≪「**報告書詳細版**」は巻末の付録USBメモリに収録しています ≫

第６部

クラウドコンピューティング基盤の構築と運用（概要版）

WIDEクラウドワーキンググループ

第１章　はじめに

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。

第２章　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[98]。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[99]。

第３章　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 Group1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PhishTank2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[100]。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。

第４章　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。