

データベースに基づく通信ネットワーク管理のための モデル構築*

加藤 泰規† 河野 浩之† 長谷川 利治†

†京都大学大学院工学研究科応用システム科学専攻

近年、通信ネットワークをめぐる技術は著しい変化を遂げており、その管理はますます複雑になっている。この種のシステム管理を支える基盤としてデータベースが利用されることが多く、より知的な処理を柔軟に実現するための、データベース性能の向上は重要なものとなっている。本稿では、通信ネットワークの統計情報が格納される RMON などの MIB に対して、どの様なデータ発掘が有効であるかについて述べる。まず、統計データの格納されたテーブルに対して相関ルールを求めるアルゴリズムを用いて、ネットワーク障害を特定するルールを導き、複数のネットワーク間において障害が波及する状況を示すルール集合を求める。また、障害原因の特定、性能上のボトルネック発見などを与えるルール集合からネットワーク管理モデルを構築し、ネットワーク管理システムや管理者にフィードバックすることを可能とする。

Construction of network management model using data mining techniques

Yasunori KATO† Hiroyuki KAWANO† Toshiharu HASEGAWA†

†Department of Applied Systems Science, Kyoto University

Recent years, very complex techniques are required to manage the advanced communication networks. Especially, the technology of database systems is important for constructing the network management systems. And intelligent query processing is essential to the management of huge complex system. In this paper, we describe what kinds of data mining algorithms are effective to analyze statistical data in MIBs including RMON. We derive a set of rules which specify network faults and fault propagation by the algorithm of mining association rules in MIBs. And using these rules, we construct a network management model which enable to specify cause of network fault and detect a bottleneck in performance.

*連絡先: 〒606 京都市左京区吉田本町 京都大学大学院工学研究科応用システム科学専攻 河野 浩之
Tel: (075) 753-5513, Fax: (075)761-2437, E-mail: kawano@kuamp.kyoto-u.ac.jp

1 はじめに

現在、高度情報化社会のインフラストラクチャとなるべき ATM ネットワークなどの超高速ネットワークが整備されており、マルチメディア技術の利用を促進する安定したシステム構築を行うことが重要な課題である。しかしながら、通信ネットワークの技術変化は著しく、その規模も絶えず拡大していることから、効率的な管理・運用技術をいかに確立するかが、重大な問題として広く認識されている。そのためには、基礎データとなるネットワークに関する情報を的確に収集し、分析することが重要となる。

現在、各ネットワーク機器に取りつけられた管理エージェントによって、ネットワークの統計情報などのデータを、通信ネットワーク管理システムは収集する。異なるベンダーにより提供されるネットワーク機器により多少の違いがあるものの、この種のネットワーク管理のためのデータは、特に加工されない生データとして MIB(Management Information Base) とよばれるデータベースに蓄えられる。そこで、これらのデータを管理・運用に活用するために、データベースシステム上で何らかの知的なデータ操作が要求される [1, 2]。しかしながら、既存のデータベースシステムは、記憶領域に関する厳しい制約の範囲内で高速な検索処理の実現を目指しており、柔軟な加工が難しいものとなっているため、これまで知識ベース構築の観点からのアプローチが多かった。

しかしながら、一般に、この種の知的処理をデータベースシステム上で実現することは重要であり、大量かつ複雑な構造をもつ生データに対して計算コストの高い問合せ処理を効率良く実行する機構を、データベースシステムのレベルに実装することはシステム設計上も望ましいと言える。

この種のデータベースに蓄えられたデータ集合に内在する知識を効率的に獲得する研究が盛んに行われており、データベースからの知識発見 (KDD:Knowledge Discovery in Databases) または、データ発掘 (data mining)

などとよばれている [8, 3, 5]。なお、これらの研究で提案されたアルゴリズムにより得られる記述は、データ間の従属性・データの要約などである。

本稿では、MIB に蓄積されたパケット数などの属性をもつデータに対して知識発掘の典型的なアルゴリズムを適用することによって、待ち行列の解析知識に基づく管理や、ネットワーク運用のためのエキスパートシステム構築を効率的に可能とするためのネットワーク管理モデルの構築が可能となることを示す。一例として、この種の処理によって発見されるネットワーク障害が、他のネットワークへと波及する状況をルールとして表現できることについて述べる。このように、データ発掘を可能とすることにより、ネットワークから収集されるデータに応じたネットワーカトポロジーの改善を、ネットワーク管理システムやネットワーク管理者は効率的に実行することができる。

