様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 １６ 日現在

機関番号：１７５０１
研究種目：若手研究(B)
研究期間：2013～2014
課題番号：２５８７０５５８

研究課題名（和文）ネットワーク管理における多種多様なデータを横断的に処理可能なシステムの研究

研究課題名（英文）A Study for a multiple log data cross-processing system in network operations

研究代表者
池部　実（Ikebe, Minoru）

大分大学・工学部・助教

研究者番号：５０６１３６５０

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：本研究は、ネットワーク管理者のログデータの分析による障害原因追求や異常検知などをサポートするために、ネットワーク管理における多種多様なデータを横断的に処理可能なシステムを構築することを目的とする。目的を達成するため、ログデータを属性名と属性値の組で管理し、JSON形式で保存するログデータ管理システムを構築した。様々なログデータを取り扱うために、DNS、Web、ハニーポット、RADIUS、pcapなどのログデータを分析して、収集用プログラムを開発した。これらのログデータを開発したシステム上で管理し、送信元IPアドレスによって複数ログを横断して検索し、攻撃者の挙動を調査することができた。

研究成果の概要（英文）：This research supports analysis for trouble shooting and anomaly detection by network administrators. Therefore, I propose a multiple log data cross-processing system for network operations. This system manages log data as pairs of attribute name and attribute value. And, the proposed system stores log data of JSON format.
I have been developing log data collector programs for several kinds of log data (ex. DNS, Web, honeypot, RADIUS and pcap). The log management system manages those log data. The network administrator can analyze the behavior of attackers by the source IP address on crossing multiple log data.

研究分野：情報ネットワーク

キーワード：ネットワーク運用　ログデータ　メタデータ　ログ分析　ネットワークセキュリティ

１．研究開始当初の背景

大学や企業における情報システムは、多数のサーバをベースとして運用され、高い信頼性が求められている。管理者は、大規模、複雑化するサーバ、ネットワーク運用において、大学内、自社内 LAN 上の多数のサーバを管理する。また、仮想化技術をベースとしてプライベートクラウド環境を大学内、自社内に構築し、多数の仮想化されたサーバを動作させている。さらには、Amazon や Google などの提供するパブリッククラウド環境で、大学の業務システムやメールシステムを動作させる例も少なくない。

サーバ、ネットワーク管理作業においては、判断や作業を完全に自動化することは困難であり、人間の判断や作業が不可欠となる。各管理者は、判断のために各サーバのログデータを参照することにより、障害の原因特定や異常検知などに必要な情報を得ている。

図 1 に示すように LAN 上に設置した各サーバのログ、クラウドコンピューティング環境上に設置した各サーバのログ、L2/L3 スイッチのログ、ルータのログ、トラフィックデータなどネットワーク管理において分析する必要のあるデータは多岐にわたる。このようにシステムの大規模化、複雑化により、多種多様なログデータが出力され続け、その中から管理者がすぐに必要なログを探し出し、分析することが困難になっている。

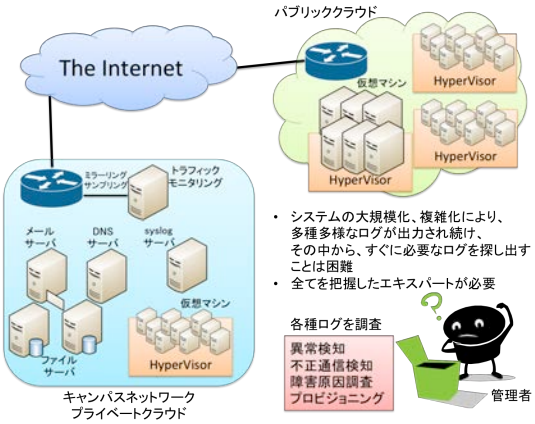

図 1 多種多様なログデータが大量に出力

２．研究の目的

本研究では、ネットワーク管理者をサポートするために、ネットワーク管理における多種多様なデータを横断的に処理可能なシステムを構築することを目的とする。例えば、図 2 に示すように、管理者がトラフィックの急激な増加にトラフィックモニタツールを確認して気づいたとする。まず、管理者は tcpdump を用いてパケットデータから急増しているトラフィックの原因を調査する。次に、トラフィック急増の原因が DNS サーバや Web サーバに対するアクセスであった場

