2　インターネットのセキュリティ技術
2 Technologies for Security of the Internet

2-1　インターネットリスク分析モデル
2-1 A Study on Process Model for Internet Risk Analysis

中尾康二　丸山祐子　大河内一弥　松本文子　守山栄松
NAKAO Koji, MARUYAMA Yuko, OHKOUCHI Kazuya, MATSUMOTO Fumiko, and MORIYAMA Eimatsu

要旨

近年、インターネットの汚染状況、リスク状況を的確に把握することにより、ネットワークを安心安全に管理運用する必要性が強く認識されている。筆者らは、ISP（インターネットサービスプロバイダー）との連携により、該リスク状況を可能な限りの実時間で認識するために、必要なデータの収集蓄積、分析前処理及び高度分析処理にかかわるプロセスモデルを検討した。さらに、具体的なシステム化に向けた考察、Telecom-ISAC との連携などについても考察を加える。

Security Incidents such as Scans and Probes, Computer Intrusions, Malicious Software (Viruses, Worms, etc), Computer Sabotage and Damage (DoS Attacks, etc.) are well recognized in our IT environments today. One of the key solutions to protect against the above security incidents and to minimize damages, "INTERNET RISK" should be efficiently observed by means of incident monitoring and analysis in conjunction with supports from ISPs (Internet Service Providers). This paper discusses the requirements of the incident analysis and proposes the model for Internet Risk Analysis. This activity was carried out with a support of Telecom-ISAC Japan.

[キーワード]
ネットワークセキュリティ，データマイニング，ログ分析，分析モデル
Network security, Data mining, Log analysis, Analysis model

1　まえがき

現在の我が国の IT 戦略において推進されてきた国内ブロードバンドインターネット環境の普及に伴い、情報通信基盤はますます重要な社会インフラとなっており、その安全性を確保することが大きな課題となっている。

更なるネットワーク基盤の高度化及び情報通信技術の発展を促進するためには、中核の要素技術であるネットワークのセキュリティ対策やサイバー攻撃対策を実施することが急務である。このため、ネットワークにおけるリスク状況を把握し、的確な対策に基づくリスク回避やリスク低減を行うことが期待される。ネットワークのリスク状況の把握については、ネットワーク上で収集できるセキュリティーログやトラフィックデータに基づく高度な傾向分析、頻度分析などのインシデント分析が重要であり、これらの分析結果から、現状のネットワークに対する危険レベル及びネットワーク設備に対する影響度を導出し、情報通信基盤の安全性を確保することが必要である。

検討の第1段階として、筆者らは、ISP(インターネットサービスプロバイダー)との連携により、該リスク状況を可能な限りの実時間で認識するために、必要なデータの収集蓄積、分析前処理及び高度分析処理にかかわるプロセスモデルを検討した。

本稿では、検討を進めたデータ収集蓄積、分析前処理及び高度分析処理にかかわるプロセスモデルを報告するとともに、今度の具体的なシステム化に向けた考察、Telecom-ISACとの連携などについても考察を述べる。

2 研究の背景と要求条件

インターネットにおける各種ログ監視、ログ分析については、侵入検知システム(IDS)やファイヤーウォール(FW)などが排出するログ情報をベースに研究、商用化が進められている。本章では、現状の研究動向及び求められる要求条件について述べる。

2.1 ログ分析にかかわる現状の活動

IDSやFWのログ情報を収集し、それらの情報に対して統計的な分析を行い、ネットワークのリスク状況を把握する試みが多く見られる。セキュリティベンダーの提供するSOC(セキュリティオペレーションセンタ)では、商用ビジネスとして、広域的に収集したログを分析し、その結果を利用者に提供するビジネスを進めている。そこで分析されている対象は、IDS/FWログで、分析の手法も統計分析及び簡単な相関分析を実施している。

統計分析では、ネットワークでのパラメータ(ポート番号、対地など)を用いて、傾向分析を実施している。また、相関分析は、同種のインシデントが発生している場合に、同種であることの相関分析により把握し、不要なログ情報の解析を避けるためのフィルター機能として利用している。

