

ネットワーク変動模擬情報を基にしたネットワーク状態評価の一検討

藤野 友也[†] 今村 誠[†] 高山 茂伸[†]

[†] 三菱電機株式会社 情報技術総合研究所 〒247-8501 神奈川県鎌倉市大船 5-1-1

E-mail: [†] {Fujino.Tomoya@cj, Imamura.Makoto@bx, Takayama.Shigenobu@db}.MitsubishiElectric.co.jp

あらまし ネットワーク観測情報を基にしたネットワーク状態の評価は、変調発生 of 早期対処や、予防のための対策を行う上で重要である。しかし一般に、キャリアネットワーク内の回線や設備の状態を直接観測することはできない。そのため、一般に死活監視としてホスト毎に観測されている ping の応答時間など、管轄下にあるホストにて確認可能な限られた情報から、通信が経由するネットワークの状態を評価することが求められる。本論文では、キャリアネットワーク内に発生が想定される変動事象について模擬し、模擬事象が発生する前後の ping 応答時間の統計値の変化を基に、C5.0 アルゴリズムにより構成した決定木による変動事例の識別を検討し、83% の適合率で識別が可能となる結果を得た。

キーワード 決定木, ネットワーク, ping, 応答時間, シミュレーション

A Study to Evaluate Network Status with Simulated Network Anomalies.

Tomoya FUJINO[†] Makoto IMAMURA[†] and Shigenobu TAKAYAMA[†]

[†] Information Technology R&D Center, Mitsubishi Electric Corporation

5-1-1 Ofuna, Kamakura-shi, Kanagawa, 247-8501 Japan

E-mail: [†] {Fujino.Tomoya@cj, Imamura.Makoto@bx, Takayama.Shigenobu@db}.MitsubishiElectric.co.jp

Abstract A network-status evaluation is one of the most important tasks to prevent instability or accidents in customers' communications environments. In general, most telecommunication companies can not observe links or nodes in career network directly. We want to evaluate a status of network with limited information such as a ping response time, which we can observe from our managed hosts. In this paper, we show a study for identification of anomaly events in career network by a decision tree, which is constructed by the C5.0 algorithm with simulated ping response time. We find that the precision ratio is 0.83 to identify the anomaly events with the decision tree.

Keyword Decision Tree, Network, Ping, Response Time, Simulation

1. はじめに

通信事業者が管理するネットワークは、様々な業務形態に属するホストを構成要素として含み、様々な通信インフラ形態が混在する環境にある。このようなネットワーク環境においても、提供するサービスの品質を維持するために、発生する変動の早期検知による迅速な対処が必要である。その変動がネットワークにて発生しているか、ホスト側で発生しているかによって、初期段階で実施すべき対処が分かれる。

自らが通信路設備(キャリア網)を持つ通信事業者(キャリア)は、ネットワークを構成する各装置におけるトラフィックやルーティング履歴、装置の自己診断ログの直接取得などにより、ネットワーク内の変調の全容を確認することができる。これらの情報を元として品質劣化の原因箇所の特定を行う品質劣化箇所の推定技術が研究されている[1]。しかし、自らが通信路設備を持たない通信事業者においては、ネットワークの

状態評価のために使用できる情報は、自らが管理するホストの情報や、管理ホストが行う通信におけるパケットの経路情報、および ICMP (Internet Control Message Protocol) により得られる情報などに限定される。ICMP を利用した情報の中でも、死活監視に使用する ping 応答は、ホスト管理を業務とする通信事業者が、最低限の情報として標準的に観測されている指標である。

本論文では、実測では確認が難しい、キャリア網内に発生が想定される、回線及びキャリア装置における変動事象の発生を模擬し、管理ホストにて観測される ping 応答時間の変動から、ネットワーク内で発生した変動事象を識別することを目的とする。本目的を実現する方式として、変動事例発生前後における ping 応答時間の統計値を入力として、C5.0 アルゴリズムにより構築した決定木を用いて、キャリア網内にて発生した異常事例を模擬結果から識別する実験を行った結果を

