Cubic Gantt Chart によるネットワークトラフィックデータの可視化

中村昌平[†1]　岡田義広[†1, 2]

概要：著者らは、プロジェクト管理や生産管理における工程管理に用いられる表であるガントチャートの機能を３次元化した Cubic Gantt Chart とよぶ可視化ツールを提案している。本論文では、ネットワークトラフィックデータを可視化した例を示しその有用性を述べる。Cubic Gantt Chart は３つの軸を持っており、それぞれの軸には任意の属性を割り当てることができる。近年、スマトーフォンや SNS の普及によりネットワークの使用も増加している一方で、サイバー攻撃も増加している。これらのサイバー攻撃の兆候を予知する手段の一つとしてダークネットのトラフィックデータを可視化解析するという手法が挙げられる。本論文では時間、送信元 IP、宛先 Port と送信元 IP、宛先 IP、宛先 Port のそれぞれ三つの属性について可視化を行った例を示し、サイバー攻撃と考えられる通信の発見に有用であることを述べる。

キーワード：データビジュアライゼーション、ネットワークデータ、ネットワークセキュリティ

Visualization for Network Traffic Data Using Cubic Gantt Chart

SHOHEI NAKAMURA[†1]　YOSHIHIRO OKADA[†1, 2]

Abstract: The authors have proposed a new visualization tool called Cubic Gantt Chart, 3D version of Gantt Chart. This paper treats Cubic Gantt Chart as a visualization tool for network traffic data. Cubic Gantt Chart has three axes in a 3D space, and then, it can be used for visualization of three-attributes data. Recently, while the Internet has become popular accordingly to the growth of smart phones and SNS users, cyber-attacks have also been increasing. One of the ways to detect any symptom of these cyber-attacks is to analyze dark-net traffic data by visualization. This paper shows several visualization examples of dark-net traffic data using Cubic Gantt Chart for three-attributes of each packet, i.e., time, source IP and destination Port or source IP, destination IP, destination Port, and then, this paper describes the usefulness of Cubic Gantt Chart as a visualization tool for network traffic data.

Keywords: Visualization, Network Data, Network Security

1. はじめに

近年、ネットワークの使用率は増大し、それに伴いサイバー攻撃も増加している。ノートンレポートによれば、サイバー攻撃による被害総額は 1130 億ドルで被害を受けた人は 100 万人を超える。日本では約 10 億ドル、アメリカでは 380 億ドルとなっている[1]。最初のコンピュータウイルス Brain が登場して約 30 年、日々新しいウイルスの開発やサイバー攻撃が行われている。Brain は不正なソフトウェアのコピーを警告するだけではなく、コンピュータウイルスの脅威をも知らせる結果となった。サイバー攻撃は SQL インジェクション、クロスサイトスクリプティング、Heart Bleed といったソフトウェアやシステムの脆弱性を狙ったものが多い。脆弱性が発見されると修正パッチが配布されるが、脆弱性が発見されてからその修正パッチが配布される間に行われるゼロデイ攻撃という脅威も存在する。その他の脅威としては標的型攻撃やフィッシングといったものがあげられる。ウイルスを含んだメールをユーザが開いてしまうとパソコン上の重要なデータが盗まれてしまう。これはウイルス対策ソフトだけでは防げない攻撃である。攻撃者はいついかなる時もユーザの PC に対してウイルスを感染させようとしている。これらの理由により今日ではサイバーセキュリティはとても重要なものになっている。

ネットワークトラフィックデータを取得し、解析することはセキュリティにおいてとても有益な手法である。しかし、ネットワークデータはとても大きく解析するのはとても難しい作業である。そのための一つの手法が可視化による解析である。われわれは、すでに、学習者の学習履歴データの可視化ツールとして、Cubic Gantt Chart[2]を提案している。本論文では、ネットワークトラフィックデータの可視化ツールとして Cubic Gantt Chart を使用した可視化結果を紹介し、有用性を述べる。

Cubic Gantt Chart は多次元データを可視化する上で有効なツールである。ユーザはデータの持つ属性のうち３つを選択し、3 次元空間の X, Y, Z 軸にそれぞれ割り当てることで、X-Y, Y-Z, Z-X 平面といった任意の方向からその属性の振る舞いを理解することができる。本論文では、X, Y, Z 軸

