(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5961183号
(P5961183)
(45)発行日　平成28年8月2日(2016.8.2)　(24)登録日　平成28年7月1日(2016.7.1)

(51)Int.Cl.　F I
G06F 21/56 (2013.01)　G06F 21/56 320

請求項の数 10 (全 37 頁)

(21)出願番号	特願2013-542187(P2013-542187)
(86)(22)出願日	平成23年12月1日(2011.12.1)
(65)公表番号	特表2014-504399(P2014-504399A)
(43)公表日	平成26年2月20日(2014.2.20)
(86)国際出願番号	PCT/US2011/062957
(87)国際公開番号	WO2012/075336
(87)国際公開日	平成24年6月7日(2012.6.7)
審査請求日	平成26年10月27日(2014.10.27)
(31)優先権主張番号	61/418,580
(32)優先日	平成22年12月1日(2010.12.1)
(33)優先権主張国	米国(US)

(73)特許権者　508041127
シスコ　テクノロジー，インコーポレイテッド
アメリカ合衆国，カリフォルニア州　９５１３４－１７０６，サンノゼ，ウェスト・タスマン・ドライヴ　１７０
(74)代理人　100101454
弁理士　山田　卓二
(74)代理人　100081422
弁理士　田中　光雄
(74)代理人　100100479
弁理士　竹内　三喜夫

最終頁に続く

(54)【発明の名称】文脈上の確からしさ、ジェネリックシグネチャ、および機械学習法を用いて悪意のあるソフトウェアを検出する方法

(57)【特許請求の範囲】
【請求項１】
ソフトウェア・アプリケーションが悪意のあるものか否かについて判断するためのコンピュータにより実行される方法であって、
前記ソフトウェア・アプリケーションから特徴ベクトルを抽出するステップと、
特徴ベクトルをクライエント側構成部品からサーバー側構成部品に送信するステップと、
前記特徴ベクトルに少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報を、前記サーバー側構成部品から受信するステップと、
前記ソフトウェア・アプリケーションに関するメタデータを抽出し、前記ソフトウェア・アプリケーションがインストールされ得るシステムに付随する文脈情報を収集するステップと、
メタデータおよび文脈情報をサーバー側構成部品に送信するステップと、
文脈情報は、クライエント側構成部品がアクセスしたウェブサイトおよびクライエント側構成部品の地理的位置を含み、
メタデータおよび文脈情報に少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報をサーバー側構成部品から受信するステップと、
前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値を計算す

るステップと、
ジェネリック・フィンガープリント値をサーバー側構成部品に送信するステップと、
ジェネリック・フィンガープリント値に少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報をサーバー側構成部品から受信するステップと、
サーバー側構成部品から受信した、メタデータ、文脈情報、およびジェネリック・フィンガープリント値から得られた前記情報に基づいて、前記ソフトウェア・アプリケーションに関して処理を実行するステップとを有することを特徴とする方法。
【請求項２】
ソフトウェア・アプリケーションが悪意のあるものか否かについて判断するためのコンピュータにより実行される方法であって、
サーバー側構成部品において、
i）前記ソフトウェア・アプリケーションからの特徴ベクトル、
ii）前記ソフトウェア・アプリケーションに関するメタデータおよび前記ソフトウェア・アプリケーションがインストールされ得るシステムに付随する文脈情報であって、クライエント側構成部品がアクセスしたウェブサイトおよびクライエント側構成部品の地理的位置を含む文脈情報、および
iii）前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値、
を含む情報をクライエント側構成部品から受信するステップと、
特徴ベクトルをクライエント側構成部品から受信すると、機械学習により導出された分類アルゴリズムを特徴ベクトルに適用するステップと、
前記ソフトウェア・アプリケーションに関するメタデータおよびクライエント側構成部品に関する文脈情報をクライエント側構成部品から受信すると、メタデータおよび文脈情報を調査するステップと、
前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値をクライエント側構成部品から受信すると、ジェネリック・フィンガープリント値が悪意のあるものとみなすべきか否か判断するステップと、
クライエント側構成部品に関して、前記ソフトウェア・アプリケーションが悪意のあるものとみなすか否かについて判断するステップと、
特徴ベクトル、メタデータ、文脈情報、およびジェネリック・フィンガープリント値に基づいて得られた情報であって、前記ソフトウェア・アプリケーションが悪意のあるものとみなすか否かについての情報をクライエント側構成部品へ送信するステップとを有することを特徴とする方法。
【請求項３】
メタデータは、従来式のシグネチャおよびジェネリックシグネチャからなるグループから選択されたものであることを特徴とする請求項２に記載の方法。
【請求項４】
クライエント側構成部品およびサーバー側構成部品は、別個の遠隔コンピュータデバイス上に構成されていることを特徴とする請求項２に記載の方法。
【請求項５】
クライエント側構成部品は、文脈情報を連続的に収集することを特徴とする請求項２に記載の方法。
【請求項６】
ソフトウェア・アプリケーションが悪意のあるものか否かについて判断するための指令を含む非一時的コンピュータ可読媒体であって、
前記コンピュータ可読指令は、
前記ソフトウェア・アプリケーションから特徴ベクトルを抽出する指令と、
特徴ベクトルをクライエント側構成部品からサーバー側構成部品に送信する指令と、
前記特徴ベクトルに少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報を、前記サーバ

ー側構成部品から受信する指令と、
前記ソフトウェア・アプリケーションに関するメタデータを抽出し、前記ソフトウェア・アプリケーションがインストールされ得るシステムに付随する文脈情報を収集する指令と、
メタデータおよび文脈情報をサーバー側構成部品に送信する指令と、
文脈情報は、クライエント側構成部品がアクセスしたウェブサイトおよびクライエント側構成部品の地理的位置を含み、
メタデータおよび文脈情報に少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報をサーバー側構成部品から受信する指令と、
前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値を計算する指令と、
ジェネリック・フィンガープリント値をサーバー側構成部品に送信する指令と、
ジェネリック・フィンガープリント値に少なくとも部分的に基づいて、前記ソフトウェア・アプリケーションが安全なものか、または悪意のあるものかについての判断に関する情報をサーバー側構成部品から受信する指令と、
サーバー側構成部品から受信した、メタデータ、文脈情報、およびジェネリック・フィンガープリント値から得られた前記情報に基づいて、前記ソフトウェア・アプリケーションに関して処理を実行する指令とを有することを特徴とする非一時的コンピュータ可読媒体。

【請求項７】
ソフトウェア・アプリケーションが悪意のあるものか否かについて判断するための指令を含む非一時的コンピュータ可読媒体であって、
前記コンピュータ可読指令は、
サーバー側構成部品において、
i）前記ソフトウェア・アプリケーションからの特徴ベクトル、
ii）前記ソフトウェア・アプリケーションに関するメタデータおよび前記ソフトウェア・アプリケーションがインストールされ得るシステムに付随する文脈情報であって、クライエント側構成部品がアクセスしたウェブサイトおよびクライエント側構成部品の地理的位置を含む文脈情報、および
iii）前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値、
を含む情報をクライエント側構成部品から受信する指令と、
特徴ベクトルをクライエント側構成部品から受信すると、機械学習により導出された分類アルゴリズムを特徴ベクトルに適用する指令と、
前記ソフトウェア・アプリケーションに関するメタデータおよびクライエント側構成部品に関する文脈情報をクライエント側構成部品から受信すると、メタデータおよび文脈情報を調査する指令と、
前記ソフトウェア・アプリケーションのジェネリック・フィンガープリント値をクライエント側構成部品から受信すると、ジェネリック・フィンガープリント値が悪意のあるものとみなすべきか否か判断する指令と、
クライエント側構成部品に関して、前記ソフトウェア・アプリケーションが悪意のあるものとみなすか否かについて判断する指令と、
特徴ベクトル、メタデータ、文脈情報、およびジェネリック・フィンガープリント値に基づいて得られた情報であって、前記ソフトウェア・アプリケーションが悪意のあるものとみなすか否かについての情報をクライエント側構成部品へ送信する指令とを有することを特徴とする非一時的コンピュータ可読媒体。

【請求項８】
メタデータは、従来式のシグネチャおよびジェネリックシグネチャからなるグループから選択されたものであることを特徴とする請求項７に記載の非一時的コンピュータ可読媒体。

【請求項９】
クライエント側構成部品およびサーバー側構成部品は、別個の遠隔コンピュータデバイス上に構成されていることを特徴とする請求項７に記載の非一時的コンピュータ可読媒体。
【請求項１０】
クライエント側構成部品は、文脈情報を連続的に収集することを特徴とする請求項７に記載の非一時的コンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、汎用演算デバイス（汎用コンピュータデバイス）のセキュリティに関し、とりわけ汎用演算デバイス上にある悪意のあるソフトウェア（マルウェア）を検出することに関する。
【背景技術】
【０００２】
当業者には知られているように、毎日数万個もの新たな悪意のあるソフトウェアが確認されている。これらのプログラムは、汎用演算デバイスのセキュリティを危険にさらすものである。可能性のあるセキュリティ侵害は、これに限定されるものではないが、システムからの情報窃取、他の不正目的によるシステムへの侵入（スパムメールの送信等）、および一般的に、他の悪意のある行為目的による（オーナ以外の第三者による）システムの遠隔操作を含むものである。
【０００３】
悪意のあるソフトウェアを検出するための１つの定評のある技術は、次のステップを含むものである。
ａ）ソフトウェア・アプリケーションが悪意のあるものであることを、（たとえば、人が手作業でアプリケーションを解析して、１つまたはそれ以上の悪意のある挙動の存在を特定することにより）複数の独立した手段を用いて確立するステップ。
ｂ）こうしたソフトウェアのハッシュまたはフィンガプリント値（明確な特徴値）を算出するステップ。ハッシュ値とは、２つの異なるアプリケーションは、圧倒的に高い確率で識別可能なフィンガプリント値を含むとの考え方に基づいて、ソフトウェア・アプリケーションの基本的なバイナリコンテンツを取り込み、比較的に短い文字列（ストリング）を生成する数学的変換値である。このフィンガプリント値をまたはハッシュ値を計算するための共通の関数は、限定するものではないが、ＳＨＡ－２５６、ＳＨＡ－１、およびＭＤ５等が含まれる。フィンガプリント値をまたはハッシュ値の他、この変換値を説明するために当業者が用いる別の用語はシグネチャである。本発明の説明に際し、ハッシュ値、フィンガプリント値、およびシグネチャの用語は、互いに置換可能なものとして用いる。これらの用語は、互いに同義語ではないが、本発明を説明する上では、これらの相違は重要ではない。
ｃ）汎用演算デバイスを操作するエンドユーザがハッシュ値にアクセスできるように、ハッシュ値を公開するステップ（たとえば既知の悪意のあるアプリケーションに関するブラックリストにハッシュ値として公表することができる。）。
ｄ）公表されたフィンガプリント値と、システムに搭載（インストール）された新しいソフトウェア・アプリケーションのフィンガプリント値とを、汎用演算デバイスを用いて比較するステップ。
ｅ）前者のフィンガプリント値と後者のフィンガプリント値が一致するか否か、所定のポリシに基づいて、一連のステップを実行する（たとえば新しいソフトウェア・アプリケーションのイントールを禁止する。）。
【０００４】
上記説明した手法には、ソフトウェア・アプリケーションが悪意のあるものであることが事前に特定された場合に限り機能するという問題がある。すなわち、この手法は反応性

