ネットワークログ分析による異常検知の可能性について

関谷 勇司[1]

概要：近年、IoT やクラウドといった技術の普及により、様々なデータ資源や計算資源がネットワークを介してアクセスし、利用できるようになっている。その一方で、サイバーセキュリティ対策が重要な課題となっており、情報漏えい等の事故も頻発している。このような状況に対応するためには、ネットワーク管理者とセキュリティ管理者が連携し、インシデントの兆候を的確に発見し対策することが求められている。しかし、現在のネットワーク運用やサイバーセキュリティ対策は、管理者の知識や経験に頼って行われる場合が多く、インシデント調査の手法が体系的に確立していない。そこで本論文では、先行研究に基づき各種のログ情報を利用することで、攻撃の兆候やインシデントの調査を体系的に行える手法を提案した。また、組織における実運用の一助となることを目指し、本手法を利用した分析事例を示した。最後に、さらなる分析と体系化のための課題点を示し、研究の方向性を述べた。

Discussing the possibility of abnormality detection by network log analysis

YUJI SEKIYA[1]

1. はじめに

近年、IoT (Internet of Things) やビッグデータといったキーワードに示されるように、様々なデバイスがインターネットに接続され、ネットワークを利用してデータを収集したり分析したりする手法が普及し始めている。様々なデバイスがインターネットに接続されているということは、攻撃対象になる可能性がある。特に近年は重要なデータの詐取や制御系システムの乗っ取り等を狙ったサイバー攻撃が増加し、その被害も深刻となっている。

このような攻撃に対して、既存のサイバー攻撃検知・防御システムでは、攻撃や侵入の事象を事後に検知する仕組みとなっており、被害を検知し対策を行う段階においては、既に情報が漏れた状態となっている。また、このような攻撃を検知・防御するシステムは、そのシステム自体の導入と運用、ならびに検知結果の利用に関して、ネットワークやサイバーセキュリティに関する専門的な知識を必要とする。また、当然ながらネットワークを利用した侵入や攻撃の場合には、その侵入経路や攻撃手法を発見し分析するにあたってセキュリティの知識のみならずネットワークの知識が必要となる。つまり、早期に発見して迅速に対策を行うためには、ネットワーク管理者とセキュリティ管理者が協調して対策に当たらなければならない。つまり、ネットワークの正常性とセキュリティを確保するためには、ネットワーク監視やセキュリティ機器を導入したとしても、それらから的確に情報を読み取る知識と経験を有したネットワーク管理者ならびにセキュリティ管理者が必要となり、さらにそれら担当者の連携が必要となる。

このような背景を受け、本論文ではネットワーク管理者やセキュリティ管理者の一助となるべく、知識や経験などの属人的な能力に頼るのではなく、体系化された手法としてサイバーセキュリティの向上に役立つ手法を提案した。具体的には、(1) ネットワーク運用とサイバーセキュリティの向上に役立つログ情報の定義と収集手法、(2) 各種情報を利用した攻撃の分析基盤の構築と事例、(3) サイバー攻撃の予知可能性、の 3 点について提案手法とその成果を述べ、実運用に役立つネットワークログ収集と分析の体系化について議論を行った。

2. ログ情報の定義と収集

本節では、ネットワーク管理者の観点から見たネット

[1] 東京大学情報基盤センター

ワーク運用管理に有用なログ情報と、サイバーセキュリティ担当者の観点から見たセキュリティ事象の特定と分析に有用なログ情報に関して定義する。

2.1 ログ情報の必要性

ネットワークの正常運用のみならずサイバーセキュリティの向上のためには、ネットワーク正常性監視が重要となる。セキュリティを脅かす攻撃や侵入行為が行われている際には、それがネットワークでの通信挙動として現れる場合が多い。具体的には、流量として現れる場合、もしくは通信先の多様性として現れる場合がある。流量の場合には SNMP を用いて取得できるトラフィック流量が情報として有用であり、通信先の場合にはフロー技術にて取得できる通信先 IP アドレスやポート番号の分散にて判別できる。つまり、ネットワークの定常状態がわかっていなければ異常状態を検知することも難しく、セキュリティ的な異常事象を検知することも難しくなる。

セキュリティ的に疑わしい事象が発生した場合には、その事象が本当にインシデントなのかを調査する必要がある。この場合、該当する機器の調査から得られる情報は有用であるが、ネットワークやセキュリティ機器にて記録されたログ情報も有用な情報となる。事象が発生したと思われる近隣の時間やそれ以前の時間に、被疑ホストからどのような通信が行われていたか、また DNS に対してどのような名前解決が行われていたか等の情報は、調査にあたって非常に有用な情報となる。この際、ネットワーク管理者とセキュリティ管理者の連携が必要となり、さらに個々のネットワーク機器やセキュリティ機器から収集したログ情報を収集し、統合して分析する必要がある。A この調査の一連の課程こそが知識と経験に基づいた属人的な手法となりがちであり、この課程を体系化することが管理者への大きな一助となる。