†1 九州大学, ,福岡県福岡市西区元岡 744 番地
Kyushu University, Motooka 744, Nishi-Ku, Fukuoka, 819-0395, JAPAN
†2 九州先端科学技術研究所, 福岡県福岡市早良区百道浜 2 丁目 1 番 22 号 福岡 SRP センタービル 7 階
Institute of Systems, Information Technologies and Nanotechnologies (ISIT), Fukuoka SRP Center Building 7F, Momochihama 2-1-22, Sawara-ku, Fukuoka, 814-0001, JAPAN

にそれぞれ時間、送信元IP、宛先Portを割り当てた場合と、送信元IP、宛先IP、宛先Portを割り当てた場合について可視化を行った。

本論文の構成は次のとおりである。まず、2章では可視化手法やサイバー攻撃に関する関連研究を紹介する。3章では今回使用したダークネットとネットワークデータについて説明する。そして4章では提案する可視化ツールCubic Gantt Chartの基本的な機能について説明する。5章では3章で説明したネットワークデータを提案した可視化ツールで実際に可視化した例を紹介する。そして最後の6章でまとめと今後の展望を述べる。

2. 関連研究

この章では関連研究について紹介する。関連研究として記述する内容はデータの可視化手法について、サイバー攻撃についてである。可視化手法については、これまでに様々なものが考案されてきた。[3]　ここでは、一般的に使われている可視化手法とネットワークデータについての可視化手法それぞれについて紹介する。サイバー攻撃についても様々な攻撃が行われてきている。ここでは、協調型攻撃について主に説明したい。

2.1 データの可視化

これまでにデータの可視化手法は様々開発されてきている。例えば、Cubic Gantt Chartの2次元版であるGantt Chart[4]、階層化されたデータを可視化するのに多く用いられるTreemap[5]、そして多次元属性を可視化するのに多く用いられるParallel Coordinate[6]などである。このように多くの可視化手法が提案されているが、時間に依存したデータを可視化することは容易なことではない[7]。

Treemap[5]は可視化ツールの中でもよく使用されるものの一つである。これはBen Shneidermanによって提案された可視化手法である。この可視化ツールは入れ子構造となった四角形によって階層化されたデータを表示する。それぞれの四角形の面積が対応するノードのもつ値に比例している。そのため、ユーザは四角形の大きさを見るだけで値の関係性を理解することができる。四角形の色は別の情報を表現するために用いることができる。

図 1　Treemap (D3.jsのサンプルから引用)

Figure 1 Treemap (from D3.js example)

Parallel Coordinate[6]は多次元データを可視化する際に最も多く使用される可視化ツールである。これはAlfred Inselbergによって提案された。Parallel Coordinateは2次元平面上に平行に多数の軸を描く。その一つ一つの軸に一つの属性をそれぞれ割り当てる。それぞれの軸の点を結ぶことによって折れ線グラフを作成する。一つの折れ線が一つのデータを表すことになる。これによって、ユーザはそれぞれのデータの全属性間の関係性を俯瞰的に理解することができる。

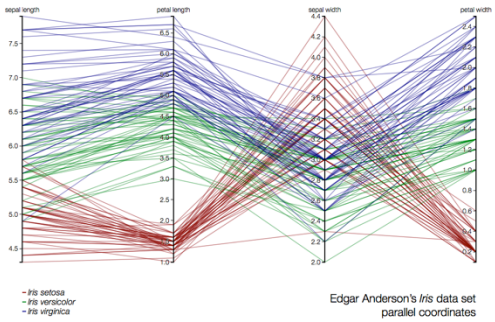

図 2　Parallel Coordinate (D3.jsから引用)

Figure 2 Parallel Coordinate (from D3.js example)

Time-tunnel[8, 9]はAkaishiらによって提案された時系列データを可視化するツールである。Time-tunnelはData-wing, Time-plane, Time-barとよばれる主に3つの部品から構成される。Time-barは細長い円柱のような形をしており、時間軸を表している。Data-wingはシートのような形をしている。一つの時系列データがData-wing上に折れ線グラフとして表示される。それぞれのData-wingはTime-barに接続しており、Time-barを軸にして回転するようになっている。これはData-wingを回転させて重ね合わせることで、時系列データを比較するためである。Time-planeもシートのような形である。Time-planeはData-wingに対して垂直に接続する。これは、ある任意の時間ですべてのデータのレーダチャートを表示するために使用される。