アプローチである。アプリケーションの基本的な作用が依然として悪意のあるものであっても、そのアプリケーションに表面的な変化を加えたものは、異なるフィンガプリント値を有することになる。換言すると、アプリケーションがブラックリストに含まれているものとは表面的には異なるように見えても、その基本的な作用は同一のものである（犯人がかつらやサングラスを用いて異なる変装をしても、基本的には同一人物であるという場合と同様である。）。ファイルが修正されると、対応するフィンガプリント値が変化し得る。フィンガプリント値が変化すると、事前に確立したアプリケーションのブラックリストのものとは合致せず、このアプリケーションは、反応性シグネチャ式プローチを用いる任意のアンチマルウェア技術による検出から逃れることができる。

【０００５】

マルウェア事例が最近急増しているのは、マルウェア作成者が、まったく新しいマルウェアアプリケーションを作成するのではなく、より数少ないアプリケーションに些細な変化を頻繁に加えることに起因するものと思われる。

【０００６】

この問題に対応するための当該技術における１つの手法は、ジェネリックシグネチャとして知られるものを開発するステップを含む。ソフトウェア・アプリケーションの基本的なバイナリコンテンツに表面的な変更を加えたとしても、このジェネリックシグネチャは不変量として設計されるものである。すなわち悪意のある第三者がバイナリコンテンツに表面的な変更を限定的に加えた場合に限り、得られるハッシュ値は変化しない。たとえばジェネリックシグネチャを生成する１つの手法は、以下の通りである。第１に、ファイルの構造的特性（たとえば異なるセクションのデータサイズ、シンボルの個数、さまざまなセクションのエントロピ等）を抽出する。第２に、これらの変数を標準化し、または複数のバケツに放り込む（複数のカテゴリに分類する）。たとえばデータサイズが０～１００であるとき、バケツ（カテゴリ）１に属するものとする。またデータサイズが１００～２００であるとき、バケツ（カテゴリ）２に属するものとする。以下同様である。このとき我々は、シグネチャを生成するために、オリジナルファイルを用いるのではなく、シグネチャの基礎として標準化された構造的特徴を用いることができる。この考え方によれば、ファイルに表面的な変更を加えても、ファイルの基本的構造に対しては、ほとんど、またはまったく変化を与えず、標準化し、カテゴリに分類した後において、まったく変化が見られない。

【０００７】

その結果、単一のジェネリックシグネチャを用いて、所与の基本的なマルウェアの脅威だけでなく、多少の変更を加えたマルウェアの脅威をも検出することができる。シグネチャの概念をより明確にしやすくする物理的類似性を検討するために、あなたが犯人について説明しようとしている場合を想像されたい。きわめて独特な特徴（たとえば毛髪の色、目の色、最後に見たときの服装）を特定することにより、犯人について説明することができる。しかし犯人はかつらを着けているか、またはカラーコンタクトを着用している場合、毛髪や目の色等の特徴は役に立たない。代わりに、構造的な属性（たとえば犯人の伸長、体重、体型、人種等）に焦点を当てた場合、犯人が変装をしたとしても、これらの属性は同じである。さらに、これらの属性を標準化した場合（たとえば犯人の身長は、およそ６フィートであると表現するのではなく、正確に６フィート２インチであると表現し、または犯人の体型はきわめて独特であると表現するのではなく、犯人は体重が最も重い人と表現した場合）、犯人が厚底の靴やぶかぶかの服装を着用していた場合であっても、犯人を特定できる可能性が高い。

【０００８】

しかし当業者には知られているように、ジェネリックシグネチャには欠点がある。限定するものではないが、その欠点とは以下の通りである。

ａ）ジェネリックシグネチャを生成するためには、手作業による介入が必要となる（たとえばコンピュータウィルスの分析者が、ソフトウェア・アプリケーションのバイナリコンテンツを直接調べて、そのアプリケーションに対してわずかに変更された場合には、シグ

ネチャの値が変わらないように、シグネチャの算出方法を特定する必要がある。）。上述の犯罪分析者の例においては、興味のある属性、およびその属性が取り得る値の範囲を正確に特定する必要があるかも知れない。
ｂ）ジェネリックシグネチャは、誤検出しやすくなる傾向がある（すなわち、実際には安全なアプリケーションであっても、悪意のあるものと誤って判断しやすくなる。）。ジェネリックシグネチャは、単一の基本的なソフトウェア・アプリケーションを特定するだけでなく、それに関連する他のアプリケーションを特定するように設計されるので、適法なアプリケーションの基本的なバイナリコンテンツが、シグネチャの基礎となる悪意のあるアプリケーションのものと類似性を有するため、適法なアプリケーションが誤って悪意のあるものと特定されるリスクがある。上述の犯罪分析者の例においては、犯人の特徴があまりにも不明瞭である場合、身長が６フィートでずんぐりした人はすべて、犯人の特徴を有することになる。

【０００９】

したがって、上記課題に対処するように悪意のあるアプリケーションを検出する方法、構成部品、およびシステムを開発する必要性が存在する。本発明は、こうした必要性に対し、以下の方法を提供するものである。
ａ）手作業により解析する度合い、およびシステムにおける誤検出を低減するために、自動化によりジェネリックシグネチャを利用するための改善された方法、
ｂ）システムにおける他の最近の（悪意のある）動作等の文脈情報を利用するための方法であって、システム上で作動する特定のソフトウェア・アプリケーションが悪意のあるものか否かに関して、より正確な実態像を形成する方法、
ｃ）懸案のアプリケーションを評価するための機械学習モデルを開発すべく、コーパスを教育して、機械学習技術を利用する方法、および
ｄ）上記方法ａ）～ｃ）のうちの２つまたはそれ以上の方法を含む方法。

【発明の概要】

【００１０】

本発明の１つの態様によれば、クライエントシステムからの文脈情報とともに、所与のソフトウェア・アプリケーションが悪意のあるものか否かについて判断するためのより攻撃的な検出エンジンを用いたシステムを提供する。このシステムは、以下のフェーズを有する。第１に、クライエントは、ソフトウェア・アプリケーションに遭遇し、ソフトウェア・アプリケーションが悪意のあるものか、安全なものかについての素性（特性）について判断することを所望する。クライエントは、限定するものではないが、アンチマルウェア技術、機械学習特徴属性等の当該技術において用いられた（ＳＨＡ-２５６のような）従来式のフィンガプリント、ジェネリックシグネチャ等、そのアプリケーションに関するメタデータを抽出する。クライエントは、たとえば最近の感染履歴、システム上で実行中のアプリケーション、アクセスしたウェブサイト等の追加的な文脈情報を収集する。この文脈情報は、適当な場合には、当該技術分野において知られた任意の技術を用いて符号化される。次に、文脈情報とともにアプリケーションに関する情報は、（必要ならばネットワークを介して）サーバ側構成部品に送信される。（この構成部品は、遠隔サーバーコンピュータである必要はなく、ロジック回路自体はクライエントコンピュータ上にあるものであってもよい。ただし、本発明を明確とするために、クライエントコンピュータから送信される情報を処理する独立した構成部品である方がイメージしやすい）。サーバコンピュータは、文脈情報およびアプリケーション情報の両方を調査して、たとえばアプリケーションは実行に際して安全である旨のアプリケーションに関する判断を行う。サーバコンピュータは、クライエントコンピュータがなすべきことについての提案を符号化した応答をクライエントコンピュータに返信する。最終的にクライエントコンピュータは、サーバコンピュータの機能として、ローカルポリシーに基づいて、取るべき処理を判断する。

【００１１】

本発明の別の態様によれば、クライエント側構成部品が提供される。連続的に文脈情報を収集し、任意的には、文脈情報をサーバーに送信し、サーバーの支援を受けて、所与の

ソフトウェア・アプリケーションが脅威を含むものか否かについて判断する。この判断に際しては、脅威を特定する従来式の技術とともに文脈情報を利用する。文脈情報は、限定するものではないが、システム上に最近インストールされたアプリケーション、システム上で発見された最近の脅威についての情報および脅威が発見された時、クライエントコンピュータが最近アクセスした任意のウェブサイト、クライエントコンピュータがある地理的位置またはクライエントコンピュータのインターネットプロトコル・アドレス（ＩＰアドレス）、およびクライエントコンピュータの識別子等を含むものであってもよい。クライエントコンピュータの識別子を用いて、サーバーコンピュータから見て同一のクライエントコンピュータによる異なるトランザクション（処理）をリンクさせるために、クライエントコンピュータを特定することができる。

