NECOMA

MATATABI:マルチレイヤ脅威分析基盤

マルチレイヤ分析とクロスレイヤ分析

我々は、単に複数データセットを対象として分析をするクロスレイヤ分析と、分析結果の相関をとり新たな分析を見いだすマルチレイヤ分析を区別しています。

ビッグデータ技術

並列・分散計算技術や、大容量データ格納技術を組み合わせ、従来のデータベース技術では困難とされてきた高速なデータ検索・分析を可能とする技術の総称です。

ビッグデータ基盤による脅威分析

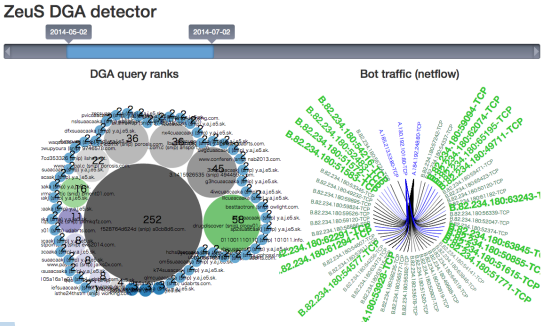
多種多様なサイバー攻撃による被害が拡大していく中で、その対策を講じるために攻撃そのものを理解する事は非常に重要です。攻撃は点描のようなもので、場当たり的な観測では攻撃の概要をとらえる事は一般に困難です。

NECOMA プロジェクトの挑戦として、「マルチレイヤ分析」と呼ばれる手法を採用する事で、分析対象とする情報を広くし攻撃の理解を深め、その結果得られる知見をデータベース化する事を掲げています。このために多種・多量の観測データを、容易に分析するための技術基盤としてビッグデータ技術を適用しています。

Apache Hadoop による実装

マルチレイヤ分析では、分析対象のデータ量が膨大となるため、処理技術に工夫が必要となります。我々は大規模データの分散処理を実現するオープンソースソフトウエア Apache Hadoop や、Facebook Presto-db を利用した脅威分析基盤 MATATABI を設計・構築し、運用しています。

更に分析結果を公開し、可視化し全体俯瞰等を行うために、ポーランド NASK で開発されている n6 SDK を用いた情報交換等を実現し、知識ベースとして利用可能としています。



クラウド型ではなく、ローカル型

MATATABI は、Google Cloud Platform や Amazon EC2 等で提供されるクラウド型データ解析サービスではなく、分析対象の観測データを保持している各組織内にて構築する事ができる、データ分析ソフトウェアです。そのため、プライバシ侵害の恐れのあるデータを組織外に持ち出す事なく脅威分析が可能です。また、大量のデータを高速に分析する必要があれば、Apache Hadoop の運用と同様に計算ノード台数を増加させる事により、容易に規模を拡大させる事が可能です。

複数の分析手法による脅威検知

「より多くの情報を分析すると、より多くの攻撃を検知できる」という仮定の下、バックボーンネットワークのトラフィック、DNS 問い合わせ記録、スパムメール、ユーザの行動追跡、フィッシングサイトデータベースなど、大量に計測される情報を一元的に、適切な速度で分析する仕組みが必要とされます。分析の着眼点として、NECOMA プロジェクトでは複数のデータセットを対象とする事で、より深い分析を実現する手法（クロスレイヤ分析）と、分析結果同士の相関をとる事で、より攻撃の実体を詳しく洗いだす手法（マルチレイヤ分析）をとっています。

それぞれの分析は各攻撃や観測データ毎に利用シーンを変えて、幅広い分析を可能としています。

NECOMA

MATATABI:マルチレイヤ脅威分析基盤

MATATABI ソフトウェアの利用方法 (Docker イメージ)

MATATABI は、導入手順を簡素化するために、Docker[*1] というソフトウェアを利用して提供しています。ここでは、Ubuntu 14.04-01 x86 64-bit 環境を利用した MATATABI の導入手順を紹介します。

まず、Docker をインストールします。

```
% sudo apt-get install docker.io
```

sudo コマンドが実行できる権限が必要です。

その後、Docker hub[*2] で公開されている MATATABI イメージをインストール(ダウンロード)します。

```
% docker pull necoma/matatabi:1.0
```

これで、MATATABI のインストールは終わりです。コンテナ技術を利用する事で、複雑な依存関係のあるプログラム群や、それに付随する設定ファイルなどを一つのイメージに封じ込める事ができ、大変便利です。

実行してみましょう。docker run コマンドを利用すると、MATATABI が動作している OS の起動と初期化を行います。

```
% docker run -i -t necoma/matatabi:1.0
```

起動が成功したら、presto[*3] というコマンドを実行して、インストール時に生成してあるサンプルデータを SQL 文で検索する処理を実行してみましょう。

```
root@a6a76dd13275:/# presto
presto:default> select date,ipaddr,qname,cname,typename from querylog;
     date     |     ipaddr      |      qname       | cname | typename
--------------+-----------------+------------------+-------+----------
  01-Jan-2010 | 10.0.0.6#64473: | mail.example.org | IN    | A
  01-Jan-2010 | 10.0.0.6#26439: | mail.example.org | IN    | A
  01-Jan-2010 | 10.0.0.6#18020: | mail.example.org | IN    | A
  01-Jan-2010 | 10.0.0.6#49146: | mail.example.org | IN    | A
  01-Jan-2010 | 10.0.0.6#30733: | mail.example.org | IN    | A
  01-Jan-2010 | 10.0.0.6#52651: | mail.example.org | IN    | A
```

この例では、DNS の名前問い合わせ(query)の記録をとっているデータに対し、特定の情報のみを querylog という名前のテーブルに対して問い合わせる SQL 文を実行しています。これを応用して、複数のテーブルに対し SQL の JOIN 句などで横断的にデータを検索する事などにより、クロスレイヤ解析を分散計算環境にて高速に実施する事ができます。

*1 Docker は、コンテナ技術と呼ばれる仮想化手法を利用して、アプリケーションの配布方法を自動化するソフトウェアです。
https://www.docker.com

*2 Docker Hub は Docker イメージを公開・共有するためのオンラインレポジトリです。
https://registry.hub.docker.com/

*3 presto は Facebook 社がオープンソースとして開発している分散 SQL クエリ処理エンジンのソフトウェアです。Presto は Apache Hive 等と比べ、処理内容によっては高速な処理が可能です。
https://prestodb.io/

NECOMA の研究活動ならびに成果についての詳細は、Web サイト www.necoma-project.jp をご覧ください。

【お問い合わせ先】
奈良先端科学技術大学院大学　情報科学研究科
インターネット工学研究室内
NECOMA プロジェクト事務局
Mail: fp7-necoma-pr@is.naist.jp
Web: http://www.necoma-project.jp

NECOMA
Nippon-European Cyberdefense-Oriented Multilayer Threat Analysis