様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　２１　日現在

機関番号：　３１３０３
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３３０１１０

研究課題名（和文）ネットワークロギングシステムの異常検知アーキテクチャの設計と構築

研究課題名（英文）Design and development of the network architecture for detecting anomalies in network logging system

研究代表者

角田　裕（Tsunoda, Hiroshi）

東北工業大学・工学部・准教授

研究者番号：３０４００３０２

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：本研究では，ネットワークの運用やセキュリティ管理上の要であるログの重要性に着目し，ログを収集するロギングシステムの異常検知を実現するアーキテクチャと要素技術について研究開発を実施した．具体的には，各ホストのログ収集に関する設定を突合し，ログ収集経路をネットワーク地図上に可視化することで，ログ収集の収集漏れにつながる設定の不整合の存在を管理者に対してわかりやすく提示可能であることを示した．そして，ロギングプロセスのシステムコールを追跡することによるログ出力数の計測方法の効果と課題を確認した．また，ログ出力数を遠隔から計測し，ロギングシステムの挙動を監視するためのデータモデルを設計した．

研究成果の概要（英文）：Nowadays, log information plays a vital role in the network operation and security management area. Thus, this research tackled the development of the anomaly detection architecture and element technologies for a logging system. In this study, we presented that the prototype application for monitoring configuration information of logging applications in a network, analyzing monitored configuration and visualizing the results on network maps. The visualization will help network administrators to detect defects in their logging system easily. We also discussed the method to measure the number of log messages sent and received by a logging application by tracing some system calls called by the logging application. Besides, we designed the data model for monitoring the number of log messages for collecting that information efficiently.

研究分野：　情報ネットワーク

キーワード：　ロギング　セキュア・ネットワーク　ネットワーク管理　情報システム

１．研究開始当初の背景

社会基盤としての重要性を増し続けるネットワークには，常に非常に高い信頼性が求められ，高度な運用管理・セキュリティ管理が必要とされている．それらの管理においては情報の収集と解析が根幹であり，情報が網羅できなければ障害やセキュリティ問題の見逃しや対応遅れにつながる．網羅されるべき情報の中でも重要なもののひとつが各機器の OS やアプリケーションが出力するログである．なぜなら，誰が・いつ・何をしたのかなど豊富な情報を含んでいるためである．

ログ収集（ロギング）には事実上の標準プロトコルとして *Syslog* が長年にわたり広く用いられている．Syslog を利用するロギングシステムでは，ネットワーク内の各機器は，ログを出力する *Originator*，Originator のログを *Collector* へ中継する *Relay*，ログを受信し蓄積する Collector のうち 1 つ以上の役割を担当する．Originator から送信されたログは，直接または 1 台以上の Relay を経由して Collector へ到達し，ネットワーク管理者は Collector に蓄積されたログを各種管理に活用する．これら Originator・Relay・Collector のすべてが正しく設定され動作していなければ，Collector に到達しないログが存在することになり，情報の網羅性が担保できない．

しかし，Syslog に関する設定はロギングシステムを構成する各ホストで独立して実施する必要がある．そのためロギングシステムの全体像を把握することが困難であり，ホスト間での設定の不整合が容易に起こりえることが潜在的な課題となっている．

また，設定の不備・不整合が発生していたり，Syslog による通信を担当するプロセスが異常終了したりした場合，本来の Collector 上にログが存在しない状態に陥る．しかし，管理者にとっては，そもそもログが発生していないのか，何らかの問題によってログが収集できなかったのか，を明確に判別する手段は用意されていない．従って，ログが提供する情報は各種障害の検出や診断に欠かせないにも関わらず，結果としてロギングシステム自身の異常や障害の検出・診断に対してはログ情報は有効に機能しないことが本質的な問題である．

また，近年の仮想化技術の発達によって無数の仮想マシンや仮想ネットワークを活用できるようになっている．それらは必要に応じて柔軟に配置され，構成が変更され，除去される．すなわち近年のネットワークは大規模化と複雑化が進むと同時に動的な性質が急速に強まっており，ロギングシステムにはそのようなネットワーク管理の根幹としてより一層の信頼性が求められることは確実である．