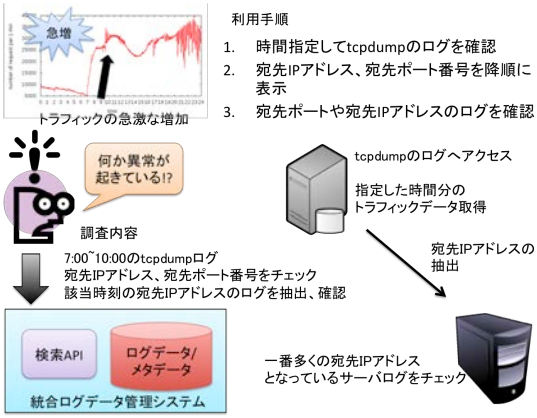

図 2 ログデータの横断的活用

合、通常のトラフィックあるいは、攻撃のトラフィックなのかを判断しなければならない。通常トラフィックであるのか、攻撃のトラフィックであるのかを判断するために、管理者は DNS サーバや Web サーバのログを分析する必要がある。そこで本研究では、この一連の作業を管理者が手軽に実行可能なシステムを構築することを目的とする。

本研究を実現することにより、分散するログデータへのリアルタイムアクセス、異なるサーバのログデータに対する横断的な処理が可能となる。

３．研究の方法

本研究を遂行するために以下の 3 つの点について研究を実施した。

(1) 分散するログデータの効率的な収集および、ログデータからのメタデータ抽出手法

本手法では、管理者が運用すべき対象のサーバがインターネット上に分散して存在するため、それらのサーバが出力するログデータを集約する。ログデータは、テキストデータやバイナリデータの場合がある。テキストデータの場合は、そのまま取り扱うことができるが、バイナリデータの場合は通常の UNIX コマンドでは取り扱うことができない。そのため、バイナリのログデータについてはデータを読み込み、書き込まれている情報をテキストデータに変換して取り扱う。また、ログデータ集約のタイミングでログデータのもつ時刻情報、ログデータ送信元 IP アドレス、ログデータ送信元ホストの役割や、ログデータに含まれる要素名をメタデータとして抽出する仕組みを実装した。ログデータがファイルに追記されたタイミングで、ログデータからメタデータを抽出し、ログデータとメタデータをログデータ管理サーバへ転送する。ログデータを転送する際、syslog サーバへログデータの 1 レコードを転送するのではなく、JSON(JavaScript Object Notation)データ形式に変換し、ログデータのプログラム処理性を確保した上で、データを転送・集約する仕

組みを構築した。データ転送、集約する仕組みは、オープンソースソフトウェアの fluentd を用いた。

(2)広域ネットワークにおけるログデータ、メタデータの分散管理手法

ログデータ、メタデータは JSON データ形式で統一しているが、ログデータの種類により保有する属性情報、属性値は異なる。そのため、データを管理するために JSON データをそのまま格納できる NoSQL(Not only SQL)を採用し、すべてのデータを属性情報と属性値の組として扱い、スキーマレスでデータを管理する。この手法ではクラウドコンピューティングにおけるマルチテナントアーキテクチャの考え方を導入し、複数のログデータ、メタデータをひとつのスキーマで管理できる。

(3)メタデータを用いたログデータの検索、該当箇所の抽出、結合可能なスクリプト言語の設計・開発

多種多様なログデータが出力され続け、その中から管理者がすぐに必要なログを探し出し、分析することが困難になっている。そのため、本手法では管理者が指定した属性情報と属性値を用いて、ログデータを検索し、実際のログデータの内容を参照する手段を提供する。また、ログデータを参照する際、ファイル全体を参照するのではなく、絞り込みや異なるファイルを連結して処理する。そのために、メタデータの属性名と属性値を用いた検索により、異なるログデータを統一的に操作できる UNIX 環境上で動作するスクリプト言語を設計し、開発する。例えば、攻撃元と疑わしい送信元 IP アドレスをもとにして、異なるログデータから grep コマンドで、該当する行を抽出して、分析・集計する。これらの動作は本来分散するログデータを集約して分析するが、本提案システムを利用するとテキストデータ、バイナリデータの差を意識することなく、メタデータとスクリプト言語から一括で操作できるようになる。

