(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2009-37545
(P2009-37545A)
(43)公開日　平成21年2月19日(2009.2.19)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 21/22	(2006.01)	G06F	9/06	660N	5B276

審査請求 未請求　請求項の数 20　ＯＬ　（全 24 頁）

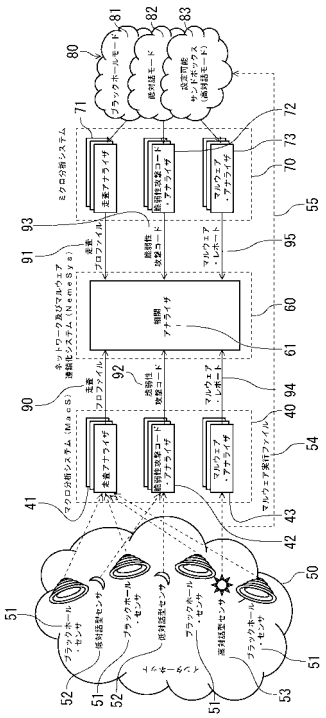
(21)出願番号　特願2007-203281 (P2007-203281)
(22)出願日　平成19年8月3日 (2007.8.3)

(71)出願人　301022471
独立行政法人情報通信研究機構
東京都小金井市貫井北町４－２－１
(74)代理人　100130111
弁理士　新保　斉
(72)発明者　中尾　康二
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内
(72)発明者　吉岡　克成
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内
(72)発明者　井上　大介
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内

最終頁に続く

(54)【発明の名称】マルウェアの類似性検査方法及び装置

(57)【要約】
【課題】　マルウェアのミクロ分析及びマクロ分析の相関に基づくマルウェアの検査精度を向上させると共に、効率よくマルウェアの検査を行う技術を提供すること。
【解決手段】　ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査方法を提供する。ミクロ分析側ではアドレス走査情報検出手段７１、脆弱性攻撃コード検出手段７２、マルウェア分析手段７３がそれぞれ設定可能サンドボックス８０を用いて複数のレベル８１～８３の分析を行うと共に、マクロ分析側でも入力したマルウェアに対して各レベルのセンサ５１～５３により走査層、脆弱性攻撃コード層、マルウェア層について観察し、アドレス走査情報検出手段４１、脆弱性攻撃コード検出手段４２、マルウェア分析手段４３が分析する。両者の結果は挙動比較手段６１によって同一性の比較、又は相関関係を算出する。
【選択図】　図３



【特許請求の範囲】
【請求項１】
ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査方法であって、
コンピュータの第１アドレス走査情報検出手段が、該第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出工程、
コンピュータの第１マルウェア分析手段が、該第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析工程、
コンピュータの第２アドレス走査情報検出手段が、該第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出工程、
コンピュータの第２マルウェア分析手段が、該第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析工程、
コンピュータの挙動比較手段が、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較工程
を有することを特徴とするマルウェアの類似性検査方法。
【請求項２】
ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査方法であって、
コンピュータの第１アドレス走査情報検出手段が、該第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出工程、
コンピュータの第１脆弱性攻撃コード検出手段が、該第１の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第１脆弱性攻撃コード検出工程、
コンピュータの第１マルウェア分析手段が、該第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析工程、
コンピュータの第２アドレス走査情報検出手段が、該第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出工程、
コンピュータの第２脆弱性攻撃コード検出手段が、該第２の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第２脆弱性攻撃コード検出工程、
コンピュータの第２マルウェア分析手段が、該第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析工程、
コンピュータの挙動比較手段が、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較工程
を有することを特徴とするマルウェアの類似性検査方法。
【請求項３】
前記第１アドレス走査情報検出工程において前記第１アドレス走査情報検出手段が、及び前記第２アドレス走査情報検出工程において前記第２アドレス走査情報検出手段が、そ

れぞれ、
　各宛先ポートのアクセス回数の比率か、連続するパケットのソースポート番号間の差異の平均値か、用いられるプロトコルの比率か、ＴＣＰ（Transmission Control Protocol）におけるＴＣＰフラグの比率か、単位時間あたりのパケットの平均個数かの少なくともいずれかの統計情報を検出する
　請求項１又は２に記載のマルウェアの類似性検査方法。
【請求項４】
　前記第１脆弱性攻撃コード検出工程において前記第１脆弱性攻撃コード検出手段が、及び前記第２脆弱性攻撃コード検出工程において前記第２脆弱性攻撃コード検出手段が、それぞれ、
　脆弱性攻撃コードに含まれるシェルコード内の命令のシーケンスを検出する
　請求項１ないし３のいずれかに記載のマルウェアの類似性検査方法。
【請求項５】
　前記第１マルウェア分析工程において前記第１マルウェア分析手段が、及び前記第２マルウェア分析工程において前記第２マルウェア分析手段が、それぞれ、
　マルウェア本体を逆アセンブリし、そのアセンブリコードを取得する
　請求項１ないし４のいずれかに記載のマルウェアの類似性検査方法。
【請求項６】
　前記第１マルウェア分析工程において前記第１マルウェア分析手段が、及び前記第２マルウェア分析工程おいて前記第２マルウェア分析手段が、それぞれ、
　マルウェアの実行によるファイル又はレジストリのアクセスログか、プラットフォーム上のＡＰＩ（Application Program Interface）ログか、複数のサーバへのアクセスログか、マルウェアにより生成されるパケットログかの少なくともいずれかのログを取得する
　請求項１ないし５のいずれかに記載のマルウェアの類似性検査方法。
【請求項７】
　前記挙動比較工程においてコンピュータの挙動比較手段が、
　前記第１アドレス走査情報検出手段で検出された走査シグネチャと、前記第２アドレス走査情報検出手段で検出された走査シグネチャとの同一性を照合する
　請求項１ないし６のいずれかに記載のマルウェアの類似性検査方法。
【請求項８】
　前記挙動比較工程においてコンピュータの挙動比較手段が、
　前記第１アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率と、前記第２アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率との類似性を相関関係式により算出する
　請求項１ないし７のいずれかに記載のマルウェアの類似性検査方法。
【請求項９】
　前記挙動比較工程においてコンピュータの挙動比較手段が、
　第１脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードと、前記第２脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードとを、それぞれ要約処理し両者の同一性を照合する
　請求項１ないし８のいずれかに記載のマルウェアの類似性検査方法。
【請求項１０】
　前記挙動比較工程においてコンピュータの挙動比較手段が、
　第１マルウェア分析手段で検出したマルウェアのコードと、前記第２マルウェア分析手段で検出したマルウェアのコードとを、それぞれ要約処理し両者の同一性を照合する
　請求項１ないし９のいずれかに記載のマルウェアの類似性検査方法。
【請求項１１】
　ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査システ

ムであって、
該第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出手段と、
該第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析手段と、
該第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出手段と、
該第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析手段と、
上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較手段と
を少なくとも備えたことを特徴とするマルウェアの類似性検査システム。
【請求項１２】
ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査システムであって、
該第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出手段と、
該第１の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第１脆弱性攻撃コード検出手段と、
該第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析手段と、
該第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出手段と、
該第２の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第２脆弱性攻撃コード検出手段と、
第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析手段と、
上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較手段と
を少なくとも備えたことを特徴とするマルウェアの類似性検査システム。
【請求項１３】
前記第１アドレス走査情報検出手段及び前記第２アドレス走査情報検出手段が、それぞれ、
各宛先ポートのアクセス回数の比率か、連続するパケットのソースポート番号間の差異の平均値か、用いられるプロトコルの比率か、ＴＣＰ（Transmission Control Protocol）におけるＴＣＰフラグの比率か、単位時間あたりのパケットの平均個数かの少なくともいずれかの統計情報を検出する
請求項１１又は１２に記載のマルウェアの類似性検査システム。
【請求項１４】
前記第１脆弱性攻撃コード検出手段及び前記第２脆弱性攻撃コード検出手段が、それぞれ、
脆弱性攻撃コードに含まれるシェルコード内の命令のシーケンスを検出する
請求項１１ないし１３のいずれかに記載のマルウェアの類似性検査システム。
【請求項１５】
前記第１マルウェア分析手段及び前記第２マルウェア分析手段が、それぞれ、
マルウェア本体を逆アセンブリし、そのアセンブリコードを取得する

