氏　　名	阿部 博
学位の種類	博士(情報科学)
学位記番号	博情第 411 号
学位授与年月日	平成 31 年 3 月 22 日
論文題目	大規模ネットワーク環境における高速なログ解析基盤と異常検知に関する研究
論文審査委員	主査　篠田陽一　北陸先端科学技術大学院大学　教授
	丹康雄　同　教授
	知念賢一　同　特任准教授
	岡部寿男　京都大学　教授
	植原啓介　慶應義塾大学　准教授

論文の内容の要旨

Network administrators usually collect and store logs generated by servers, networks, and security appliances to stably manage the network systems they operate. When they find network troubles and security incidents, they investigate the contents of log information to identify the source of problems. The size of the system to store and search the log messages tends to be larger when the size of the target managed network becomes more substantial.

The large-scale networks usually integrate the multi vendor's network equipment, such as routers/switches and security devices. And recent virtualization technology such as Virtual Machine, NFV, Containers cause of output the large log data. Network administrators need to search the keyword as soon as possible to solve the network troubles and incidents. But the high-speed log search system usually make cluster systems up and its hard to manage, they need to have a massive management cost to maintain the search system.

And the next problems is to detect anomaly event on a vast scale network. Network administrators monitor their networks using log data manually or using monitoring systems pay attention to the values of target devices threshold. But it's hard to integrate the automatic anomaly detection system if the network is so vast, tune thresholds of too many target devices are also difficult.

In this research, we proposed a fast log storing and searching system "Hayabusa" which is optimized for the time-dimensional search operation. We usually use a general-purpose database to store log data such as RDBMS storing the parsed data or Hadoop ecosystems to store original data. Our method treats log data like time series data, but the data is not parsed and store original data at high speed by the time structured directory design. And search speed is also fast because of the design of directory, particular database structure for full-text search and core scale processing mechanism. We also propose a simple distributed system which adds scalability to the existing Hayabusa system. The time required to perform a full-text search over 14.4 billions

of records data is about 7 seconds, which is fast enough for daily operations of administrators managing a vast scale network. The distributed Hayabusa had some problems about request management and storage capacity, and we improved these problems using request management system and network storage. And we also released the distributed Hayabusa system as the Open Source Software.

About the network anomaly, we propose a method to detect abnormalities by analyzing the total amount of syslog data which collected in the event network using the Bollinger Bands algorithm used in stock trading. We performed anomaly detection by lightweight calculation in the statistical method using real data of syslog that collected by ShowNet constructed by Interop Tokyo. And we evaluated the effectiveness of the Bollinger Bands algorithm.

This research result such as Hayabusa and detection of anomalies using the Bollinger band will be a proposal to contribute to reducing the operational burden for network administrators.
And since Hayabusa is released as open source software, many people can benefit from using a fast log retrieval system.

Key Words: time serise data, log search engine, distributed system, anomaly detection, network management

論文審査の結果の要旨

ネットワーク管理者は日々ネットワークの健全性を評価し、トラブルが発生した場合には様々な角度から分析を行いその原因を特定し、安定したネットワーク運用を実現している。障害が発生した場合に、管理者は可能な限りトラブルの原因を早急に調査し解決しなければならない。一方、管理者はトラブル解決のために高速なログ検索システムを利用しなければならないが、管理者自身が運用コストのかかるクラスタを管理することによる、ネットワークを安定的に運用するために費やせる時間の減少は問題となる。本研究では、この問題を解決すべく、多数のマルチベンダ機器が出力する大量の時系列ログを高速に蓄積し、高速に検索可能なシステムが提案されている。また、多数のマルチベンダ機器が混在する大規模なネットワークにおいて、管理者にとって未知のログが多数出現する過酷な環境下であっても時系列で収集されるログの総量から異常を読み取り、効率的に通知可能なアルゴリズム適用が提案されている。

第 1 章では、大規模なネットワークを安定的に管理するための問題点として、安定的なネットワーク運用のためのデータ収集の難しさと、収集した大量なデータ群から必要な情報を高速に抽出する必要性について述べた。また、大規模ネットワークでの異常検知について、大量なログの扱いの難しさについて述べた。第 2 章では、大量のログを扱う汎用的なシステムの解説と、時系列データベースについての分析と考察を行った。また、異常検知手法に関して古くから研究される統計を用いた手法に関して考察を行った。第 3 章では、

大量の時系列ログデータを処理するための新しい提案として Hayabusa を設計・実装した。そして、ログを時系列データとして扱うために適した形で保存し、高速に処理をするためのアーキテクチャを実現した。第 4 章では、Hayabusa 自体の分散システム化を行った。データの複製と、Remote Procedure Call を用いることで Hayabusa のシンプルさを崩さずに分散版の Hayabusa を設計、実現することができた。結果として、144 億行のログを約 7 秒でフルスキャンし集計可能なシステムとして分散システム化することに成功した。第 5 章では、分散版 Hayabusa の課題を解決するために、ネットワークストレージとリクエスト管理機構を導入する改良を行った。改良を加えた分散版 Hayabusa の新アーキテクチャを OSS として幅広く公開したことにより、全世界のユーザに利用される可能性を実現した。第 6 章では、ネットワーク全体の異常を検知するために、大規模なネットワークイベントで収集した実データを用いて、異常検知アルゴリズムを設計した。結果として、大規模なイベントネットワークであったとしても、ログの統計値はある一定範囲で推移することがわかった。第 7 章にて、本論文にて提案された Hayabusa と異常検知機構を用いることによる社会的意義と学術的意義を示し、結論を総括している。

以上、本論文は Society5.0 に代表される情報化社会の安定な存続のための安定したネットワーク運用を支える大規模なログ記録・検索に関して、学術的な考察を通した実践的な解を与えるものであり、提案されているシステムはすでに複数の運用実績があるなど工学的な有用性も高い。また、本人の該当分野における知識や知見そして研究開発の能力は高く、博士（情報科学）の学位に相応しいと判断する。