【００１２】

本発明の別の態様によれば、構成部品は、クライエントまたはサーバーのいずれかに搭載することができるものであり、クライエントにより送信された文脈情報を用いたロジック回路を有し、このロジック回路は所与のソフトウェア・アプリケーションが悪意のあるものか否かについて判断を行うことができる。同様に、サーバーは、他のクライエントにより記述されたアプリケーションの文脈とともに、懸案のアプリケーションについて他のクライエントにより問い合わせされた頻度およびタイミング等の複数のクライエントから収集可能な追加的な文脈情報を用いることができる。一旦、判断がなされると、対応する提案が判断され、クライエントに送信される。

【００１３】

本発明に係る別の態様によれば、（クライエントシステム上で実行される）基本的な方法は、クライエントから文脈情報を収集して、所与の懸案のアプリケーションが脅威であるか否かの判断を支援することができる。基本的情報の具体例は、クライエントに対する最近のセキュリティイベント（たとえば他の悪意のあるソフトウェアすなわちマルウェアの検出等）、または特定の「リスクのある」ソフトウェア・アプリケーション（ピアツーピア・ファイル共有アプリケーション等）がシステム上の有無を含む。

【００１４】

本発明に係る別の態様によれば、ユーザシステム上の当該ソフトウェア・アプリケーションに付随する文脈情報とともに、所与の懸案のソフトウェア・アプリケーションに関するデータを調査し、当該ソフトウェア・アプリケーションについての判断（たとえば当該ソフトウェア・アプリケーションが悪意のあるものであって、侵入を防ぎ、排除すべきであるか否かについての判断）を行う方法が提供される。この方法は、一連の単純なルールを用いてもよい。たとえば、最近過去１時間以内に１０件の脅威が発見され、別の脅威検出システムに基づいて（たとえば機械学習技術を用いて、またはジェネリックシグネチャを用いて）、現在のアプリケーションが６５％の確率で悪意のあるものである場合、当該アプリケーションは悪意のあるものと判断する（他の情報がない場合には、真陽性判断の確率がたった６５％であるということは、決定的な判断を行う上で不十分なものであるが、最近１０件の脅威という文脈情報により、アプリケーションが悪意のあるものであるという可能性は遥かにより大きくなる。）。この方法は、機械学習技術を採用して、一連のルールを生成し、追加的なルールを効率的に符号化するより一般的な（よりジェネリックな）モデルを生成することができる。

【００１５】

本発明の態様によれば、所与のソフトウェア・アプリケーションに関するジェネリックフィンガプリント値を計算するとともに、同一のジェネリックフィンガプリント値を有するアプリケーションが悪意のあるものとみなすべきか否かを判断することができる。この場合、一連の予定された処理が当該ソフトウェアに対してなされることになる。

【００１６】

本発明の別の態様によれば、以下のステップを実行するサーバー側構成部品を提供される。第１に、ソフトウェア・アプリケーションを数学的に変形して、ジェネリックフィンガプリントを生成する。第２に、当該ソフトウェア・アプリケーションのジェネリックフ

ィンガプリントを記録する。第３に、汎用演算デバイス上で実行可能な１つまたはそれ以上のステップを行い、ジェネリックフィンガプリントが悪意のあるものとみなすべきか否か判断する。第４に、クライエント側構成部品にその情報を送信する。

【００１７】

本発明の別の態様によれば、クライエント側構成部品が提供され、これは、第１に、ソフトウェア・アプリケーションに関するジェネリックフィンガプリント値を計算し、第２に、ジェネリックフィンガプリント値をサーバー側構成部品に送信し（またはサーバー側構成部品に関連するデータおよび局在的な処理についての知識を有する場合、これらのステップを局在的に反復し）、第３に、サーバー側構成部品により提供される予定された一連の処理を行い、その処理とは、（１）ジェネリックフィンガプリント以外の方法によりソフトウェア・アプリケーションが安全であるとみなされた場合にソフトウェア・アプリケーションを無視する処理、（２）ソフトウェア・アプリケーションが安全でないとみなされた場合にソフトウェア・アプリケーションをシステムから排除する処理、および（３）ソフトウェア・アプリケーションのさらなる処理および解析のために、異なるサーバー側構成部品にソフトウェア・アプリケーションを送信するｄ）処理を含む。

【００１８】

本発明の別の態様によれば、所与のソフトウェア・アプリケーションが計算されたジェネリックシグネチャを有する対象であるか否かを特定する方法が提供される。本発明に係る１つの実施形態において、この方法は、サーバー側構成部品において実行され、（電子回路、専用ロジック回路等の）ハードウェアと、（汎用コンピュータシステムまたは専用マシン等上で実行されるような）ソフトウェア、またはこれらの組み合わせを備えたロジックを処理することにより実行される。ただし、この方法をどこで、どのように実行するかの選択は、本明細書に開示するものに限定するものではなく、当業者にとっては、こうした選択肢は数多く存在することは明らかである。

【００１９】

本発明の別の態様によれば、主として、算出されたジェネリックシグネチャ値に基づいて、所与のジェネリックシグネチャを有するアプリケーションが悪意のあるもの（または安全なものである）とみなすべきであるか否かを判断する方法が提供される。１つの実施形態において、この方法は、サーバー側構成部品において実行され、（電子回路、専用ロジック回路等の）ハードウェアと、（汎用コンピュータシステムまたは専用マシン等上で実行されるような）ソフトウェア、またはこれらの組み合わせを備えたロジックを処理することにより実行される。ただし、この方法をどこで、どのように実行するかの選択は、本明細書に開示するものに限定するものではなく、当業者にとっては、こうした選択肢は数多く存在することは明らかである。

【００２０】

本発明の別の態様によれば、機械学習技術を用いて、ソフトウェア・アプリケーションが悪意のあるものとして特定するシステムが提供される。このシステムは、以下のフェーズを有する。第１に、学習データのコーパスを用いてモデルを構築する学習フェーズがある。このモデルは、ソフトウェア・アプリケーションを数学的に変換することにより求めることができる特徴ベクトルを入力値として取り込む。第２に、クライエントシステムが、悪意のある可能性を有するソフトウェア・アプリケーションから特徴ベクトルを抽出し、このモデルを用いてソフトウェア・アプリケーションを直接的に評価するか、またはバックエンド・サーバーに評価させるためにサーバーに送信することができる特徴抽出フェーズがある。第３に、抽出された特徴ベクトルがモデルに適応されて、懸案のアプリケーションが悪意のある可能性が高いか、または安全なものである可能性が高いかについて診断する評価フェーズがある。任意的には、２値的な分類（仕分け）だけでなく、この特徴（悪意のあるものか、安全なものか）の可能性を示すスコア（たとえば０～１００までのスコアであって、スコア０は確実に安全なものであることを示し、スコア１００は確実に悪意のあるものであることを示す。）を生成する。第４に、この判断に基づいて、適当なポリシーを適用することができる。本発明の別の態様によれば、１つまたはそれ以上のサ

ーバー側構成部品は、学習フェーズを実行することができる。１つの実施形態において、モデルを求めるために用いられるデータは、サーバー側構成部品と通信する実際のクライエントシステムの処理履歴から直接的に収集することができる。学習させることができる方法は、限定するものではないが、サポートベクターマシン、ニューラルネットワーク、決定木、単純ベイズアルゴリズム、ロジスティック回帰、および教師あり学習、半教師あり学習、および教師なし学習によるその他の技術を含むものである。本発明に係る学習態様または「モデル導出」態様は、ソフトウェア・アプリケーションを分類するための方法を提供できるものである限り、任意の上記技術を用いて実施することができる。学習が完了し、１つのモデルが導出されると、サーバー側構成部品は、そのモデルを用いて新たなソフトウェア案件の特徴ベクトルを評価するモジュールを自働的に構成することができる。

【００２１】

本発明の別の態様によれば、以下のステップを実行するクライエント側構成部品が提供される。すなわちクライエント側構成部品は、第１に、ソフトウェア・アプリケーションから関連する特徴ベクトル値を抽出するステップと、第２に、任意的には、これらの値をローカルモデルと比較して、ソフトウェア・アプリケーションが悪意のあるものか、安全なものであるかを判断するステップと、第３に、任意的には、小さいバイト数で符号化できるように、特徴ベクトルを圧縮するステップと、第４に、（圧縮した、または圧縮しない）特徴ベクトルをサーバー側構成部品に送信するステップと、第５に、サーバー側構成部品の応答に基づいたポリシーを適用するステップとを実行する。サーバー側構成部品の応答に基づくポリシーは、限定するものではないが、１つまたはそれ以上のオプションを含む。第１に、アプリケーションが確実に悪意のあるものと判断された場合、クライエント側構成部品は、そのアプリケーションをシステムから排除し、ユーザがインストールしようとする試みを妨害してもよい。第２に、アプリケーションが確実に悪意のあるものと判断されたのではないが、その可能性が高い場合、クライエント側構成部品は、そのアプリケーション自体のコピーをサーバー側構成部品に送信し、さらに拡張して処理を行い、詳細に解析を行ってもよい。本発明の別の態様によれば、以下のステップを実行するサーバー側構成部品が提供される。すなわちサーバー側構成部品は、第１に、（クライエント側構成部品から送信された）特徴ベクトルを受信するステップと、第２に、任意的には、クライエント側構成部品により圧縮された特徴ベクトルを解凍するステップと、第３に、特徴ベクトルを評価し、アプリケーションが悪意のあるものである可能性を判断するステップと、第４に、クライエント側構成部品がどのように応答すべきかを示す任意的な指令とともに、悪意のある可能性に関する情報をクライエント側構成部品に送信するステップとを実行する。本発明の１つの実施形態によれば、異なるサーバー側構成部品からの応答をどのように処理すべきかの実際のポリシーを、クライエント側構成部品自身に記憶させ、サーバー側構成部品単純な応答を送信してもよい。本発明の別の態様によれば、ソフトウェア・アプリケーションが悪意のある虞があるか否かを判断するモデルに学習させるための方法が提供される。この方法は、実際の領域内利用データを潜在的に活用することができる。本発明の別の態様によれば、クライエント側構成部品が、システム上の関連する文脈情報とともに、ソフトウェア・アプリケーションからの特徴ベクトルを抽出し、任意的には、この情報を圧縮し、そして圧縮した情報をサーバー側構成部品に送信するための方法が提供される。本発明の別の態様によれば、サーバー側構成部品が提供される。潜在的に圧縮された特徴ベクトルを受信し、圧縮された場合には解凍し、モデルに対して特徴ベクトルを評価し、この評価結果を悪意のあるアプリケーションを特定するための他の方法で得られた結果と比較し、そしてクライエント側構成部品にその素性を提供するものであってもよい。