この調査課程を体系化したものとして、SIEM (Security Information Event Management) という概念が存在する。SIEM は、単一の機器ではなく複数の機器から得られる情報を組み合わせて攻撃や侵入を検知する手法であり、セキュリティ事象の調査と特定を行う管理者への大きな一助となる手法である。しかし、実際には SIEM はセキュリティ企業がそれぞれの経験や知識を基に、専任のセキュリティ担当者を雇用して実現しているサービスであり、一般的に体系化された手法ではない。そこで本研究では、SIEM のように多種のログを調査する基盤を提供することでシステムとして調査課程を体系化し、様々なレベルの管理者であっても一定レベルで調査と分析が行える手法を提供することを目指した。

なお、本論文にて扱うログ情報は、ネットワーク機器やセキュリティ機器といった機械が生成する情報を「インフラログ情報」と定義し、人間の意図や挙動を記録した SNS 等の情報を「ソーシャルログ情報」と定義する。

2.2 ネットワーク管理に有用なログ情報

ネットワーク管理者は、ネットワーク機器から収集できる情報を取得し日々のネットワーク管理に利用している。これらは「インフラログ情報」であり、数種類の基本的な情報がネットワーク管理に有用と定義できる。

最も一般的な情報はトラフィック流量に関する情報であり、SNMP[1] というプロトコルを利用して収集されている。SNMP を用いることで、機器のインタフェース単位のトラフィック流量であったり、何かパケット入出力にエラーが発生した場合のカウント値を収集することで、ネットワークの正常性を監視する。

さらに、より詳細なトラフィック情報を集める手法として、NetFlow[2] や sFlow[3] に代表されるフロー技術が広く利用されている。また、最近は IPFIX[4] というプロトコルも利用され始めている。NetFlow や IPFIX は、機器にて処理されるパケットを TCP や UDP のセッションという単位でまとめ、宛先と通信元に関する IP アドレスやポート番号、通信量などの情報を記録するプロトコルである。さらに、IPFIX ではパケットのペイロード部分を抽出し記録する拡張も実装されている。一方、sFlow はパケットをサンプリングにて抽出し、そのパケットに関する IP アドレスやポート番号、パケットサイズ等の情報を記録する。現在組織内のネットワーク構成に広く利用されている、

また、通信を直接処理するものではないが、ネットワークに付加的な機能を提供する機器として DHCP サーバや DNS サーバの存在があげられる。これらは通信を行うにあたって必要となる付加的な情報を端末に提供する役目を担っている。これらの機器からのログ情報を取得することで、どの端末がどの IP アドレスを取得し、どの通信先に対して名前解決を行ったかという情報を記録できる。さらに DHCP は、組織内においては 802.1X[5] といった認証機構と組み合わせて利用される場合も多く、その場合は端末の利用者であるユーザ情報もあわせて記録することが可能となる。他にも、LDAP や ActiveDirectory 等のディレクトリサービスと呼ばれるプロトコルを利用することで認証を行い、組織内の特定の資源に通信を許可する制御が行われている。これらの認証情報も有用なログ情報である。これらの情報は、多くの場合 syslog[6] と呼ばれるプロトコルにて転送され、記録される。

2.3 セキュリティ調査・分析に有用なログ情報

セキュリティ機器には、多くの種類が存在する。一般的にファイアウォールと呼ばれる、一定条件に従って通信を遮断する機器や、IDS/IPS と呼ばれる不正侵入検知ならびに防御のための機器が普及している。

これらセキュリティ機器からの情報は「インフラログ情

報」であり、何か異常や攻撃と思われる通信を検知した場合、その事象の重要度とともに通信に関する情報が syslog を用いて送出される。セキュリティ管理者は、これら機器が判定し送出したログ情報をもとに、攻撃や侵入と思われる通信を発見し対策を行う。しかし、一般的にこれら機器が検知し送出するログ情報はかなりの量に及ぶことが多く、誤検知や過検知と呼ばれる間違った情報も含まれている。そのため、セキュリティ管理者はログ情報の中から組織にとって本当に危険と思われる情報を見つけ出し、その真偽を確認する作業を強いられる。この際に、セキュリティ機器からのログ情報のみならず、ネットワーク機器からのログ情報が必要となる場合が多い。