以下、2 章に、通信ネットワーク管理におけるデータベースの概要について述べる。3 章は、待ち行列理論を用いて障害を発見するための基本的な処理について述べる。4 章では、障害の波及状況を発見する手法について述べる。5 章では、4 章の結果を用いたネットワーク管理モデルの構築手法について述べ、ネットワーカトポロジーの改善方法について示す。6 章は、結論と将来の課題について述べる。

2 通信ネットワークの管理

通信ネットワーク管理は、大きく次の 5 種類に分けられる [9]。

- 性能管理 (performance management)
- 障害管理 (fault management)
- 課金管理 (accounting management)
- 構成/ネーム管理 (configuration and name management)
- 機密管理 (security management)

この中でも、性能管理は、ネットワークの安定性を考える上で非常に重要である。例えば、急

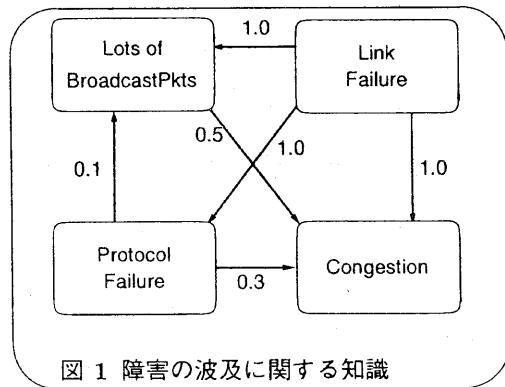
速に成長しているインターネットのように相互接続されているネットワークの性能管理を行うには、その特性を把握し QoS (Quality of Service) の適切な決定は欠かせない。しかしながら、流れるデータは、制約の緩いテキストデータのみではなく、動画像・音声・ファイルのように異なる制約をもつものが混在しているため、適切な管理は非常に難しい。もはや、ネットワーク管理者の経験のみに頼るには限界に達していると言え、この種の管理を的確にサポートする知的処理が要求されている。

以後、ネットワーク管理の中でも性能管理と障害管理に焦点をあてながら、ネットワーク管理プロトコルについて述べる。まず、ネットワーク管理者が、ネットワークの性能を低下させないための管理手順は以下の通りである。

【手順】

1. ネットワークとサーバを監視して統計を収集する。
2. 統計を分析して、ネットワーク中のどこに原因があるかを検出する。
3. ネットワークトポロジーを調整したり、デバイス並びにサーバーのパラメータを調整し、性能低下を招いた原因を取り除く。

この種の一連の動作を遂行する上で、ネットワーク管理者は、障害管理のために図 1 のような知識をもっており、ネットワーク障害の波及状況を推測することとなる [6]。



ネットワーク管理者は、図 1 の知識ネットワークを、状況によって絶えず調整していくことが必要である。また、それに伴うネットワークの監視による統計データの収集のため、どのような種類のデータを集めることができるかななど、ネットワーク管理プロトコルの性質についても深く知る必要がある。しかしながら、ネットワークの変化は激しく、この種の知識の質を適切に維持しながら、的確な推論を行うことは非常に難しい。

ここで、どのようなデータを元にした知的処理が必要であるかを明らかにするために、ネットワーク管理プロトコルの中でもインターネット標準の SNMP プロトコル (Simple Network Management Protocol) について簡単に説明する。

SNMP の管理情報ベースについては MIB-II が定義されており [4]、物理的なインターフェースに関する情報や、各インターフェース上のイベント統計などを含む interface グループのほか、system, at, ip, icmp, tcp, udp, egp, transmission, snmp などの各グループがある。これらの中でも、性能評価に有用であると考えられる interface グループの属性について以下に示す。

ifIndex	(I.1)
ifDescre	(I.2)
ifType	(I.3)
ifMtu	(I.4)
ifSpeed	(I.5)
ifPhysAddress	(I.6)
ifAdminStatus	(I.7)
ifOperStatus	(I.8)
ifLastChange	(I.9)
ifInOctets	(I.10)
ifInUcastPkts	(I.11)
ifInNUcastPkts	(I.12)
ifInDiscards	(I.13)
ifInErrors	(I.14)
ifInUnknownProtos	(I.15)
ifOutOctets	(I.16)
ifOutUcastPkts	(I.17)

ifOutNUcastPkts	(I.18)
ifOutDiscards	(I.19)
ifOutErrors	(I.20)
ifOutQlen	(I.21)
ifSpecific	(I.22)