４．研究成果

(1)ログ収集・管理システム構築

本研究での基本となるログ収集・管理システムは、オープンソースソフトウェアである fluend、MongoDB、ElasticSearch をベースとして構築した(図 3)。ログを集中管理する方法を採用しているが、データベースが単一障害点となるため、ネットワークセグメントごとなどに分割し、エリアを担当するログ管理サーバを設置した。ログ管理サーバは複数エリアのログデータベースの内容を、複製して保持する(図 4)。

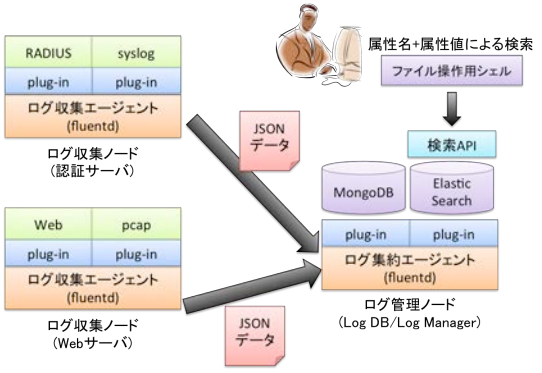

図 3 ログ収集・管理システムの概要

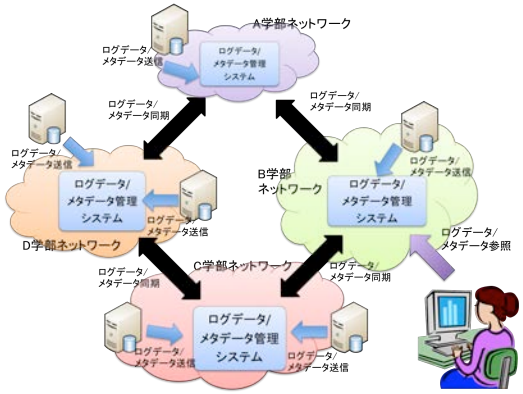

図 4 ネットワーク単位でのログ収集・管理システムの設置

(2)アプリケーション・サーバごとのログ収集システム構築

(1)で開発したログ収集・管理システムでログデータを収集し、メタデータの属性名と属性値から構成される JSON データ形式に変換するプログラムを複数のアプリケーション向けに実装し、適用した。従来の多くのログデータは printf スタイルがデファクトスタンダードであり、ユーザが読むことを前提で出力されていた。アプリケーション向けのログ収集プラグインを実装することにより、ログデータを JSON 形式で管理することにより、ユーザの可読性を確保しつつ、プログラムによる解析性を確保した。

①MAC アドレスによる端末認証システムにおける認証スイッチはベンダ、機種によってログの出力形式・方法が異なる。syslog へ出力する方法や RADIUS の Accounting メッセージとして出力する方法がある。RADIUS の Accounting メッセージもスイッチの機種により出力する RADIUS 属性の数もまちまちである。異なるベンダ・機種の認証結果のログを取り扱うために、変換スクリプトを作成し、保持していない属性については補完するようにして、ログ管理システムへ保存する際には、すべての認証ログデータで同様の属性を保持するようにした(図 5、図 6)。運用している RADIUS サーバには約 8000 件の MAC アド

レスが登録されている。認証スイッチから出力された認証に関わるログを抽出し、約 105 時間分で 73364 件のログが記録されていた。1秒あたり約0.19件のログが記録されていたことになる。このログを対象として、ログ転送(JSON 形式への変換)およびログ管理システムへの蓄積を実行したところ、処理時間は 718 秒であり、1 秒あたり 102.2 件のログを処理でき現在のログ出力速度に対しては対応可能であることを確認した。