多くの場合は、これらの機能を提供する利用者への情報提供が主で、ワーム/ウィルスの蔓延度の認識、不正侵入の発見などが主な目的となる。

近年、日本のJPCERT/CC[1]、IPA[2]、@Police[3]、Telecom-ISAC Japan[4]、韓国のKRCERTでは、広域的にIDSを配備させ、多地点でのインシデント情報を収集し、それらの傾向分析を実施している。現状のログ分析においては、次節で示す要求条件が抽出できる。

2.2 ネットワークリスク分析における要求条件

上述のログ分析の現状をかんがみ、インターネットのリスク分析をより広域に高精度に更に効果的に実施するためには、以下のような要求条件(要件)を考慮した分析を実施する必要がある。

(1) モニター対象に関する要件

現状のモニター対象は、IDS/FWの排出するインシデントログ情報であるが、ISPとの連携をとることにより、トラヒック情報をも収集し、インシデントログ情報と統合的な分析を進める必要がある。

(2) モニター領域・地域に関する要件

現状では主に利用者サイド(企業、一般利用者など)におけるモニターを実施しているため、モニター領域が限定されている。さらに、広域に網羅的なモニターを実施するためには、ネットワーク事業者であるISP(インターネットサービス提供者)などにモニター点を拡大する必要がある。

(3) 大容量処理に関する要件

インシデントログ情報のみならず、トラヒック情報を収集することを想定すると、収集される情報が膨大となる。これらの大量な情報を効率的に処理できるような大容量処理能力が必要である。単純に大容量化するだけでなく、不要な情報を落とすためのフィルター処理の検討も必要になる。

(4) モニターデータの分析精度にかかわる要件

大容量のモニターデータを高精度に、しかも迅速に分析するためには、分析手法についても階層的な分析の検討や他の分析手法(コード静的解析、動的解析など)との組合せ分析などの手法検討が必要である。

(5) モニターデータの機密性にかかわる要件

収集されるモニターデータの機密性にかかわる課題は、個人情報保護や通信事業者の約款などと関連して、その取扱いに充分な注意が必要

である。技術的には、分析に不要である情報を事前に削除して収集させるなどのフィルター機能の検討が必要である。

3　インターネットリスク分析のためのプロセスモデル

2 で述べたリスク分析のための要件を満足することを目的として、具体的なシステム構築の中で分析プロセス設計を進めている。設計したリスク分析プロセスモデルは、図 1 に示すとおり、以下の構成要素により成り立つ。

3.1　モニターデータソース

ISP（インターネットサービスプロバイダ）の安定運用を支援する Telecom-ISAC Japan との連携を前提に、広域なモニター点から、IDS、ファイヤーウォール等のインシデントログデータ及び複数 ISP におけるトラヒックログ情報を収集する。これらのモニターデータにより、インシデントログとトラヒック情報のそれぞれの分析のみならず、両者の情報の相関をとることにより、更なる高精度な分析を目指す。ここで扱うモニターデータは静的なオフラインデータ及び動的なオンラインデータの両方を扱うことを前提として設計する。なお、静的なオフラインデータについては、問題点が ISP に関連する箇所で発生し、それを局所データのモニターにより分析解決を実施することを想定している。

3.2　フィルター、ダイジェスト処理モジュール

モニターされたデータを分析がしやすい形式にフィルター処理又はダイジェスト処理を実施する。すなわち、分析に必要なデータ項目が欠落することなく、不要なデータをフィルター処理により削除し、より軽量で扱いやすいデータにダイジェスト化する処理をつかさどるモジュールが本部分である。ダイジェスト化としては、TCP（UnixTime、IP アドレス（元／先）、ポート番号（元/先）、ペイロード長、フラグ（IP/TCP）、HTTP メソッド等）、UDP（UnixTime、IP アドレス（元／先）、ポート番号（元／先）、ペイロード長等）、ICMP（UnixTime、IP アドレス（元／先）、タイプ、コード、ペイロード長等）などのパラ