2.4 攻撃の動機分析に有用なログ情報

セキュリティ機器を用いたログ収集からでは得られにくい情報として、攻撃の動機があげられる。散発的に行われる愉快犯による攻撃も数多く存在するが、組織的かつ計画的に情報を詐取しようとする攻撃も見受けられる。また、政治的思想や経済的な目的を持った大規模なサービス妨害攻撃も発生している。このような攻撃には攻撃者の意図が存在しており、そのきっかけとして社会的な事件や記念日といったものが考えられる。これらの攻撃動機につながる情報をネットワーク上から集めるにあたっては、Web 上の記事や書き込み、SNS への書き込み、セキュリティ企業や組織が公開しているブラックリスト IP アドレスやドメイン名、CVE[*1] 等の脆弱性公開ページ等が有用であり、日々のクローリングによって収集する。

3. ログの収集と分析

前節にて、ネットワークの正常性を確認しセキュリティに関連する事象を調査・分析するにあたって有用なログ情報、ならびにその動機の分析に有用な情報を定義した。本節では、それらの収集と分析を体系的に行う基盤について提案する。

先行研究として、NECOMA Project[*2] があげられる。これは日欧の組織が共同研究にて行った、多種多様なデータを用いたサイバー脅威の検知と防御に関する研究である。この研究では、機器が送出する各種テキスト形式のデータやバイナリ形式の「インフラログ情報」や、Web サイトや SNS から抽出されたサイバーセキュリティ脅威に関する「ソーシャルログ情報」を統合して蓄積し、分析するための基盤である、MATATABI システム [7] を提案した。

MATATABI システムは Apache Hadoop[*3] とその上で動作する Apache Hive[*4]、ならびに Facebook Presto[*5] といったオープンソースソフトウェアを利用して構築されたデータ蓄積と分析のためのシステムである。機器が送出するインフラログ情報は、数分間に一回のバッチジョブによりデータ形式に応じた変換プラグインを経由して、MATATABI システム上の hdfs に蓄積される。MATATABI システムに蓄積された情報は、同様にバッチジョブによって SQL データベースのようなテーブルスキーマに変換され、Presto を用いて横断的に情報を検索することが可能となる。例えば、ある時間にあるホストから発生した通信を調べるにあたっては、時間範囲と src IP address を指定することで、そのホストから発生した通信の他にもそのホストが宛先となった通信、ならびにそのホストが DNS を利用して検索した名前、それに関連したセキュリティ事象のログ等が統合的に検索され、結果として表示される。MATATABI システムでは、管理者が知識や経験に基づいて関連付けして個々の機器から検索している情報を、統合的に検索する手法を提供することで、調査手法の体系化を実現した。

この MATATABI システムを用いて、DNS への UDP 増幅攻撃や毒入れ攻撃を検知した例として、論文 [8][9] があげられる。これらの論文では DNS の名前解決情報とトラフィック情報を統合して分析し、問い合わせと応答のクエリ数の不一致や不正な DNS フラグパターン等から攻撃の兆候を発見する手法を紹介した。その他にも、C&C (Communication and Control) サーバへの通信に用いられる、DGA (Domain Generation Algorithm) と呼ばれる名前生成を行うマルウェアに感染したホストの発見や実際の通信挙動の解析 [10] が事例としてあげられる。

さらに、Web 記事や SNS から収集できる情報を用いて、攻撃の可能性が高まっていることを予測できないかと考え、検証を行った [11]。この結果として、過去に発生した実際のサイバー攻撃に関してそれを示唆する何らかの情報が Web や SNS 上に示唆されていた場合が多いことは判明したが、どのような攻撃がいつ発生するのか、という実運用に役立つ情報を分析するには至っていない。しかし、このようなソーシャルログ情報からサイバー攻撃の動機を発見し、実際の攻撃種別とその時期を予測することが有用であると考え、機械学習や深層学習を用いてサイバー攻撃を予測する NML : Network Muscle (not Machine) Learning Project[*6] を立ち上げ、手法の確立に向けた研究を行っている。

これらの事例を通して、ログ情報を定常的に収集し統合的に蓄積しておくことで、セキュリティ的に疑わしい事象が発生した場合に容易に調査と分析を行うことが可能であ

[*1] https://cve.mitre.org/
[*2] http://www.necoma-project.eu/
[*3] http://hadoop.apache.org/
[*4] https://mapr.com/products/product-overview/apache-hive/
[*5] https://prestodb.io/
[*6] https://www.sekiya-lab.info/?page_id=416

ることがわかった。その一方で、提案手法によるログ収集と分析からできる異常検知やサイバーセキュリティ対策の限界もわかった。

まず、リアルタイム性の欠如があげられる。ログ情報を収集し統合するという手法では、断続的に送出されるインフラログをリアルタイムに処理することが難しい。複数種類のログ情報を統合して検索するための仕組みを提供するためには、何らかのフィールドをキーとしたインデキシングが必要であり、そのためのデータ変換とインデキシングに一定時間を要するためである。特にインフラログ情報は大容量になりがちであり、分散ストレージや分散処理のための Hadoop 等のソフトウェアを用いると、その処理速度の点から現在進行形で発生している事象を分析することが難しい。この解決策としては、論文?において、複数種類のログ情報を共通のテキスト形式に変換し、それをリアルタイムに近い時間範囲にて蓄積してインデキシングする手法を提案した。しかしこの手法も規模性の面で問題が残っており、さらなる研究開発が必要である。