報告する。

2. 課題と従来技術

キャリア網を原因とする変調が事前に識別出来ない環境では、変調が発生した場合にはホスト・端末を確認し、ホスト・端末側に問題が無いと確認された後に、キャリア側へ問い合わせるといった手順が必要となる。このため、障害回復までの時間の遅延と、余分な現地出勤が発生することによるコストの損失が発生する。このような損失を軽減するために、応用上 80%以上の識別精度にて、ネットワーク状態の変調事象を識別し、初期段階で原因箇所がキャリア網か否かの切り分けを行うことが求められる。

通信の始点・終点にあるホストにおける通信パケットの観測結果から、通信経路の状態を推定する研究として、Paxson は、パケットロスおよびパケット到着遅延について原因別の比率を報告し、適切な監視指標を提案する方式を提案している[2]。Thottan らは、特定ホストのトラフィックの増加に対して、自己相関回帰によって予測されるトラフィック量と乖離するパターンから、障害の種類を推定する方式を提案している[3]。Duffield らは、2 点間通信の観測により、リンク遅延や別のパケット通信との独立性を評価する方式を提案している[4]。しかし従来技術は、発生した変調事例が把握可能である前提において学習および精度評価を行う必要があった。一般に、キャリア網における回線・装置の変調について、キャリア以外の事業者が情報を得ることは難しく、観測データに基づく学習が難しいという課題がある。また同時に、キャリア網内で、どの種類の変調事象が発生しているかを識別する方式も、実データから確認することが困難であった。

本論文では、これらの課題を解決するため、キャリア網にて発生しうる変調事象を個別に模擬し、ping 応答時間の統計値変化を判定基準として、想定した個別の変調事象を識別する。識別方式として、C5.0 アルゴリズム[5]により構成される決定木により評価する。

決定木とは、複数項目を持つ多次元データに対して、特定の項目の値を基に複数のグループに分割することを繰り返す、木構造の分類器である。C5.0 アルゴリズムでは、決定木を構成する際、情報理論におけるエントロピーに対応する情報利得が、分岐後に最大となるように分岐条件を採用する特徴を持つ。また C5.0 アルゴリズムは、分岐先が 2 つに限定されない決定木を生成する。

3. ネットワーク模擬方式

3.1. 変調事例

本論文では、設定誤りやネットワーク経由の侵入な

ど人為的活動によるものは評価対象とせず、自然発生的に生じる変調現象を評価対象とする。

キャリア網を原因とするネットワーク変調事象として、表 1 に示す 9 種類の変調事例に分類して考える。

表 1 回線/キャリア装置 変調事例

変調箇所	変調事例		番号
回線	通信品質低下	過負荷	N1
		スループット減少	N2
		パケットロス	N3
		遅延	N4
キャリア装置	通信断絶	接続不可	D1
		タイムアウト	D2
	通信品質低下	発信過多	D3
		スループット減少	D4
		遅延	D5

番号 N1 で示す回線/通信品質低下/過負荷の変調事例は、キャリア網内のルーティング経路上のリンクにて、外環境に起因する過大な通信が発生している事例を示す。番号 N2 で示す回線/通信品質低下/スループット減少の変調事例は、キャリア網内のルーティング経路上のリンクのバンド幅が、外乱などの影響にて低下する事例を示す。番号 N3 で示す回線/通信品質低下/パケットロスの変調事例は、キャリア網内のルーティング経路上のリンクの通信において、外乱などの影響にてパケットの劣化消失が発生する事例を示す。番号 N4 で示す回線/通信品質低下/遅延の変調事例は、キャリア網内のルーティング経路のリンク上にて遅延が発生する事例を示す。