ネットワーク管理者は、属性(I.10)や(I.16)からシステムが送受信した総オクテット数を、属性(I.22)から送出パケットキューの長さなどを調べて輻輳状態について把握し、属性(I.7)と(I.8)を比べることによりネットワークがダウンしていないかを調べたりすることが可能である。また、RFC1271で新しく定義されたRMON(Remote Network Monitoring)と呼ばれる機能により、サブネットワーク単位の効果的かつ効率的な監視の実現を試みている[10]。

なお、RMON MIBには、各サブネットワークに関する低レベルの利用率やエラー統計を保持する Statistic グループや、Statics グループ中の利用可能な情報から、一定期間ごとの統計サンプルを記録するための History のほか、Alarm, Host, HostTopN, Matrix, Filter, Capture, Eventなどの各グループがある。以下に、ネットワークの性能管理に特に必要な Statistics グループの属性について示す。

etherStatsIndex	(S.1)
etherStatsDataSource	(S.2)
etherStatsDropEvent	(S.3)
etherStatsOctet	(S.4)
etherStatsPkts	(S.5)
etherStatsBroadcastPkts	(S.6)
etherStatsMulticastPkts	(S.7)
etherStatsCRCAlignErrors	(S.8)
etherStatsUndersizePkts	(S.9)
etherStatsOversizePkts	(S.10)
etherStatsFragments	(S.11)
etherStatsJabbers	(S.12)
etherStatsCollisions	(S.13)
etherStatsPkts64Octets	(S.14)
etherStatsPkts65to127Octets	(S.15)
etherStatsPkts128to255Octets	(S.16)
etherStatsPkts256to511Octets	(S.17)
etherStatsPkts512to1023Octets	(S.18)

etherStatsPkts1024to1518Octets	(S.19)
etherStatsOwner	(S.20)
etherStatsStatus	(S.21)

上記属性からわかるように、RMONは、イーサネット上のパケット数や、パケット長分布を統計データとしてもつ。なお、現在RMONのStatisticsグループが調べることのできるネットワークはイーサネットだけである。このため、FDDI(Fiber Distributed Data Interface)などのインターフェースに関しては従来どおりMIB-IIから統計情報を得ることにする。

各ネットワーク管理エージェントは、これらの統計情報を基本的には、30秒ごとに調べ、そのデータは各エージェントが保持する。また、ネットワーク管理アプリケーションは、図2に示したように、エージェントをポーリングすることによりこれらの情報を得ることが可能となっている。

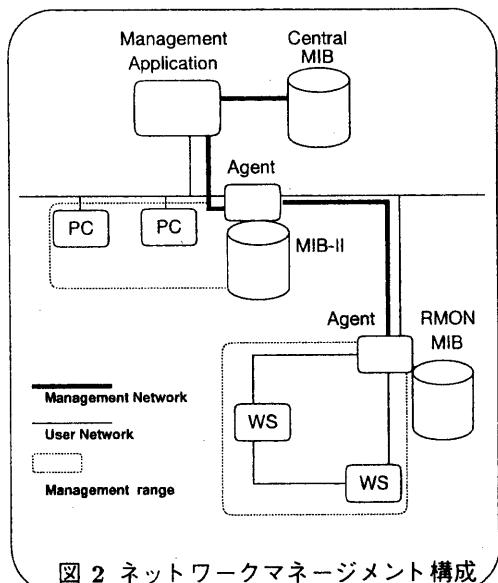


図 2 ネットワークマネージメント構成

3 障害の発見のための基本的処理

ネットワークの性能を低下させる主な原因として、ネットワークの帯域幅の不足のために起こるトラヒックの著しい混雑、すなわち輻輳(congestion)が考えられる。

そこで、本稿では、RMONで収集が可能な

イーサネットとその他のインターフェースを区別して処理する。すなわち、FDDIなどのイーサネット以外のインターフェースはinterfaceグループから、イーサネットはStatisticsグループから、それぞれ関係するデータを収集する。

