ハニーポットのアクセスログを可視化するシステムの設計

2016SE021 伊藤和哉
指導教員：沢田篤史

1　はじめに

サイバー攻撃への対策を考える上で実際の攻撃を観測することは重要であり，攻撃者の情報を取得するハニーポットの運用・開発・研究が進んでいる．ハニーポットを利用したセキュリティ対策は，取得したアクセスログを可視化し，それを元に分析し対策するのが一般的である．従来型のハニーポットの分析基盤は，リアルタイム性が十分に考慮されていない．攻撃分析が場当たり的に行われていることで，取り返しのつかなくなってしまう問題がある [6]．

本研究の目的は，ハニーポットのアクセスログを分析するシステムのためのアーキテクチャ設計である．本研究では，MVVM アーキテクチャ [5] を基礎に Hadoop[1] と Pipes and Filters[4] を組み合わせアーキテクチャを設計する．提案するアーキテクチャの妥当性を確認するために，ログを可視化する簡単なプロトタイプシステムを構築する．これを通じ，提案する手法により，リアルタイムにログ分析を行うツールの実現が可能であることを示す．

2　ハニーポットのログ分析における問題点

田崎らの研究 [8] では，ネットワーク運用とサイバーセキュリティの向上に役立つ攻撃の分析基盤の構築を述べている．攻撃の分析基盤である MATATABI システム [8] は，Hadoop とその上で動作する Hive，Facebook Presto といったオープンソースソフトウェアを利用して構築されたデータ蓄積と分析のためのシステムである．MATATABI システムは，数分間に一回のバッチジョブによりデータ形式に応じた変換プラグインを経由して，システム上の HDFS に蓄積される．MATATBI システムに蓄積された情報は，同様にバッチジョブによって SQL データベースのようなテーブルスキーマに変換され，Presto を用いて横断的に情報を検索することが可能となる．しかしこのシステムにはリアルタイム性への対応が不十分だと考えられる．田崎らの手法では，ログの適切な可視化，ログの安全な管理は考慮されているが，断続的に送出されるアクセスログをリアルタイムに処理するのことが難しい．

3　ハニーポットのログをリアルタイムに可視化するシステムのアーキテクチャ

3.1　設計指針

本研究は MVVM アーキテクチャ [5] を基礎に Hadoop[1] と Pipes and Filters[4] を組み合わせたアーキテクチャの設計をする．これにより大量のログを安全に保管し，リアルタイムな分析をおこなうためのソフトウェア基盤を構築する．

大量のログを安全かつ長期間保存するために Hadoop を利用する．Hadoop のファイルシステム (HDFS) はデータを分散して保存するシステムであり，耐故障性を考慮して設計されているので，ログを安全に保存できる．大量に流れてくるアクセスログを迅速に処理するために Pipes and Filters アーキテクチャを利用する．ハニーポットからのアクセスログをストリーミング処理し，ログデータベースに保存する際にアクセスログを検索，可視化に適したデータ構造になるようフィルタで加工する．

3.2　アーキテクチャ設計

設計したアーキテクチャの静的構造を図 1 に示す．

各コンポーネントについて説明する．正規表現フィルタはアクセスログを JSON 化する．抽出フィルタはアクセスログの必要な部分だけを抽出する．Model は HDFS をログデータベースとして使用する．ViewModel は Elasticsearch[2] のインデックス構造を利用することで，Hadoop のデメリットであるリアルタイム性を補える．View は Kibana[3] を利用することで，検索したデータを分析に適した形で可視化できる．Kibana では更新されるダッシュボードを保存できるので，数秒間隔で Elasticsearch のインデックスを更新すれば，アクセスログをリアルタイムで可視化できる．

図 1 を元に動的振る舞いを説明する．Honeypot のアクセスログを正規表現フィルタと抽出フィルタが加工する．２つのフィルタで加工をおこなうと，アクセスログは構造を持ち，可視化したい項目だけを抽出できる．加工したデータをデータシンクであるログデータベースが分散して保存する．

検索器はユーザの入力したインデックス名を元に検索する．検索結果データベースは加工データを取得し，グラフ作成器に検索結果を送信する．グラフ作成器は検索結果を元にグラフ作成し，ディスプレイにグラフ送信する．ディスプレイはユーザにグラフを表示する．

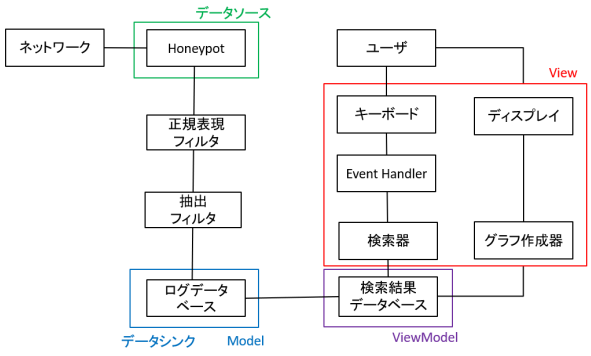

図 1　：静的構造

4　実装

提案したアーキテクチャの妥当性について確認することを目的におこなった簡単な分析アプリケーションの構築について説明する．本研究では実際にハニーポットを設置してアクセスログを収集する実験はおこなうことができなかった．そこでハニーポットのアクセスログの代わりとして IDS 及び IPS の研究で扱われる代表的なデータセット KDDCUP99DataSet[7] を使用して，可視化がリアルタイムにおこなえることを確認する．