番号 D1 で示すキャリア装置/通信断絶/接続不可の変調事例は、キャリア網内の装置への接続が拒否される事例を示す。番号 D2 で示すキャリア装置/通信断絶/タイムアウトの変調事例は、キャリア網内の装置への接続処理において、他の通信の待ち状態が多く、待ち状態のままタイムアウトとなる事例を示す。番号 D3 で示すキャリア装置/通信品質低下/発信過多の変調事例は、キャリア網内の装置からの発信イベントが過大に発生し過負荷状態である事例を示す。番号 D4 で示すキャリア装置/通信品質低下/スループット減少の変調事例は、キャリア網内の装置を経由時に、外乱などにより処理可能なパケット数が減少する事例を示す。番号 D5 で示すキャリア装置/通信品質低下/遅延の変調事例は、キャリア網内の装置を経由時に、外環境からの通信に起因する待ち状態が多く、遅延が発生する事例を示す。

最後に、いずれの変調事例も発生しない正常事例についても評価対象とし、番号を U1 とする。

回線およびキャリア装置を原因とした変調事例 N1～N4およびD1～D5のそれぞれの変調事例を識別することで、応用上、早期段階で変調の影響範囲や対策を判断することを可能とする。

3.2. ネットワークモデル

ネットワークの模擬のために、トポロジ、フローおよびイベントを定義する。

ping を発信するホストを基準ホスト、ping を受信し応答を返すホストを目標ホストと呼ぶことにする。模擬するネットワークモデルとして、図 1 に示すように、基準ホストから目標ホストまでのキャリア網内にて、ping 応答のルーティング経路上に存在するルータや、ルータを結ぶリンクにて変調事象が発生するモデルを作成し、評価を行う。

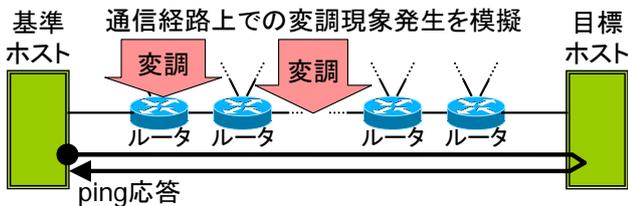


図 1 ネットワークモデル概念図

キャリア網などの大規模ネットワークでは、表 2 に示すように、リンク、トラフィック、パケットサイズに関して、それぞれの特性項目にて対応する特性値を平均的に持つことが報告されている[6]。この特性値を基準にしてモデルを構成する。

表 2 大規模ネットワークの特性値

分類	特性項目	特性値
リンク	各ルータの他のルータとの平均リンク数	4
	ホストから最初のホップルータまでのリンクの種類	50%は 56kb 50%は 10Mbps
	ルータ間リンクのバンド幅	50%は 10Mbps 40%は 100Mbps 5%は 622Mbps 5%は 2.4Gbps
トラフィック	ホストからルータへのリンクのビジー率	50%
パケットサイズ	リンク上の平均パケットサイズ	500bits

3.3. トポロジ定義

以降では、基準ホストから目標ホストまでのルーティング経路上にあるルータを、メインルータと呼ぶ。

メインルータに隣接する、メインルータでないルータを外環境ルータと呼ぶ。また、メインルータ同士をつなぐリンクをメインリンクと呼ぶ。

表 2 に示す特性の通り、各ルータの他のルータの平均リンク数は 4 である。そのため、メインルータから他のメインルータ(もしくは基準ホスト・目標ホスト)へ 2 つ、外環境ルータへ 2 つのリンクを持つ、図 2 の構成を基本単位として、ネットワークトポロジを構成する。

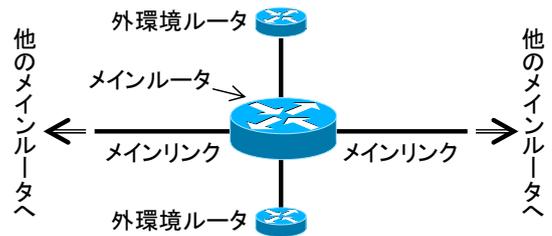


図 2 ネットワークトポロジの基本構成要素

図 3 に示す通り、基準ホストから目標ホストまでの通信経路をメインパスと呼んで、トポロジ全体像とする。キャリア内のメインルータ以外のネットワーク変動は、外環境の影響を外環境ルータの出力のみで模擬することで、トポロジを単純化する。

