多地点断片ダークネットのための統合データ解析ツールの開発

廣津登志夫 豊橋技術科学大学†
塩野祐輔 豊橋技術科学大学†
福田健介 国立情報学研究所‡
菅原俊治 早稲田大学§

1 はじめに

現在のインターネット環境では，機密情報の意図しない流出やサービス提供サイトに対する DDoS 攻撃等によるサービスの停止など問題が生じている．このようなインターネットにおける攻撃に対処するためには，攻撃情報を観測・収集して，それらを解析・モデル化・類型化することで，実際に行われている様々な攻撃に対処することが重要である．来るべき情報爆発時代のネットワークにおいては，有用な情報だけ無くこのようなスキャン・侵入・DDoS(通信不能) 等の各種の攻撃も増大することを予想せざるをえず，また守る対象である情報空間自体も爆発的に広がるため，それを十分に守りうるスケーラブルな監視・運用環境が非常に重要になる．

このようなスケーラブルな監視・運用環境を目指して，現在，分散協調型監視アーキテクチャ(図 1) に関する研究をすすめている．ここで目指しているのは，従来の Darknet[1] のように /8 規模の大規模アドレス空間を監視するのではなく，各組織が割り当てられたネットワークのアドレス空間のうち使われていない一部のアドレス空間をを用いて小規模な断片 Darknet を構成し，複数の断片 Darknet が協調することで全体として広いアドレス空間の監視を実現するものである．ここで，断片 Darknet 間を連携・協調させて攻撃情報を収集するというのは従来の Darknet に無い視点であり，それぞれの断片 Darknet がどのように協調を行うのが良いかはよくわかっていない．例えば，それぞれの拠点で所有しているネットワーク空間のどの部分をどの程度の広さ監視空間に割り当てれば良いか (つまり，どのようにサンプリングすれば良いか) は分散協調型の監視アーキテクチャの設計運用上は非常に重要な問題となるが，従来は適切なサンプリング空間を設定するという観点では解析されておらず，収集した攻撃性情報を様々な観点から解析し特徴を把握する必要が生じる [2]．本研究ではそのような

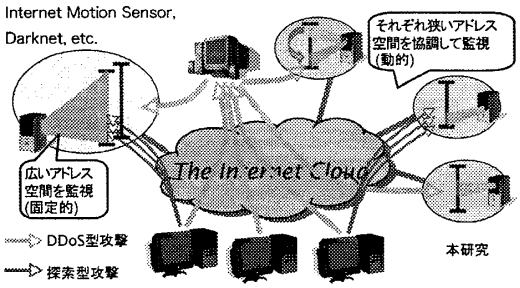

図 1: 分散協調型監視アーキテクチャ

トラフィック情報の多様な解析を助ける統合データ解析ツールを開発したのでその設計と実装について述べる．

2 攻撃性情報の解析

攻撃性トラフィックデータから特徴量や傾向を把握する際には，通常以下のような手順を踏む．

1. tcpdump[3] や pcap ライブラリ，script 言語等を使用して，特定の条件 (時刻，プロトコル情報等) に合ったデータだけ抽出するプログラムを作成し，抽出したデータに関する統計情報を求める．
2. 得られた情報を表やグラフにして，特徴的な挙動を把握する．
3. 特徴的な部分や挙動がよくわからない部分だけを取り出す条件 (ルール) を考えて，1. から繰り返す．

このとき，情報抽出プログラムはルールの部分の多様性は多いが，処理としては比較的類似したものであることが多く，また，統計処理も典型的に使われるものは限られる．さらに 3. の段階では，得られたグラフ等の視覚的表現の中の特定部分 (特徴的な部分や挙動不明の部分) に関する情報を取得したり，その部分の情報を取り出すルールを取得したいという要求が大きい．

従来よく使われているツールとしては tcpdump や wireshark が挙げられるが，これらのツールは収集したトラフィックデータに対して，特定の条件を設定して抽

Design and Development of the Collective Analysis Tool for Fragmented Darknets
†Toshio Hirotsu and Yuusuke Shiono, Toyohashi University of Technology
‡Kensuke Fukuda, National Institute of Informatics
§Toshiharu Sugawara, Waseda University