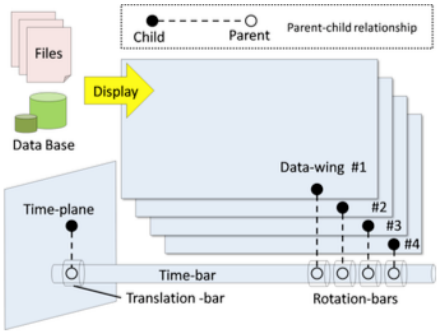

Figure 1: Component structure of Time-tunnel.

図 3　Time-tunnel（[9]から引用）

Figure 3 Time-tunnel (from [9])

ネットワークデータを可視化するものとして代表的なオープンソースツールGephi[10]を紹介する。Gephiは Force

Atlas2 という力指向のアルゴリズムを採用した force layout の一種である。Gephi ではリアルタイムに大規模なグラフを可視化することができ、またフィルタリング機能も備えている。

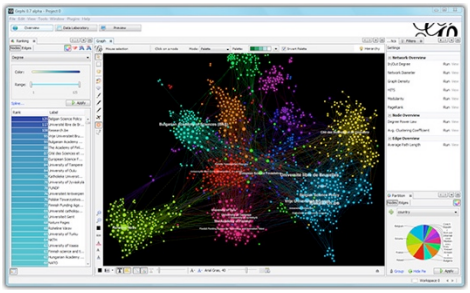

図 4　Gephi 公式サイトから引用
Figure 4 Gephi official site

2.2 サイバー攻撃（分散型攻撃）

サイバー攻撃についても様々な手法が登場している。すでに挙げた脆弱性をついた攻撃である SQL インジェクションやクロスサイトスクリプティングやある特定の人物に向けての標的型攻撃やフィッシングなどがある。このほかにも DoS 攻撃、その派生形の DDoS 攻撃なども存在する。

DoS 攻撃（サービス不能攻撃）とは、TCP Flood 攻撃や UDP Flood 攻撃といったものが存在する。いずれも通信リクエストを攻撃対象の PC に送る。このとき、リクエストを受け取った PC は SYN/ACK と呼ばれる通信を許可するリクエストを通信を許可する送信元に対して送信する。正常な通信であれば、送信元の PC は ACK を返して通信が開始される。しかし、DoS 攻撃では ACK を返さずに攻撃する PC のリソースが枯渇するまでコネクションリクエストを送信する。このようにして、攻撃対象の PC がサービスを提供できないようにする。

DDoS 攻撃とは Distributed DoS 攻撃と呼ばれる分散型の攻撃である。攻撃方法としては DoS 攻撃と同じように大量のリクエストを送信することで攻撃対象の PC のサービスを提供できなくするという点では同じであるが、攻撃対象の PC にリクエストを送信する PC が複数台あるという違いがある。この攻撃には、ボットネットを用いられることが多い。ボットネットとはマスタと呼ばれる攻撃者が使う PC とウイルスに感染したスレーブという複数台の PC によって構成される。攻撃者が操作するマスタがスレーブに対して攻撃する PC に対して一斉に通信を開始するように指示を出す。これによって攻撃する PC のサービスを停止させる。

3. ダークネットデータ

本研究は、総務省による「国際連携によるサイバー攻撃の予知技術の研究開発」プロジェクトの一つとして実施している。本プロジェクトで収集しているダークネットトラフィックデータを対象に Cubic Gantt Chart での可視化を試みた。

ダークネットとは、インターネット上で到達可能かつ未使用の IP アドレス空間のことを指す[11]。通常は、ホストが割り当てられていないアドレス空間に対して通信を行うことはない。しかし、ダークネットのデータを監視してみると、実際には相当数のパケットが到達している。井上氏ら[11]によれば、それらのパケットは

- リモートエクスプロイト型マルウェアが次の感染対象を探査するためのスキャン
- マルウェアが感染対象の脆弱性を攻撃するためのエクスプロイトコード
- 送信元 IP アドレスが詐称された DDoS 攻撃を被っているサーバからの応答であるバックスキャッタ