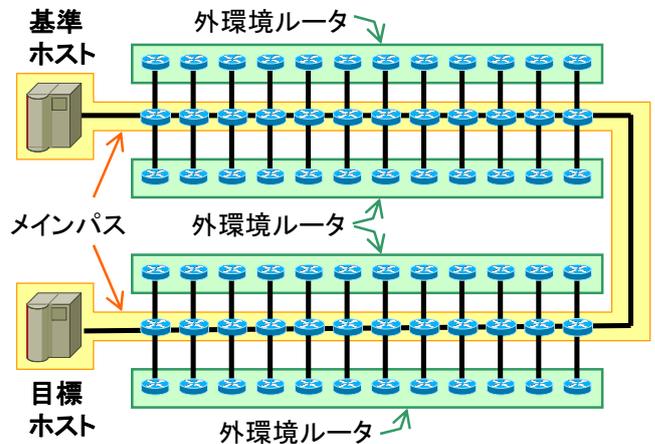


図 3 トポロジ全体像

各リンクのバンド幅については、表 2 に示す特性から、以下の通りとする。

- 基準ホストからメインルータへのリンクは 10Mbps とし、目標ホストとメインルータを結ぶリンクは 10Mbps と 10kbps を、試行事例ごとにランダムに割り当てる。
- ルータ同士を結ぶリンクのバンド幅について、50%は 10Mbps、40%は 100Mbps、5%は 622Mbps、5%は 2.4Gbps に設定する。各々のバンド幅が設定される

リンクは、試行事例ごとにランダムに指定する。

3.4. フロー定義

パケットフローとして、評価対象としてのフローと、それ以外のフロー(外環境としてのフロー)の 2 種類を定義する。

(1) 評価対象としてのフロー

基準ホストから目標ホストへの ping 応答を模擬する。

(2) 外環境としてのフロー

ある外環境ルータから他の外環境ルータへの通信により、キャリア内の他の通信トラフィックを模擬する。外環境としてのフローの始点と終点は、図 3 の上側の外環境ルータから下側の外環境ルータへの総当たり通信として定義する。

3.5. イベント定義

表 3 に示す変調事例を模擬するイベントを定義する。

表 3 変調事例に対応する属性値の変更

変調事例		属性の変更内容	
正常時	U1	なし	
回線	過負荷	N1	外環境ルータ間の通信で過大な通信を模擬し、ビジー状態にする
	スループット減少	N2	メインリンクのバンド幅を 1/10 に下げる
	パケットロス	N3	メインリンクのノイズを模擬し、ロスを発生させる
	遅延	N4	メインリンクの遅延時間を 10 倍にする
キャリア装置	接続不可	D1	ノードへの接続を不可とする
	タイムアウト	D2	受信キューのサイズを 1/10 に減らし、受信キューを通信待ちパケットで満杯にする
	発信過多	D3	送信イベントを追加送信する
	スループット減少	D4	インタフェースの最大パケット数を 1/10 に下げる
	遅延	D5	受信キューのサイズを 10 倍にし、90%を通信待ちパケットで埋める

事例 U1 を模擬する際は、正常事例であるため属性の変更は行わない。

変調事例 N1 を模擬するイベントとして、外環境ルータ間での通信で過大な通信を模擬し、ビジー状態にすることで、回線の過負荷を模擬する。変調事例 N2

を模擬するイベントとして、メインリンクのバンド幅を 1/10 に下げることで、回線のスループット減少を模擬する。変調事例 N3 を模擬するイベントとして、メインリンクのノイズを模擬し、ロスを発生させることで、回線のパケットロスを模擬する。変調事例 N4 を模擬するイベントとして、メインリンクの遅延時間を 10 倍にすることで、回線の遅延を模擬する。