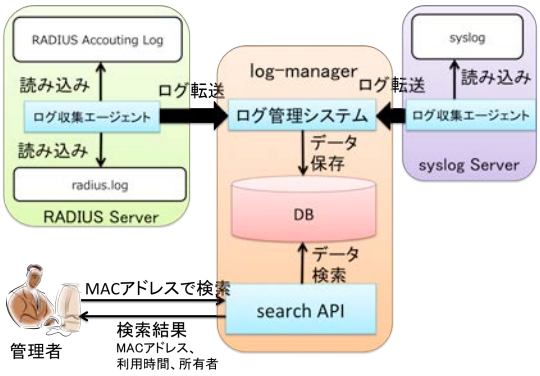

図 5 RADIUS サーバや syslog サーバからの認証ログの抽出、認証ログの管理のフロー

```
mapping":{
    "mac_authlog":{
        "properties":{
            "authlog_id":{"type":"long"},
            "switch_type":{"type":"string", "index":"not_analyzed"},
            "date":{"type":"date", "format":"date_optional_time"},
            "timestamp":{"type":"long"},
            "user_name":{"type":"string", "index":"not_analyzed"},
            "macaddr":{"type":"string", "index":"not_analyzed"},
            "vlanid":{"type":"integer", "index":"not_analyzed"},
            "nas_identifier":{"type":"string},
            "nas_ip_address":{"type":"string", "index":"not_analyzed"},
            "acct_status_type":{"type":"string", "index":"not_analyzed"},
            "calling_station_id":{"type":"string", "index":"not_analyzed"},
            "acct_authentic":{"type":"string", "index":"not_analyzed"},
            "nas_port_id":{"type":"string", "index":"not_analyzed"},
            "nas_port":{"type":"integer"},
            "service_type":{"type":"string", "index":"not_analyzed"},
            "acct_session_time":{"type":"integer"},
        }
    }
}
```

図 6 認証ログのメタデータ情報(JSON スキーマ)

②学外と学内の境界に設置しているスイッチからポートミラーしたパケットを tcpdump にて pcap ファイル(トラフィックログデータ)を保存する。pcap ファイルはバイナリデータで保存されており、tcpdump コマンドや Wireshark アプリケーションを用いて通常はファイルを読み込み、そこから情報を読み取る。バイナリデータであるので、通常は、UNIX コマンドである grep や awk などで pcap ファイルは扱うことはできない。そのため、pcap ファイルを JSON 形式に変換するスクリプトを作成した。これまでに、蓄積している pcap ファイルをログデータ管理システムへバルクデータ転送で挿入した(図 7)。

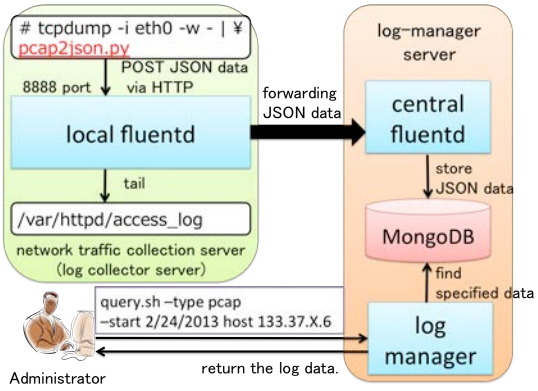

図 7 パケットデータ(pcap ファイル)の管理

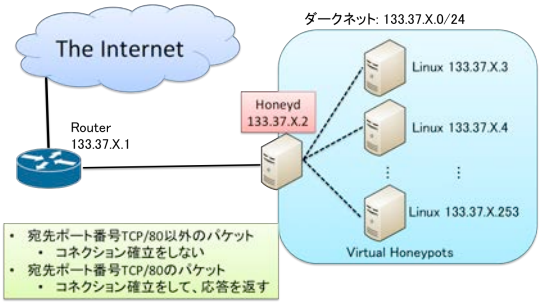

図 8 ハニーポットの設置環境

図 9 honeyd が出力するログ

③ハニーポットソフトウェアである honeyd を用いてダークネット宛てのパケットを観測している。honeyd が出力するログをログ管理システムで取り扱うようにプラグインを作成した。図 8 に示した環境にハニーポットを設置し、honeyd サーバにログ収集エージェントを仕込みログを収集した。図 9 に示した honeyd のログをログ管理システムで管理している。

④ DNS サーバソフトウェアである bind9 のキャッシュ DNS サーバへの問い合わせログをログ管理システムで取り扱うためにプラグインを開発し、DNS サーバの問い合わせログを取り扱い可能とした。

(3) 異なるアプリケーションログデータの突合による分析

収集したログデータを照合することにより、管理者による分析を実施した結果を示す。

①SMTP サーバと権威 DNS サーバのログの照合

SMTP サーバのメールログ、権威 DNS サーバに対する MX レコードの問合せログを照合し、それぞれの送信元 IP アドレスを問い合わせることにより、spam ボットの検出を試みた。それぞれの送信元 IP アドレスが一致するものを取り出し、ヒルベルト空間充填曲線によって可視化し、spam ボットが存在する可能性が高い IP アドレスブロックを検出した。