次に、定常状態の判定手法があげられる。インフラログ情報からどの観点に基づいて定常状態の数値を算出し、その数値をどの程度逸脱すれば異常値と判断するかは提案手法では自動的にはできず、やはり人間の知識や経験にもとづいて値を設定するしかない。この解決策としては、機械学習や深層学習を用いて、定常状態の学習による異常状態の検知が可能であると考える。定常状態を学習させるためには、深層学習が有効であると考える。これは前述の NML Project において研究を進めている。

最後に、攻撃の予測があげられる。攻撃手法やその時期の予測は、攻撃が行われた後に対応するリアクティブな対策ではなく、事前に予防策を講じるというプロアクティブな対策を可能とする。防戦一方となりがちなサイバーセキュリティ対策において、プロアクティブな対策はこれから必要となる対策法であり、そのためにはより詳細な攻撃の動機分析が必要と考える。この研究も前述の NML Project において進めている。

4. おわりに

本論文では、ネットワークの正常性とセキュリティを確保するために複数のログ情報を利用する手法を体系的に提案した。現在のネットワーク運用とセキュリティの確保は、管理者の知識や経験に依存した経験則的な手法により行われており、様々な組織にて一定のサイバーセキュリティ対策レベルを確保するためには、体系的な手法が必要となる。そこで本提案手法では、まず収集すべきログ情報を定義し、その情報を統合的に蓄積し、解析するためのシステムについて紹介した。また、そのシステムを用いた実際の分析事例を紹介し、組織のログ情報管理とセキュリティ調査を一定レベルで体系化する手法を提案した。最後に、残る課題点について述べ、今後の研究の方向性を示した。

謝辞

本論文を執筆するにあたり、先行研究となった NECOMA Project にて共同研究を行った各組織の研究者、ならびに NML Project にて共同研究を行っている東京大学の岡田和也助教、石原知洋助教、奈良先端科学技術大学院大学の宮本大輔特任准教授、IIJ-II 社の島 慶一研究員、阿部 博研究員、Preferred Networks 社の土井 裕介氏、浅井 大史氏に感謝します。

参考文献

[1] D. Harrington, R. Presuhn, and B. Wijnen : An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks, Internet Engineering Task Force, Request For Comments 3411, December 2002.

[2] B. Claise, Ed. : Cisco Systems NetFlow Services Export Version 9, Internet Engineering Task Force, Request for Comments 3954, October 2004.

[3] P. Phaal, S. Panchen, and N. McKee : InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks, Internet Engineering Task Force, Request For Comments 3176, September 2001.

[4] B. Claise, Ed., B. Trammell, Ed., and P. Aitken : Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, Internet Engineering Task Force, Request for Comments 7011, September 2013.

[5] Mick Seaman : 802.1X-2010 - Port Based Network Access Control, IEEE Standards Association, February 2010.

[6] R. Gerhards : The Syslog Protocol, Internet Engineering Task Force, Request For Comments 5254, March 2009.

[7] Hajime Tazaki, Kazuya Okada, Yuji Sekiya and Youki Kadobayashi : MATATABI: Multi-layer Threat Analysis Platform with Hadoop, In Proceedings of International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014), September 2014.

[8] Tomohiro Ishihara, and Yuji Sekiya : Case-based study and Discussion of threat analysis on DNS traffic using MATATABI platform, IA2016 Workshop on Internet Architecture and Applications 2016, IEICE Tech. Report, vol. 116, no. 282, IA2016-48, p. 99-102, November 2016.

[9] Tomohiro Ishihara, Hajime Tazaki, Kazuya Okada, Daisuke Miyamoto, and Yuji Sekiya : DNS Traffic Analysis Platform with Hadoop Framework, IEICE Technical Report, vol. 113, no. 502, ICSS2013-80, pp. 131-135, March 2014.

[10] Yuji Sekiya : MATATABI : Cyber Threat Analysis and Defense Platform using Huge Amount of Datasets, APNIC 40, APOPS Technical Session, Jakarta, Indonesia, September 2015

[11] Munkhdorj Baaatarsuren, and Yuji Sekiya : Cyber attack prediction using social data analysis, IOS Press, Journal of High Speed Networks, vol. 23, no. 2, pp. 109-135, DOI 10.3233/JHS-170560, April 2017.