変調事例 D1 を模擬するイベントとして、メインリンクに対応するノードへの接続を不可とすることで、キャリア装置の接続不可を模擬する。変調事例 D2 を模擬するイベントとして、メインリンクに対応するノードの受信キューのサイズを 1/10 に減らし、受信キューを通信待ちパケットで満杯にすることで、キャリア装置のタイムアウトを模擬する。変調事例 D3 を模擬するイベントとして、メインリンクに対応するノードからの送信イベントを追加送信することで、キャリア装置の接続不可を模擬する。変調事例 D4 を模擬するイベントとして、メインリンクに対応するノードのインタフェースの最大パケット数を 1/10 に下げることで、キャリア装置のスループット減少を模擬する。変調事例 D5 を模擬するイベントとして、メインリンクに対応するノードの受信キューのサイズを 10 倍にし、90%を通信待ちパケットで埋めることで、キャリア装置の遅延を模擬する。

4. 評価

4.1. 模擬実験方式

ネットワーク模擬ツールとして、イベント駆動型のネットワークシミュレータである NS-2 (Network Simulator version 2) [7]を使用する。

ネットワーク模擬では、正常事例 U1 および 9 種類の変調事例(N1~N5, D1~D5)のそれぞれについて 20 回ずつ模擬し、ping 応答時間を出力として得る。200 分間の模擬時間のうち、30 分時点で変調事例が発生する模擬設定にて、模擬を行う。

また、それぞれの変調事例とそれに対応する ping 応答時間の模擬結果を一定時間幅で統計値 (平均値, 最大値, 最小値, 標準偏差および件数) に集計した結果を入力として、C5.0 アルゴリズムによる決定木を作成する。このとき、模擬した変調事例と、模擬結果の ping 応答時間から決定木によって判定された変調事例とを比較した精度を比較する。模擬処理の実行環境を表 4 に示す。

表 4 模擬処理環境

OS	Linux Fedora 12
CPU	Intel Xeon E5620 (4 コア) × 4
MEM	23.6GB
HDD	824.9GB (SATA 6Gb/s, 7200rpm)

4.2. 模擬結果

10 種類の正常・異常事例の模擬結果の代表として、正常事例 U1 と発信過多事例 D3 における 10 分毎の ping 応答時間(以降では RTT, Round Robin Time と呼ぶ)の平均値の推移を図示する。30 分時点の縦線が、変調事例の模擬発生時刻を示す。

(1) 正常事例 U1 模擬

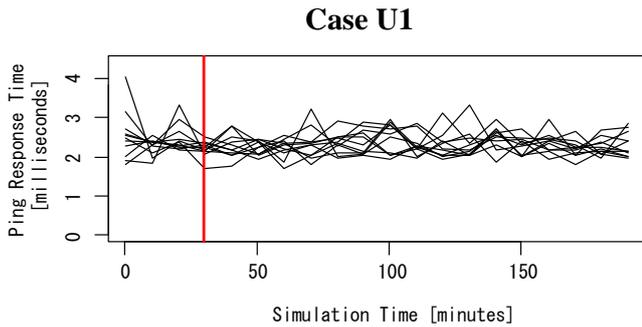


図 4 正常事例模擬結果

図 4 に示す正常事例模擬(20 試行)では、30 分時点で通信環境を変化させるイベントは発生しておらず、安定した特性を示す。

(2) 変調事例 D3 模擬

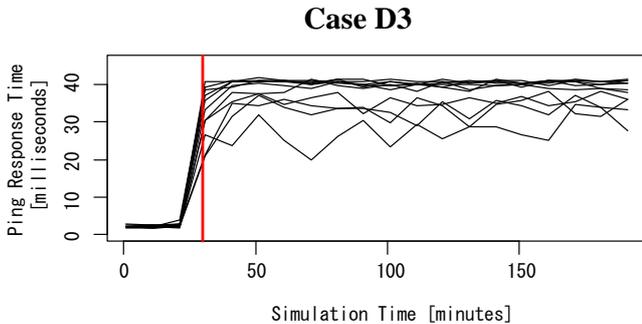


図 5 異常事例 D3 模擬結果

図 5 に示す変調事例 D3 模擬(20 試行)では、30 分時点で、発信過多のイベントを模擬したことにより、RTT 平均値は 2ms 台から 20~40ms 台に増大している。