請求項１１ないし１４のいずれかに記載のマルウェアの類似性検査システム。
【請求項１６】
前記第１マルウェア分析手段及び前記第２マルウェア分析手段が、それぞれ、
マルウェアの実行によるファイル又はレジストリのアクセスログか、プラットフォーム上のＡＰＩ（Application Program Interface）ログか、複数のサーバへのアクセスログか、マルウェアにより生成されるパケットログかの少なくともいずれかのログを取得する
請求項１１ないし１５のいずれかに記載のマルウェアの類似性検査システム。
【請求項１７】
前記挙動比較手段が、
前記第１アドレス走査情報検出手段で検出された走査シグネチャと、前記第２アドレス走査情報検出手段で検出された走査シグネチャとの同一性を照合する
請求項１１ないし１６のいずれかに記載のマルウェアの類似性検査システム。
【請求項１８】
前記挙動比較手段が、
前記第１アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率と、前記第２アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率との類似性を相関関係式により算出する
請求項１１ないし１７のいずれかに記載のマルウェアの類似性検査システム。
【請求項１９】
前記挙動比較手段が、
第１脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードと、前記第２脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードとを、それぞれ要約処理し両者の同一性を照合する
請求項１１ないし１８のいずれかに記載のマルウェアの類似性検査システム。
【請求項２０】
前記挙動比較手段が、
第１マルウェア分析手段で検出したマルウェアのコードと、前記第２マルウェア分析手段で検出したマルウェアのコードとを、それぞれ要約処理し両者の同一性を照合する
請求項１１ないし１９のいずれかに記載のマルウェアの類似性検査システム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明はマルウェアの類似性を検査する方法と装置に関し、特に２種類以上の異なる程度の検査を組み合わせて効率的にマルウェアの特定を行う技術に関する。
【背景技術】
【０００２】
ボットのような高度に組織化され、精巧になったマルウェアの最近の異常発生により、それらを検出し、解析し、対応する技法の必要性が高まっている。多様な商業的なプロジェクト、学問的なプロジェクト、または政府支援によるプロジェクトが研究を進めている。（非特許文献１～１３参照）。
【０００３】
従前のステップとして、これらのプロジェクトの多くは、ネットワーク・イベント監視に基づいて、特定のポート番号でのアクセスの急激な増加等、統計的なデータの提供に集中している。これらの活動の過程で、世界的に公表されている未使用のＩＰアドレス（非特許文献１、２、１４参照）の集合であるダークアドレス空間を監視することが一般的な手法である。
【０００４】
これらのアドレス空間では、多様なマルウェアをひき付けるためにハニーポットを設置したり、あるいはマルウェアが感染先を探索するために行う走査、ＤＤｏＳ攻撃のバックスキャタ等を含む入信パケットを監視する（ブラックホール監視）センサを用意する（非

特許文献１５～１８参照）。

【０００５】

別の一般的な方法は、実際のネットワークに設置されたＩＤＳログとＦＷログを分析することである。ネットワーク・イベント監視に基づいたこれらの巨視的観察をマクロ分析と呼ぶ。マクロ分析は、世界的に分散されたセンサによって、インターネット上のマルウェア動作の過程で（走査等の）巨視的な挙動を把握するために使用する。しかし、それは巨視的なレベルの遠隔観察に基づいて、攻撃者の挙動及び攻撃者とセンサ間の環境に関する明示的な情報なしに実行されるため、多くの場合、結果にある程度の不確実性を残す問題がある。

【０００６】

他方、実際のマルウェア実行ファイルを分析することには別の課題がある。マルウェアの構造を解析する際、マルウェア実行ファイルを逆アセンブルするためにリバース・エンジニアリング技法が適用される（非特許文献１９、２０参照）。

また、マルウェア・コードが実際には閉じられた（アクセスが制御された）実験環境で実行されるサンドボックス分析はその挙動を観察することができる（非特許文献１９、２１～２３参照）。

【０００７】

マルウェア・コード自体を目標としたこれらの微視的分析をミクロ分析と呼ぶ。ミクロ分析は、閉じられた実験環境において実行されるため、実際のネットワークでのマルウェアの活動を観察することはできないが、マルウェアの詳細な構造及び挙動を明らかにする。

前述したマクロ分析とミクロ分析が研究され、多様な分析システムに配備されているが、これらの活動から得られた知識は効果的かつ効率的にリンクされておらず、セキュリティ・インシデントの根本原因の特定をさらに難しくしている。

【０００８】

これらの課題を解決するため、本件発明者は戦術的緊急対応のためのネットワーク・インシデント分析センタ（ｎｉｃｔｅｒ）を開発されている（非特許文献３、１９、２３、２４参照）。ｎｉｃｔｅｒは、マクロ分析による実ネットワーク上の攻撃の観察及びミクロ分析による「実験室内の」マルウェア分析を関連付けることで、観察された攻撃について、その考えられる根本原因、すなわちマルウェアとを結び付けることを可能とする、マクロ－ミクロ相関分析を実現している。

【０００９】

しかしながら、相関分析はおもにブラックホール監視によって観察されるマルウェアの走査挙動の関連付けによってだけ行われてきたため、マクロ分析とミクロ分析の間のリンクは詳細な攻撃の挙動の正確な識別を保証するほど強力ではなかった。

なお、このような２つ以上のマルウェアによる挙動の類似性を検査する技術として、非特許文献２４、２５が開示されており、本発明出願時未公開の特許文献１、２において開示されている。

【００１０】

【非特許文献１】Bailey, M., Cooke, E., Jahanian,F., Nazario, J., Watson, D.: The internet motion sensor: a distributedblackhole monitoring system, The 12th Annual Network and Distributed SystemSecurity Symposium (NDSS05), 2005年

【非特許文献２】Moore, D.: Network telescopes:tracking Denial-of-Service attacks and internet worms around the globe, 17thLarge Installation Systems Administration Conference (LISA '03), USENIX, 2003年

【非特許文献３】Nakao, K., Yoshioka, K., Inoue,D., Eto, M., Rikitake, K.: nicter: an incident analysis system usingcorrelation between network monitoring and malware analysis, The 1st JointWorkshop on Information Security (JWIS06), pp. 363 - 377, 2006年

【非特許文献４】Yegneswaran, V., Barford, P.,Plonka, D.: On the design and use o

f Internet sinks for network abusemonitoring. Recent Advances in Intrusion Detection (RAID 2004), LNCS 3224, pp146 - 165, 2004年
【非特許文献５】Honeypot project,http://www.leurrecom.org/
【非特許文献６】Internet Motion Sensor,http://ims.eecs.umich.edu/
【非特許文献７】IT Security Center, Information-TechnologyPromotion Agency, Japan, https://www.ipa.go.jp/security/index-e.html
【非特許文献８】Japan Computer Emergency ResponseTeam Coordination Center, http://jpcert.jp/isdas/index-en.html
【非特許文献９】National Cyber Security Center,Korea, http://www.ncsc.go.kr/eng/
【非特許文献１０】REN-ISAC: Research and EducationNetworking Information Sharing and Analysis Center, http://www.ren-isac.net/
【非特許文献１１】SANS Internet Storm Center,http://isc.sans.org/
【非特許文献１２】Telecom Information Sharing andAnalysis Center, Japan, https://www.telecom-isac.jp/
【非特許文献１３】@police,http://www.cyberpolice.go.jp/english/obs_e.html
【非特許文献１４】Bailey, M., Cooke, E., Jahanian,F., Myrick, A., Sinha, S.: Practical darknet measurement, 2006 Conference onInformation Sciences and Systems (CISS '06), pp. 1496 - 1501, 2006年
【非特許文献１５】Alata, E., Nicomette, V.,Kaaniche, M., Dacier, M.: Lessons learned from the deployment of ahigh-interaction honeypot, 6th European Dependable Computing Conference(EDCC-6), pp. 39 - 44, 2006年
【非特許文献１６】Leita, C., Dacier, M.,Massicotte, F.: Automatic handling of protocol dependencies and reaction to0-day attacks with ScriptGen based honeypots, 9th International Symposium onRecent Advances in Intrusion Detection (RAID2006), pp. 185 - 205, 2006年
【非特許文献１７】Provos, N.: Honeyd A virtualhoneypot daemon. 10th DFN-CERT Workshop, 2003年
【非特許文献１８】Provos, N.: A virtual honeypotframework. 13th USENIX Security Symposium, pp 1 - 14, 2004年
【非特許文献１９】Nakao, K., Matsumoto, F., Inoue,D., Baba, S., Suzuki, K., Eto, M., Yoshioka, K., Rikitake, K., Hori, Y.:Visualization technologies of nicter incident analysis system, IEICE TechnicalReport, vol.106, no.ISEC-176 pp.83 - 89, 2006年
【非特許文献２０】Isawa, R., Ichikawa, S.,Shiraishi, Y., Mohri, M, Morii, M.: A virus analysis supporting system; forautomatic grasping virus behavior by code-analysis result, The ComputerSecurity Symposium 2005 (CSS2005), vol. 1, pp. 169 - 174, 2006年
【非特許文献２１】Hoshizawa, Y., Morii, M., Nakao,K.: A proposal of automated malware behavior analysis system, Information andCommunication System Security, IEICE, ICSS2006-07, pp. 41 - 46, 2006年
【非特許文献２２】C. Willems, T. Holz, and F.Freiling, "Toward Automated Dynamic Malware Analysis Using CWSandbox"Security & Privacy Magazine, IEEE, Volume 5, Issue 2, pp. 32 - 39, 2007年
【非特許文献２３】Nakao, K., Matsumoto, F., Inoue,D., Baba, S., Suzuki, K., Eto, M., Yoshioka, K., Rikitake, K., Hori, Y.:Visualization technologies of nicter incident analysis system, IEICE TechnicalReport, vol.106, no.ISEC-176 pp.83 - 89, 2006年
【非特許文献２４】Yoshioka, K., Eto, M., Inoue, D.,Nakao, K.: Macro-micro correlation analysis for binding darknet traffic andmalwares, The 2007 Symposium on Cryptography and Information Security(SCIS2007), 2F2-2, 2007年