現在多くのネットワーク管理システムでは単一属性に対する単純なグラフ化のみを提供しており、複数の属性間に関連した分析を容易に行うための処理機構をもたない。そこで、これらのデータに対してより複雑な規則の発見を可能とするために必要となる典型的な処理例を以下に述べる。

interface グループに対する処理

まず、interface グループからの輻輳の発見について考える。属性(I.5)は、インターフェースの現在のデータ転送能力(bps)であり、属性(I.16)は、インターフェースが送出した総オクテット数である。これらの属性を用いて、ネットワークの利用率Uを次式で定義する。

$$U = \frac{(I.16) \times 8}{(I.5) \times 30} \quad (1)$$

Uは、ネットワークの現在の転送スピードに対する実際に流れたデータの量をあらわしている。従って、Uが大きいほどネットワークは混雑していると考えられる。管理者は、Uに関する閾値TUを決め、その閾値を越えたものを輻輳状態と見なすことが可能である。この閾値は、各インターフェースごとに異なることが予想されるが、属性(I.3)により、インターフェースの種類を識別することが可能である。

その他、リンクの故障については属性(I.7)と(I.8)の比較により、プロトコル故障、過度のブロードキャストパケットについては、属性(I.15)と(I.18)に関する閾値として、 $T_{I.15}$ と $T_{I.18}$ を設けることによってそれぞれ障害であることを判定することが可能である。

Statistics グループに対する処理

Statistics グループでは、イーサネットに関する詳細なデータを得ることが可能である。属性(S.5),(S.6),(S.7)は、それぞれ、ユニキャスト、ブロードキャスト、マルチキャストのパケットの数であり、属性(S.13)は、衝突(collision)

が起きたパケット数である。ここで、パケットの送信成功確率 P_s を総パケット数を用いて定義する。

$$P_s = \frac{(S.5) + (S.6) + (S.7)}{(S.5) + (S.6) + (S.7) + (S.13)} \quad (2)$$

我々は、パケットが頻繁に衝突をおこすのは、ネットワークの輻輳が原因であると考え、ネットワーク管理者が定めた成功確率の閾値 T_{P_s} を下回る場合、輻輳状態であると見なすことが可能である[7]。また、過度のブロードキャストパケットについては、(S.6)と(S.7)の和の閾値 T_{BMP} を設けることによって判定することが可能である。リンクの故障などの障害は、Statistics グループで判定することができないが、Interface グループから判定できる。

以上のように、MIB データベースにおける複数の属性をグループ化したデータ処理を効率良きリアルタイム性を保ちながら実現する処理機構を、データベースシステムが備えることは有用である。

4 障害の波及の発見

本節では、ネットワーク障害の波及状況を把握するために、データ発掘で提案された相關ルールを求めるアルゴリズムをMIBに適用し、図3のネットワークを例として用いながら、依存関係について論じる。

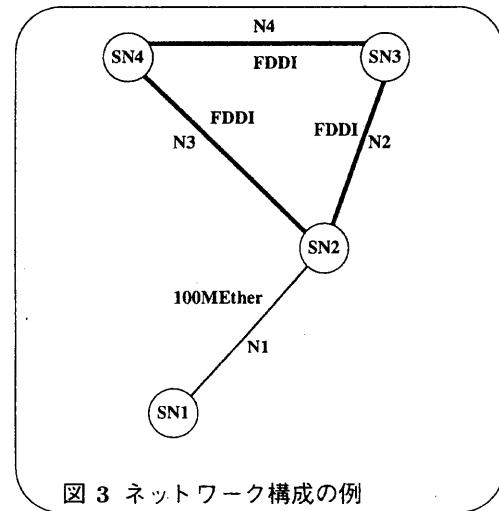


図3 ネットワーク構成の例

図中の円は、各サブネットワークを表しており、それぞれゲートウェイを通して他のサブネットワークと結ばれている。 N_1, N_2, \dots で示したFDDIや100Mイーサネットによって、ネットワークは結ばれる。ネットワークにおける統計データは各エージェントのMIBに蓄えられ、必要なデータのみが管理アプリケーションにより表1の中央MIBへと収集され保管される。