出を行ったり単純な表示をするには適しているが，処理結果から特定部分を指定してさらに処理を繰り返すという部分の処理に対しては特に機能的な支援はない．これに対して，「解析」「表示」「抽出」という手順を繰り返す解析において，表示された視覚的表現を直接指示することによって，指示部分のトラフィック情報が取得できると攻撃の特徴や傾向の把握を助けることができる．また，表示したデータを視覚的に指示することで次の処理に使うデータを抽出できると，あたかも電卓で計算処理を指定しながら答えを求めるかのように容易にトラフィックデータの解析処理を進めることができる．

3 設計と実装

前節に述べたような処理を実現する上で重要なのは，解析し表示した後に表示情報から視覚的に範囲指定してデータを抽出するフェーズで，その指定範囲に該当する解析の元データを取り出さねばならないことである．[4]においては，データの保存期間や解析に必要とされる用途に応じて，長期的なデータは用途に適した形での程度集約をかけてある程度記憶容量を抑制するようなことが行われているが，ここで対象としている断片 Darknet の連携のためのデータ解析では，そのような集約は困難である．もちろん，解析ツールの側面からは処理するデータ量が小さい方が応答性等の点で利点があるのは明らかであるが，着目すべき情報自体がよくわからず収集した攻撃性情報を様々な観点から解析することで特徴を把握するという目的からは，特定の集約によって必要とされる情報が隠れてしまうことを避けるために，元のデータに戻りうる機能を持たせる必要がある．そこで，多地点にまたがる断片 Darknet のデータを解析することも考慮して，解析ツールを以下の三つの部分から構成した．

バックエンド 元のトラフィック情報の全てを持つデータストレージ．今回は，SQL データベースに「パケットの取得時刻」「宛先アドレス・ポート」「送信元アドレス・ポート」を蓄積した．

データマップ 解析して集計した各々の結果 (すなわち表示するデータ) と (解析や集計の) 元データを取り出すためのキーの対応を保持する．

フロントエンド GUI ベースで表示すると共に，ユーザの範囲指示を読み取り，データマップの情報から元データを取り出す処理を行う．

解析ツールはまず指定した範囲のデータをバックエンドから取り出し，解析・集計結果を表示する．そして GUI 操作で指定されたら，その範囲に該当するデータをデータマップの情報を元にしてバックエンドに要求する．そしてバックエンドから得られたデータに，ユーザが望む (多くの場合その前とは別の) 解析・集計を施すことで，段階的に情報の解析を行うことができる．以上の機能を持ったツールを QT4 と PostgreSQL を用いて作成した．そのスクリーンショットを図 2 に示す．

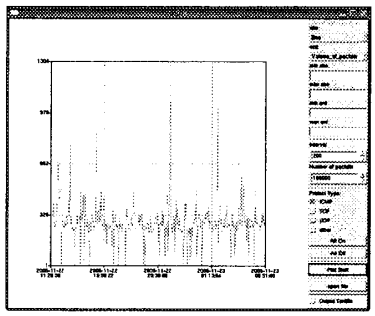

図 2: スクリーンショット

4 まとめ

本稿では，断片ダークネットのためのトラフィックデータ解析ツールの設計と実装について述べた．現状では，7 千万パケットの処理に 500 秒以上の時間がかかっているが，特定の目的のために解析処理プログラムを作成する手間と比べると処理の効率は向上している．処理時間の解析からこの時間の殆どはデータベースから必要なデータを取り出すために使われていることが判明しており，データベース構造の見直しによる高速化でさらに応答性を高めることが今後の課題である．

謝辞

この研究は科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」の支援を受けている．

参考文献

[1] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of Internet background radiation", *4th ACM SIGCOMM conference on Internet measurement*, pp. 27–40 (2004).

[2] 廣津登志夫，福田健介，栗原聡，明石修，菅原俊治，"断片アドレスを用いた分散協調インターネット監視に関する一考察"，情報処理学会 OS 研究会研究報告，Vol. 2007, No. 83 (2007-OS-106), pp.39–45, (2007).

[3] V. Jacobson, C. Leres and S. McCanne, "tcpdump", available via anonymous ftp to ftp.ee.lbl.gov, June 1989.

[4] E. Cooke, A. Myrick, D. Rusek and F. Jahanian, "Resource-Aware Multi-Format Network Security Data Storage", *ACM SIGCOMM LSAD'06*, pp. 177–184 (2006).