【非特許文献２５】衛藤 将史、薗田 光太郎、吉岡 克成、井上 大介、竹内 純一、中尾 康二、「スペクトラム解析を用いたマルウェアの類似性検査手法の提案」 IEICE、SCIS2007、２００７年
【特許文献１】日本特許出願２００７年１２０７０号
【特許文献２】日本特許出願２００７年１２０７１号
【発明の開示】
【発明が解決しようとする課題】
【００１１】
本発明は上記従来技術が有する問題点に鑑みて創出されたものであり、その目的はマルウェアのミクロ分析及びマクロ分析の相関に基づくマルウェアの検査精度を向上させると共に、効率よくマルウェアの検査を行う技術を提供することである。
【課題を解決するための手段】
【００１２】
本発明は、上記の課題を解決するために、次のようなマルウェア類似性検査方法を提供する。
請求項１に記載の発明によれば、ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査方法を提供する。
【００１３】
本方法はコンピュータ上に実装されるものであって次の各工程からなる。
（Ｓ１）コンピュータの第１アドレス走査情報検出手段が、第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出工程
（Ｓ２）コンピュータの第１マルウェア分析手段が、第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析工程、
（Ｓ３）コンピュータの第２アドレス走査情報検出手段が、第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出工程、
（Ｓ４）コンピュータの第２マルウェア分析手段が、第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析工程、
（Ｓ５）コンピュータの挙動比較手段が、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較工程。
【００１４】
請求項２に記載の発明によれば、上記請求項１の各工程に加え、脆弱性攻撃コードに係る類似性検査方法を提供する。すなわち、次の各工程からなる。
（Ｓ１）コンピュータの第１アドレス走査情報検出手段が、第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出工程、
（Ｓ０２）コンピュータの第１脆弱性攻撃コード検出手段が、第１の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第１脆弱性攻撃コード検出工程、
（Ｓ２）コンピュータの第１マルウェア分析手段が、第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析工程、
（Ｓ３）コンピュータの第２アドレス走査情報検出手段が、第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出工程、

（Ｓ０４）コンピュータの第２脆弱性攻撃コード検出手段が、第２の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第２脆弱性攻撃コード検出工程、
（Ｓ４）コンピュータの第２マルウェア分析手段が、第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析工程、
（Ｓ５）コンピュータの挙動比較手段が、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較工程。

【００１５】
請求項３に記載の発明によれば、上記の第１アドレス走査情報検出工程において第１アドレス走査情報検出手段が、さらに第２アドレス走査情報検出工程において第２アドレス走査情報検出手段が、それぞれ、各宛先ポートのアクセス回数の比率か、連続するパケットのソースポート番号間の差異の平均値か、用いられるプロトコルの比率か、ＴＣＰ（Transmission Control Protocol）におけるＴＣＰフラグの比率か、単位時間あたりのパケットの平均個数かの少なくともいずれかの統計情報を検出することを特徴とする。

【００１６】
請求項４に記載の発明によれば、上記の第１脆弱性攻撃コード検出工程において第１脆弱性攻撃コード検出手段が、さらに第２脆弱性攻撃コード検出工程において第２脆弱性攻撃コード検出手段が、それぞれ、脆弱性攻撃コードに含まれるシェルコード内の命令のシーケンスを検出することを特徴とする。

【００１７】
請求項５に記載の発明によれば、第１マルウェア分析工程において第１マルウェア分析手段が、さらに第２マルウェア分析工程において第２マルウェア分析手段が、それぞれ、マルウェア本体を逆アセンブリし、そのアセンブリコードを取得することを特徴とする。

【００１８】
請求項６に記載の発明によれば、第１マルウェア分析工程において第１マルウェア分析手段が、第２マルウェア分析工程おいて第２マルウェア分析手段が、それぞれ、マルウェアの実行によるファイル又はレジストリのアクセスログか、プラットフォーム上のＡＰＩ（Application Program Interface）ログか、複数のサーバへのアクセスログか、マルウェアにより生成されるパケットログかの少なくともいずれかのログを取得することを特徴とする。

【００１９】
請求項７に記載の発明によれば、上記の挙動比較工程においてコンピュータの挙動比較手段が、第１アドレス走査情報検出手段で検出された走査シグネチャと、第２アドレス走査情報検出手段で検出された走査シグネチャとの同一性を照合することを特徴とする。

【００２０】
請求項８に記載の発明によれば、上記の挙動比較工程においてコンピュータの挙動比較手段が、前記第１アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率と、前記第２アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率との類似性を相関関係式により算出することを特徴とする。

【００２１】
請求項９に記載の発明によれば、挙動比較工程においてコンピュータの挙動比較手段が、第１脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードと、第２脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードとを、それぞれ要約処理し両者の同一性を照合することを特徴とする。

【００２２】
請求項１０に記載の発明によれば、挙動比較工程においてコンピュータの挙動比較手段が、第１マルウェア分析手段で検出したマルウェア本体のコードと、第２マルウェア分析手段で検出したマルウェア本体のコードとを、それぞれ要約処理し両者の同一性を照合することを特徴とする。

【００２３】
　また、本発明は、次のようなマルウェアの類似性検査システムを提供することもできる。
　すなわち、請求項１１に記載の発明によれば、ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査システムを提供することができる。
【００２４】
　本システムには、第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出手段と、第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析手段と、第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出手段と、第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析手段と、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較手段とを少なくとも備えたことを特徴とする。
【００２５】
　請求項１２に記載の発明によれば、ネットワーク上で他のコンピュータに対して不正処理を行う第１のソフトウェアの処理結果から得られる第１の挙動情報と、検査対象の第２のソフトウェアの処理結果から得られる第２の挙動情報とを比較して両者の類似性を検査するマルウェアの類似性検査システムを提供することができる。
【００２６】
　該システムにおいて、第１の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第１アドレス走査情報検出手段と、第１の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第１脆弱性攻撃コード検出手段と、第１の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第１マルウェア分析手段と、第２の挙動情報のうちネットワークアドレスの走査に係る挙動の統計情報を検出して記憶手段に格納する第２アドレス走査情報検出手段と、第２の挙動情報のうち脆弱性攻撃コードを検出して記憶手段に格納する第２脆弱性攻撃コード検出手段と、第２の挙動情報のうちマルウェア本体のコード及び挙動の少なくともいずれかを分析しその結果を記憶手段に格納する第２マルウェア分析手段と、上記第１の挙動情報に関する各工程による情報と、上記第２の挙動情報に関する各工程による情報との相互の同一性又は類似性を所定の比較式に基づいて比較処理する挙動比較手段とを少なくとも備えたことを特徴とする。
【００２７】
　請求項１３に記載の発明によれば、上記の第１アドレス走査情報検出手段と第２アドレス走査情報検出手段が、それぞれ、各宛先ポートのアクセス回数の比率か、連続するパケットのソースポート番号間の差異の平均値か、用いられるプロトコルの比率か、ＴＣＰ（Transmission Control Protocol）におけるＴＣＰフラグの比率か、単位時間あたりのパケットの平均個数かの少なくともいずれかの統計情報を検出することを特徴とする。
【００２８】
　請求項１４に記載の発明によれば、上記の第１脆弱性攻撃コード検出手段と第２脆弱性攻撃コード検出手段が、それぞれ、脆弱性攻撃コードに含まれるシェルコード内の命令のシーケンスを検出することを特徴とする。
【００２９】
　請求項１５に記載の発明によれば、上記の第１マルウェア分析手段と第２マルウェア分析手段が、それぞれ、マルウェア本体を逆アセンブリし、そのアセンブリコードを取得することを特徴とする。