表1 中央MIB

	P_s	0.28	0.24	0.25	...
N_1 (Sta)	BMP	500	1200	5000	...
	(S.5)	10000	15000	10000	...
	(S.6)	200	400	2000	...
	(S.7)	300	800	3000	...
	(S.13)	4000	5000	5000	...
	A	0.13	0.16	0.08	...
N_2 (In)	(I.5)	10M	10M	10M	...
	(I.16)	5M	6M	3M	...
	(I.6)	1	1	1	...
	(I.7)	1	1	1	...
	(I.15)	200	300	100	...
	(I.18)	300	800	400	...
:	:	:	:	:	:

表1の属性(I.6),(I.7)は、それぞれインターフェースの望まれるステータスと現在のオペレーションステータスを表しており、up(1), down(2), testing(3)のいずれかを値としてもつ。また、StatisticsとInterfaceグループから、それぞれデータを収集していることを(Sta),(In)で示した。

ここで、データのサンプリング間隔を標準の30秒に1回とし、1週間分のデータを蓄えて中央MIBに保存するときに必要なサイズを考える。属性値1つの大きさは、32bit(4byte)であることを考えると、管理するべきネットワーク1つに対するサイズは、

$$7 \times 4 \times 2880 \times 7 \approx 560KB$$

となり、30個のオブジェクトでは約17MB必要になる。また、1週間単位の周期的変動を考えると、数ヶ月以上のデータが必要となる。よって1つの特性に対して数百MB必要とし、複数のネットワーク機器の処理を行うためには、かなり大規模な処理を必要とする。

この中央MIBからサブネットワークを含めたネットワーク間の障害の波及状況を発見するために、閾値を用いて正常(Normal)か障害(Fault)かを判断する。ここで、図1に表されている各障害を以下のように表す。

F(1): Lot of BroadcastPkts

F(2): Link Failure

F(3): Protocol Failure

F(4): Congestion

また、正常・障害の各状態を、表2の中央MIBの各タプルによって表した。

表2 障害の発見

N_1		N_2				...
F(1)	F(4)	F(1)	F(2)	F(3)	F(4)	...
N	N	F	N	N	F	...
F	N	F	N	F	F	...
N	N	F	F	F	F	...
N	N	N	N	N	F	...
:	:	:	:	:	:	:

本稿で提案するネットワーク管理モデルは、あるネットワークでの障害が他のネットワークへどのように波及するかを示すものであり、ネットワーク障害の波及状況を条件付き確率を用いて表す。また、あるネットワークで特定の障害が起こる確率も、管理のために必要である。そこで、ネットワーク N_p の $F(i)$ のフィールドである $N_p.F(i)$ を用いて、これらの確率を次式のようく表記する。

$$P_{N_p.F(i)}^{N_q.F(j)} = Pr(N_p.F(i) = F | N_q.F(j) = F) \quad (3)$$

$$P_{N_p.F(i)} = Pr(N_p.F(i) = F) \quad (4)$$

式(3)は、 N_q で $F(j)$ が起こったとき同時に N_p で $F(i)$ が起こる確率を表しており、式(4)は、観測データ中 N_p で $F(i)$ が起こる確率を表している。また、式(3)は次式のような形式のルールでも表すことが可能である。

$$(N_q.F(j) \text{ is } F) \rightarrow (N_p.F(i) \text{ is } F)(\alpha\%) \quad (5)$$

式(3)と式(4)の確率を、観測データを組向アルゴリズムによって処理することにより求める。ネットワーク障害の波及は直接的には物理的に接続しているものにのみ起こることを利用し、実際の計算では、 N_p と N_q が、物理的に接続しているもののみを計算する。また、性質

が顕著に現れているものだけを発見するため、全ての条件付き確率を求めることはせず、条件付き確率 $P_{N_p, F(i)}^{N_q, F(j)}$ の閾値 T_{CP} を定め、閾値を越えるものだけを出力することにする。これを用いて、サブネットワークを含めた全てのネットワークに対して $P_{N_p, F(i)}$ と $P_{N_q, F(i)}$ を求めるアルゴリズムを適用すると $P_{N_p, F(i)}$ は、表3で表され、 $P_{N_p, F(i)}^{N_q, F(j)}$ は、図4のルール集合で表される。このアルゴリズムは、データベースを1回走査するだけである、実用的である。