NS2 による ping 応答時間の模擬結果に対する、20 事例の RTT 変動のグラフによる視認による評価結果を、表 5 に示す。

変調事例 N1 では、正常事例である U1 と大きな差は見られない。変調事例 N2 では、30 分時点でのスループット減少の変調事例発生により、RTT 平均値が 2ms

台から 5~7ms 台に増大している。変調事例 N3 では、ping 応答の成立件数が半数になっており、RTT 平均値は 2ms 前後で安定している。変調事例 N4 では、遅延発生に伴い、RTT 平均値が 0.2ms ほど増大しているが、これは 0.1ms 遅延が増大したリンクを往復していることに起因する。

変調事例 D1 では、イベント後に ping 応答が成立せず、欠測値となっている。変調事例 D2 では、正常事例と大きな差異は見られないが、ping 応答の成立数は 20% 程度減少している。変調事例 D3 では、発信過多イベントのメインルータからの発生により、RTT 平均値は 2ms 台から 20~40ms 台に増大し、試行毎の RTT 平均値のばらつきも増加している。変調事例 D4 では、正常事例である U1 と大きな差は見られない。変調事例 D5 では、全体的に 0.1ms ほど RTT 平均値が増大しており、一時的に 4.0ms 台の過大な RTT 平均値が発生している。

表 5 概要評価結果

変調事例		評価結果
正常	U1	評価の基準。
過負荷	N1	正常事例(U1)と大きな差異は見られない。
スループット減少	N2	30 分時点でのスループット減少の変調事例発生により、RTT 平均値が 2ms 台から 5~7ms 台に増大。
パケットロス	N3	ping 応答の成立件数が半数になっている。RTT 平均値は 2ms 前後で安定。
遅延	N4	遅延発生に伴い、RTT 平均値が 0.2ms ほど増大。0.1ms 遅延が増大したリンクを往復しているためで妥当な結果。
接続不可	D1	イベント後、ping 応答が成立せず。
タイムアウト	D2	正常事例と大きな差異は見られない。ping 応答の成立数は 20% 減少。
発信過多	D3	発信過多イベントの発生により RTT 平均値は 2ms 台から 20~40ms 台に増大し、ばらつきも増加。
スループット減少	D4	正常事例(U1)と大きな差異は見られない。
遅延	D5	全体的に 0.1ms ほど RTT 平均値が増大している。一時的に過大な RTT 平均値(4.0ms 台)が発生。

4.3. 識別精度評価

本節では、変調事例を模擬したイベントの発生直前と直後の一定期間内に発生した ping 応答時間の統計値を基に、C5.0 アルゴリズムにより構築された決定木により、U1, N1~N4, D1~D5 を識別する精度について評価した結果を示す。

運用上よく設定される ping 応答集計間隔である 1 分～10 分における条件別の精度を比較する。ping 応答時間統計値を集計する単位期間について、60 秒から 600 秒まで、60 秒刻みで変更し、適合率による精度の差異を評価した結果を表 6 および図 6 に示す。

表 6 変動模擬前後の単位時間毎の適合率

時間幅 (秒)	適合率	時間幅 (秒)	適合率
60	0.641	360	0.749
120	0.774	420	0.739
180	0.769	480	0.749
240	0.744	540	0.754
300	0.831	600	0.764

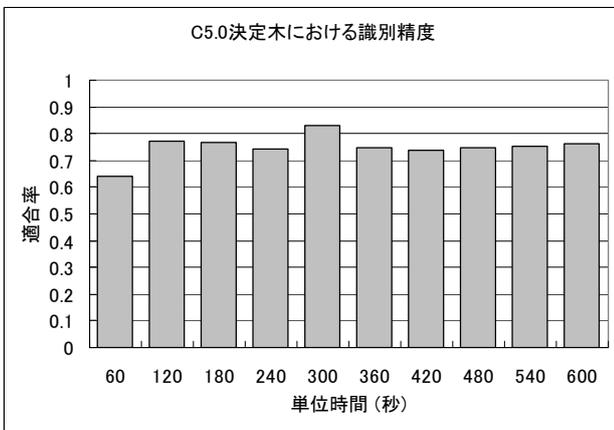


図 6 単位時間毎の識別精度(適合率)