【００３０】
請求項１６に記載の発明によれば、上記の第１マルウェア分析手段と第２マルウェア分析手段が、それぞれ、マルウェアの実行によるファイル又はレジストリのアクセスログか、プラットフォーム上のＡＰＩ（Application Program Interface）ログか、複数のサーバへのアクセスログか、マルウェアにより生成されるパケットログかの少なくともいずれかのログを取得することを特徴とする。
【００３１】
請求項１７に記載の発明によれば、上記挙動比較手段が、第１アドレス走査情報検出手段で検出された走査シグネチャと、第２アドレス走査情報検出手段で検出された走査シグネチャとの同一性を照合することを特徴とする。
【００３２】
請求項１８に記載の発明によれば、上記挙動比較手段が、第１アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率と、第２アドレス走査情報検出手段で検出された宛先ポートのポート番号の比率との類似性を相関関係式により算出することを特徴とする。
【００３３】
請求項１９に記載の発明によれば、上記挙動比較手段が、第１脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードと、前記第２脆弱性攻撃コード検出手段で検出した脆弱性攻撃コードとを、それぞれ要約処理し両者の同一性を照合することを特徴とする。
【００３４】
請求項２０に記載の発明によれば、上記挙動比較手段が、第１マルウェア分析手段で検出したマルウェア本体のコードと、第２マルウェア分析手段で検出したマルウェア本体のコードとを、それぞれ要約処理し両者の同一性を照合することを特徴とする。
【発明の効果】
【００３５】
本発明は、以上の構成を備えることにより、次の効果を奏する。
すなわち、本発明の特徴であるマルウェアのミクロ分析とマクロ分析の結果から相関度を求める際に従来のように走査のプロファイルを比較するだけでなく、脆弱性攻撃コード（Exploit code：脆弱性実証コードとも言う。)やマルウェア自体のコード、挙動をも比較対象とすることにより高精度の相関分析が実現できる。例えば、走査の挙動が類似した２つの異なるマルウェアに対しても、脆弱性攻撃コードやマルウェア自体のコード、挙動の相関度を求めることで正確な識別をおこなうことができる。
本発明では、走査に係る挙動（走査層）とマルウェア本体のコード及び挙動（マルウェア層）とを組み合わせ、走査層における検出にはより簡易なホストを用いてネットワーク上の広域に配備することを可能にすると共に、マルウェア層の検出には高精度なホストを設置し、さらに正確な検出に寄与することができる。
【００３６】
特に、マルウェア層の検出用ホストは、実際にマルウェアに感染させて挙動を調べるために、十分に管理していないと他のホストを攻撃する恐れがある。そのため、極めて慎重な運用が必要であり、多くを配備することはできない。一方、走査層の検出用ホストは他のホストに攻撃または感染するような高度の能力を持たないため、感染する恐れがなく、大量に配備することができる。
【００３７】
そして、走査層の検出用ホストを広域に配置することで、ネットワーク上の全体的なマルウェアの走査の傾向や分布を把握した上で、さらに、それぞれの走査が具体的にどのマルウェアによってもたらされたものであるかを高い精度で推定することができる。これによりネットワーク上で実際に活動中のマルウェアの種類や比率を推定することができる。
【００３８】
また、脆弱性攻撃コードを比較する工程（脆弱性攻撃コード層）を設けることで、マルウェア層の検出ほど高度なホストを要さず、また感染の危険性を減らしながら、走査層の

みによる相関分析に比して精度の高い検出を行うことができる。
このように各層の検出用ホストをネットワーク上に段階的に配備することで、高精度なマルウェアの検出に寄与することができる。
【発明を実施するための最良の形態】
【００３９】
以下、本発明の実施形態を、図面に示す実施例を基に説明する。なお、実施形態は下記に限定されるものではない。
まず、本発明の概要を説明する。本発明は、正確かつ実践的な分析をすでに提案している技術に加え、マクロ－ミクロ相関分析の精度を大幅に高めるためにマルチレイヤの観察に基づいた新規の分析方法を提案するものである。
【００４０】
本発明の概念では、マルウェアの活動はマルウェアの根本的な伝搬ステップに従って複数の層で監視、分析する。すなわち、マルウェアは最初にネットワーク上で走査することによってそのターゲットを探す。攻撃されやすいホストを見つけると、マルウェアはそのターゲット・ホストの権限を握るために脆弱性攻撃コードを送信する。その後、マルウェアはターゲット・システム上で自らをコピーし、次の伝搬に備える。
【００４１】
このような特性に着目し、本発明ではこれらの３つのステップを対応する３つの層、すなわち走査層、脆弱性攻撃コード層、及びマルウェア層で観察、分析する。
その後、その３つの層におけるマクロ分析とミクロ分析両方からのすべての観察と分析の結果が相関アナライザによって効果的に収集され、相互に関連付けられ、その結果ネットワーク攻撃とその原因を高い精度で効果的に特定することができる。
これによって取得されるマルウェアの内部の挙動及びネットワーク挙動は、マルウェア駆除ツールやＩＤＳ（Intrusion Detection System）のためのシグネチャを生成する等の、マルウェアのさらなる活動を軽減し、防止するための処置を講じるために活用できる。
【００４２】
次にマルウェア活動のマルチレイヤ観察に基づいた新規の分析概念、マクロ－ミクロ相関分析について説述する。マクロ－ミクロ相関分析の基本的な考え方は、ネットワーク上で実際に起こっているマルウェア活動の観測をベースにした広域的な観察（マクロ分析）を、隔離された「実験室」で実行される詳細なマルウェア分析（ミクロ分析）の結果に結び付けることである。
相互関連の精度を高めるために従来のシステムを高度化するものとして、本発明のマルチレイヤ観察を提案する。
【００４３】
すなわち、マルウェアの基本的な伝搬ステップに従って、マルウェア活動を走査層、脆弱性攻撃コード層、及びマルウェア層という３つの層で観察、分析する。その３つの層におけるマクロ分析とミクロ分析両方からのすべての観察と分析の結果は相関アナライザによって効果的に相互に関連付けられ、その結果ネットワーク攻撃とその根本的な原因ははるかに高い確実性をもって効果的に結び付けられる。
【００４４】
ここで、図１には開発中のシステムの概要を説明する。本システムは、インフラストラクチャとエンドユーザの両方に多大な損害を生じさせることのあるワーム、ウィルス、及びボット等の伝搬するマルウェアを検出、分析することを目的として開発されている。
図１に示すように４つのサブシステム、つまりマクロ分析システム（ＭａｃＳ）（１０）、ミクロ分析システム（ＭｉｃＳ）（１１）、ネットワーク・マルウェア関連付けシステム（ＮｅｍｅＳｙｓ）（１２）、及びインシデント処理システム（ＩＨＳ）（１３）から成り立っている。
【００４５】
ＭａｃＳ（１０）はインターネット（２０）でのマルウェア活動の過程で走査等の巨視的な挙動を把握するためにネットワーク上で分散されたセンサ（２１）・・を配備する。

本発明ではこのようなネットワークから検出される検査対象のマルウェアを第２のソフトウェアと呼んでいる。

本実施例では観察のために複数の／１６ダークネットと／２４ダークネットがあり、その中で入信パケットだけを検知する幅広い範囲のブラックホール・センサ、ＴＣＰ　ＳＹＮパケットとＩＣＭＰエコー要求等の特定の入信パケットに対応する低対話型センサと、マルウェア・サンプル自体を含む多様な情報を引き出すために攻撃者と多種多様な対話が可能な高対話型センサ(ハニーポット)を配備している。

【００４６】

図２では、２００７年３月に発明者らによって／１６ブラックホール・センサの１つによって観察された入信パケット数（実線）と一意のＩＰアドレスの数（点線）を示している。平均では、１日あたりほぼ３５０万個のパケット、及び１５万以上の一意のＩＰアドレスが観察される。

これらのセキュリティ・イベントは追加分析のためにＭａｃＳ（１０）内の後述する多様なアナライザに送信される。ＭａｃＳ（１０）のアナライザは、重大なセキュリティ・インシデントの初期的または予兆的事象であるインシデント候補（ＩＣ）を検出する。

例えば、変化点アナライザ（非特許文献２６参照）は、特定のポート番号での走査頻度データ等の時系列データの急速な変化を検出し、ＩＣ警報（１４）を発行する。ＩＣ警報（１４）は、人間のオペレータが手動による詳細な分析を開始するようにＩＨＳ（１３）に送信してもよい。

【非特許文献２６】Takeuchi, J., Yamanishi, K.: Aunifying framework for detecting outliers and change points from non-stationarytime series data, IEEE Transactions on Knowledge and Data Engineering, Vol. 18,No. 4, pp.482 - 489, 2006年

【００４７】

このＩＣ警報（１４）はＮｅｍｅＳｙｓ（１１）にも送信され、本発明に係る相関分析をトリガする。

一方、すべての分析結果（１５）は相関分析のためにＮｅｍｅＳｙｓ（１１）に送信される。さらに、本システムは、人間のオペレータが観察されたトラフィックを直感的に理解するための複数のトラフィック・ビジュアライザを配備することもできる（トラフィック・ビジュアライザについては非特許文献２３参照）。

【００４８】

他方、ＭｉｃＳ（１２）は、微視的にマルウェアを分析する。ＭｉｃＳ（１２）では、後述する高対話型センサによって取り込まれる、あるいは入力として他のソース（２２）から受け取られるマルウェア実行ファイル（２３）を採取する。

予め分析しておくマルウェアを本発明では第１のソフトウェアと呼んでいる。従って、第１のソフトウェアは他のソース（２２）から取得してもよいし、インターネット（２０）から収集してもよい。

