大規模ネットワークログデータにおけるトラフィック可視化手法の提案

佐々木 聡志　　土井 章男
岩手県立大学 ソフトウェア情報学部

1. はじめに

近年，インターネット接続の導入が非常に容易になってきており，誰でもインターネットに接続することが可能になった．しかし，利用者が増える一方で，様々な問題が起こっているのも事実である．その一つとして，セキュリティ対策不足が考えられる．それによって引き起こされる障害の例として，ネットワーク障害や不正アクセスなどがある．その第一の対策として障害が起きている位置を素早く特定し，その箇所を対処することが重要である．それによって，被害拡大を最小限に抑えることが可能となる．

そこで本研究では，時系列ネットワークログデータから，障害箇所を特定するトラフィック可視化手法を提案する．また，岩手県立大学で取得したネットログデータに本手法を適用して，本手法の有効性を検討する．

2. 提案手法

一般にネットワークログデータは数万行以上あり，ログファイルをそのままの状態で直接的に見ることは現実的ではない．そのためネットワークログファイルをなんらかの手法で可視化する必要がある．代表的な可視化ツールとしては，Hyperbolic Tree[1]， Cone Tree[2]などがあるが，直接ネットワークログの可視化には使用できない．

今回使用するネットワークログデータのフォーマットは，図 1 のようになっている．本手法では，送信元 IP アドレスと送信先 IP アドレスを Z 方向に配置し，それらの IP アドレスに送信状態を表す直線で連結する．

日	時間	IP アドレス送信元	IP アドレス送信先	データ量
18-Jul-03	0:00:01	172.18.2.129	cba.182.169.36	152
18-Jul-03	0:00:03	172.19.4.112	cba.181.141.76	786
18-Jul-03	0:00:03	172.19.4.112	gb.209.169.202	767
18-Jul-03	0:00:03	172.19.4.112	gg.44.238.249	1565

図１．ネットワークログデータ

さらに，時系列を考慮して，IP 間のトラフィックデータに対して，データが発生した箇所にそのデータ量を表示し，これにより異常なデータが転送されていないかを認識可能にする．また，これらのデータ量に着目した重要度を定め，その重要度により，透明度の設定や色の強調表示を行う（図 2）．

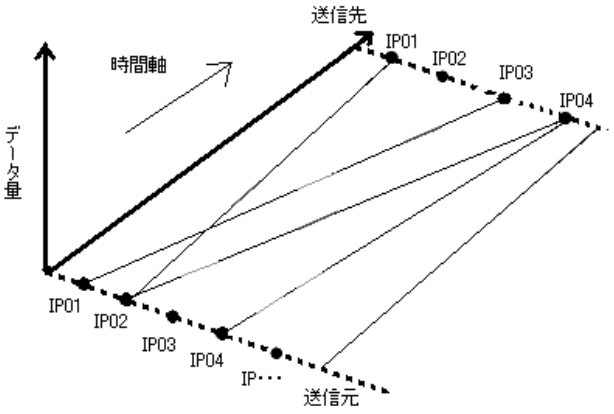

図 2．トラフィック可視化手法

3. 実装と適用

実装は，Windows XP 上にて MFC および OpenGL を用いて行った．

データ量の配置方法は，IP アドレス経路上にデータを配置することを前程に X 軸である時間系列の概念を考慮し，2 次元に置き換えたとき送信元 IP アドレスを(x_1, y_1)，送信先 IP アドレスを(x_2, y_2)，t を時間として，式（1）を用いてデータを配置した(図 3)．

ネットワークログデータは，2003 年 7 月 18 日に取得した岩手県立大学の 1 日分のデータである．図 4 の各軸は，X 軸が時間，Y 軸がデータ量，Z 軸が IP アドレス軸となっている．また，送信元 IP アドレス（手前）から送信先 IP アドレス（奥）の IP アドレス送信経路を X-Z 平面に描画

A proposal of the traffic visualization technique in large-scale network log data
Satoshi Sasaki and Akio Doi・Iwate Prefectural University Faculty of Software and Information Science

$$\begin{cases} x = x_1 + (x_2 - x_1)t \\ y = y_1 + (y_2 - y_1)t \end{cases} \quad (1)$$