などが含まれているという。これらのパケットはインターネット上で不正な活動をしているということができる。ダークネットに来るパケットは、ポートスキャンや DDoS 攻撃のもののほか、Peter 氏らによれば、ファイル共有のためのものも有る[12]。

ダークネットには上で述べたパケットが大量に届く。これらのトラフィックを解析することにより、現在流行しているマルウェアや攻撃の動向などを把握することが可能になると考えらえる。

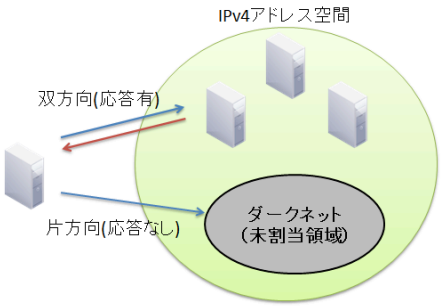

図 5　ダークネットイメージ図（[13]より引用）
Figure 5 Image of Dark-net (from [13])

4. Cubic Gantt Chart

この章では、提案する Cubic Gantt Chart[2]について説明する。Cubic Gantt Chart は Gantt Chart を 3 次元化したものである。提案する可視化手法の詳細を説明する前に、通常の Gantt Chart について説明する。

4.1 Gantt Chart

Gantt Chart[4]とは 1910 年代に Henry Gantt によって考案

された棒グラフの一種である。Gantt Chart はプロジェクト管理や生産管理などにおける工程管理に用いられる表である。縦の軸にそれぞれの仕事の種類を書き出し、横の軸に時間の流れを記述する。それぞれの仕事は四角形によって表現される。左側と右側の辺はそれぞれ四角形によって表されている仕事の始まりと終わりの時間を示している。ユーザは Gantt Chart を見ることで次に示すことを一目で理解することができる。

- どのような仕事が存在するのか
- それぞれの仕事がいつ始まり、いつ終わるのか
- それぞれの仕事はどのくらいかかるのか
- 仕事が他の仕事と重なっている時間が存在するか
- プロジェクトがどのぐらい進んでいるのか
- いつプロジェクトが始まり、いつ終わるのか

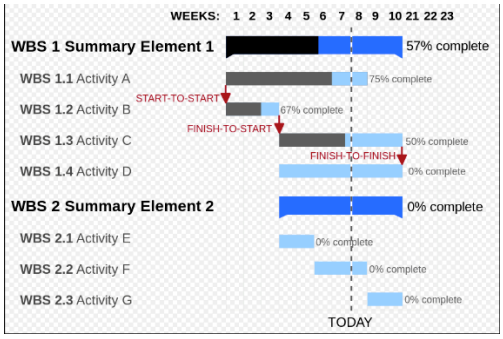

図 6　Gantt Chart（ウィキペディアから引用）

Figure 6 Gantt Chart (from Wikipedia)

4.2 Cubic Gantt Chart

この提案する可視化手法は２つの属性を扱う Gantt Chart に比べ､軸を一つ増やして３次元空間を利用しているため、３次元の属性を扱うことができる。３つの属性はそれぞれ３次元空間の X, Y, Z 軸に割り当てる。この可視化ツールを X-Y 平面、Y-Z 平面、Z-X 平面といった任意の方向から見ることによって、ユーザはデータが持つ対応する属性の関係性を理解することができる。

ユーザは透視投影と正方投影を選択して可視化ツールを見ることができる。透視投影は全体を俯瞰するときに適している。一方、正方投影は X-Y 平面、Y-Z 平面、Z-X 平面を見ることによって、ある二つの属性についての関係性を理解するのに適している。