【００４９】

次に、ＭｉｃＳ（１２）は、マルウェアの詳説された構造及び動作を明らかにするために、マルウェアの挙動分析とマルウェアのコードの分析によって該実行ファイル（２３）を分析する。

ＭｉｃＳ（１２）のアナライザもＩＣ警報（１６）を発行することができる。例えば、アナライザは、未知の挙動のマルウェア検出時にそれを発行する。すべての分析結果（１７）は相関分析のためにＮｅｍｅＳｙｓ（１１）に送信される。

【００５０】

ＮｅｍＳｙｓ（１１）は、観察された攻撃をさらに正確なレベルで特定するために、ＭａｃＳ（１０）とＭｉｃＳ（１２）からの結果を相互に関連させる。ＮｅｍｅＳｙｓ（１１）は、ＭａｃＳ（１０）とＭｉｃＳ（１２）の両方から、分析結果（１５）（１７）を受け取る。それがＭａｃＳ（１０）またはＭｉｃＳ（１２）からＩＣ警報（１４）（１５）を受け取ると、ＭａｃＳ（１０）で観察された第２のソフトウェアによる攻撃を、ＭｉｃＳで分析されたマルウェア（第１のソフトウェア）と結び付けるために相関分析を開始

する。
【００５１】
ＩＨＳ（１３）は、多様な分析結果とオペレータ（３０）の間のインタフェースとなる。オペレータはＩＨＳ（１３）を介してセキュリティ・イベントと分析結果を取り出すことができる。想定されるどのインシデントが検出され、より深く、あるいは手作業で分析されるべきかをオペレータが理解できるようにＭａｃＳ（１０）とＭｉｃＳ（１２）の中のアナライザによって発行されるＩＣ警報（１４）（１６）も管理する。
これは、オペレータ（３０）に対してシステムの最終的な出力として発行されるインシデント・レポート（３１）を作成するためのシステムでもある。
【００５２】
セキュリティ・イベントの獲得、ＭａｃＳ（１０）とＭｉｃＳ（１２）それぞれにおける分析、及びＮｅｍｅＳｙｓ（１１）による相関分析を含むすべての手順が、想定されるセキュリティ・インシデントが迅速に検出、分析できるように、完全に自動化され、リアルタイムで処理される点は本発明の特徴である。
【００５３】
前述したように、マルウェアにはその活動を特徴付ける複数の典型的な伝搬ステップがある。最初に、マルウェアは攻撃されやすいホストを求めてネットワーク上で走査する。攻撃されやすいホストを検出した後、マルウェアはそのターゲット・ホストの制御を奪取するために脆弱性攻撃コードを送信する。最後に、マルウェアは自らをターゲット・ホスト上でコピーし、さらなる伝搬に進む。
【００５４】
本発明は、このようなマルウェアの特徴を捕らえて、実際のネットワークにおける活動をより高精度に検出できるように、マルウェア本体に加えて、マルウェアによる走査及び脆弱性攻撃コードの間のリンクを検出する。
そのために本発明ではマルウェア活動が３つの層、すなわち走査層、脆弱性実行コード層、マルウェア層で観察する新しい概念を提案する。
【００５５】
図３には本発明に係るマルウェアの類似性検査システムの構成図を示す。図１に示した従来提案済みのシステムの各構成を基礎としているが、新たに各層ごとの分析技術を加えるための構成である。そのため図面中では符号を改めている。
ＭａｃＳ（４０）では、走査及び脆弱性攻撃コードは、広範囲のネットワーク（例えばインターネット）（５０）で分散されているブラックホール・センサ（５１）と低対話型センサ（５２）によってそれぞれ観察される。
【００５６】
さらに、マルウェア・サンプルであるマルウェア実行ファイル（５４）は高対話型センサ（５３）によって取得される。その結果、それらは対応するアナライザ、つまり走査アナライザ（４１）、脆弱性攻撃コード・アナライザ（４２）、及びマルウェア・アナライザ（４３）によって分析される。
分析結果はＮｅｍｅＳｙｓ（６０）の相関アナライザ（６１）に通知される。
【００５７】
なお、本実施例では、マルウェア・サンプルを第２のソフトウェアとしてマルウェア・アナライザ（４３）に入力すると共に、第１のソフトウェアとして後述する設定可能サンドボックス等に送信している。（図中の通知線５５）
【００５８】
一方、ＭｉｃＳ（７０）では、設定可能サンドボックス（８０）と呼ばれる新規の構成を用いることを特徴としている。サンドボックスとは、マルウェアの実行ファイルについて、その挙動の分析のために観察できるように実行される実験環境である。
設定可能サンドボックス（８０）は、そのネットワーク環境が３つの焦点を当てられているネットワーク挙動、つまりブラックホール・ネットワークでの走査（ブラックホールモード）（８１）、脆弱性攻撃コードの送信（低対話モード）（８２）、それら自体のタ

ーゲット・ホストでのコピーを撤回するために適する高対話モード（８３）の３つの実行モードを用いる。

【００５９】

ＭａｃＳ（４０）と同様に、取得された走査及び脆弱性攻撃コード、マルウェア本体は対応するアナライザ（７１）（７２）（７３）によって分析され、分析結果はＮｅｍｅＳｙｓ（６０）の相関アナライザ（６１）に送信される。

【００６０】

ＮｅｍｅＳｙｓ（６０）では、相関アナライザは３つの層のＭａｃＳ（４０）とＭｉｃＳ（７０）の両方からすべての分析結果を受け取る。受け取った分析結果はマルウェア知識プール（ＭＮＯＰ）と呼ばれる図示しない記憶手段のデータベースに記憶される。ＭＮＯＰ内で分析結果をリンクするために、相関アナライザ（６１）が走査プロファイル（９０）（９１）、脆弱性攻撃コード・サンプル（９２）（９３）、及びマルウェア・サンプル（９４）（９５）のために後述する方法で相関関係を求める。

【００６１】

このような相関関係の導出によって、相関アナライザ（６１）はＭａｃＳ（４０）で観察された攻撃するマルウェアを、ＭｉｃＳ（７０）で分析されたマルウェアと結び付けることができる。

【００６２】

以下、図３及び図４を用いて各層の分析手法を詳述する。

（走査層）

走査層の役割は、広範囲のブラックホール・ネットワークでマルウェアのネットワーク走査を分析することである。マルウェアによるネットワーク走査は広範囲のブラックホール・センサによって見られるときに特徴的である。

それらがどの宛先ポート上を走査するのか、それらがソースポート番号をどのようにして選ぶのか、それらがどれほど速く走査できるのか、それらがどのようにして宛先ＩＰアドレスを選ぶのか等の走査挙動は、観察された走査の発生元を特定するために有用な情報である。したがって、実際のネットワークの監視によって観察されるものと比較するためにサンドボックス内のそれぞれの取り込まれたマルウェアの走査を観察し、要約する。

【００６３】

本発明に係る第１アドレス走査情報検出手段は、設定可能サンドボックスのブラックホールモード（８１）と走査アナライザ（７１）とからなる。第２アドレス走査情報検出手段は、ブラックホール・センサ（５１）と走査アナライザ（４１）とからなる。

走査アナライザ（４１）（７１）は未処理のパケット・データを採取し、ソースＩＰアドレスでそれをスライスし、ソースＩＰアドレスごとにパケットの基本的な統計を計算する。

【００６４】

ここでいう統計には、各宛先ポートのアクセス比率、連続するパケットのソースポート番号の間の差異の平均値、プロトコル（ＴＣＰ／ＵＤＰ／ＩＣＭＰ）とＴＣＰフラグ（ＴＣＰを使用している場合）の比率、時間単位あたりのパケット平均数等を含むことができる。これらのいずれか、又は任意の組み合わせとしてもよい。

【００６５】

さらに、公知の方法で、所定の分類規則のセットに基づいて走査挙動を複数のタイプに分類する。（非特許文献２７参照）。

次に、走査タイプと宛先ポート・セットの文字列連結の要約計算方法（ＭＤ５：MessageDigest 5）である走査シグネチャが、走査を特定するために計算される。前記の基本的な統計と走査シグネチャの集合を走査プロファイルと呼ぶ。（非特許文献２４に記載）

