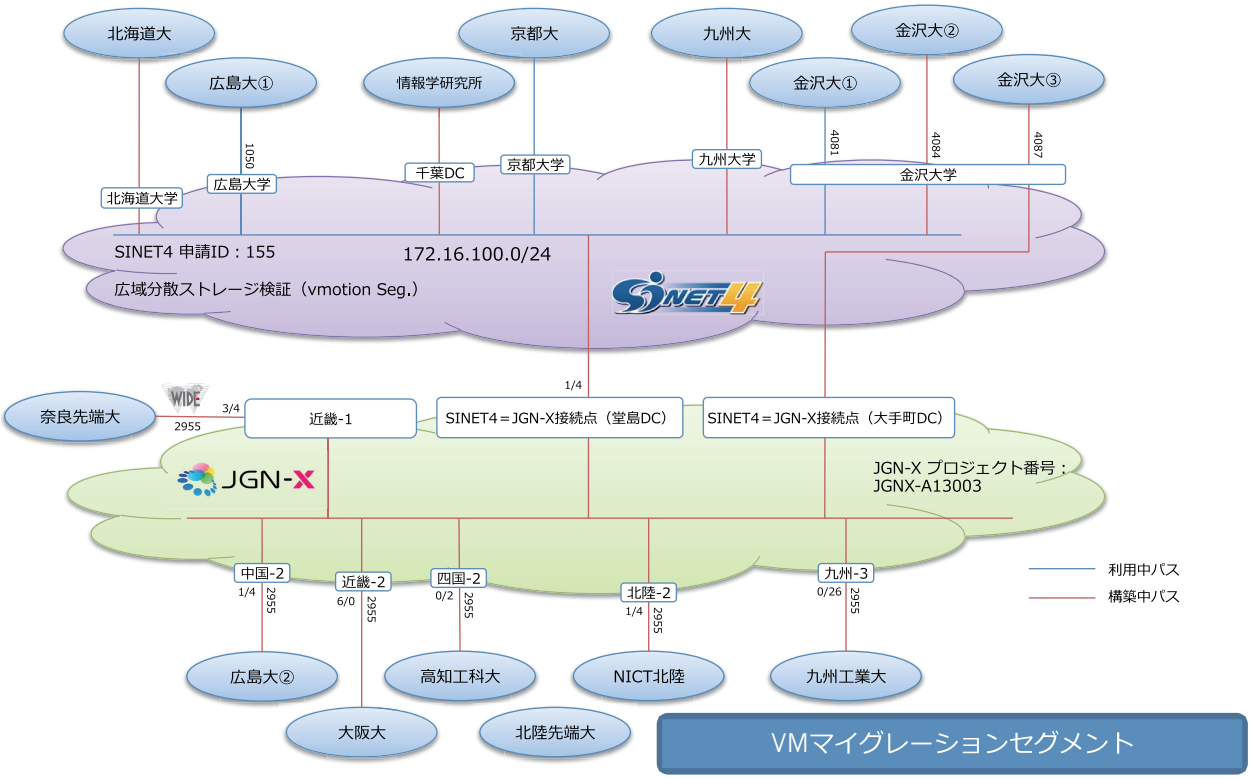

図 7　distcloud の SINET4 と JGNX からなるネットワークトポロジ
Fig. 7　A network topology of distcloud with SINET4 and JGN-X

の太平洋横断ライブマイグレーション実験から [9]，JGN-X との接続も行っている (図 7).

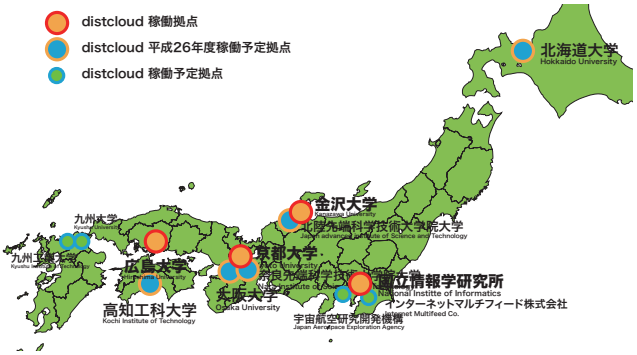

図 8　広域分散仮想化環境 distcloud の接続拠点
Fig. 8　Access Points of wide-area distributed virtualization infrastructure “distcloud”