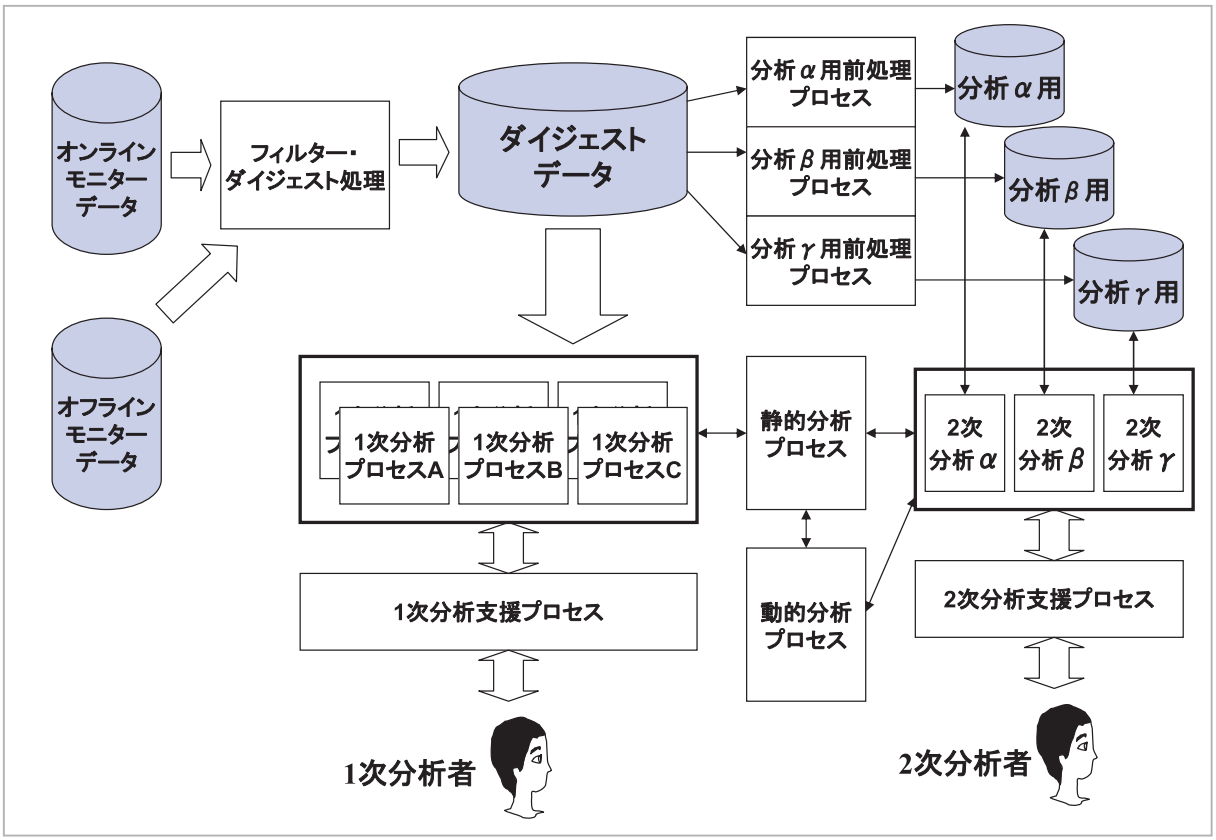

図1　インターネットリスク分析モデル

メータを保有するデータに圧縮する。また、フィルター処理としては、データ量が大きい不要なペイロードを削除することによりスリム化させること及び同様な攻撃が多数存在するなどデータ間の相関処理により正規化処理が挙げられる。これらのダイジェスト化手法については、分析(1次、2次とも)の手法及びその結果から、スパイラル的な検討が必要である。本ダイジェスト処理、フィルター処理により、データ量を9割程度の減少を目指す。