表3 データより得られた $P_{N_p, F(i)}$

N_1		N_2				\dots
F(1)	F(4)	F(1)	F(2)	F(3)	F(4)	\dots
0.01	0.05	0.09	0.02	0.06	0.1	\dots

- $(N_3, F(2) \text{ is } F) \rightarrow (N_4, F(4) \text{ is } F)(95\%)$
- $(N_2, F(2) \text{ is } F) \rightarrow (SN_3, F(4) \text{ is } F)(92\%)$
- $(N_1, F(4) \text{ is } F) \rightarrow (N_2, F(4) \text{ is } F)(80\%)$
- $(N_3, F(4) \text{ is } F) \rightarrow (N_2, F(4) \text{ is } F)(75\%)$
- $(N_1, F(2) \text{ is } F) \rightarrow (N_1, F(4) \text{ is } F)(100\%)$
- $(N_1, F(1) \text{ is } F) \rightarrow (N_1, F(4) \text{ is } F)(40\%)$
- $(N_1, F(3) \text{ is } F) \rightarrow (N_1, F(4) \text{ is } F)(30\%)$

図4 データより得られた $P_{N_p, F(i)}^{N_q, F(j)}$

5 ネットワーク管理モデルの構築

本章では4章で得られた結果を用いて、図5のように表されるネットワーク管理に有用なモデルを与える。

このモデルは、各ネットワークに対応する **Box** と障害に対応する **Frame** とそれらを結ぶラベルづけされた **Arrow** からなる。**Box** は、障害の数だけ **Frame** を持つおり、**Frame** 中には、そのネットワークで各々の障害が起こる確率が表示されている。例えば、図中で **Box** N_1 の **Frame** F(1) には、0.01 が入っているが、これは、 N_1 で、F(1) が起こる確率が 0.01 であることを示している。また **Arrow** は、**Frame** と **Frame** を結んでいるが、これは障害の波及を表している。なお、図中で N_1 の F(4) から、 N_2 の F(4) へ 0.80 とラベルづけされた **Arrow** があるが、これは、 N_1 に F(4) が起こったとき

の 80% は、 N_2 でも同時に F(4) が起こっていることを示している。

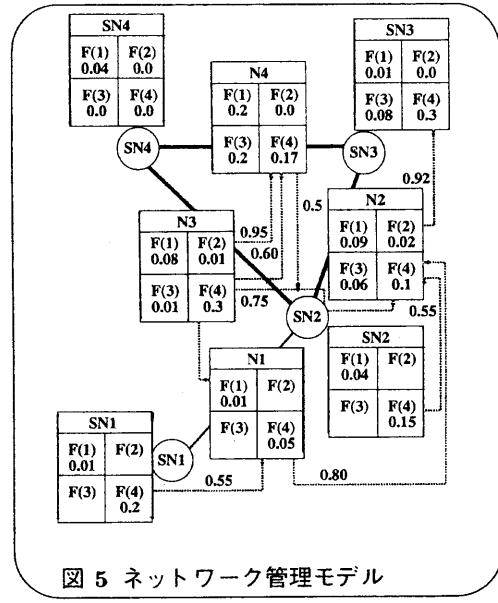


図5 ネットワーク管理モデル

以上のルールを用いることによって、ネットワーク管理者は、性能管理とともに障害管理に関する次の事項の把握が可能となる。

- 各ネットワークにおける障害の起こる確率
- ネットワーク障害の波及の状況

図5の例で考えると、 N_3 は、輻輳の起こる確率も高く、また、 N_3 で輻輳がおこったとき、高い確率で他のネットワークで同時に輻輳が起こっている。以上から N_3 は、このネットワーク全体のボトルネックになっている可能性が高いということが分かる。このことから、ネットワーク管理者は、 N_3 の MIB に対して詳細なデータの解析を行い、必要に応じて、ネットワークトポロジーの変更を行うこととなる。