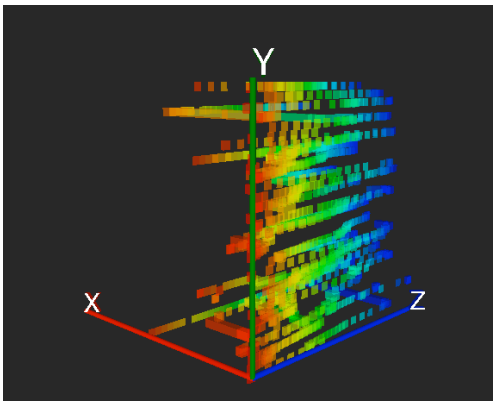

図 7 透視投影での Cubic Gantt Chart

Figure 7 Perspective view of Cubic Gantt Chart

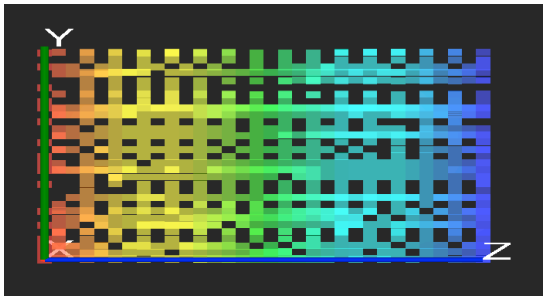

図 8 正方投影での Cubic Gantt Chart

Figure 8 Orthogonal view of Cubic Gantt Chart

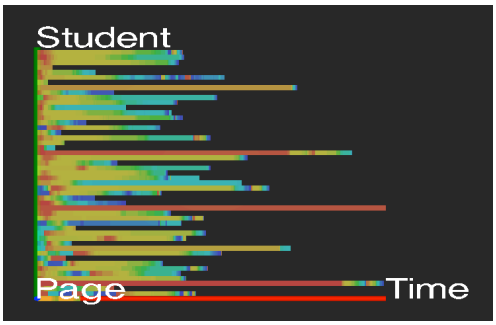

図 9 正方投影での生徒と時間の関係

Figure 9 Orthogonal view toward plane of Student and Time

このグラフに配置されているそれぞれの voxel は半透明である。これは配置された voxel が重なることで色が濃くなり情報の大きさを表すためである。正方投影を選択することで、ユーザはヒートマップと同じようにそのデータの対応する属性値の大きさを見ることができる。それぞれの voxel は着色されている。e-learning の学習ログを可視化した際には Cubic Gantt Chart の X, Y, Z 軸それぞれに時間、生徒の ID、ある一回の授業で使用されるスライドのページ番号を割り当てた。赤色の voxel ははじめのページ番号を表し、青色の voxel は最後のページ番号を表しており、その間は赤から青へとグラデーションしている。そのため、ユ

ーザは Y-Z 平面、Z-X 平面を見ることで、生徒がどのように授業で使用されたスライドを閲覧していったかを見ることができる。

それぞれの voxel の長さは２次元版 Gantt Chart の時間のような連続値であれば、対応する属性方向にその値と同じ長さとなっている。離散値であればその値が対応する属性場所に意味をもたせて長さは１としている。例えば、X 軸が離散値であるとすると、最初の場所に一人目の生徒の情報を、二番目の場所に二人目の生徒の情報を表すといったようにである。２次元版の Gantt Chart は縦軸に仕事の種類である離散値を、横軸に時間の流れである連続値を表していることと同じである。

4.3 開発環境

提案する可視化ツールは数あるプログラミング言語の中から Javascript を選択した。大きく２つの理由がある。１つは、web ベースの可視化ツールとすることでウェブブラウザが使用できれば端末の OS やプラットフォームに依存することなく、また特別なソフトウェアをインストールすることなくして使えるからである。２つ目の理由としては Three.js, jQuery.js といったライブラリ群が多く存在することである。