4.2　初歩的な検証

JGN-X 上の ASR で提案手法の検証を行う前に，実験的な環境を仮想化基盤内に作成し，動作検証を行った．Cisco System 社から提供される All-in-one VM 環境は ova 形式で提供されている．この環境を展開し起動すると Ubuntu ベースの環境が立ち上がり，3 つの onePK 対応仮想ルータが相互接続された仮想ネットワークが提供される．これらのトポロジを図 9 に示す．

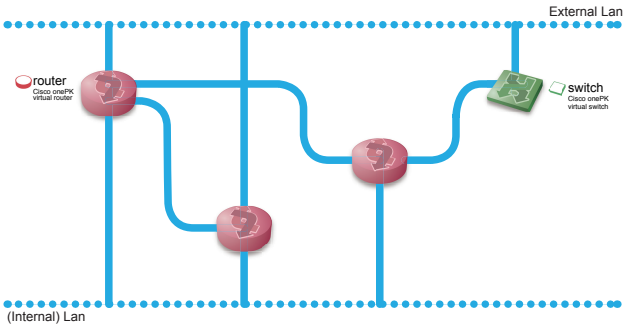

図 9　All-in-one VM における仮想ルータのトポロジ
Fig. 9　A topology of virtual routers in All-in-one VM

これらの仮想ルータに接続し onePK による操作を行うことができる．初歩的な検証においてはコントローラ (Ubuntu) から仮想ルータのインターフェイスをダウンさせてトポロジを変化させることを確認した．onePK において，API を用いてネットワーク機器を操作し，操作を完了するまでには以下の手続きが必要となる．

(1) Network Application オブジェクトの生成
(2) Network Application オブジェクトから生成される Network Element オブジェクトの生成
(3) 機器と TLS 通信するための Session Config の生成
(4) Session Config を用いた Network Element オブジェクトへの接続 (セッションの確立)
(5) Network Element オブジェクトの操作

(6) Network Element との接続の切断 (操作の完了)

Network Element オブジェクトには様々なメソッドがあり，インターフェイスの情報を取り扱う Network Interface オブジェクトを生成するためのメソッド (Python の API においては `get_interface_by_name`) が用意されてある．この Network Interface オブジェクトにはインターフェイスの状態を保持する変数が用意されており，この変数は no shutdown / shutdown の二値を持つことができる．この変数は Network Interface の `shut_down` メソッドを通してのみ変更が許されている．仮想ルータの複数のインターフェイスに振られた複数の IP アドレスは `get_address_list` メソッドにより取得することができる．これを用いて任意の IP アドレスが振られたインターフェイスに対して明示的に `shut_down` メソッドを実行することでリンクの切断を行うことが可能であることを確認した．

また onePK では API を用いて操作した設定を，Network Element オブジェクトとの接続が切断されたあと，指定した時間の後に揮発させることが可能である．そのため onePK コントローラが障害により各ルータとのネットワーク疎通がなくなった場合においても，状況を元に戻すことが可能となっている．この機能を用いて，前述のリンクの切断を行った後，onePK コントローラと仮想ルータとを接続するコントローラ側のインターフェイスを OS 上で切断することにより Network Element オブジェクトとの接続を強制的に切断した．この切断後，指定した時間 (ここでは 30 秒後) にネットワークトポロジが `shut_down` メソッドの実行前のトポロジに回復することを確認した．

4.3 SC14 での実証実験

JGN-X を用いた distcloud での耐災害性・耐障害性検証実験を今年 11 月に米ルイジアナ州ニューオーリンズで開催される SuperComputing 2014[*13] で行う．distcloud のアクセスポイントである広島の ASR1004 のポートを大阪にある仮想計算機から制御し，データセンター障害状態を作り出し，一拠点が切り離された状態においても distcloud が正常に稼働し続けることを示す (図 10)．現在，PRAGMA[*14] の米国参加拠点を distcloud の拠点とし，VM の太平洋横断マイグレーションと，超長距離マイグレーションをすることによりメリットのあるアプリケーションについてその有用性を示すデモンストレーションを行う予定である．