【００２２】

本発明の別の態様によれば、２つまたはそれ以上のジェネリックシグネチャ、文脈情報、または機械学習導出モデルが、クライエントアプリケーションおよびサーバーアプリケーションのいずれか一方または両方において適用され、ソフトウェア・アプリケーション

が悪意のあるものか否かを判断する。この実施形態によれば、クライエントアプリケーションは、（ｉ）ソフトウェア・アプリケーションから特徴ベクトルを抽出するステップと、（ii）ソフトウェア・アプリケーションに関するメタテータを抽出し、ソフトウェア・アプリケーションがインストールされる可能性のあるシステムに関する文脈情報を収集するステップと、（iii）ソフトウェア・アプリケーションのジェネリック・フィンガープリント値を計算するステップのうちの２つまたはそれ以上のステップを実行し、得られたデータに関する情報をサーバーコンピュータに送信してもよい。サーバーアプリケーションは、上記情報を処理した後、真偽判断または関連情報をクライエントアプリケーションに返信し、クライエントアプリケーションは、サーバー側構成部品から受信した情報に基づいて、ソフトウェア・アプリケーションに対する処理を行ってもよい。

【００２３】

関連する実施形態によれば、サーバーアプリケーションは、（ｉ）ソフトウェア・アプリケーションから得た特徴ベクトル、（ii）ソフトウェア・アプリケーションに関するメタデータ、およびソフトウェア・アプリケーションがインストールされる可能性のあるシステムに関する文脈情報、ならびに（iii）ソフトウェア・アプリケーションのジェネリック・フィンガープリント値のうちの２つまたはそれ以上を、クライエントアプリケーションから受信するものであってもよい。受信した情報にも依存するが、サーバーアプリケーションは、特徴ベクトルに関する情報をクライエントアプリケーションから受信すると、機械学習により導出された分類アルゴリズムを特徴ベクトルに適用し、メタデータおよび文脈情報をクライエントアプリケーションから受信すると、ソフトウェア・アプリケーションに関するメタデータ、およびクライエントシステムに関する文脈情報を精査し、そして／またはソフトウェア・アプリケーションのジェネリックシグネチャを受信すると、そのジェネリックシグネチャが悪意のものあるものとみなすべきか否かを判断する。上記ステップが完了すると、サーバーアプリケーションは、ソフトウェア・アプリケーションが悪意のものあるものとみなすべきか否かについて判断を行い、ソフトウェア・アプリケーションが悪意のものあるものとみなすべきか否かについての判断に関する情報を、クライエントアプリケーションに送信する。

【図面の簡単な説明】

【００２４】

添付図面を参照して、本発明に係る好適な実施形態について以下説明する。

【図１】本発明に係るジェネリックシグネチャを用いた実施形態におけるクライエントコンピュータの処理手順を示すフローチャートである。

【図２】本発明に係る態様に基づいて、ファジー・フィンガープリントが決定的に悪意のあるものであるか否かを判断する方法を示すフローチャートである。

【図３】本発明に係る態様に基づいて、ファジー・フィンガープリントが悪意のある可能性があるものであるか否かを判断する方法を示すフローチャートである。なお、この方法に係るステップは、ファジー・フィンガープリントが決定的に悪意のあるものであるか否かを判断する方法のステップとほとんど同一である。実施化における相違点は、選択される数値パラメータＭおよびＣの値にある（アプリケーションが悪意のある可能性があるものであるか否かを判断するのではなく、アプリケーションが決定的に悪意のあるものであるか否かを判断するためには、少なくともＭの値を大きくし、少なくともＣの値を小さくすることを必要とする。）。これらのパラメータに対する適当な値は当業者により特定されることが期待される。

【図４】本発明の実施形態に係るジェネリック・フィンガープリント生成モジュールを含むクライエント側構成部品を示す構成図である。

【図５】本発明の実施形態に係るジェネリック・フィンガープリントに対して（悪意ありとする）確からしさを判断するための履歴データを解析するモジュールを含むサーバー側構成部品を示す構成図である。

【図６】本発明の機械学習の実施形態に係る機械学習方法を示すフローチャートである。

【図７】本発明の機械学習の実施形態に係るクライエントコンピュータによる特徴抽出方

法を示すフローチャートである。
【図８】本発明の機械学習の実施形態に係るサーバーコンピュータによる評価方法を示すフローチャートである。
【図９】本発明の機械学習の実施形態に係る特徴ベクトル抽出モジュールを含むクライエント側構成部品の構成図である。
【図１０】本発明の機械学習の実施形態に係る特徴ベクトル評価モジュールおよび特徴ベクトル抽出モジュールを含むサーバー側構成部品の構成図である。
【図１１】本発明の実施形態に係る、懸案のアプリケーションが悪意のあるものであるか否かを判断するための文脈情報属性を収集する方法のステップを示すフローチャートである。
【図１２】本発明の実施形態に係る、文脈情報属性を用いて悪意のあるアプリケーションを特定する方法のステップを示すフローチャートである。
【図１３】本発明の実施形態に係る文脈情報収集モジュールを含むクライエント側構成部品の構成図である。
【図１４】本発明の実施形態に係る文脈情報確信モジュールを含むサーバー側構成部品の構成図である。
【図１５】本発明の実施形態に係る例示的なコンピュータシステムの構成図である。
【発明を実施するための形態】
【００２５】
以下の発明の詳細な説明において、本発明をより完全に説明するために、数多くの詳細な特徴について説明する。しかし、当業者には明らかなように、特定の詳細な特徴を備えていなくても、本発明を実施することができる。換言すると、本発明を明確に説明するために、広く知られた構造体およびデバイスについては、詳細には説明せず、ブロック図の形態で示す。
【００２６】
以下の発明の詳細な説明のいくつかの部分は、コンピュータメモリ内のデータビットに対する演算に関するアルゴリズムおよび象徴的表現を用いて説明する。他の当業者に対して本発明の実質的内容を最も効率的に教示するために、こうした説明および表現は、データ処理技術における当業者が用いる意義で行われる。複数のステップは、物理量の物理的操作を必要とするものである。必ずという訳ではないが、通常、これらの物理量は、記憶、転送、組み合わせ、比較、その他の操作の対象となり得る電気的または磁気的な信号の形態を有する。時には、主に共通の利用を理由に、これらの信号を、ビット、値（バリュー）、構成要素、シンボル、キャラクタ、用語、数字等として参照する。
【００２７】
ただし、これらの用語および類似する用語のすべては、適当な物理量に付随するものであり、これらの物理量に付与された単なる便宜上のラベルに過ぎないという点に留意されたい。以下の詳細な説明において特に言及がない限り、明細書全体において、「処理ステップ（processing）」、「演算ステップ（computing）」、「計算ステップ（calculating）」、「判断ステップ（determining）」、「表示ステップ（displaying）」等の用語を用いて、コンピュータシステムまたは同様の電子演算デバイスを説明するが、これは、コンピュータシステムのレジスタおよびメモリ内で（電子的な）物理量として表されるデータを操作し、コンピュータシステムのメモリまたはレジスタ、あるいは他の情報記憶デバイス、通信デバイス、または表示デバイス内の物理量として同様に表される他のデータに変換するものである。
【００２８】
本発明は、同様に、処理を実行する装置に関する。この装置は、必要な目的のために特に構成されるものであってもよいし、コンピュータに記憶されたコンピュータプログラムにより選択的に起動され、再構成される汎用コンピュータを含むものであってもよい。こうしたコンピュータプログラムは、コンピュータシステムバスに接続されるコンピュータ判読可能記憶媒体に記憶され、コンピュータ判読可能記憶媒体として、限定するものでは

ないが、フロッピーディスク、光ディスク、ＣＤ－ＲＯＭ、光磁気ディスク、リード・オンリー・メモリ（ＲＯＭ）、ランダム・アクセス・メモリ（ＲＡＭ）、ＥＰＲＯＭ、ＥＥＰＲＯＭ、磁気カード、光カード等の任意のタイプのディスク、または電気的指令を記憶するのに適した任意のタイプの媒体が含まれる。

【００２９】

以下の説明は、本質的に、任意の特定のコンピュータまたは他の装置に関するものではない。本明細書の開示内容に基づいたプログラムを有する、さまざまな汎用システムを用いることができ、必要な方法ステップを実行するために、より特化した装置を構成することは有用であることを示すことができる。これらのさまざまなシステムに必要な構成は、以下の説明から明らかとなる。さらに本発明は、任意の特定のプログラム言語を用いて表現されるものではない。さまざまなプログラム言語を用いて、ここに記載される教示内容に係る本発明を実行することができる。マシン判読可能媒体は、マシン（たとえばコンピュータ）で判読可能な形態を有する情報を記憶または通信するための任意の機構を有する。マシン判読可能媒体は、たとえばリード・オンリー・メモリ（ＲＯＭ）、ランダム・アクセス・メモリ（ＲＡＭ）、磁気ディスク記録媒体、光記録媒体、フラッシュメモリデバイス、あるいは電気的、光学的、音響学的、または他の形態の伝播信号（たとえば搬送波、赤外線信号、デジタル信号等）を記憶する媒体であってもよい。