②Web サーバとハニーポットのログの照合

学内で稼働している Web サーバと Honeyd で取得した TCP/80 番ポート宛のアクセスログを照合することにより、送信元の挙動を調査した(図 10)。その結果、Shellshock に対する攻撃の様子を確認することができた。Shellshock とは UNIX 系 OS で用いられる bash(Bourne Again Shell)に発見された脆弱性である。HTTP ヘッダの User-Agent:以降に“{ :;}; 任意のコマンド”を記述することで、脆弱性をもつサーバに対して任意のコマンドを実行することができる。Shellshock を狙った攻撃は 2014 年 9 月 25 日から 10 月 7 日までの期間に観測された。Web サーバ、Honeyd ともに Shellshock の攻撃を観測し、連続した IP アドレスにおける Shellshock を攻撃するための協調動作を Web サーバ、Honeyd のログの分析結果より発見した(図 11)。

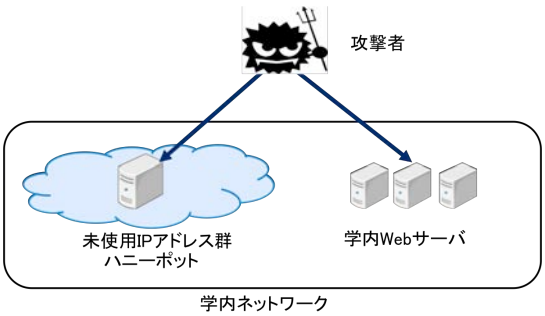

図 10 ハニーポットと Web サーバのアクセスログ照合による同一送信元の挙動調査

③IDS と SSH パスワードクラッキング攻撃検知システムのログ照合による検知条件へのフィードバック

scan 攻撃や DoS 攻撃を検知する「不正通信検知システム(IDS)」、SSH サーバへの不正侵入のための攻撃を検知する「SSH パスワードクラッキング攻撃検知システム(SCRAD)」をこれまで運用・開発してきた。攻撃者の挙動として、ネットワーク全体または一部に対しての scan 攻撃後に、発見した SSH サーバへのパスワードクラッキング攻撃を仕掛けていることを発見した。そこで 2 つのシステムでの検知結果をログ管理システムで管理し、あ

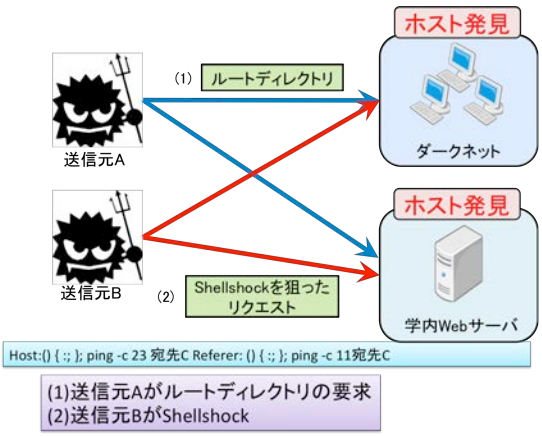

図 11 発見した Shellshock 攻撃の挙動

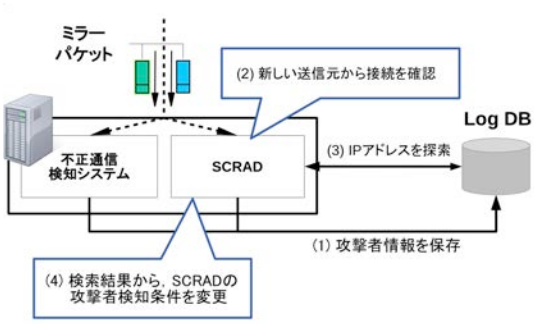

図 12 IDS、SCRAD での攻撃者検知ログの収集

```
'{
 "from":0,"size":1,
 "query":{"term":{"attacker-ip-address":4176167024} },
 "filter":{
  "range":{
   "detection-time":{"from":138001799,"to":1380593799}
  }
 },
 "sort":[{"attack-type":{"order":"asc"}}]
}'
```