5. 可視化ツールとその可視化例

この章では前章で説明した可視化ツール Cubic Gantt Chart を用いてネットワークトラフィックデータを実際に可視化した例を紹介する。この可視化ツールでは X, Y, Z 軸にそれぞれ時間、送信元 IP、宛先 Port をそれぞれ割り当てている。配置される voxel の Y 軸の場所は 0 には送信元 IP が 0.0.0.0 を割り当て、１には 0.0.0.1 を割り当てる。Z 軸の Port に関しても同様である。しかし、従来の Cubic Gnatt Chart では縦、横、高さの軸に大きな数を割り当てて可視化を行うと一つひとつの voxel が潰れてしまい見ることが困難になってしまった。そのため、提案した Cubic Gantt Chart を階層的にすることにした。送信元 IP が 0.0.0.0 から 0.255.255.255 のものを Y 軸の 0 の点に配置し、1.0.0.0 から 1.255.255.255 のものを Y 軸の１の点に配置するといったように voxel を配置した。Z 軸に関しても同様に Port 番号が 0 から 255 のものを Z 軸の 0 の点に 256 から 511 のものを Z 軸の１の点に配置した。この方法により配置された voxel が小さくて見えなくなる問題を解決した。配置された voxel をクリックすることで、その voxel に含まれるデータを詳細表示し、ドリルダウン式にネットワークデータの振る舞いを見ることができるように可視化を行った。色情報は、今回は TCP, UDP の通信に対して黄色、赤色を割り当てることにした。もともとはデータ数で青色から赤色へとグラデーションする仕様としていたのだが、通信量の大きさをどこから大きいとするかを明確に定めることが困難であったため、色の使い方を変更した。

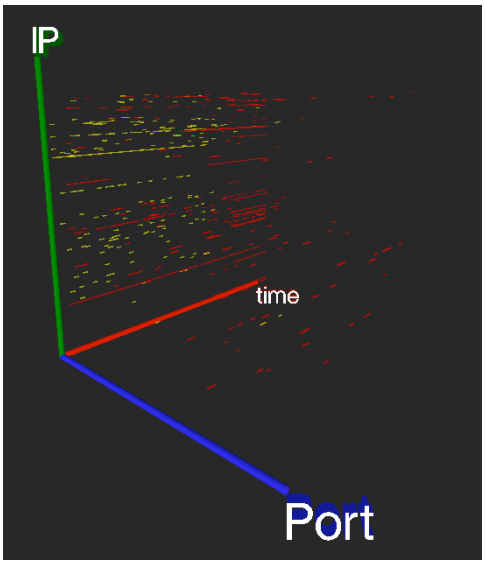

図 10　2013 年 4 月 27 日のデータを可視化したもの
Figure 10 Visualization example (data is Spr. 27th in 2013)

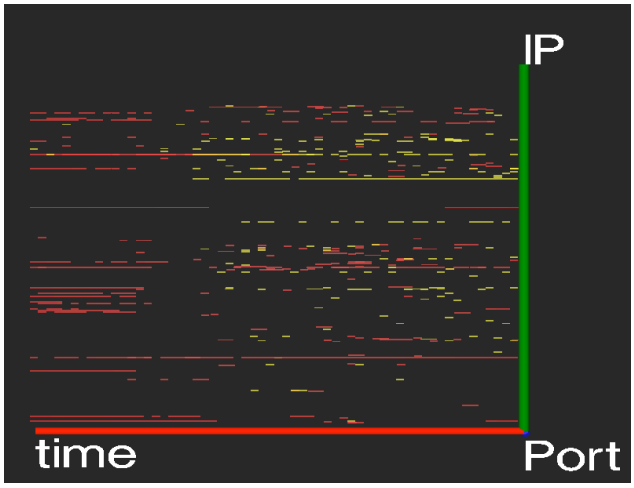

図 11　図 10 を X-Y 平面で正方投影したもの
Figure 11 Orthogonal view toward X-Y plane of Fig 10

図 10, 11 に関しては、2013 年 4 月 27 日の 1 時間分のデータを可視化したものである。ここで見られた特徴としては、時間と送信元 IP との関係を見るとある送信元 IP から連続的に限定的なポートに対して通信が行われていたことが分かる。図 11 に見られる時間軸と平行になって現れている線である。Voxel をクリックすることでドリルダウン式にネットワークデータの振る舞いを理解することができる。今回は図 12 に示した voxel の詳細なデータを見ることにした。

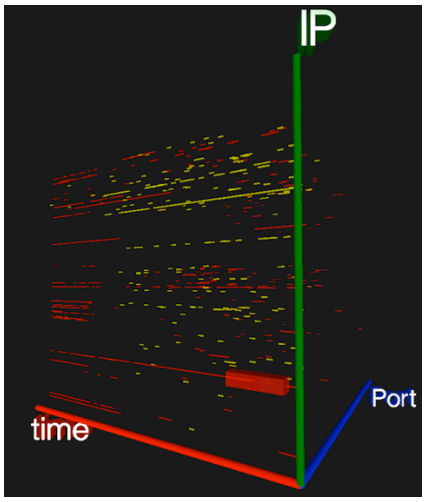

図 12　選択された voxel
Figure 12 Selected voxel

詳細なデータを確認すると、Port 番号 2623 に対して送信元 IP が 46.271.66.246 からのアクセスということが見ることができた。この送信元 IP が直接 Port 番号 2623 に攻撃をしてきているのか、マルウェアによって操作されているのか判断はできないがこのホストからの通信は危険であること考えられる。