【００３０】

以下の説明において、当業者に一般に知られた用語について言及する。当該技術分野において、マルウェアの用語は、悪意のあるソフトウェア・アプリケーションを意味する。こうしたアプリケーションは、数多くの不正な目的を有することがある。たとえば、マルウェアを用いて、数多くの不正行為を実行することができる。これらの不正行為は、限定するものではないが、被害者のマシンからデジタル信号を窃取する行為、被害者のマシンを犯行に用いて、他の不正行為を行う行為（たとえば未承諾の電子メールメッセージまたはスパムを送信する行為）、被害者のマシンを遠隔制御する行為、および被害者のマシンの正常な動作を阻害する行為を含む。当該技術分野において、コンピュータウィルスは、一般に、悪意のあるソフトウェアの１つの具体例であると考えられている。コンピュータウィルスに加え、当該技術分野における他のタイプのマルウェアは、トロイの木馬、ワーム、ダウンローダ、ミスリーディングアプリケーションを含む。

【００３１】

アプリケーションの悪意は主観的なものと理解され、しばしばユーザは依存し、通常、明確に定義された一連のルールを有する。本明細書の目的のため、悪意のあるアプリケーションは、ユーザに不快感を与えるアプリケーションを意味するものと理解すべきである。

【００３２】

当該技術分野において、誤検出（偽陽性検出）の用語は、本来は適法なアプリケーションが誤って悪意のあるものと認定される状況を意味するものである。同様に、正検出（真陽性検出）の用語は、悪意のあるアプリケーションが正しく悪意のあるものとして認定される状況を意味するものである。偽陽性率とは、安全なアプリケーションがアンチマルウェア技術により間違って悪意のあるものと認定される確率（尤度）を意味する。真陽性率とは、悪意のあるアプリケーションがアンチマルウェア技術により正しく悪意のあるものとして認定される確率（尤度）を意味する。したがって、アンチマルウェア・ソフトウェアは、偽陽性率を低く維持して、高い真陽性率を実現することを目的とする。したがって、一般に、２つの確率の間には、逆のトレードオフの関係（二律背反）が成立する。あるアンチマルウェア技術がきわめて攻撃的で、数多くの脅威を検出するものである場合、偽陽性率が高くなる可能性がより大きくなる。したがって、アンチマルウェア技術が控えめで、より少数の脅威を検出するものである場合、偽陽性率が低くなることがある。当該技術分野において、真陽性率はときどき（正）検出率と呼ばれることがある。しかし、偽陽性率および偽陽性率は、データサンプルを用いて、一般に、概算により求められたものであることに留意されたい。アンチマルウェアの供給者は、偽陽性検出率と真陽性検出率と

の間の好適なトレードオフを提供しようとする。適法で重要なビジネスアプリケーションを誤って悪意のあるものと判断すると、顧客に対して甚大な財務上の損害を与えかねない。したがって、偽陽性検出はきわめて不適当である。いくつかの実施例では、偽陽性検出があまりにも適切でないため、供給者は、偽陽性検出率をきわめて小さく維持するために、低い真陽性検出率に甘んじることになる。

【００３３】

当該技術分野において、比較的に短い値の順列（シーケンス）であるシグネチャを用いて、アプリケーションが悪意のあるものか否かを特定することができる。最も一般的な変形において、ソフトウェア・アプリケーション全体に対して付与される変換値として、シグネチャを計算することである。当該技術において、シグネチャは、通常、マルウェアの既知の一部分について計算される。シグネチャは、クライエントのシステム内に転送されるか、またはサーバー内に記憶される。クライエントシステムがソフトウェアの新規な部分を確認すると、そのソフトウェア上でシグネチャを計算し、クライエントシステムのローカルデータ記憶媒体をチェックし、またはサーバーシステムに問い合わせ（照会）することにより、そのシグネチャがマルウェアソフトウェアの既知の一部分と一致するか否か判断する。当該技術分野において、シグネチャは、固有のものであってもよいし、一般的なものであってもよいことを理解されたい。２つのソフトウェア・アプリケーションが同一の固有のシグネチャを有する場合、相当の確率（尤度）で、これら２つのアプリケーションが完全に同一のものである。当該技術分野における固有のシグネチャの一つの具体例は、ＳＨＡ－２５６のアプリで算出するハッシュ値である。

【００３４】

一般的なシグネチャは、固有のシグネチャとはことなるものであり、所与のアプリケーションの一連の変形アプリケーションは依然として同じシグネチャを有する可能性がある。あるアプリケーションについて検討すると、これに表面的に変更がなされた場合、固有のシグネチャが元のアプリケーションに対して計算されたものと相当に異なるのに、このアプリケーションに対するジェネリックシグネチャは、依然として、元のアプリケーションのものと同一である場合がある。当該技術分野におけるジェネリックシグネチャの一例はＰＥハッシュ（pehash）である。当該技術分野におけるジェネリックシグネチャの別の例は、ＳＳディープハッシュ（ssdeep）である。

【００３５】

当該技術分野において、フィンガプリントの用語は従来式のシグネチャと関連付けられ、ファジー・フィンガープリントの用語はジェネリックシグネチャと関連付けられることがよくある。ファジー・フィンガープリントは、ソフトウェア・アプリケーションを入力値とし、（好適には短い）象徴的なシーケンスを出力する変換関数である。理想的には、ファジー・フィンガープリントは２つの素性を有する。第１には、２つのアプリケーションは（一方のアプリケーションに表面的なわずかな変更を加えて他方のアプリケーションを作成した場合）、本質的にきわめて類似しており、これらのアプリケーションのそれぞれのファジー・フィンガープリント値は同一となるはずである。第２には、２つのアプリケーションが相当に異なる場合には、理想的には、これらのアプリケーションのファジー・フィンガープリント値は異なるはずである。こうした素性は理想的なものであり、両方の素性が複数の事例において成立しない場合であっても、ファジー・フィンガープリント値は、なお有用である。ファジー・フィンガープリント値は、ジェネリックシグネチャの一例であり、ジェネリックシグネチャを計算するすべてのアプローチにより、対応するファジー・フィンガープリント値が得られる訳ではない。特に、ファジー・フィンガープリント値を用い、あるアプリケーションのファジー・フィンガープリント値が複数の既知の悪意のあるソフトウェア・アプリケーションに与えられた複数のファジー・フィンガープリント値と一致するか否かを確認して、そのアプリケーションが悪意のあるものか否かを判断することができる。わずかに異なるアプリケーションが同一のファジー・フィンガープリント値を有し得るので、ファジー・フィンガープリント値として機能し得る。当該技術分野におけるファジー・フィンガープリント値の一例は、ＰＥハッシュ値（PEhash）で

ある。当該技術分野におけるファジー・フィンガープリント値の別の例は、ＳＳディープハッシュ値（SSdeep）である。

【００３６】

当該技術分野において、コンビクション（確信）の用語は、ソフトウェアの一部がクライエントシステムに対して悪意のあるものと判断された状況を意味するものである。

【００３７】

当該技術分野において、デジタルシグネチャ（デジタル署名）の用語は、公開鍵暗号作成法を用いて、１つのファイルから比較的に短い文字列を計算するための標準的な技術を意味するものである。ファイルから文字列を計算する暗号化するためには、いわゆる個人鍵（プライベート鍵）を利用する必要がある。公開鍵認証を用いて、ファイルの意図したデジタル署名が正しく計算されたか否か判断することができる。個人鍵の知識がなければ、復号化すべき署名を有効に計算することが導出不可能なものとして、署名方式のセキュリティを確保するものである。当該技術分野においては、このデジタル署名と、上記説明した悪意のあるアプリケーションを検出するタイプのシグネチャをともに「署名（シグネチャ）」と呼ぶが、両者を混同してはならない。

【００３８】

以下の説明において、機械学習の技術分野における当業者に知られた用語を用いる。最も単純な態様において、機械学習技術を用いて、オブジェクトを複数のセットからなる１つに分類することができる。アンチマルウェアの解決手段に関連して、機械学習技術を用いて、所与のソフトウェア・アプリケーションが悪意のありそうか、安全なものでありそうか判断し、分類に対する（悪意ありとする）確からしさ（判断確信度）を反映するスコア（得点）を生成することができる。以下の説明において、本発明の詳細説明が不明瞭となるのを避けるために、機械学習技術に関連する用語体系について、ソフトウェア・アプリケーションが悪意のあるものか、安全なものであるかについての分類のアプリケーションを参照しつつ説明する。第１に、機械学習アプローチは、当該技術分野においては学習フェーズ（training phase）として知られるものを含む傾向がある。悪意のあるものか、安全なものであるかについての分類ソフトウェア・アプリケーションに関連して、まず学習コーパス（training corpus）を構築する。学習コーパスは、通常、一連のソフトウェア・アプリケーションを有する。この一連のアプリケーションのそれぞれには、任意的には、その素性（特性）に関して、たとえば「安全なもの」、「悪意のあるもの」、または「未知のもの」等の「ラベル」が付与されている。このラベルは、手動による解析、またはその他の独立した、おそらくはより高価な手段を用いて判定することができる。ラベル付けされたデータを入手することはより高価であるものと理解されるが、未知のサンプルはより少数であることが好ましい。