5. 災害・障害の深刻度分類

本プラットフォームは今後，次の二点の開発を進める．

(1) 過去に生じた災害による障害を模倣した障害を発生させる機能．

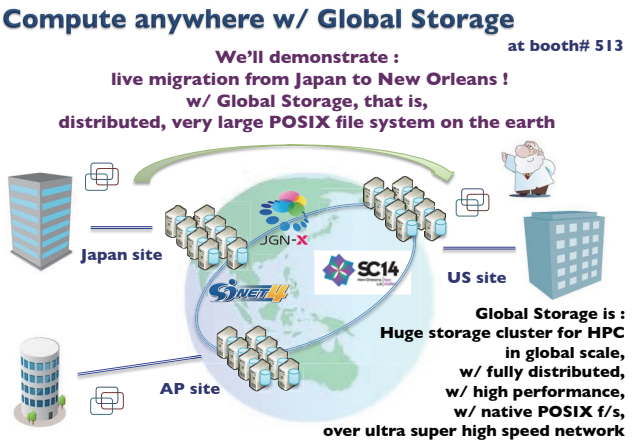

図 10 SC14 における検証実験

Fig. 10 A diagram of verification experiments on SC14

(2) まだ生じていないがこれから発生することが予想される障害を発生させる機能．

本プラットフォームが対象とする分散システムにおいて，どのような状況においても耐災害性・耐障害性を実現することは，設備投資コストが無尽蔵に提供される場合を除いて現実的ではない．そのため分散システムを維持する担当者は，発生する災害・障害の規模をあらかじめ想定し，その想定に対する耐災害性・耐障害性を実現しようとする．分散システムを構成する機器やそれらを繋ぐネットワークを対向グラフのノードとリンクで表すとき，全てのノードとリンクは分散システム全体に対して異なる「致命性」を持つ．この致命性は，そのノードやリンクが機能を停止したとき，あるいは機能が提供する品質が低下したとき，分散システム全体に与える影響の大きさで表される．全てのノードやリンクが異なる致命性を持つため，災害や障害の「規模」と，分散システムが受ける影響の間には相関がなく，「どこで」「何に対して」生じるかにより影響度が異なり，また発生する影響の種類も異なる．

災害や障害により受ける影響で代表的なものの一つとして，split-brain 問題 (ネットワークパーティション問題) が挙げられる．split-brain 問題とは，並列分散環境を構成する拠点間の通信が障害により途絶することで環境が分断され，分散システムが提供するサービスが途絶する，あるいは一貫性を喪失してしまう問題である．分散システムを構成するリンクの通信障害を引き起こす代表的な原因は電力供給の途絶によるものである．地震や津波，豪雨や洪水といった災害によりこの途絶が生じ易い．欧州や米国においては陸続きの国土を持つ国が多いため，インターネットを構成する自律システム間の TE での対応が 2005 年頃から提案されている [10]．しかし日本やニュージーランド，インドネシアのような島嶼国においてはこれらの災害が発生する確率が高く，また島嶼を跨ぐ国土であるため回線敷設コストが国内において不均一となっている特殊な問題点を

*13 `http://sc14.supercomputing.org/`
*14 the Pacific Rim Applications and Grid Middleware Assembly `http://www.pragma-grid.net`

表 2　災害・障害の影響度分類例
Table 2　Examples of classification for influence of disaster and disorder.