Cubic Gantt Chart は任意の連続値、離散値をそれぞれの軸に割り当てることができる。別の可視化例として送信元 IP、宛先 IP、宛先 Port を割り当てたものを紹介する。図 13, 14, 15 は 2015 年 5 月 28 日の 1 時間分のデータを可視化したものである。図 13 では先ほどの例と同じ属性を割り当てたものである。そして図 14, 15 では時間の属性を宛先 IP に変更して可視化したものである。

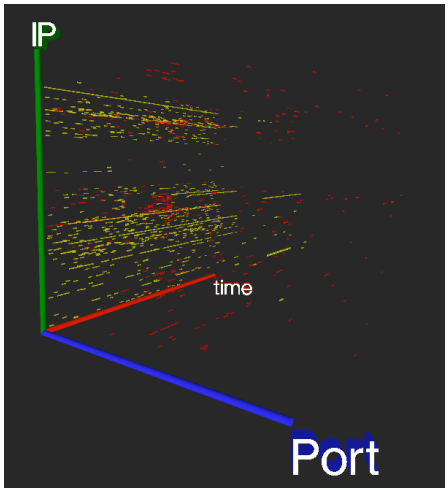

図 13　2014 年 5 月 28 日のデータを可視化したもの
Figure 13 Perspective view (data is May. 28th 2014)

図 13 の例もある特定のポートに対して連続的にアクセスが来ていることが分かる。

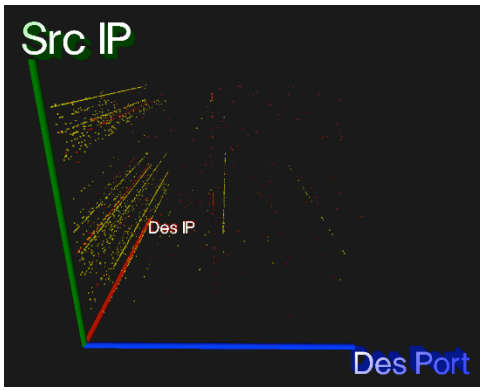

図 14 図 13 の時間属性を宛先 IP に変更したもの
Figure 14 Visualization example by changing Time attribute to Destination IP

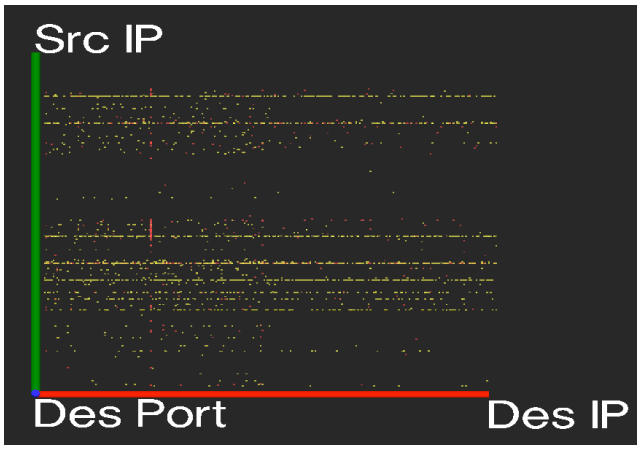

図 15　図 14 を X-Y 平面で正方投影したもの
Figure 15 Orthogonal view toward X-Y plane of Fig 14

図 14, 15 を見ると明確にあるポートに対してホストスキャンをしていることが分かる。脆弱性のあるポートに対して、その対策が施されていないホストを探していると考えられる。また図 15 の中段やや左に赤い線が縦に伸びている。これは送信元 IP 群から当該のホストに対して同じ宛先 Port でアクセスが来ていることが分かった。詳細なデータを見ると送信されたパケットは多くないため、DDoS 攻撃とは言えないが、それに近いことが行われていると考えられる。