２．研究の目的

本研究の目的は，ロギングシステム自身に発生する異常の検知を支援するアーキテクチャの創出に資することであり，次に示す要素技術の研究開発を進めた．なお，本研究では Syslog を利用したロギングシステムを前提とする．

(1) ロギング設定の分析と可視化方式の検討
ホスト間のロギング設定の不備に起因するログの収集漏れを未然に防止または発見するための研究項目である．各ホストの設定情報を収集・集約し，突合して分析・可視化する手法について検討する．

(2) ロギングに関する統計情報の監視方式の検討
各ホストが送受信するログの数などの統計情報の観測手段と，その情報を集約するための方式を検討する．ここで監視した情報が次の異常検知に活用する基本情報となる．

(3) ロギング動作の分析と異常検知方式の検討
統計情報に基づいて個々のホストのロギング動作を分析し，ホストやネットワークにおける異常を検知する方式について検討する．

３．研究の方法

(1)ロギング設定の分析と可視化方式の検討

本研究項目における課題は多数のホストからのロギング設定の効率的な収集と，その結果の分析およびわかりやすい可視化である．

各ホストのロギングに関する設定を効率的に収集するために，本研究項目では研究代表者らが以前に提案・実装した Syslog CMIB (Syslog Configuration Management Information Base) モジュールを活用することとする．なお，MIB モジュールとは，ネットワーク管理プロトコル SNMP (Simple Network Management Protocol) で使用する管理情報定義を指す．

次に，各ホストから収集した設定情報を突合し，設定の不整合を検出するアルゴリズムを開発する．研究代表者らは以前の研究において，最も基本的な syslog 実装である BSD syslog について設定情報の分析を実施した知見がある．その知見を活用して，他の syslog 実装の設定情報の解析を試みる．ネットワーク地図上に可視化する方式を検討する．

(2) ロギングに関する統計情報の
監視方式の検討

まずロギングの動作に関する統計情報の監視方式について調査・検討する．Syslog には複数の実装があり，統計情報を出力する機能を備えた実装も存在している．まずは，既

存の実装が有する統計情報の出力機能およびその活用可能性を調査する．そして，既存実装の統計情報出力機能が不十分な場合には，それを補う機能を開発する．

(3) ロギング動作の分析と異常検知方式の検討

ロギングシステム全体の動作の分析には，複数のホストの統計情報を始めとする必要な情報を各ホストから収集し，一元的に管理する必要がある．これまでの知見により，効率的な収集には SNMP の活用が有用であることから，まずは統計情報等を管理情報として格納するための MIB モジュールを設計する．

次に，各ホストから収集した情報を基づき，特に最も基本的で扱いやすい情報であるログの出力数に着目した異常検出の基本アルゴリズムについて検討する．

４．研究成果

(1)ロギング設定の分析と可視化方式の検討

研究室 LAN 内に実験用のロギングシステムを整備し，システムを構成する各ホストに Syslog CMIB モジュールを実装した．そして，SNMP によるロギング設定の収集を実施し，その分析と可視化を行った．

本研究項目で使用した Syslog CMIB モジュールで定義されている管理情報とその構造を図 1 に示す．これは，Syslog の設定ファイル *syslog.conf* の各行を 1 つの管理情報として格納するための定義となっている．この MIB モジュールを各ホスト上で実装することで，syslog.conf の内容を SNMP を利用して遠隔から取得できた．

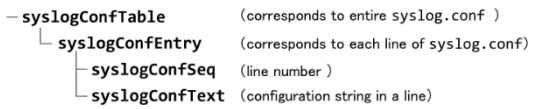

図 3　Syslog CMIB の管理情報

設定の分析については，Syslog 実装のひとつである Rsyslog の設定情報の解析にも新たに対応した．近年では，多くの Linux ディストリビューションで Rsyslog が標準で採用されているため，この対応によって本研究項目の成果の活用範囲が大きく広がると言える．