【００３９】

さらにコーパスは、最終的には機械学習技術を適用する現実世界のシナリオを表現するものであることが好ましい。たとえばソフトウェア・アプリケーションの分類に関して、コーパス内のアプリケーションが通常、エンドユーザのコンピュータシステム上で確認される可能性のあるアプリケーションを反映し、とりわけ機械学習技術を用いて分類されるコンピュータシステム上のファイルを反映することが好ましい。学習フェーズの最初のプロセスにおいて、特徴ベクトルが各ソフトウェア・アプリケーションから抽出される。特徴ベクトルは、コーパス内のソフトウェア・アプリケーションの顕著な特徴を表現する一連の値である。これらの値は、ソフトウェア・アプリケーションが悪意のあるものか、安全なものか、いずれの傾向が大きいかについて判断する上で、特に関連性が深い（有意義な）ものであることが期待される。

【００４０】

たとえば１つの特徴値は、ファイルがデジタル署名されたものであるか否かを示す単一のバイナリデジット（０または１）であってもよい。この特徴は、実際には、違法なソフトウェア・アプリケーションがデジタル署名されることはほとんどないので、関連性が深いものとなり得る。別の有意義な特徴は、ソフトウェア・アプリケーションを含むファイ

ルサイズである。この特徴は、悪意のあるソフトウェア・アプリケーションが安全なソフトウェア・アプリケーションよりファイルサイズが小さい傾向があるので、関連性が深いものとなり得る。重要なことは、悪意のあるソフトウェア・アプリケーションか、安全なものかを判断する上で、任意の単一の特徴が決定的な証拠とはなり得ないが、複数のこうした特徴値を確認することが決定的な証拠となり得る点に留意するである。数多くのアプリケーション事例において、機械学習システムで利用される特徴の種別は、完全自動化手段を用いて求められるというより、しばしば特定の専門家の判断により決定される。たとえば、ファイルがデジタル署名されたか否かを知ることが有用な情報であると判断する上で、専門家の判断を必要とする場合がある。

【００４１】

特徴ベクトルが学習コーパスから抽出されると、これらのベクトルは、任意のファイル自体に付随するラベルとともに、「学習フェーズ」を実現するアルゴリズムに供給される。学習フェーズの目的は、自動的に「モデル」を算出することにある。モデルは、入力値が特徴ベクトルで、出力値が分類である数学的関数を効率的に符号化するものである。マルウェアを検出するための機械学習技術を用いることに関連して、モデルの出力値は、「安全なもの」または「悪意のあるもの」というバイナリラベルとなり得る。特定の機械学習モデルは、ラベルの確からしさ（悪意ありと判断する確信度）を反映するスコアを生成することもできる。たとえば出力値は（「悪意尤度」が０．９５）の形態を有する符号化であってもよく、この場合、モデルは特徴ベクトルが９５％の確率で悪意のあるソフトウェア・アプリケーションであると確信するということを意味するものである。機械学習アルゴリズムは、理想的には、分類器を生成する必要があり、分類器は、学習例で得られたラベル（確からしさ）に合理的に一致し、新たなアプリケーション事例に対して生成される合理的な尤度（確率）を有するものである。実際には、モデルは、素性が知られていないアプリケーション事例に対して評価されることが予定されているので、合理的な尤度を生成することは重要なことである。

【００４２】

当該技術分野における特定の機械学習アルゴリズムは、単純ベイズアルゴリズム、人工ニューラルネットワーク、決定木（Decision Trees）、サポートベクターマシン、ロジスティック回帰、近傍法、等を有する。分類器の用語を用いて、モデルを記述することもできる。たとえばサポートベクターマシン分類器を意味する場合もある。一旦、分類器／モデルが確立されると、これを用いて、コンピュータまたはコンピュータネットワークに実際に提供されたソフトウェア・アプリケーションに関する新しい事例を評価することができる。

【００４３】

マルウェアの検出に関して、クライエントシステムは、まずソフトウェア・アプリケーションに付随する特徴ベクトルを抽出し、その特徴ベクトルをモデルに適応させて、その素性、および任意的には尤度（確からしさまたは悪意ありと判断する確信度の値）を求める。最後に、この情報に基づいて特定のポリシー（手順）を適用する。実際の分類化プロセスは、クライエントコンピュータで局在的に行う必要はない。むしろ遠隔サーバーコンピュータ上で実行することができ、この場合、クライエントコンピュータは、特徴ベクトルの符号をサーバーコンピュータに送信する。そしてサーバーコンピュータは、分類器を用いて、特徴ベクトルを評価し、関心のあるソフトウェア・アプリケーションの良否について対応する判定を行う。最終的な分類に関するポリシーは、尤度（確からしさまたは悪意ありと判断する確信度の値）を含む場合には、複雑なものとなり得る。たとえばシステムがきわめて危険な状態にあるか、またはきわめて影響を受けやすい情報を含むとき、ソフトウェア・アプリケーションが安全であるという可能性が高くなければ、そのソフトウェア・アプリケーションを遮断してもよい。他方、システムが影響を受けやすくないとき、反対の手順を取ることができる。すなわち特に、ソフトウェア・アプリケーションが悪意のあるものである可能性が高い場合に、そのソフトウェア・アプリケーションを遮断する。

【００４４】
以下の説明において、当業者に知られたログ情報（コンピュータの履歴情報）の概念を用いる。ログ情報とは、所与のシステム上のトランザクション（変更処理）およびアクション（動作）に関する履歴情報である。たとえばシステムがウェブサーバーである場合、ログには、システムに接続された複数のクライエントコンピュータに関する状況、接続回数、および処理内容が含まれる。ログを用いて、所与のシステム上で起こったことについて、合理的な概要を構成することができる。サーバー側構成部品を用いて、所与のソフトウェア・アプリケーションの素性を所望するクライエントコンピュータを支援するアンチウィルス・システムに関連して、イベント履歴エントリは、必ずしも限定するものではないが、以下のものを含む。すなわちイベント履歴エントリは、同じクライエントコンピュータからの個別のトランザクションをリンクするために用いられるクライエントコンピュータの識別子、クライエントコンピュータが特定のアプリケーションに対する素性を要求した時刻を特定するタイムスタンプ、（インターネットプロトコルまたはＩＰアドレスにより特定される）クライエントコンピュータの特定、素性が要求されたファイルの説明（たとえばＭＤ５やＳＨＡ－２５６等のファイルフィンガプリントにより符号化されたファイルの説明）、アプリケーションに付随する任意のアンチウィルス・フィンガプリント（限定するものではないが、従来式のフィンガプリントおよび一般的なフィンガプリントを含む）、懸案のソフトウェア・アプリケーションの属性（限定するものではないが、懸案のソフトウェア・アプリケーションの機械学習特徴ベクトル）、懸案のソフトウェア・アプリケーションに関してその素性の判定を支援する文脈データ（コンテキストデータ）、サーバー側構成部品の応答（限定するものではないが、ソフトウェア・アプリケーションに最終的に付与された素性、ソフトウェア・アプリケーションが以前には存在しなかったか、フィールドで共有されるようになったソフトウェア・アプリケーションに関する追加的に記述する予備的素性、ソフトウェア・アプリケーションについてサーバーコンピュータがクライエントコンピュータに対して行った提案、および最終的な素性を補充するプロセスで用いられたさまざまな副次的な技術により付与された素性、および応答が有効である時間を示す反応に対する保持時間または生存時間（ＴＴＬ））がある。
【００４５】
サーバーコンピュータに対する複数のクエリ（問い合わせ）は、複雑で多面的なものとなり得るので、履歴エントリには、クエリのタイプを特定するエントリが含まれることがある。たとえば、サーバーコンピュータに対する１つのクエリにおいて、クライエントコンピュータは、基本的なフィンガプリントのみを有することがある。同一ファイルに対する後続のクエリにおいて、クライエントコンピュータは、追加的な情報を含むことがある。これら２つのクエリは（解析するとき、同一のクライエントコンピュータが同一のファイルに対して２つのクエリを行った事実をリンクさせやすくできるが）、異なるクエリの種別により独立して記録してもよい。履歴は、複数のクライエントコンピュータから送信される複数の履歴エントリを含む。開示される発明において、展開される機械学習技術を履歴データ用いて学習させる。
【００４６】
開示される発明の目的のため、クライエントコンピュータ上で実行される２組のアプリケーションを区別することは有益である。「懸案のアプリケーション」なる用語は、クライエントコンピュータシステム上のユーザまたはアンチマルウェア構成部品が当該アプリケーションの素性（特性）に興味がある場合において、クライエントコンピュータシステム上に搭載されているソフトウェア・アプリケーションまたは搭載しようとしているソフトウェア・アプリケーションを参照するために用いられる。懸案のアプリケーションに加えて、本明細書は、たとえば懸案のアプリケーションの実行中に別に実行する可能性のある他のタイプのソフトウェア・アプリケーションについて記載する。こうしたソフトウェア・アプリケーションは、限定するものではないが、ウェブブラウザ、ピアツーピア・ファイル共有アプリケーション、銀行取引アプリケーション、またはＰＤＦリーダを含むものであってもよい。悪意のあるアプリケーションは、ピアツーピア・ネットワークを介し

て送信されることが多いので、懸案のアプリケーションの実行中に、ピアツーピア・ファイル共有アプリケーションを実行させようとすることは、懸案のアプリケーションが悪意のあるものである可能性を若干増大させるものと指摘されている。同様のラインの中で、銀行取引アプリケーションを実行させようとする場合、懸案のアプリケーションが悪意のあるものか否かにかかわらず、懸案のアプリケーションの侵入を妨害し、またはその処理を一時停止させることは有意義である。なぜなら、懸案のアプリケーションが悪意のあるものであるリスクが小さい場合であっても、金銭的なデータが危険に晒され、または窃取されたときのコストに見合うものではないためである。こうした検討事項は、懸案のアプリケーションが悪意のある可能性に連動する単なる信号である点に留意すべきである。個別に事案を検討すると、こうした信号は、当該アプリケーションに対してアクションを取る根拠を与えるには十分でない可能性がある。しかし、懸案のアプリケーションに関する情報に加え、こうした複数の信号が、懸案のアプリケーションが悪意のあるイベントであるか否かについてのより決定的な証拠を提供することができる。こうした信号を特徴ベクトルの属性として見ることにより、機械学習方法を上記信号に適用することができる。