図 13 SCRAD による攻撃者情報の問合せ

らたに攻撃が発生した際、ログ管理システムへ過去に検知した攻撃者か問合せることにより、検知条件を変更するシステムを構築した。2 つのシステム側から直接ログ管理システム側へログデータの送信、問合せを実行している。

(4)アプリケーションログの解析による必要なコマンドの設計

(2)(3)で示したように、いくつかのアプリケーションから出力されるログデータを JSON 形式にてログ管理システムへ蓄積するように開発を進めてきた。これらのログデータをメタデータから操作可能にするためのスクリプト言語を設計するために、まず、それぞれのログデータの基本的な解析手法を分析した。ログ解析は、grep や awk など UNIX コマンドを組み合わせることが多く、それらの基本的な機能を洗い出した。異なるログデータを照合する場合などには、UNIX コマンドだけではうまくいかないケースも存在するため、メタデータを付与してログデータを管理

し、メタデータとスクリプト言語の組み合わせによるログデータ操作の有用性を確認した。しかしながら、開発を予定していたスクリプト言語については設計段階に留まり、今後継続して開発を進めていく。

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　10 件）
(1) 池部実, 宮崎桐果, 吉田和幸: ハニーポットによる大分大学におけるダークネット宛通信の分析, 情報処理学会インターネットと運用技術研究会 (IOT), Vol.2015-IOT-29, No.17, pp.1-8, 2015 年 5 月, 別府国際コンベンションビューロー(大分県別府市)
(2) 小刀稱知哉, 清水光司, 池部実, 吉田和幸: 複数の攻撃検知システムの連携による攻撃者検知手法の運用結果, 火の国情報シンポジウム 2015, pp.1-8, 2015 年 3 月, 佐賀大学(佐賀県佐賀市)
(3) 小刀稱知哉, 池部実, 吉田和幸: 複数の攻撃検知システムの連携による攻撃者検知手法の提案と評価, インターネットコンファレンス(IC)2014, pp.105-114, アステールプラザ(広島県広島市)
(4) 宮崎桐果, 小刀稱知哉, 池部実, 吉田和幸: 大分大学の未使用 IP アドレスに対する TCP/80 番ポートへの通信の解析, 第 67 回電気・情報関係学会九州支部連合大会, pp.89-89, 2014 年 9 月, 鹿児島大学(鹿児島県鹿児島市)
(5) 山口舞子, 松井一乃, 渡辺拳竜, 池部実, 吉田和幸: メール送信時における MX レコード問合せタイミングの調査, 第 67 回電気・情報関係学会九州支部連合大会, pp.88-88, 2014 年 9 月, 鹿児島大学(鹿児島県鹿児島市)
(6) 池部実, 吉田和幸: MAC アドレスによる利用者認証における認証ログの統合・分析システムの提案と実装, 情報処理学会マルチメディア, 分散, 協調とモバイル(DICOMO2014)シンポジウム, pp.190-196, 2014 年 7 月, 月岡温泉　ホテル泉慶(新潟県新発田市)
(7) 渡辺拳竜, 松井一乃, 池部実, 吉田和幸: 権威 DNS サーバのクエリログの可視化による攻撃の発見と分析, 情報処理学会インターネットと運用技術研究会 (IOT), pp.1-6, 2014 年 5 月, ホルトホール大分(大分県大分市)
(8) 金高一, 松井一乃, 加来麻友美, 池部実, 吉田和幸: ハニーポットを用いたアドレスハーベスタと spam 送信者の spam 活動の調査, 情報処理学会インターネットと運用技術シンポジウム(IOTS)2013, pp.25-32, 2013 年 12 月 , 広島大学(広島県東広島市)
(9) Tomoya KOTONE, Naomi NAKAMOTO, Minoru IKEBE and Kazuyuki YOSHIDA: Proposal of a detection method for SSH attack based on SYN packets transmission interval, 2013 International Workshop on ICT, pp.1--pp.4, Dec. 2013, 別府亀の井ホテル（大分県別府市)
(10) Minoru IKEBE and Kazuyuki YOSHIDA: An Integrated Distributed Log Management System with Metadata for Network Operation, The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2013), 5th International Workshop on Virtual Environment and Network-Oriented Applications (VENOA2013), pp.747--pp.750, Jul. 2013, 台中（台湾)

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
池部　実　(Ikebe, Minoru)
大分大学・工学部・助教
研究者番号：50613650

(2)研究分担者
(　　)

研究者番号：