適合率が 83% と最も高い、イベント発生前後 300 秒における統計値集計結果を基にした決定木による、模擬事例(U1, N1～N4, D1～D5) ごとの適合率、再現率および総合的な精度を示す f 値(適合率と再現率の調和平均)を表 7 に示す。

表 7 模擬事例毎の精度 (単位時間 300 秒)

模擬事例	適合率	再現率	f 値
U1	0.94	0.85	0.89
N1	0.64	0.70	0.67
N2	0.95	0.95	0.95
N3	0.76	0.80	0.78
N4	0.90	0.95	0.92
D1	0.89	0.85	0.87
D2	0.81	0.85	0.83
D3	0.93	0.87	0.90
D4	0.71	0.75	0.73
D5	0.83	0.75	0.79

正常事例と比較して、変動による影響が観測値から視認できない N1, D4 については識別精度がやや低い、それ以外の異常事例については 75%～95% の適合率・再現率を示している。

結果として、300 秒を集計単位とした統計値を基にした変動事例の識別において、全体としての適合率が 83% を示し、応用上必要な精度を確保していることを確認した。

5. まとめ

本論文では、実測データとして情報を得ることが難しいキャリア網内の回線・装置の変調事象について、発生しうる 9 種類の変調事象の模擬を行うモデルを作成し、それぞれの変調事象発生前後の ping 応答時間模擬結果を取得。模擬事象発生前後の一定時間幅における ping 応答時間の統計値を基に、C5.0 アルゴリズムによる決定木を用いた変動事例を識別する方式を提案し、精度評価を行った。

その結果、時間幅を 300 秒とした場合の統計値の集計結果を用いた決定木により、応用上の要件を満たす 0.831 の適合率により、正常事例および 9 種類の変調事例のそれぞれの識別ができることを確認した。

今後は、今回得られた決定木における分岐条件毎に、その分岐条件が精度へ与える影響を検証し、精度向上の検討を行う。また今回の模擬結果について、実網における ping 応答時間観測結果と関連付けて評価を行うための追加実験の実施を行う。この際、表 3 のように 1/10 や 10 倍などに設定した、変動事例発生時のキャリア網の属性パラメータの変動幅について、実網の観測結果と符合する形へ再調整を行う。既存のネットワーク監視システムへの本判定方式のフィードバックを通して、実証を行っていく。

参考文献

- [1] 菅野裕宣, 宮脇豊, 加藤大世, 池田稔, “大規模 IP ネットワークにおける高精度な障害切分けシステムの開発”, NTT DoCoMo テクニカル・ジャーナル, Vol. 18, No.1, pp. 21-26, 2010.
- [2] V. Paxson, “End-to-End Internet Packet Dynamics”, IEEE/ACM Transactions on Networking, Vol.7, No.3, pp. 277-292, June 1999.
- [3] M. Thottan and C. Ji, "Using Network Fault Predictions to Enable IP Traffic Management", Journal of Network and Systems Management, Volume 9 Issue 3, pp. 327-346, 2001.
- [4] N. G. Duffield and Francesco Lo Presti, “Network tomography from measured end-to-end delay covariance”, IEEE/ACM Transactions on Networking, Vol. 12, No. 6, pp. 978-992, 2004.
- [5] Rulequest Research, “Data Mining Tools See5 and C5.0”, <http://www.rulequest.com/see5-info.html>.
- [6] G. Riley and M. Ammar, "Simulating Large Networks: How Big is Big Enough?", Proc. of the First International Conference on Grand Challenges for Modeling and Simulation, January 28-31, 2002.
- [7] S. McCanne and S. Floyd, “ns Network Simulator”, <http://www.isi.edu/nsnam/ns/>.