6. まとめと今後の課題

本論文では、われわれが既に提案しているガントチャートの機能を 3 次元化した Cubic Gantt Chart とよぶ可視化ツールによってダークネットトラフィックデータの可視化を行った結果を報告した。時間、送信元 IP、宛先 Port の 3 属性と送信元 IP, 宛先 IP, 宛先 Port の 3 属性の関係について可視化することに成功した。全体のネットワークの振る舞いを確認し、怪しいと思われる通信をドリルダウン式に確認していくことでどこからの通信でどのような Port に対し

て通信を行っているのかを確認できるようになった。

今後の課題としては、別の場所にあるダークネットのトラフィックデータについて、今回の結果と同じ時間帯で可視化を行ってみることである。今回使用したデータは違う場所のダークネットで同じ時間帯のデータの数が少なく、可視化を行うことができなかった。これらのデータを取得し、可視化を行うことで、同じような送信元 IP から同じような Port に対して通信をしていることが可視化できればその送信元 IP 群はボットネットであると判断できると思われる。また、現在この可視化ツールは一度に複数のウインドウを開いて同期させて動かすことができない。様々な場所にあるダークネットのトラフィックデータを可視化する際に一度に複数のウインドウを開き見ることができた方がユーザにとってデータの比較をする際、分かりやすいものになる。今後、複数の可視化ツールの同時表示・操作機能も実装したい。

謝辞

本研究の一部は、総務省による「国際連携によるサイバー攻撃の予知技術の研究開発」プロジェクトの支援を受けたものである。

参考文献

[1] ノートンレポート http://www.symantec.com/content/ja/jp/about/presskits/2013_Norton_Report.pdf

[2] Shohei Nakamura, Kosuke Kaneko, Yoshihiro Okada, Chengjiu YIN, Hiroki Ogata : Cubic Gantt Chart as Visualization Tool for Learning Activity Data, ICCE2015 e-Book workshop

[3] Herman, Ivan, Guy Melançon, and M. Scott Marshall. "Graph visualization and navigation in information visualization: A survey." Visualization and Computer Graphics, IEEE Transactions on 6.1 (2000): 24-43.

[4] James M. Wilson. “Gantt charts. A centenary appreciation” European Journal of Operational Research 149 (2003) 430 – 437

[5] Shneiderman, B. (1992). Tree Visualization With Treemaps: A 2-D Space-Filling Approach. ACM Transactions on Graphics, 11(1), pp.92-99.

[6] A. Inselberg and B. Bimsdale, “Parallel Coordinates: A Tool for visualizing Multidimensional Geometry.” Proc. IEEE Visualization 1990, IEEE CS Press, pp 361-378, 1990

[7] Aigner, Wolfgang, et al. "Visual methods for analyzing time-oriented data." Visualization and Computer Graphics, IEEE Transactions on 14.1 (2008): 47-60.

[8] Akaishi, M. and Okada, Y. : Time-tunnel: Visual Analysis Tool for Time-series Numerical Data and Its Aspects as Multimedia Presentation Tool, Proc. of 8th International Conference on Information Visualization (IV04), IEEE CS Press, pp. 456-461, London UK, July 2004.

[9] Okajima, S., Okada, Y.: Interactive Analysis of Multidimensional Data on the Web by Using Time-tunnel, Proc. of 5th International Conference on Web Information Systems and Technologies (WEBIST 2009), INSTICC Press, pp. 415-418, March 23-26, 2009.

[10] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy : “Gephi: An Open Source Software for Exploring and Manipulating Networks” In Proceedings of 3rd International AAAI Conference on Weblogs and Social Media, 2009. 361-362

[11] 情報セキュリティ技術動向調査（2008 年下期） https://www.ipa.go.jp/security/fy20/reports/tech1-tg/2_07.html

[12] Biddle, Peter, et al. "The darknet and the future of content distribution." ACM Workshop on Digital Rights Management. Vol. 6. 2002.

[13] 高柳涼, 岡田義広 “Treemap と Edge Bundling を利用したダークネットデータの可視化システムの提案” 情報処理学会研究報告