3.3 ダイジェストデータDB

上記のフィルター、ダイジェスト処理の結果は、以後の分析(1次及び2次)において利用される。データベース(DB)としては、各々のダイジェスト情報のパラメータによって分析が容易な構成を有する。本DBにて扱われるデータ量は、複数(数点を想定)のモニター点から送られてくるモニターデータを含むため、インシデント情報(約1GB/日)、トラヒックデータ(約500GB/日)と想定すると、数日で数TBのデータ量が蓄積されることになり、データ保存期間、保存手法などの検討が必要である。

3.4 1次分析処理プロセスモジュール

1次分析処理の目的は、現状のインターネットのリスク状況(傾向など)を極めて迅速に把握することである。この目的を達成するために、本モジュールは、蓄積されたダイジェスト情報を用いて、必要なパラメータにより迅速な分析(1次分析と呼ぶ)を実施するプロセス群により構成される。具体的には、本分析では、IPアドレス、ポート、対地、アプリケーションなどのパラメータに基づく基礎統計処理を実行する。これらの1次分析結果は、次節で述べる1次分析者に対しViewerにて表示され、1次分析者とのインターラクションにより、必要なパラメータの組合せなどによる更なる1次分析を進めることも可能としている。この1次分析処理の結果、早期に対策を行う必要があることが判明した場合は、Telecom-ISACの指揮の下、ISPにおいて緊急の第1次措置が取られることとなる。

3.5 1次分析支援プロセスモジュール

本モジュールは、上記の1次分析処理を1次分析者の立場から分析支援を実施するためのモジュールである。初期1次分析の結果、更なるパラメータの組合せによる統計分析が必要な場合は、1次分析者から分析処理モジュールに再分析要求を上げ、1次分析で必要となる統計分析を実施することが可能となる。具体的な本手法については、別稿[5]に述べる。

3.6 2次分析用前処理プロセスモジュール

上記の1次分析処理を更に高精度化するために、2次分析処理を実施する。2次分析の元となるデータは、ダイジェストデータDBに格納されるダイジェストデータを加工し、以降のデータマイニング手法などの分析に適した形に前処理されるものである。具体的には、ダイジェストデータを様々な観点から集計し、分析に使用する新たなデータ項目を作成する処理である。このように作成したデータ項目により、データに内在する特徴を明らかにし、また以降の分析をより効果的に行うことができる。

2次分析用前処理の一例として、(1)ある特定の時間単位でデータのサマリを作成する、(2)あるデータ項目をキーにしてデータサマリを作成する、という処理があげられる。図2は、あるサーバへのアクセスを、HTTPメソッドに着目して1時間ごとのアクセス数を記録し、その経時変化をプロットしたものである。単にログを見ただけでは分からないウィルスの亜種の特徴が明らかになっていることが確認できる。

本前処理の後に行われる2次分析の成否は、その前処理によって決定されると言っても決して過言ではなく、したがって、本処理には、十分な時間をかけて、分析の入力となるデータ項目を検討する必要がある。なお、ここで必要となる前処理プロセスモジュールは、各2次分析処理プロセスに特化した前処理であり、2次分析手法の数だけ必要となる。