本発明のアドレス走査情報である走査プロファイルはＸＭＬフォーマットで出力され、記憶手段に格納される。

【非特許文献２７】Suzuki, K., Baba, S., Takakura,H.: Analyzing traffic directed to unused IP address blocks, IEICE TechnicalReport, vol.105, no.530, IA2005-23,

pp.25 - 30, 2006年１月
【００６６】
ＭａｃＳ（４０）の走査層では、多様なブラックホール・センサ（５１）を複数の／１６ネットワークと／２４ネットワークを含む広範囲のダークアドレス空間に展開する。入信トラフィックは走査アナライザ（４１）に入力され、各ソースＩＰアドレスの走査プロファイルがＮｅｍｅＳｙｓ（６０）に出力され、挙動比較手段である相関アナライザ（６１）で分析される。
【００６７】
ＭｉｃＳ（７０）の走査層では、マルウェア・サンプルはブラックホール・サンドボックス（８１）で実行され、取り込まれたトラフィックは走査アナライザ（７１）に入力される。ブラックホール・サンドボックス（８１）は、ＭａｃＳ（４０）のブラックホール・センサ（５１）と同じ挙動となる動作モードを実現している。
【００６８】
すなわち実行されるマルウェアによって生成されるあらゆる走査に応答しないように構成され、その結果、結果として生じる走査プロファイルはＭａｃＳ（４０）で観察される走査プロファイルと比較できる。走査アナライザ（７１）はＮｅｍｅＳｙｓの相関アナライザ（６１）に対し、マルウェア実行ファイルごとの走査プロファイルを出力する。
【００６９】
（脆弱性攻撃コード層）
脆弱性攻撃コード層の役割は、各マルウェアがどのような種類の脆弱性攻撃コードを使用するのかを探査することである。脆弱性攻撃コードは、ターゲット脆弱性が悪用された後に実行されるシェルコードを含んでいる。
そこで各分析システム（４０）（７０）の脆弱性攻撃コード・アナライザ（４２）（７２）は、ターゲット脆弱性を活用するために必須であるシェルコード内の命令のシーケンスを検出する。
【００７０】
これらの命令はシェルコードの識別子として抽出され、使用される。脆弱性攻撃コード層では、実際のネットワーク監視によって観察されるものと比較するためにサンドボックス内の各マルウェアによって生成される攻撃パケットのペイロードでこれらの命令を観察し、抽出する。
【００７１】
本発明に係る第１脆弱性攻撃コード検出手段は、設定可能サンドボックスの低対話モード（８２）と脆弱性攻撃コード・アナライザ（７２）とからなる。第２脆弱性攻撃コード検出手段は、低対話型センサ（５２）と脆弱性攻撃コード・アナライザ（４２）とからなる。
【００７２】
脆弱性攻撃コード・アナライザ（４２）（７２）における脆弱性攻撃コード検出技法は、ＩＤＳを使用することについて公知の技術がある。最近の研究では、自己書き換えをおこなうポリモーフィック型の脆弱性攻撃コードの検出及び分類を可能にしている（非特許文献２８参照）。
【非特許文献２８】Payer, U., Teufl, P., Lamberger,M.: Hybrid engine for polymorphic shellcode detection, Detection of Intrusionsand Malware, and Vulnerability Assessment (DIMVA2005), LNCS 3548, pp 19 - 31,2005年
【００７３】
さらに本実施例では高速処理を達成するために、ＡＳＨＵＬＡと呼ばれている最先端の脆弱性攻撃コード検出技術を活用することもできる。（非特許文献２９参照）。
ＡＳＨＵＬＡはパケットのペイロードから遠隔システムの権限を握るために必須である前述された命令のシーケンスを検出し、抽出する。この命令は脆弱性攻撃コードの実際の一部であるため、脆弱性攻撃コードを検出するためのパターン・マッチングのためのＩＤＳシグネチャとして使用することもできる。

【非特許文献２９】Nogawa. H.: Shellcode detection:fight against polymorphism, The Joint Information Security Workshop on InternetMonitoring and Analysis (ISWIMA), 2006年

【００７４】

ＭａｃＳ（４０）の脆弱性攻撃コード層では、複数の／２４ネットワークのダークアドレス空間に低対話型センサ（５２）を配備する。多くのマルウェアはそのターゲットの権限を握るために脆弱性攻撃コードを送信する前に、走査し、ＴＣＰ接続を確立する等の特定のステップに従う。

【００７５】

そこで本発明の低対話型センサ（５２）はマルウェアから脆弱性攻撃コードを引き出すために入信攻撃に正しく反応するように構成されている。低対話型センサ（５２）への入信トラフィックはソースＩＤアドレスでスライスされ、その後スライスされたトラフィックのそれぞれが脆弱性攻撃コード・アナライザ（４２）に入力される。アナライザ（４２）はソースＩＰアドレスごとに抽出された命令のリストを出力する。

【００７６】

一方、ＭｉｃＳ（７０）の脆弱性攻撃コード層では、マルウェア実行ファイルは低対話モードサンドボックス（８２）で実行され、取り込まれたトラフィックは脆弱性攻撃コード・アナライザ（７２）に入力される。

低対話モードサンドボックス（８２）は、実行されるマルウェアからの要求に適切に答えるように構成され、それはＭａｃＳ（４０）内の低対話型センサ（５２）と同じ挙動である。脆弱性攻撃コード・アナライザ（７２）はマルウェア実行ファイルごとに抽出された命令のリストを出力する。

【００７７】

（マルウェア層）

マルウェア伝搬の最終段階では、マルウェアのコピーをターゲット・ホスト上でダウンロードし、ホストがリブートされるときにコピーを確実に実行するようにしている。コピーは、ＦＴＰ及びＨＴＴＰのようなプロトコルを使用して特定のサーバから、あるいは以前に感染したホストからダウンロードされることもある。マルウェア層ではマルウェア・アナライザ（４３）（７３）によってサンプル自体を分析する。

【００７８】

本発明に係る第１マルウェア分析手段は、設定可能サンドボックスの高対話モード（８３）とマルウェア・アナライザ（７３）とからなる。第２マルウェア分析手段は、高対話型センサ（５３）とマルウェア・アナライザ（４３）とからなる。

【００７９】

本実施例では、マルウェア本体のコードを分析するマルウェア・コード・アナライザと、マルウェアによる挙動を分析するマルウェア挙動アナライザの２つのアナライザをマルウェア・アナライザ（４３）（７３）に備えている。

【００８０】

マルウェア・コード・アナライザは実行ファイルを逆アセンブルし、その内部の特徴と構造を明らかにするために使用される。もっとも多くのマルウェアはＡＳＰａｃｋ、ＦＳＧ、Ｐｅｔｉｔｅ、ＵＰＸ等の多様なパッキング技術を使用して難読化されている。

難読化のために、多くのマルウェアは従来の方法では逆アセンブルできない。難読化を克服するために、マルウェア・コード・アナライザはまず、隔離されたホストでマルウェア・サンプルを実行する。

【００８１】

大部分のマルウェアはアクティブとなるためにＤＬＬをロードまたはアンロードするときにメモリ上で自らを解凍（難読化の解除）するため、マルウェア・コード・アナライザは、サンプルが動的リンク・ライブラリ（ＤＬＬ）をロードまたはアンロードするたびにホストのメモリをダンプする。

【００８２】

次に、マルウェア・コード・アナライザはダンプされたコードのＰＥファイルヘッダのいくつかのフィールド（例えばＰｏｉｎｔｅｒＴｏＲａｗＤａｔａ、ＳｉｚｅＯｆＤａｔａ等）を適切に上書きする。

マルウェア・コード・アナライザは、最終的にはダンプされたコードを逆アセンブルし、アセンブリコードを取得できる。アセンブリコードを読み取ることによって、Ｗｉｎｄｏｗｓ（登録商標）　ＡＰＩの呼出シーケンスを取得する。

【００８３】

同時に、マルウェア・コード・アナライザは、作成・修正されたＵＲＬまたはＩＰアドレスであるファイルとレジストリを抽出してもよい。

最後に、すべての情報はコード分析レポートとして要約、ＸＭＬフォーマットで出力される。

【００８４】

他方、マルウェア・アナライザ（４３）（７３）のうち、マルウェア挙動アナライザは実際のインターネットに配備されている高対話型センサ（５３）におけるマルウェア実行から、あるいは実験室での設定可能サンドボックス（８３）から取得される多様なログを分析する。

【００８５】

ログは、ファイル／レジストリ・アクセス・ログ及びＡＰＩログを被害ホストの中に、サーバ・アクセス・ログを複数のタイプのサーバの中に、あるいはマルウェア・サンプルによって生成されるパケットログを含む。

これらのログに従って、マルウェア挙動アナライザは、所定の挙動定義に基づいたマルウェア・サンプルの顕著な挙動を抽出する。抽出された挙動は挙動分析レポートとして要約され、ＸＭＬフォーマットで出力される。

【００８６】

ＭａｃＳ（４０）のマルウェア層では、高対話型センサ（５３）がダウンロードされた実行ファイルを取り込むために配備されている。高対話型センサ（５３）には、それらが攻撃ホストから容易に発見され、感染し、最終的にダウンロードされるサンプルを取り込むように、脆弱性が回復されていないホストや、あるいは仮想のＷｉｎｄｏｗｓ（登録商法）を備えたシステムを用いることができる。