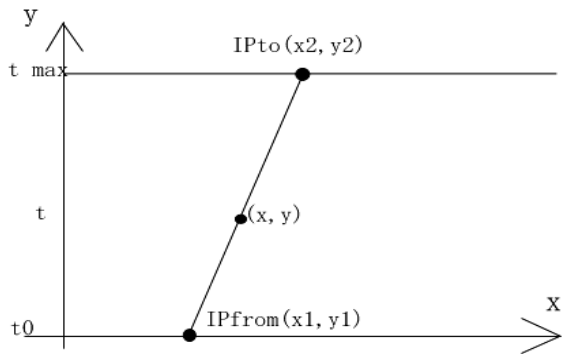

図 3. パラメータ式

している．なお IP アドレス経路は，その送信先 IP アドレスに一度でもアクセスされていると，経路線を描画する．

色設定は，閾値を与えることによって，一定量以上のデータ量のときに色を変えて表示する．また，IP アドレス経路にも透明度を与え，何度も同じ IP アドレスにアクセスしている場合には濃く表示されるようにした（図 4）．

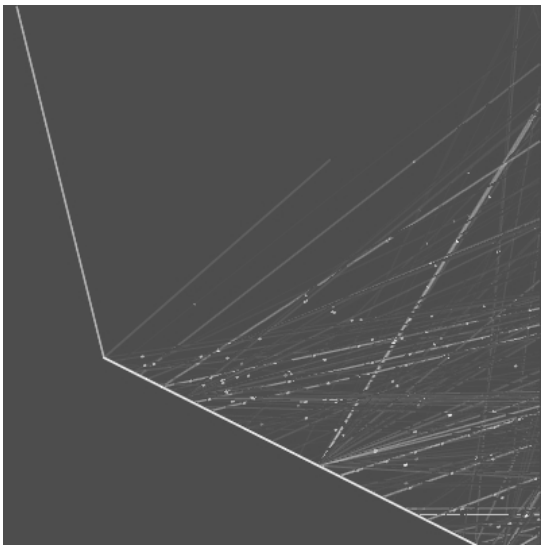

図 4. 色設定を考慮した実装図

図 4 に示されるように，ネットワークログ情報が大量の場合，各 IP アドレス経路同士が重なり合い，すべての情報が表示されていない．そこで，我々は，IBM T220（解像度 3840×2400，22 インチ 920 万画素）の超高精細液晶モニタに表示することを試みた．本装置を用いることで，普通のディスプレイでは表現することができなかった，より精細な箇所の表現が可能になった（図 5）．

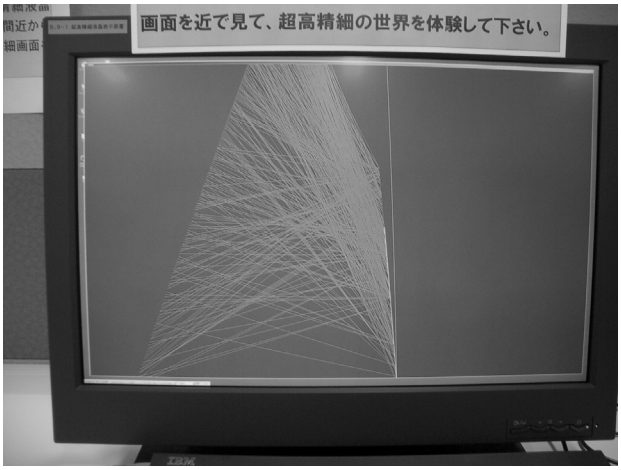

図 5. 超高精細液晶モニタによる表示例

4. 考察と今後の展開

今回，大規模なネットワークログデータの可視化手法を提案し，実際のネットワークログに適用して，その有効性を確認した．しかしながら，「情報の重なり」は，本可視化手法の問題点のひとつでもある．つまり，IP アドレス経路のアクセス数が多い場合，経路同士が重なってしまう．

これらの問題を解決するために，情報の削減やグループ化を行う必要がある．そのひとつの方法として，IP アドレス数の削減方法が考えられる．IP アドレスは，第 1 オクテッドから第 4 オクテッドで構成され，現在はネットワークログデータに存在している IP アドレス（第 4 オクテッドレベル）のすべてを表示している．そこで今後の展開として，任意のオクテッドレベルで表示可能にすることによって情報の重なりの軽減することが可能になる．また，この可視化方法の拡張案として，より多くの情報を持たせられる多次元表示方法であるパラレルコーディネーションの適用も必要かと思われる．

参考文献

[1] Surver, Information Visualization http://apollo. u-gakugei. ac. jp/~yoshiki/research/survey-iv. php

[2] G. G. Robertson, J. D. Mackinlay and S. K. Card,Cone Trees: Animated 3D visualizations of hierarchical information, In Proceedings of the ACM Conference on Human Factors in Computing Systems(CHI'91), pp. 189--194. ACM Press, 1991.