3.7 2次分析用DB

上記の2次分析前処理プロセスの結果を格納するDBであり、2次分析ごとに存在する。これらの2次分析用DBは、物理的に別々のDBモ

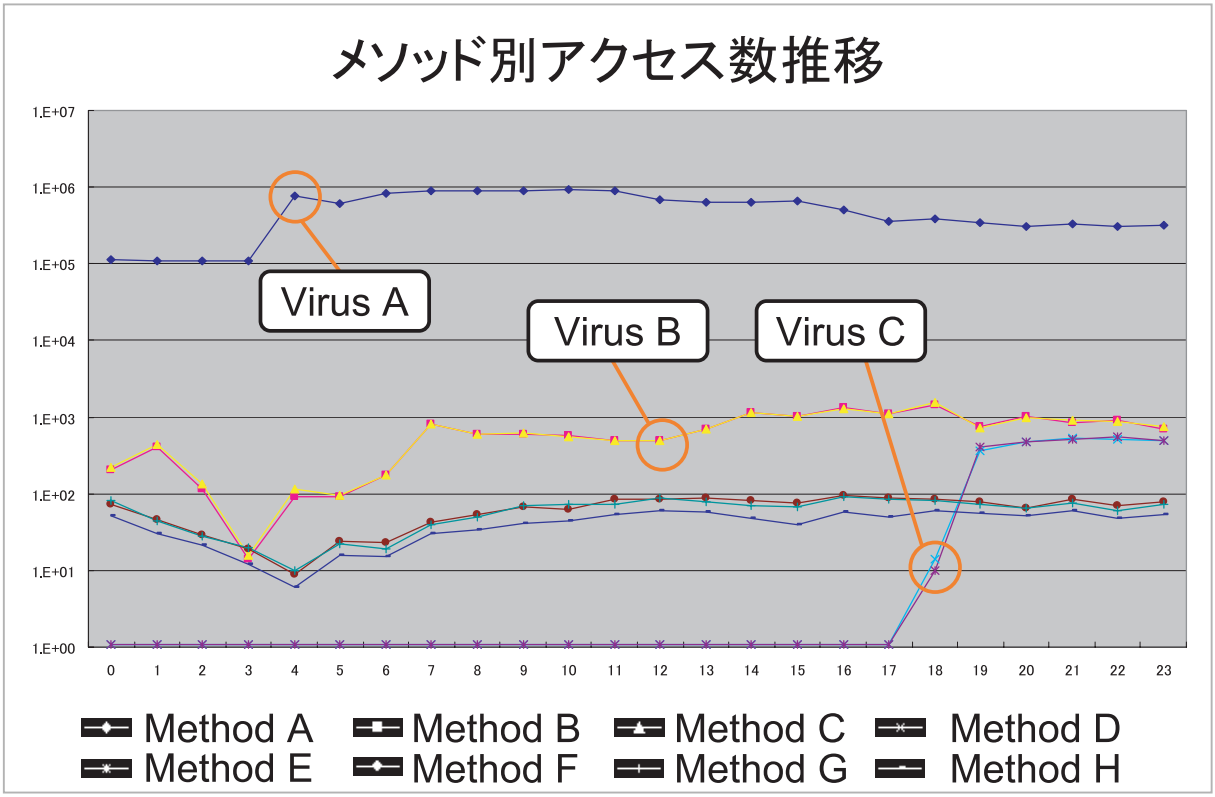

図2 HTTPメソッドを用いた前処理データ表示

ジュールに格納される必要がないが、それぞれの 2 次分析プロセスからは独立した形で存在するものである。

3.8 2 次分析処理モジュール

2 次分析処理の目的は、1 次分析処理では発見できないような高精度の分析を実施することで、微細な変化点や特徴点の変化をとらえ、インターネットのリスク状況をきめ細かく探求することである。2 次分析処理は、固定的な処理を実施するよりも、いろいろな分析手法を試すことにより、その分析精度や効果を評価でき、より有効な分析手法を見いだしていくことを前提としている。現在、異常行動検出手法及び時系列変化点検出手法の検討を進めており、これらは統計的に全体のデータの持つパタンを学習し、学習した統計的モデルから外れるデータを異常として検出する手法である[6]。特に、データの変化に適応的であるために、過去のデータを程よく忘れながら学習を行い、リアルタイムに異常を検出することを特徴とする。

・異常行動検出手法

大量のアクセス履歴データ（以下、セッションと呼ぶ）のなかから、異常なセッションを検出することを目的とし、「異常セッションの検出（全体のアクセスパタンから大きく外れる）」及び「異常行動パタンの検出」などの異常検出を行う。特に、異常行動パタン検出については、全体のアクセスパタンにはないセッションが同一パタンを持って集中発生する場合に、これを新しい行動パタンとして検出する。ここで、新しく増加/減少した行動パタンは、動的モデル選択理論[7]に基づいて時間的に変化する行動パタン数の最適な数を計算することにより、発見できる。