【００８７】

マルウェア・サンプルの収集のためには、公知のｎｅｐｅｎｔｈｅｓ（非特許文献２９）及びＡｒｇｏｓ（非特許文献３０）のようなツールも使用できる。

【非特許文献２９】P. Baecher, M. Koetter, T. Holz,M. Dornseif, F. C. Freiling, "The Nepenthes Platform: An EfficientApproach to Collect Malware," Proceedings of Recent Advances in IntrusionDetection, 9th International Symposium, RAID 2006, pp. 165-184, 2006年

【非特許文献３０】G. Portokalidis, A. Slowinska,and Herbert Bos, "Argos: an emulator for fingerprinting zero- day attacksfor advertised honeypots with automatic signature generation," Proceedingsof the 2006 EuroSys conference, pp. 15 - 27, 2006年

【００８８】

ＭａｃＳ（４０）は、マルウェア・サンプルを取り込むために脆弱なＷｉｎｄｏｗｓ（登録商標）システムの複数の高対話型センサ（５３）だけではなく、／２４ネットワークにもｎｅｐｅｎｔｈｅｓを配備することができる。ＭＤ５について新しいサンプルがセンサ（５３）によって取り込まれると、それは追加の分析の為にマルウェア・コード・アナライザとマルウェア挙動アナライザ（あわせて４３）に送信される。

【００８９】

一方、サンプルは、ＭｉｃＳ（７０）が３つの層でサンプルの分析を開始できるようにＭｉｃＳ（７０）の設定可能なサンドボックス（８０）に送信される。

最後に、各マルウェアのコード分析レポート及び挙動分析レポートからなるマルウェア

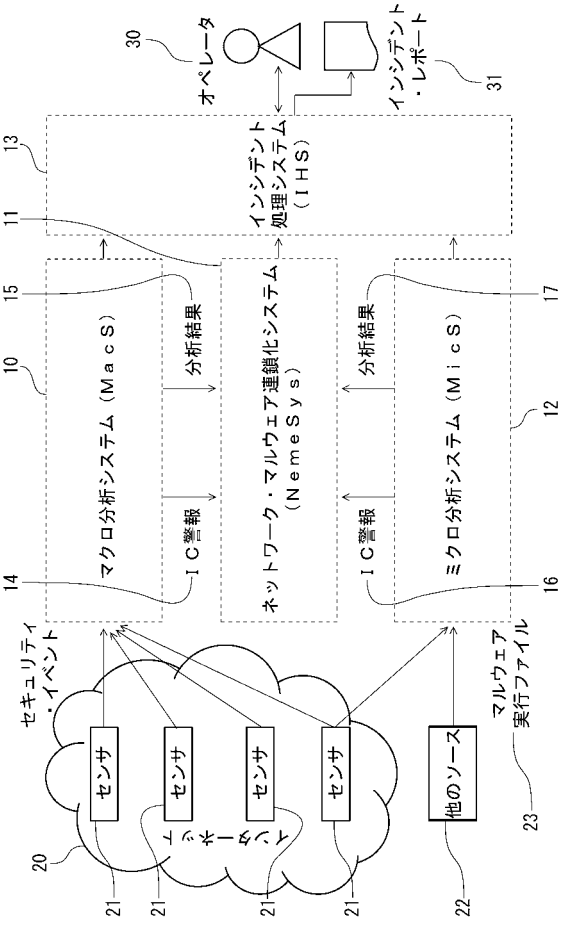
・レポート（９４）が相関アナライザ（６１）に送信される。
【００９０】
ＭｉｃＳのマルウェア層では、マルウェア・サンプルは高対話モードのサンドボックス（８３）で実行される。高対話型サンドボックス（８３）は、新しいサンプルをダウンロードできる別の被害ホストを提供する。ダウンロードされるサンプルがＭＤ５について新しい場合、それはマルウェア・コード・アナライザとマルウェア挙動アナライザ（あわせて７３）に送信される。最後に、各マルウェアのコード分析レポート及び挙動分析レポート（あわせてマルウェア・レポート９５）が相関アナライザ（６１）に送信される。
【００９１】
設定可能サンドボックス（８０）についてさらに詳述する。本発明で創出した設定可能サンドボックス（８０）は、マルウェア活動の多様な観察結果を提供する重要な技術である。
マルウェア・サンプルが入力されると、ファイル・アクセス、レジストリ・アクセス、及び通信を含むそれらの挙動を観察するために実験環境での入力サンプルを実行する。
【００９２】
最新のマルウェアは、それらが実験環境で実行されないことを確認するために様々な手法をとっている。これらの分析は多くの場合、便宜上仮想ホストで行われるためにホストＯＳとゲストＯＳ間のチャネルを検出することによって仮想ＯＳを検出する他のマルウェアがある一方、インターネットの到達可能性をチェックするために特定のウェブページにアクセスしようとするものもある。
【００９３】
したがって、本発明の設定可能サンドボックス（８０）はＡＰＩフッキング機構及びファイル／レジストリ監視機構を備える被害ホストのために実際のホストを使用する。
設定可能サンドボックス（８０）は、マルウェアを欺くためにインターネット・エミュレータと呼ばれる環境も使用する。インターネット・エミュレータは被害ホストに仮想インターネット・アクセスを提供する。
【００９４】
該インターネット・エミュレータは、入力マルウェアが実際のインターネット環境であると信じ、伝搬を開始するように、ＤＮＳ、ＦＴＰ、ＴＦＴＰ、ＨＴＴＰ、ＳＭＴＰ及びＩＲＣ等の多様なサービスを実装している。
マルウェアによって生成されるすべてのパケットはインターネット・エミュレータに送られ、次にインターネット・エミュレータはそのパケットをそのポート番号の代りにそのペイロードの観察に従って適切なサービスに転送する。
【００９５】
これは、マルウェアが特にＨＴＴＰサービス及びＩＲＣサービス低対話モードサンドボックス（８２）に対して不規則なポート番号を使用することがあるためである。設定可能サンドボックス（８０）は、３つのモード、つまりブラックホールモード（８１）、低対話モード（８２）、及び高対話モード（８３）で動作する。
【００９６】
ブラックホール・サンドボックスとも呼んでいるブラックホール・モード（８１）では、サンドボックス（８０）は、ＭａｃＳ（４０）でのブラックホール監視に類似した状況を生じさせるために最小のインターネット・サービスを提供する。
【００９７】
被害ホストから感染の結果生じるトラフィックはグローバル・アドレスでの走査、ローカル・アドレスでの走査、特定のサーバに対するアクセス等に分けられる。
被害ホストが実際のインターネット環境にあると、ブラックホール・センサによってグローバル・アドレスでの走査だけが監視できる。そのため、グローバル走査のスライスされたトラフィックが走査プロファイルを取得するために走査アナライザに送信される。
【００９８】
低対話型サンドボックスとも呼ぶ低対話モード（８２）では、サンドボックス（８０）

は、脆弱性攻撃を観察するために低相互作用監視に類似した状況を作出し、実行されている被害ホストの要求に適切に答えるように構成される。
被害ホストからの結果として生じるトラフィックは、どの脆弱性攻撃コードが入力マルウェアから送達されるのかをチェックするために脆弱性攻撃コード・アナライザ（７２）に入力される。
【００９９】
高対話型サンドボックスとも呼ぶ高対話モード（８３）では、サンドボックス（８０）は、感染しやすいホストが存在する環境をエミュレートするように構成されている。
例えば、上述したＡｒｇｏｓまたはｎｅｐｅｎｔｈｅｓのようなツールを、攻撃されやすいホストをエミュレートするために設定可能サンドボックス（８０）にインストールすることができる。
【０１００】
サンプルをダウンロードするために、設定可能サンドボックス（８０）は実際のインターネットに接続される必要がある。したがって、サンプルをダウンロードするための必要なトラフィックだけがサンドボックス（８０）から確実に出るようにするためにアクセス制御は重要である。
【０１０１】
ＭＤ５について新しいマルウェアが取り込まれるときには、それはＭｉｃＳ（７０）のマルウェア・アナライザ（７３）に出力されるだけではなく、設定可能サンドボックス（８０）にも再帰的に入力される。
【０１０２】
次に、本発明の挙動比較手段である相関アナライザ（６１）について説述する。
相関アナライザ（６１）は、ＭａｃＳ（４０）とＭｉｃＳ（７０）のすべてのアナライザから分析結果を受け取り、それらをＮｅｍｅＳｙｓ（６０）の図示しない外部記憶手段に格納されたマルウェア知識プール（ＭＮＯＰ）と呼ばれるデータベースに記憶する。相関アナライザ（６１）の役割は、それらの間でリンクを検出することによりそれらを統合し、充実させるために種々の分析結果をリンクすることである。
【０１０３】
図４に示すように、各アナライザからの分析結果の主要な属性を要約する。この図は、さまざまなアナライザからの分析結果が、それら自体を相互に関連付けるために使用できる多くの共通の属性を有していることを示している。以下では、これらの共通の属性における相関分析の手法を説明する。
【０１０４】
（走査層における相関分析）
走査を相互に関連付けるための最も簡単な方法は、プロファイルの中の走査シグネチャを比較することである。走査シグネチャは非特許文献２７に開示されるような規則によって示される走査タイプと宛先ポートのセットの連結のＭＤ５を用いるのが簡便である。
したがって、２つの走査プロファイルのシグネチャが正確に一致する場合、それは２つの走査が同じタイプであり、それらが同じポート番号に向けられていることを意味する。
【０１０５】
さらに厳密な方法は、２つの走査間の類似性を定義することである。走査プロファイルは多様な静的パラメータを含むため、それらを使用してその類似性を定義することができる。例えば、走査プロファイルは各宛先ポートに対するアクセスの比率を含む。２つのプロファイルのこれらの比率は、相関係数を使用することによって比較できる。単純な統計を使用して走査を相互に関連付けるため技術は非特許文献２４のように公知である。
【０１０６】
大きなデータセットから検索する必要があるときには、シグネチャ・マッチングは非常に高速かつ有用であるが、それは「類似する」走査を検出できない。実際に、３つの宛先ポート１３９、４４５及び１０２５での走査及び２つのポート１３９と１０２５上で走査のためのシグネチャは、たとえそれらがいくらか類似し、関連付けられていても、別のシ