【００４７】

ジェネリックシグネチャの実施形態

本発明に係る１つの実施形態において、クライエント側構成部品およびサーバー側構成部品は、以下のように機能する。サーバー側構成部品は任意的な初期化フェーズで機能し、既知の悪意のあるファイルおよび安全なファイルの両方に対するファジー・フィンガープリント値を計算する。これらの結果は、従来型データベースやカード型データベース等のデータ記憶装置に記憶される。ファジー・フィンガープリント値を計算するためのアルゴリズムは、たとえばＰＥハッシュ値（PEhash）やＳＳディープハッシュ値（SSdeep）等の当該技術分野において知られたものであってもよい。択一的には、人手を要するアルゴリズムまたは専用のアルゴリズムを利用してもよい。フィンガープリントの態様の選択（すなわちクライエント側構成部品およびサーバー側構成部品が用いる同一のアルゴリズム）が一致している限り、フィンガープリントの態様の選択が、発明の具体化に影響を与えることはない。

【００４８】

ファジー・フィンガープリント値が決定的に悪意のあるもの（たとえば、これと同一のフィンガープリント値を有する数多くの既知の悪意のあるアプリケーションが存在し、これと同一のフィンガープリント値を有する安全なアプリケーションがまったく存在しない）との十分な証拠があるとサーバー側構成部品が判断した場合、そのファジー・フィンガープリントを決定的に悪意のあるものとラベル付けすることができる。この判定を支援するために、サーバー側構成部品は、安全なものであるか、その属性に基づいて安全なものであると確信される複数のアプリケーションに関するファジー・フィンガープリントを含むデータ構造を維持することができる。データ構造において確認されたファジー・フィンガープリントを有する任意のソフトウェア・アプリケーションは、決定的に悪意のあるものとラベル付けしないことが好ましい。この素性は、クライエントコンピュータへ直接送信されるか、サーバー側構成部品自体に記憶してもよいし、（クライエントコンピュータからクエリ（問い合わせがあったとき、クライエントコンピュータが利用できるようにしても）よい。

【００４９】

いくつかの証拠はあるが、依然として決定的な証拠がないとサーバー側構成部品が判断した場合（たとえば、これと同一のフィンガープリント値を有する既知の安全なアプリケーションが存在するが、処理して得たフィンガープリント値を有する１つまたはそれ以上の悪意のあるアプリケーションが存在する場合）、サーバー側構成部品は、そのファジー・フィンガープリントを悪意のある可能性があると判断することができる。また、いくつかの証拠はあるが、依然として決定的な証拠がないとサーバー側構成部品が判断した場合（たとえば、これと同一のフィンガープリント値を有する既知の安全なアプリケーションが存在する場合）、サーバー側構成部品は、そのファジー・フィンガープリントを安全な

ものであると判断することができる。同様に、安全なアプリケーションおよび悪意のあるアプリケーションの両方が特定のファジー・フィンガープリントを有する場合、サーバー側構成部品は、そのファジー・フィンガープリントを矛盾があるもの判断することができる。

【００５０】

クライエントコンピュータは、新たなファイルに遭遇したとき、まず任意的に当該技術分野における標準的技術を用いて、そのファイルが脅威を与えるか否かについて判断することができる。そのようなステップは、任意的には、アプリケーションの従来式のフィンガープリント値（たとえばＳＨＡ－２、ＭＤ５、および当該技術分野における他の既知の技術）を計算するステップと、任意的に、ファイルが悪意のあるものであるか否かについて、（可能ならば、遠隔サーバーコンピュータの支援を受けて）判断するために用いられる他のメタデータを収集するステップを含む。

【００５１】

クライエントコンピュータは、アプリケーションのファジー・フィンガープリントを計算する。任意的には、クライエントコンピュータは、ローカル記憶媒体にあるファジー・フィンガープリントを検索して、アプリケーションが悪意のあるものであるか否か判断し、悪意のあるものと判断した場合、適当な処理を行う。択一的には、クライエントコンピュータは、遠隔サーバーコンピュータに、ファジー・フィンガープリント値、およびアプリケーションに関して収集された他のすべてのデータ（従来式のフィンガープリント値およびその他のファイルのメタデータ）とともにクエリ（問い合わせ）を送信する。

【００５２】

その後、サーバーコンピュータは、受信した情報を記録することができる。（サーバーコンピュータが当該アプリケーションに関して受信した情報とともにすでに記録したであろう情報を用いて）フィンガープリント値が悪意のあるものであると決定的に判定された場合、サーバーコンピュータは、この素性（特性）をクライエントコンピュータに通知することができる。クライエントコンピュータは、適当な処理を行うことができる（本発明の１つの実施形態において、この処理は、当該アプリケーションを完全に削除するか、ユーザのコンピュータへのインストールを阻止することを含む。）。フィンガープリントが悪意のある可能性があると判定された場合、サーバーコンピュータは、この素性（特性）をクライエントコンピュータに通知することができる。クライエントコンピュータは、適当な処理を行うことができる（本発明の１つの実施形態において、この処理は、当該ソフトウェアのコピーを、さらに解析するためサーバーに、コンピュータへ送信することを含む。）。

【００５３】

本発明に係る１つの実施形態において、サーバーコンピュータは、所与のアプリケーションが悪意のあるものであるリスクを低減するために、数多くの安全装置を整備することができる。これらの安全装置は、限定するものではないが、以下のものを含む。第１に、より直接的な手段（たとえば、安全なソフトウェア・アプリケーションの既知のホワイトリスト中にあるアプリケーションと一致する、ＳＨＡ－２５６などの従来式のフィンガープリント）を用いて、当該アプリケーションが安全であると判断された場合、サーバーコンピュータは、ファジー・フィンガープリントの素性を書き換えることができる。第２に、ファジー・フィンガープリントの利用を抑制することができる。たとえば、サーバーコンピュータは、このフィンガープリントに関する確からしさ（悪意ありと判断する確信度）を控えめな（適度な）ものに制限することができる。同様に、ファジー・フィンガープリントに基づく確からしさを、懸案のアプリケーションの出現頻度が所定の閾値より小さくなる状況に制限することができる。このシナリオにおいて、システムにパラメータＮを導入することができ、当該アプリケーションがＮ個未満のシステムに出現すると考えられる場合に限り、悪意のあるものと判断してもよい。このように制限することにより、誤検出があった場合の誤検出による損害を少なくとも抑制することができる。悪意のあるファイルは、安全なファイルより出現頻度が小さいという傾向があることが知られている。し

たがって、ファイルの出現頻度が高い場合、悪意のあるファイルであるとの判断は、より注意深く行う必要がある。第３に、ファジー・フィンガープリントを用いて、ファイルが悪意のあるものと判断することを、悪意のある可能性が若干より高い特定の分類ファイルに制限することができる。たとえば、データサイズがより小さいファイルは、より大きいファイルに比して、悪意のある可能性がより高いことが知られている。これは、悪意のある第三者にとって、より小さいファイルを、被害者のコンピュータマシンにうまく送信する可能性がより高いためである。当該技術において、デジタル署名されたファイルは、デジタル署名されていないファイルに比して、悪意のある可能性がより低いことも知られている。他のファイル属性に対しても同様の検討を行うことができる。したがって、本発明の１つの実施形態では、任意的ではあるが、特に、データサイズが所定閾値より小さく、デジタル署名されていないソフトウェア・アプリケーションに対して、確からしさ（悪意ありと判断する確信度）に基づいたファジー・フィンガープリント値を抑制することができる。第４に、特定の状況の下では、確からしさに基づいたファジー・フィンガープリントを保存することができる。本発明の１つの実施形態では、あるコンピュータマシンが特定の脅威に感染しやすい場合（たとえば過去において、このタイプの脅威に遭遇したことがある場合や、特定の脅威に付随した地理的領域内にある場合）、こうした場合にファジー・フィンガープリントを利用することができる。

【００５４】

本発明の１つの実施形態では、サーバーコンピュータは、特定のファジー・フィンガープリントが悪意のあるものか、または安全なものかに対応するか否かについて、独立して判断することができる。この場合、サーバーコンピュータは、特定のファジー・フィンガープリントを有する既知のマルウェアの集合体から、複数のソフトウェア・アプリケーションの有無等に関する第三者の情報に依拠することができる。択一的には、サーバーコンピュータは、特定のファジー・フィンガープリントを有する安全なマルウェアの集合体から、複数のソフトウェア・アプリケーションの有無等について検索することができる。最後に、サーバーコンピュータは、ユーザコンピュータ履歴データを調査して、当該アプリケーションが悪意のあるものか、または安全なものかについて、その可能性を判断することができる。とりわけ、特定のファジー・フィンガープリントを有するアプリケーションがきわめて一般的なものであるが、悪意のあるものとして知られたものではない場合、当該アプリケーションが実際に安全なものである可能性は通常高い。この場合、この同一のファジー・ハッシュ値を有するアプリケーションを悪意のあるものと判断することには、リスクがある。

【実施例１】

【００５５】

本発明に係る１つの態様を説明するために実施例１を示す。この実施例は、本発明に係る１つの実施可能なワークフローを説明するためのもので、本発明を明確に理解しやすくするためことを意図したものである。すなわち本発明を不明確にするのを回避するために実施例１では説明しないが、本発明の全体の範疇に含まれる変形例が数多くあるので、実施例１により本発明を制限する意図はまったくない。

【００５６】

実施例１において、クライエントコンピュータおよびサーバーコンピュータが設けられる。新たなソフトウェア・アプリケーションがクライエントコンピュータに提供される。クライエントコンピュータは、このファイルに対してジェネリックシグネチャおよび特定シグネチャの両方を計算し、サーバーコンピュータに送信する。サーバーコンピュータは、これら両方のフィンガプリント値を認証する。サーバーコンピュータは、これら２つの情報（フィンガプリント値）のみから、ソフトウェア・アプリケーションの最終的な良否判定（既知のブラックリストまたはホワイトリストに含まれるものであるかの判定）を行い、この素性（特性）を返答する。