これらの異常検出により、アクセス先やアクセス順序が通常のセッションとは大きく異なる脆弱性チェックやワームによる攻撃の検出及びDos攻撃発生時やワーム発生時のアクセスパタンを発見するのに有効である。

・時系列変化点検出手法[6]

トラフィック量の時系列データの急激な変化を検出することを目的とし、以下の 2 種類の異常検出を行う。1）外れ値検出、2）変化点検出。1）

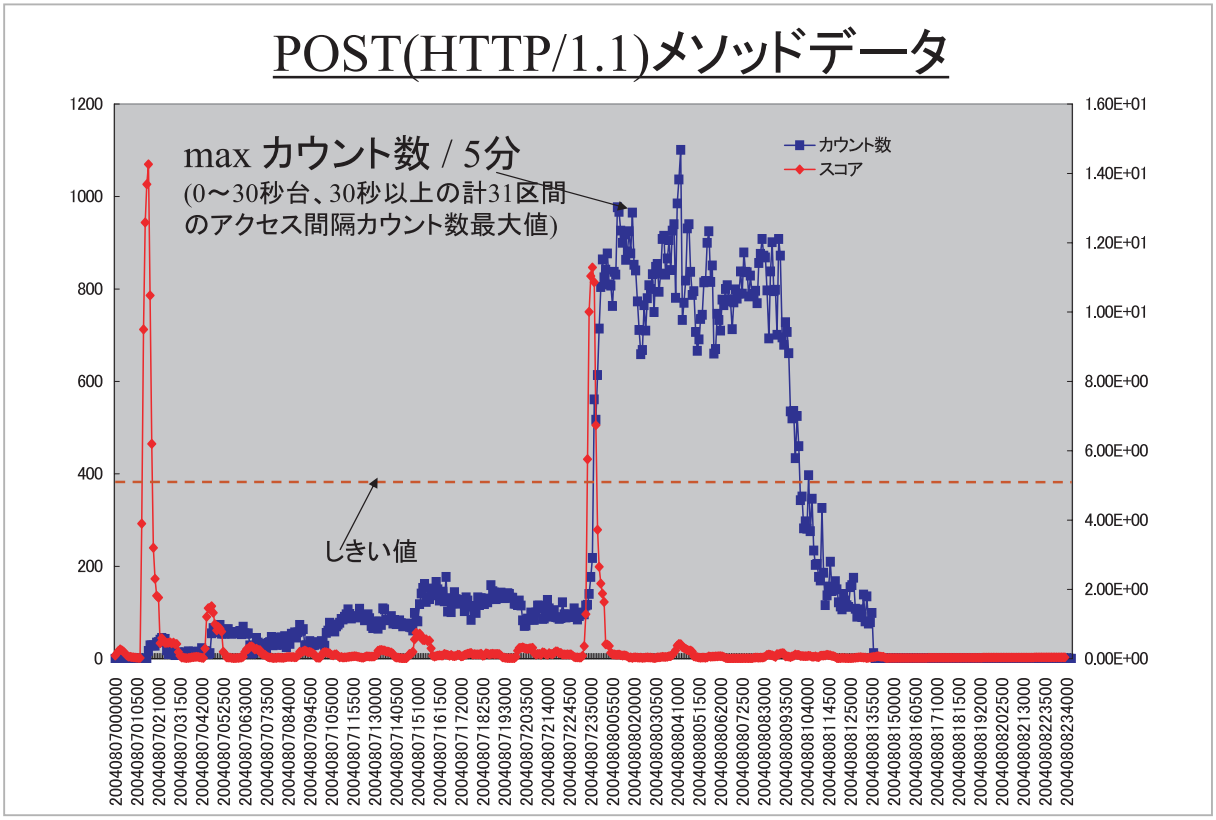

図3 変化点検出の事例（Anntinyの場合）

では、全体のトラフィック量の時系列的なパタンから大きく外れる値を検出する。2) では、トラフィック量が急激に変化し始める開始点を検出する。これらの異常検出により、Dos 攻撃やワームの発生をトラフィックの急増をいち早く検出することができる。以下、図 3 に Anntiny ワームの場合の POST (HTTP/1.1) メソッドデータから得た変化点検出の例を示す。