そして，先行研究で実施していた収集経路のループの検出に加えて，本来存在しないはずの不正な Collector や，正当な Collector へログを送信していないホストのなどの異常を発見可能であることを示した．また，それらの異常を，図 2 および図 3 に示すようにネットワーク地図上に可視化し，視覚的にわかりやすい形で管理者に対して提示できることを示した．

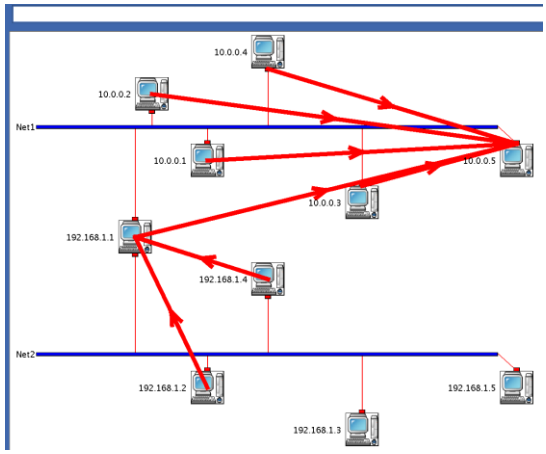

図 1 設定不備により誤った Collector へログの送信が行われている場合の表示例

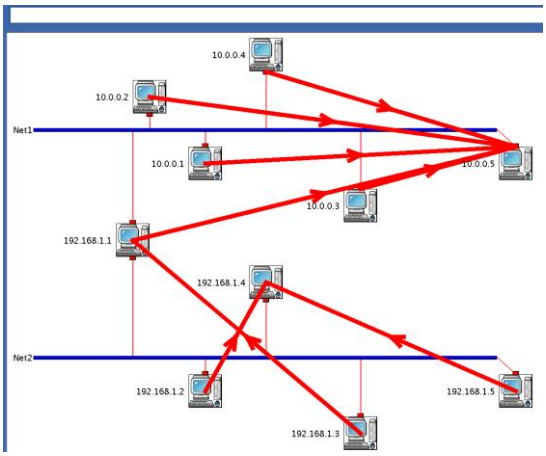

図 2 設定不備によりログの収集が行われていないホストが存在している場合の表示例

(2) ロギングに関する統計情報の監視方式の検討

主要な Syslog 実装として BSD syslog, Rsyslog, syslog-ng の 3 種について統計情報の出力機能を調査した．その結果，BSD syslog には当該機能はしないが，Rsyslog と syslog-ng には存在することを確認した．

しかし，Rsyslog と syslog-ng の当該機能を調査した結果，指定間隔で統計情報を出力可能ではあるが，起点となる時刻は syslog プロセスの起動時刻であることがわかった．つまり，起動時刻が異なる複数のホストについて同一の時間帯に出力したログの数を横断的に比較する，といった基本的な比較は当該機能を利用しても困難であることがわかった．

そこで本研究では，プロセスが発行するシステムコールに着目し，それをトレースすることでログの出力数を始めとする統計情報を監視する方式を提案した．Syslog によるログメッセージの出力・送信はシステムコールによって行われているため，ロギングデーモンのプロセスが発行するシステムコールをトレースし，その結果を利用することによってロギングデーモン自体やその機能に依存しない

ログ出力数の取得ができると考えた．そして Linux 上で strace コマンドを利用して，Rsyslog が発行するシステムコールを監視しログ出力数を計測するプロトタイプシステムを実装しその基本的な有効性を確認した．今後は，BSD syslog や syslog-ng など他の syslog 実装にも提案方式の適用対象を広げていく予定である．

(3) ロギング動作の分析と異常検知方式の検討

ログ出力数の情報を始めとする統計情報等を遠隔から取得し，複数のホストの情報を一元的に監視・管理するための管理情報のデータモデルとして Syslog MIB モジュールを設計した．図 5 および図 4 に定義した管理情報の概要を示す．当該モジュールについては，その設計と定義をまとめた文書をインターネット技術の標準化を担う組織である IETF (Internet Engineering Task Force) に提出し，意見を募っている段階である．今後は本モジュールの標準化を目指し，プロタイプ実装の公開を目標とする．