度数	障害内容	対応策
I	分散システムが 3 つ以上のパーティションに分断される	障害の原因となる複数のリンクの冗長化
II	2 つのパーティションに分断され，リンク回復後もサービス品質が低下	障害の原因となるリンクの冗長化とトラフィックエンジニアリング (TE)
III	2 つのパーティションに分断される．	障害の原因となるリンクの多重化
IV	パーティションに分断されないリンク障害	リンクの多重化
V	一箇所或いは複数輻輳によりサービス品質が低下	輻輳の原因となるリンクの帯域増強，或いは TE

持つ．対応方法として回線の冗長化が挙げられるが，広域での迂回を実現するような回線の冗長化には高いコストを要し，回線を利用するユーザはその設計に関する決定権を持たない場合もある．広域迂回が実現可能な回線においても，迂回路に様々なトラフィック要求が集中した結果，輻輳を起こして回線品質が低下することが想定される．

そのため前述の二点の開発において，特に後者においてはその障害の規模をもって分類するのではなく，分散システムのトポロジや代替コストを予め設定することにより，障害が分散システムに与える影響度や種類を推定し分類する必要がある．表 2 に分類例を示す．分散システムの維持担当者は，この分類の度数を指定することにより，指定に応じた障害が本フレームワークから提供され，同程度の影響度を持つ災害・障害への対策が十分に行われていることを確認することができる．

6. まとめ

既存の実装では，障害を発生させる箇所を予め指定し，障害を発生させることしかできない．しかし実際の防災訓練においてはかつて発生した災害を再現し，あるいは今後発生するであろう災害を模擬し，訓練を行う．この「災害の再現」と「予想災害の模擬」を実現する機構が本ソフトウェアの価値である．より広範に普及させるためにビジネスモデルの検討を行う予定である．

謝辞　本研究は平成 26 年度北海道大学情報基盤センター共同研究「Software Defined Datacenter を実現する広域分散環境の検証」，平成 26 年度国立情報学研究所共同研究「拠点間マイグレーションを実現する広域分散クラウド技術の方式検討および SINET4 等を利用した実証実験」による支援，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」および JSPS 科研費課題番号「26870325」の助成を受けました．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) および地域間インタークラウド分科会 (RICC) からの支援をいただきました．コンピュータリソースのご提供をいただいた各大学，SINET4 の回線をご提供いただいた国立情報学研究所，JGN-X の回線をご提供いただいた情報通信研究機構，および，クラスタストレージ技術である EXAGE / Storage をご提供いただいた株式会社インテック，および，アクセスサーバとして UCS をご提供いただいた Cisco Systems 合同会社に感謝します．

参考文献

[1] George Coulouris, Jean Dollimore, Tim Kindberg, Gordon Blair: Distributed Systems: Concepts and Design, 5th edition, ISBN: 9780132143011, Addison-Wesley Publishing Company (2011)
[2] 山下文男: 津波てんでんこ 近代日本の津波史, 新日本出版社, ISBN: 9784406051149 (2008)
[3] 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 (2014)
[4] 菊池豊, 岡村健志, 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, 栢分正人, 井上望美, 柴田祐輔: 地域 IX で恣意的な障害を発生させることによる耐障害性の検証, ITRC technical report 2013 (2014)
[5] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, Jonathan Turner: OpenFlow: Enabling Innovation in Campus Networks, SIGCOMM Comput. Commun. Rev., Vol. 38, No. 2, pp.69–74 (2008)
[6] D. Harrington, R. Presuhn, B. Wijnen: An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks, RFC 3411 (2002)
[7] R. Enns, M. Bjorklund, J. Schoenwaelder, A. Bierman: Network Configuration Protocol (NETCONF), RFC 6241 (2011)
[8] 柏崎礼生, 北口善明, 近堂徹, 楠田友彦, 大沼善朗, 中川郁夫, 阿部俊二, 横山重俊, 下條真司: 広域分散仮想化環境のための分散ストレージシステムの提案と評価, 情報処理学会論文誌, Vol. 55, No. 3, pp. 1140–1150 (2014)
[9] Ikuo Nakagawa, Kohei Ichikawa, Tohru Kondo, Yoshiaki Kitaguchi, Hiroki Kashiwazaki and Shinji Shimojo: “Transpacific Live Migration with Wide Area Distributed Storage”, Proc. the 2014 IEEE 38th Annual International Computers, Software and Applications Conference, Sweeden, pp. 486–492, Jul. 21-25, 2014. (DOI: 10.1109/COMPSAC.2014.71)
[10] Jason Schiller: Inter-AS Traffic Engineering Case Studies as Requirements for IPv6 Multihoming Solutions, Proc. 34th North American Network Operator’s Group (NANOG34), (2005).