3.9 2 次分析支援プロセスモジュール

本モジュールは、上記の 2 次分析処理を 2 次分析者の立場から分析支援を実施するためのモジュールである。1 次分析支援プロセス同様、初期 2 次分析の結果、更なるパラメータの組合せによる複合 (統合) 分析が必要な場合は、2 次分析者から分析処理モジュールに再分析要求を上げ、2 次分析で必要となる高度分析処理を実施することが可能となる。

3.10 静的分析・解析プロセスモジュール

本モジュールでは、ワーム、ウィルスなどの挙動をそのコードの静的分析を実施することにより、プログラム上の挙動予定を事前に確認する。1 次分析処理や 2 次分析処理などのモニターデータをベースとした分析において、本モジュールの分析結果はその基礎データとして非常に有益であると言える。

3.11 動的分析・解析モジュール

上記の静的分析におけるコード解析では、コードを実行させた状態での評価ができない。このため、本モジュールでは、ある閉領域においてワームやウィルスを実行させ、実際の挙動を分析するとともに、ネットワークへの影響度についても検証する。本モジュールは、特に 2 次分析において有効活用され、動的挙動分析を具体的に行うことで、高精度分析結果の検証をスパイラル的に実施することが可能である。具体的な本手法については、別稿 [8] に述べる。

4 考察

以上、インターネットリスク分析のためのプロセスモデルを検討したが、ログ分析のための

要求条件及び Telecom-ISAC Japan との関係について、以下に考察する。

4.1 要求条件に関する考察

⑴ モニター対象に関する要件

Telecom-ISAC との連携により、可能な限りのトラヒック情報及びインシデントログ情報の収集を行うことができる体制を確立している。

⑵ モニター領域・地域に関する要件

具体的なモニター点の拡大ができていないが、本件についても、Telecom-ISAC の協力で、複数の ISP によるモニターを進める予定である。現在、数箇所の ISP 事業者にモニターを設置すると同時に、遠端の利用者の情報を収集するための策を検討中である。本モニターについては、モニターポイントでの収集のみならず、一時蓄積を行う手法についても検討を進める。

⑶ 大容量処理に関する要件

大量なログ情報をいかに効率的に分析処理するかといった命題は大きな課題である。大量なデータをフィルターする機能は、モニター側と収集（集約）側に設置する考え方がある。大量なデータの一定の保存を考慮すると、両方にフィルター機能（一部は正規化処理相当）を保有する必要があると考える。収集されたデータについては、ある程度長期間のログ分析ができるように、一定期間の保存を考慮する。

特に、オンラインにてデータを収集する場合は、いったんすべてを蓄積することをせずに、順次（準リアルタイム）ダイジェスト処理をしていくことにより、大容量への対応をとるべきである。これらの処理については、並列処理が必須である。さらに、上記のモニターポイントでの一時蓄積を並行して考慮すると、並列処理と分散処理を同時に考慮する必要がある。

⑷ モニターデータの分析精度にかかわる要件

今回のモデル化では、目的の異なる分析については、極力並列的に分離作業とし、迅速に傾向分析を実施する 1 次分析処理と高精度な分析を実施する 2 次分析処理に分けた。しかし、ダイジェストデータについては、共通に利用するデータであるため、双方の分析処理からアクセスできるようにした。また、ワーム等の静的解析結果を双方の分析処理から参照できるように設計した。さらに、2 次分析処理については、詳細、高精度分析結果をテストベッドでするための動的分析との連携も考慮した。