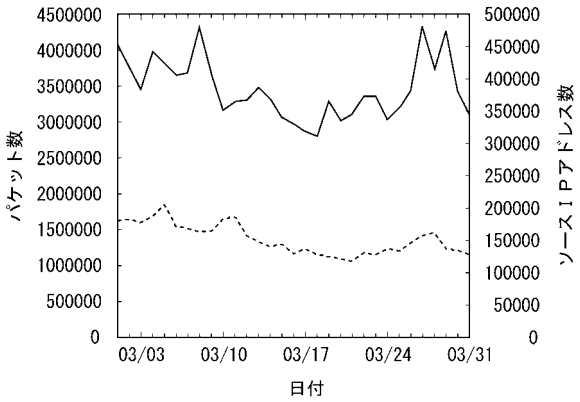
グネチャを有する。
【０１０７】
ブラックホール・センサ（５１）は多くの場合単一の攻撃者から走査パケット全体の一部だけを観察するため、前者の方法では正確に同一性を照合できない場合がある。他方、類似性の定義による相関分析は、それらに対する検索の複雑度が定義に応じて高い場合にも類似した走査を検出できる可能性が高い。
【０１０８】
（脆弱性攻撃コード層における相関分析）
脆弱性攻撃コードの相関分析は、抽出される命令を比較することによって行うことができる。走査の相互関連のケースに関しては、要約マッチングが最も容易な方法である。命令の２つのシーケンス間の類似性を測定するための方法は任意であるが、例えば上記のようにＭＤ５を用いても同一性を照合してもよい。
【０１０９】
（マルウェア層の相関分析）
２つのマルウェア・サンプルを比較する最も簡単な方法は、その要約（例えばＭＤ５）によるものである。別の簡単な方法は、既知のマルウェアにとって効果的な方法であるが、周知のアンチウィルス・ソフトウェアを用いる方法である。特に簡便で有効な方法は、それらの分析レポートを比較することである。
【０１１０】
マルウェア・コード・アナライザ及びマルウェア挙動アナライザは、ＡＰＩ、その出現順、改変されたレジストリ・キーとファイル、それらのミューテックスの名前、それらがアクセスを試みるサーバ名等を含むマルウェアについての情報を検出可能である。
上述した走査層における相互関連と同様に、マルウェア・アナライザ（４３）（７３）の分析結果を使用してそれらの間の類似性を算出することができる。
【図面の簡単な説明】
【０１１１】
【図１】従来提案されているマクロ分析・ミクロ分析の相関分析を行うシステムの構成図である。
【図２】ブラックホールセンサによる測定結果例のグラフである。
【図３】本発明に係るマルウェアの類似性検査システムの構成図である。
【図４】本発明に係るマルウェアの類似性検査システムのリンクを示す図である。
【符号の説明】
【０１１２】
４０　マクロ分析システム
４１　走査アナライザ
４２　脆弱性攻撃コード・アナライザ
４３　マルウェア・アナライザ
５０　インターネット
５１　ブラックホール・センサ
５２　低対話型センサ
５３　高対話型センサ
６０　ネットワーク及びマルウェア関連付けシステム
６１　相関アナライザ
７０　ミクロ分析システム
７１　走査アナライザ
７２　脆弱性攻撃コード・アナライザ
７３　マルウェア・アナライザ
８０　設定可能サンドボックス
８１　同、ブラックホールモード
８２　同、低対話モード

８３　同、高対話モード
９０　走査プロファイル
９１　走査プロファイル
９２　脆弱性攻撃コード
９３　脆弱性攻撃コード
９４　マルウェア・レポート
９５　マルウェア・レポート

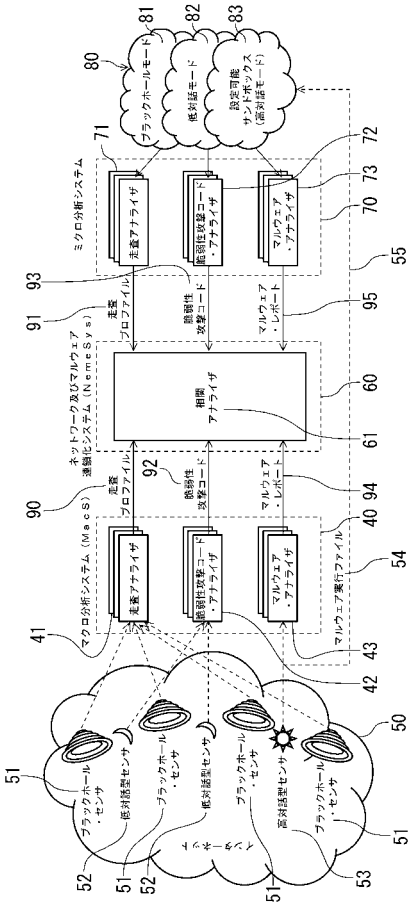
【図１】



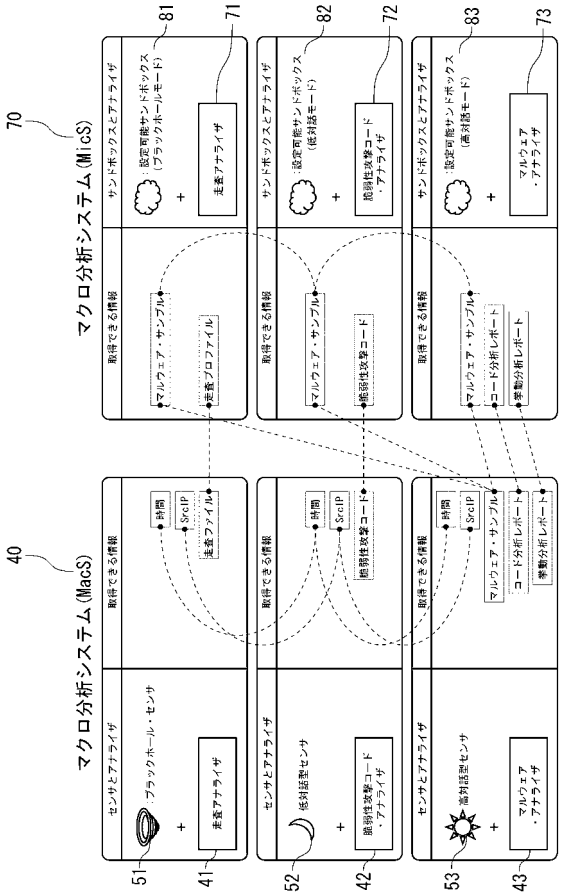
【図２】



【図３】


ブラックホールモード
低対話モード
設定可能サンドボックス（高対話モード）
ミクロ分析システム
走査アナライザ
脆弱性攻撃コード・アナライザ
マルウェア・アナライザ
走査プロファイル
脆弱性攻撃コード
マルウェア・レポート
相関アナライザ
ネットワーク及びマルウェア連動化システム（Nemesys）
マクロ分析システム（MacS）
マルウェア実行ファイル
ブラックホール・センサ
低対話型センサ
高対話型センサ
ハニーネット

【図４】


マクロ分析システム（MicS）
マクロ分析システム（MacS）
サンドボックスとアナライザ
センサとアナライザ
取得できる情報
設定可能サンドボックス（ブラックホールモード）
設定可能サンドボックス（低対話モード）
設定可能サンドボックス（高対話モード）
走査アナライザ
脆弱性攻撃コード・アナライザ
マルウェア・アナライザ
ブラックホール・センサ
低対話型センサ
高対話型センサ
時間
SrcIP
走査プロファイル
脆弱性攻撃コード
マルウェア・サンプル
コード分析レポート
挙動分析レポート

フロントページの続き

(72)発明者　衛藤　将史
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内

Ｆターム(参考) 5B276 FD05　FD08　FD09