さらに、ネットワーク管理者は、本稿で提案した手法を行う前にもつ図1のような知識の更新が可能となる。つまり、初期段階で与えられる知識は、ネットワークによっては実際のものと異なっている可能性が高い。そこで、本節で述べたモデル化によって得られた値を用いて、

これらの知識をより現実のネットワークで成立する知識に更新することが可能となることを、図6に示した。

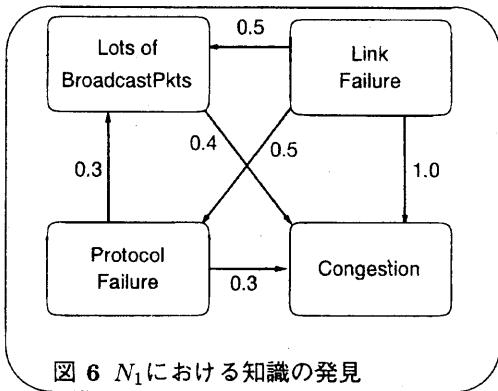


図 6 N_1 における知識の発見

6 おわりに

ネットワーク管理者にとって、ネットワーク間に内在する障害の波及について、ネットワークの位置まで含めて明確に把握することは困難であった。特に、ネットワーク規模が大きくなるにつれて、正確な状況を把握するための複雑さは指数関数的に増大するものである。

そこで、MIBにおける生データに対してデータ発掘の手法を適用し、条件付き確率の形でネットワーク障害の波及を記述することにより、より現実的な知識として導出することが可能となった。また、本稿で用いた障害波及のグラフ表現により、ネットワーク管理者の経験による知識の変化の一侧面を明確に記述することができた。本稿では、ネットワーク障害は、物理的接続が無ければ波及しないという制約を利用したため、その計算コストは極めて小さく、巨大なネットワークにおけるMIBでも十分適用できるものと考えられる。

最後に、将来の課題であるが、様々な情報ネットワークシステムにおいて膨大なデータが収集されるにつれ、多様な管理のためにデータの知的な加工の必要性が増大すると思われる。これは、本稿で適用した性能管理や障害管理だけでなく、課金管理や構成管理など、知識発掘をデータベースシステムが実装した場合の広い適用可能性を示していると考えられる。

謝辞

日頃御指導頂く大阪大学工学部情報システム工学科 西尾章治郎 教授に深謝の意を表します。本稿の一部は文部省科学研究費(08750431・08244103)のもとでの研究成果による。

参考文献

- [1] 河野 浩之, 西尾 章治郎, 長谷川 利治, “知識獲得アルゴリズムによる通信ネットワーク管理の考察,” 信学技報, AI92-59, pp.49-56, 1992.
- [2] H. Kawano, S. Nishio, J. Han and T. Hasegawa, “How Does Knowledge Discovery Cooperate with Active Database Techniques in Controlling Dynamic Environment?,” Proc. 5th Int'l Conf. on Database and Expert Systems Applications (DEXA'94), Athens, Greece, pp.370-379, 1994.
- [3] 河野 浩之, 西尾 章治郎, Jiawei Han, “データベースからの知識獲得技術,” 人工知能学会誌, vol.10, no.1, pp. 38-44, 1995.
- [4] K. McCloghrie, “Management Information Base for network management of TCP/IP-based internets: MIB-II,” RFC1213, 1991.
- [5] 西尾 章治郎, “大規模データベースにおける知識獲得,” 情報処理, vol.34, no.3, pp.343-350, 1993.
- [6] T. D. Ndousse and T. Okuda, “Computational Intelligence for Distributed Fault Management in Networks Using Fuzzy Cognitive Maps,” IEEE ICC, Dallas, U.S.A., pp.1558-1562, 1996.
- [7] K. Ori, H. Kawano, Y. Takahashi and T. Hasegawa, “An Approach for Network Diagnosis Based on Performance Analysis,” Proc. of 4th Int'l Conf. on Telecomm. Systems, pp.24-32, 1996.
- [8] G. Piatetsky-Shapiro and W. J. Frawley, *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991.
- [9] W. Stallings, “SNMP, SNMPv2, and CMIP: the practical guide to network management standards,” Addison-Wesley, 1993.
- [10] S. Waldbusser, “Remote network monitoring Management Information Base,” RFC1271, 1991.