⑸ モニターデータの機密性にかかわる要件

収集されるモニターデータの機密性にかかわる課題は、個人情報保護や通信事業者の約款などと関連して、その取扱いに充分な注意が必要である。技術的には、分析に不要である情報を事前に削除して収集させるなどのフィルター機能の検討が必要である。

Telecom-ISAC との連携部分が多々存在するが、本稿で検討した分析モデルは、上記のそれぞれの要件をほぼ満足しているものであり、具体的な今後の実装及び評価により、本機能要件の達成度を検証していく必要がある。

4.2 Telecom-ISAC Japan との連携

Telecom-ISAC は、ISP 安定運用のための連合組織であり、具体的な各 ISP における情報モニター、分析結果に基づくインシデント対策立案、必要な情報の ISP への情報広報機能、各国の同種活動との連携などの役割を担っている。

筆者らは、Telecom-ISAC との分析協力の関係から連携を図ることにし、ISAC からのモニターデータの提供に応えて、必要な高精度分析結果を返すスキームを考えている。

5 むすび

インターネットにおけるリスク状況をいかに的確に、迅速に把握するかといった命題を解決することは難しい。しかしながら、ISP の連合である Telecom-ISAC との連携を図ることにより、より生の情報をモニターし、迅速に、さらに精度の高い分析を実現することができる。本稿で提案した分析プロセスのためのモデルは、分析のための要求条件をかんがみ、これまでの思考錯誤的なトライアルの中から整理されたものであり、今後の活動の第一ステップであると認識する。

さらに、本分析活動を確実なものにするためには、よりコンスタントなモニターデータの収集、分析者（1 次及び 2 次）とのヒヤリング調整、

更なる分析手法の開拓、分析結果の自動対策化など、検討・解決すべき課題が山積である。しかしながら、脅威が日に日に増すインターネットにおいて、セキュリティを充分に考慮したISPの安心安全な運用は、これまで以上に重要となることは明白であり、その一助となる本分析研究に精力的に取り組む所存である。

最後に、本研究を進めるに当たり、貴重な助言、ご支援をいただきました横河電気　馬場様、鈴木様に感謝すると同時に、Telecom-ISAC Japanの皆様のご協力にも深く感謝します。

参考文献

1 JPCERT/CC Internet Scan Data Acquisition System (ISDAS) http://www.jpcert.or.jp/isdas/
2 IPA http://www.ipa.go.jp/security/
3 警察庁セキュリティポータルサイト@police http://www.cyberpolice.go.jp/detect/observation.html
4 Telecom-ISAC Japan https://www.telecom-isac.jp/
5 松本，馬場，井澤，中尾，“ISPのためのネットワークセキュリティ分析支援ツールの設計と実装”，SCIS2005に投稿
6 K.Yamanishi, J.Takeuchi, "A Unifying Approach to Detecting Outliers and Change-Points from Non stationary Data", The Eighth ACM SIGKDD International Conference on Data Mining and Knowledge Discovery, ACM Press, pp. 676-68.
7 Y.Maruyama, K.Yamanishi, "Dynamic Model Selection and Its Applications to Computer Security", The IEEE Information Theory Workshop 2004, (http://ee-wcl.tamu.edu/itw2004/)
8 藤長，森井，中尾，“ウィルス分析のためのテストベッド構築”，SCIS2005 投稿

中尾康二（なかおこうじ）
情報通信部門セキュリティ高度化グループリーダー
情報セキュリティ技術

松本文子（まつもとふみこ）
情報通信部門セキュリティ高度化グループ研究員
ユーザインタフェース、セキュリティログ分析

丸山祐子（まるやまゆうこ）
情報通信部門セキュリティ高度化グループ専攻研究員
データマイニング技術

守山栄松（もりやまえいまつ）
情報通信部門セキュリティ高度化グループ主任研究員
セキュリティ、移動通信

大河内一弥（おおこうちかずや）
情報通信部門セキュリティ高度化グループ専攻研究員
データマイニング技術