本研究では，Elasticsearch のインデックスに KDDCUP99DataSet を投入して可視化をおこなった．可視化した分析画面が図 2 である．項目 1 個に対して 1 つグラフを使用した．実際にインデックスを作成し，50 万件のデータ登録にかかった時間は 28.911 秒であり，グラフ作成はユーザがインデックス名を入力から 1 秒以内におこなえた．したがって，100 件のデータ登録時間は約 0.0058 秒となり，実際のハニーポットの断続的に送出されるアクセスログであればリアルタイムに処理できると考える．

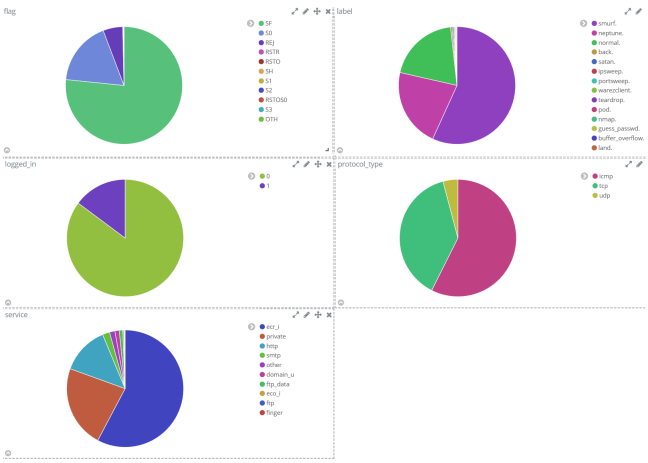

図 2　：分析画面

5　考察

5.1　アーキテクチャの妥当性

本研究では MVVM アーキテクチャを基礎に Hadoop と Pipes and Filters を組み合わせることでリアルタイムでアクセスログを可視化できる分析基盤が得られると考えた．本研究で提案したアーキテクチャの妥当性について確認するために，簡単な分析アプリケーションを構築し実装をおこなった．実際にユーザの入力に対して迅速にグラフを生成し可視化することができたので，リアルタイム分析が可能なアーキテクチャ設計であると考える．以上のことからアーキテクチャが妥当であるといえる．

5.2　関連研究との比較

田崎ら [8] の提案している分析基盤は Hadoop 上で動作するオープンソースソフトウェアを利用して，データ蓄積と分析をおこなっている．本研究で設計したアーキテクチャは，ログの加工，検索，可視化は Hadoop 上ではおこなわず，MVVM アーキテクチャを基礎に Hadoop と Pipes and Filters を組み合わせ設計した．アーキテクチャに基づき実装した結果，ユーザの入力に応じて素早く可視化をおこなうことができ，リアルタイム性を解決できたと考える．

6　おわりに

本研究では，ハニーポットのアクセスログを可視化するシステムのためのアーキテクチャ設計をおこなった．従来型のハニーポットの分析基盤では大量のアクセスログの処理に時間を要しリアルタイムな攻撃分析がなされていない．MVVM アーキテクチャを基礎 Hadoop と Pipes and Filters を組み合わせることで，ハニーポットのアクセスログをリアルタイムに処理し，ログを安全に長期保存，リアルタイム可視化するシステムの構築基盤得られると考えた．実際に設計したアーキテクチャを元に実装をおこなうことでアーキテクチャの妥当性を確かめた．

今後の課題として，実際にハニーポットを運用しアーキテクチャの考察をおこなう必要がある．本研究では，Hadoop，Elasticsearch, Kibana や Pipes and Filters を利用したが，他の構成と比べてどの程度の性能向上ができたのか検証する必要がある．

参考文献

[1] Apache，Hadoop，https://hadoop.apache.org/，2019.

[2] elastic，Elasticserch，https://www.elastic.co/jp/products/elasticsearch，2019.

[3] elastic，Kibana，https://www.elastic.co/jp/products/kibana，2019.

[4] Frank，B.，Regine，M.，Hans，R.，Peter，S.，and Michael，S.，PATTERN-ORIENTED SOFTWARE ARCHITECTURE，WILEY，1996.

[5] Gossman，J.:Introduction to Model/View/ViewModel pattern for building WPF apps，https://blog.msdn.microsoft.com/johogossman/2005/10/08/introduction-to-modelviewviewmodel-pattern-for-building-wpf-apps/，2005.

[6] IPA，高度サイバー攻撃への対処における ログの活用と分析方法，https://www.jpcert.or.jp/research/apt-loganalysis.html，2016.

[7] KDD Cup 1999 Data，https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html，2015.

[8] 田崎　創，岡田　和也，関谷　勇司，門林　雄基，“MATATABI : Hadoop によるマルチレイヤ脅威分析基盤の設計と構築，”電子情報通信学会，vol. 113，no. 502，pp. 113-118，2014.