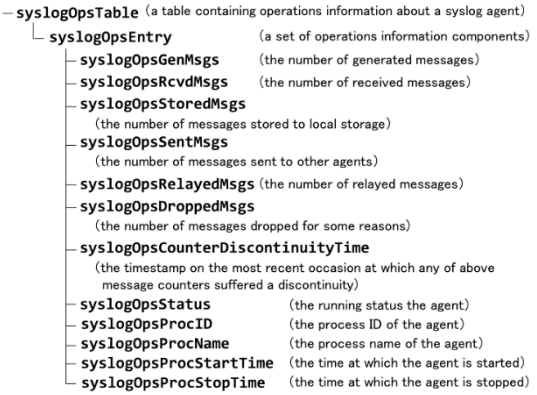

図 5 ログの出力数等の運用状況を表す管理情報

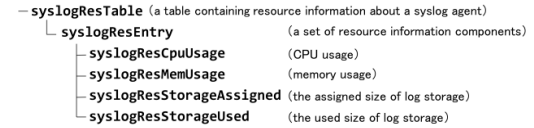

図 4 ロギングに関連のリソース情報を表す管理情報

実験用システムで収集した各ホストのログの出力数およびログ情報を精査したところ，多くのホストでログの出力数は安定していることがわかった．つまり，ホストにおいてログの出力数が他の時間帯と比較して著しく多くなったり，逆に少なくなったりすれば，ホスト上またはロギングシステムで何らかの異常が発生していると判断できる．例えば，実験では，通常時には 1 日あたり約 10 個しか出力されないある種のログが，10 分間に 8000 個以上出力されたという異常が見られた．これはプログラムのミスにより，特定のプロセスが短時間連続して呼び出されるという異常に起因するものであった．また，逆に，あるホストから通常時には定期的に一定個数発生していたログの発生が停止するという事象も見られた．この原因は当該ホストのロギングプロセスが一時的に停止していたことによるものであった．これらの結果から，ログの出力数という最も基本的な情報に対して単純なしきい値を設定するだけでも，ホストやロギングシステムの異常の把握につながる可能性があることが明らかにできた．今後は，ホストごと特性を考慮した通常時のログ出力数のモデル化手法を開発し，より効果的な異常検知手法について検討を継続する．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 7 件）

① 免田 健太朗, 角田 裕, “SNMP を利用したロギングシステムの一元的な監視に関する検討”, 平成 29 年 東北地区若手研究者研究発表会，2017 年 3 月 4 日，東北学院大学
② 佐藤 利紀，髙橋宏明，角田 裕，“ネットワークロギングの管理システムに関する考察”，平成 28 年 東北地区若手研究者研究発表会，2017 年 3 月 1 日，日本大学
③ 斎藤 康平，角田 裕，“システムコールのトレースによるログ出力数の監視と異常検知”，平成 28 年 東北地区若手研究者研究発表会，2017 年 3 月 1 日，日本大学
④ Hiroshi Tsunoda, Glenn Mansfield Keeni, “Managing syslog”, Proceedings of The 16th Asia-Pacific Network Operations and Management Symposium (APNOMS2014), 2014. 9.19, Taiwan
⑤ Hiroshi Tsunoda, Glenn Mansfield Keeni, “Monitoring syslog”, 電子情報通信学会 通信方式研究会, 2014 年 9 月 12 日，東北大学

〔その他：標準化提案文書〕

① Hiroshi Tsunoda, Glenn Mansfield Keeni, “Syslog Management Information Base”, draft-tsuno-syslog-mib (work in progress), https://datatracker.ietf.org/doc/draft-tsuno-syslog-mib/

６．研究組織
(1)研究代表者

角田 裕（Hiroshi Tsunoda）
東北工業大学・工学部・准教授
研究者番号：30400302