WIDE Technical-Report in 2018

# WIDE クラウド WG 2017 年度活動報告

wide-tr-cloud-report-summary-2017-00.pdf


WIDE Project : http://www.wide.ad.jp/

*If you have any comments on WIDE documents, please contact to board@wide.ad.jp*

Title: WIDE クラウド WG 2017 年度活動報告
Author(s): WIDE クラウドワーキンググループ
Date: 2018-01-05

# WIDE クラウド WG 2017 年度活動報告

WIDE クラウドワーキンググループ

WIDE クラウドワーキンググループは、今後のクラウド技術の研究開発を推進するために 2010 年 1 月に設立された。複数の WIDE 組織間に渡って運用される広域連邦型クラウドシステムである WIDE クラウドシステムの運用と、それを用いた研究開発を行っている。

2017 年度は、大規模ログ検索エンジンの設計と実装、また安全なウェブサービスアクセスを実現するための URL 分類手法を検討した。

## スケールアウト可能なログ検索エンジンの設計と実装

ネットワークのトラブルシューティングやセキュリティインシデントへの対応にはサーバやネットワーク機器から出力されるログの調査が重要であり、大量に出力されるログの蓄積と高速な検索は問題を早期に解決するための重要な要素である。大規模なネットワークでは出力されるログの量も多く、蓄積・検索システムの規模も巨大化しがちである。本活動では、大量に出力される機器のログを高速に蓄積し、高速に検索するためのログ蓄積・検索システム Hayabusa を開発した [1]。Hayabusa はシンプルな実装を念頭に設計されておりスタンドアロン環境で動作する。全文検索速度のみを比較すればスタンドアロン動作する Hadoop システムよりも高速である。ただし、単体動作ではその性能に限界がくることも自明であるため、これを発展させ、検索性能をスケールアウト可能なシンプルな分散システムの設計を行い評価した。PoC を用いた結果によれば、スタンドアロン環境で動作する Hayabusa の約 78 倍高速な分散処理システムを実装し、144 億レコードの syslog メッセージを約 6 秒でフルスキャンし全文検索可能なスケールアウトするシステムを実現することができた [2]。

## URL ビット列出現頻度による URL 分類

ネットワークの利用は年々拡大しており、利用者数、通信量は常に増大している。インターネットを用いた不正行為・犯罪行為も増え続けており、社会基盤の一部として運用されるようになったインターネットにおいて、ネットワーク管理者が安全なインターネットを提供することがますます重要になっている。フィッシングはそのような不正・犯罪行為に深く関係する事例の一つであり、Anti Phishing Working Group[1] の報告によれば 2016 年には 120 万件を超えるフィッシング被害が報告されている。管理者としては、利用者がフィッシングサイトに誘導される前に、それを検知し通信の警告や遮断をすることが重要である。我々は、深層学習の技術を応用し、良性の URL と悪性の URL を、その文字コードの並びのビット列を元に区別する技術を研究した。暫定的な結果ではあるものの、実際にネットワーク上で観測された生きた URL アクセスログと、フィッシングサイトの共有サイトとして知られている PhishTank[2]から入手したフィッシングサイトの URL を 95%以上の正確さで判別できることを確認した [3]。本研究は、深層学習がネットワークデータ解析にも応用可能であることを示唆しており、今後多様なデータの解析に挑戦する予定である。

## まとめ

より詳しい報告は別途配布される詳細報告書を参照して欲しい。大規模化の進むネットワークサービスを安定運用するためにはより良いシステム状態把握技術が必要である。引き続き研究開発を進めていく。

[1] https://www.antiphishing.org
[2] http://phishtank.com/

# 参考文献


[1] Hiroshi Abe, Keiichi Shima, Yuji Sekiya, Daisuke Miyamoto, Tomohiro Ishihara, and Kazuya Okada. Hayabusa: A simple and fast full-text search engine for massive system log data. In *International Conference on Future Internet Technologies (CFI2017)*, June 2017.

[2] 阿部博, 篠田陽一. スケールアウト可能なログ検索エンジンの実現と評価. インターネットと運用技術シンポジウム 2017, December 2017.

[3] Keiichi Shima, Daisuke Miyamoto, Hiroshi Abe, Tomohiro Ishihara, Kazuya Okada, Yuji Sekiya, Hirochika Asai, and Yusuke Doi. Classification of URL bitstreams using bag of bytes. In *Proceedings of First International Workshop on Network Intelligence (NI2018)*, February 2018.