【００５７】

これら２つの情報のいずれかから、決定的な判定を得ることができない場合、サーバー

コンピュータは、クエリ（検索要求）で送信されたジェネリック・フィンガープリントに関して過去に認識されたすべての特定シグネチャを調べる。（複数の異なるファイルが同一のジェネリック・フィンガープリントを有し得るので、同一のジェネリック・フィンガープリントに付随し得る複数の特定フィンガプリントを共有することがある点に留意されたい。）簡略化すると、クエリの中に次のフィンガプリント、すなわち（Ｇ,ＳＯ），（Ｇ,Ｓ１），（Ｇ,Ｓ２），（Ｇ,Ｓ３），．．．，（Ｇ,Ｓ９）が存在する場合を考慮されたい（Ｓ１～Ｓ９は異なる特定フィンガプリントであり、これらのすべてが同一のジェネリック・フィンガープリントＧに対応するものである。）。ここで、これらの特定フィンガプリントの閾値が悪意のあるものと仮定する（たとえばＳＯ～Ｓ７がすべて既知のマルウェアに対応するものと仮定する）。さらに、過去に確認されたこれらの特定フィンガプリントはいずれも既知の安全なファイル（ホワイトリストに含まれるファイル）に付随するものではないと仮定する。換言すると、特定フィンガプリントＳ８，Ｓ９は、これまで知られていない素性を有する（すなわち特定フィンガプリントＳ８，Ｓ９は、悪意のあるもの、または安全なものであるか不明で、これまで誰も良否判定を行っていない。）。この場合、バターンが現れる。１つのジェネリック・フィンガープリントＧに付随する特定フィンガプリントの大多数は、悪意のあるものであるように見える。このとき、このジェネリック・フィンガープリント自体は悪意のあるものとして識別すべきであると判断することが合理的であると考えられる。

【００５８】

サーバーコンピュータは、この一連のステップの後に、このジェネリック・フィンガープリント「Ｇ」は悪意のあるものと識別し、クライエントコンピュータに対応する応答を返信する。

【００５９】

上記判定プロセスは、リアルタイムの事象として（すなわちオンザフライで（中間ファイル出力をせず、複数の処理を行うこと））説明したが、実際には、個別に行ってもよい。換言すると、サーバーコンピュータ上のソフトウェアモジュールは、過去のクエリの履歴を周期的に検索して、過去のクエリに関連する大多数の特定フィンガプリントが悪意のあるものらしいとの理由から、悪意のある可能性の高い特定フィンガプリントを抽出するよう試みる。こうして、これらのジェネリック・フィンガープリントを、悪意のあるものと識別することができる。

【００６０】

以上のように、サーバーコンピュータが判定を求められると、オンザフライで計算を実行することを試みるのではなく、単に検索を行ってもよい。これと同時に、このアプローチは、最近の履歴がすでに解析されているので、関連情報を収集して利用することはない。

【００６１】

［機械学習の実施形態］

本発明に係る１つの実施形態において、クライエントコンピュータおよびサーバーコンピュータは以下のように機能する。初期化段階において、サーバーコンピュータは、識別器に学習させる。１つの実施形態において、サーバーコンピュータは、ファイルに対するフィンガプリントが過去に提供され、おそらくは独立した手段を用いて分類された実際に存在するユーザ履歴から学習データを直接的に取得することができる。このファイルが、たとえば既存のホワイトリストまたはブラックリストの存在により安全なものか、または悪意のあるものか、認識することができる。

【００６２】

任意的には、たとえばユーザが自然的分類およびサブ分類に属するか否かといったさまざまな条件に基づいて、履歴データを階層化し、または分割化することができ、自然的分類等は、限定するものではないが、地域グループ（すなわち同様の地域にいるユーザからなるグループ）、および関連グループ（すなわち、同一企業の社員からなるユーザ、本発明に係るシステムまたはソフトウェアを共通の提供源（たとえば共通のダウンロード・サ

ーバーコンピュータまたは共通の頒布チャンネル）から入手したユーザからなるグループ）を含むものであってもよい。所定の条件にもとづいて、学習データを階層化し、または分割化した場合、用いられる学習データを、履歴からの複数の区分または複数の階層に基づいて求めることができる。学習データを分割することの利点は、入力空間の特定領域において複数の機械学習分類器を微調整することができ、その結果として、機械学習分類器の入力空間の特定領域におけるパフォーマンス（性能）を改善することができる点にある。学習フェーズにおいて複数のパラメータが含まれる。一旦、分類器が確立されると、フィールドにおいて展開させることができる。

【００６３】

１つの実施形態において、分類器が特定する数学的関数を実行する実際のコンピュータ指令（または順次説明されるコンピュータ指令の適当な符号化）を自動的に生成することができる。１つの実施形態によれば、これらのコンピュータ指令を遠隔サーバーコンピュータに記憶させてもよい。別の実施形態において、これらのコンピュータ指令を複数のクライエントシステムに送信することができる。

【００６４】

本発明に係る別の実施形態において、クライエントシステムは、新たなソフトウェア・アプリケーションに遭遇したとき、このソフトウェア・アプリケーションに付随する特徴ベクトルを、ソフトウェア・アプリケーションが安全なものか、悪意のあるものかを独立して判定する他の任意のデータとともに抽出する。特徴ベクトルは、特定のアプリケーションの属性に限定する必要はなく、アプリケーションを実行するはシステムに関する他の属性を含むものであってもよい。アプリケーションのバイナリコンテンツに特に関連する特徴ベクトルの属性は、限定するものではないが、アプリケーションのバイナリコンテンツの素性、アプリケーションが参照するダイナミック・リンク・ライブラリ（ＤＬＬ）のリスト、バイナリコンテンツ内の特定位置の値、領域の数、シンボルの数、バイナリコンテンツの異なる領域の個数、シンボル個数、位置、およびバイナリコンテンツのデータサイズが含まれる。

【００６５】

いくつかの実施形態では、特徴ベクトルはアプリケーションが参照するダイナミック・リンク・ライブラリの符号を含む。その他の実施形態では、特徴ベクトルは、バイナリコンテンツ内の異なる領域に関する領域の数、シンボルの数、および位置を含む。その他の実施形態では、特徴ベクトルは、バイナリコンテンツのデータサイズを含む。アプリケーションに付随する特徴ベクトルの属性は、限定するものではないが、アプリケーションで用いられる登録鍵（registry key）および（通常、ウィンドウズ上で実行する脅威に対する）登録鍵に加えられたすべての変形に関する情報、アプリケーションのファイル名、利用されるネットワークポートおよびアプリケーション・プログラミング・インタフェース（ＡＰＩ）通信に関する挙動的属性、アプリケーションにより変更され、生成されるファイル、およびアプリケーションにより中止され、または開始されるサービスが含まれる。

【００６６】

いくつかの実施形態では、特徴ベクトルは、アプリケーションのファイル名および利用される登録鍵を含む。アプリケーションのジェネラルコンテキストデータに付随する特徴ベクトルの属性は、限定するものではないが、アプリケーションに新たに遭遇したときにシステム上で実行するプロセス、アプリケーションの情報源（ＣＤ　ＲＯＭ、ＵＳＢスティックメモリ、ウェブサイト）、マシンの感染履歴、マシンの地理的位置、およびマシンのＩＰアドレスが含まれる。一般に、特徴ベクトルは、複数の特徴に関する情報を含む。

【００６７】

特徴ベクトルを構成する際、上記特徴ベクトルの値は、逐語的に送信する必要はないが、機械学習アプリケーションを支援するように符号化されることに留意すべきである。たとえば、１つのソフトウェア・アプリケーションに関するすべてのダイナミック・リンク・ライブラリのリストではなく、むしろバイナリ値を用いて、ウィンソック（winsock.dll）等の特定のダイナミック・リンク・ライブラリが用いられたか否かを示すことができ

る。１つの実施形態では、特徴ベクトルに加えて、クライエントコンピュータは、ＳＨＡ－２５６等の従来式のフィンガプリント、またはＰＥハッシュ値（PEhash）やＳＳディープハッシュ値（SSdeep）（ともに当該技術分野において知られている。）、またはこれらの組み合わせ等のジェネリック・フィンガープリント値を計算してもよい。特徴ベクトルは、概略的に上記説明した機械学習技術を用いてファイルを分類する際、主として関連するものであるが、他のデータを将来の学習のために用いることができる。たとえば、最初に遭遇したファイルの素性が未知の場合、当該ファイルが後に既知の悪意のあるアプリケーションのブラックリスト上で発見される可能性がある。そのリストがＳＨＡ－２５６を用いて索引付けされると、ＳＨＡ－２５６の値および特徴ベクトルの両方をクライエントコンピュータに計算させることにより、特徴ベクトルを特定の素性に関連付けることができる。この特徴ベクトルを、その後の学習フェーズのための学習コーパスに追加することができる。

【００６８】

本発明に係る１つの実施形態において、クライエントコンピュータは、特徴ベクトルの値（データ）を求め、これを圧縮することができる。データを圧縮するための汎用技術があり、この特別の事例においては、望ましい性能パラメータ、特にクライエントコンピュータとサーバーコンピュータとの間で通信されたデータ容量に関する性能パラメータを生成する専用技術を用いることができる。

【００６９】

この特徴ベクトルを任意的に圧縮すると、本発明に係る１つの実施形態によれば、得られたデータは、遠隔サーバーコンピュータに送信される。択一的には、クライエントコンピュータは、遠隔的な検索を回避するために、サーバーコンピュータに接続される論理回路に記録してもよい。

【００７０】

本発明に係る１つの実施形態において、サーバーコンピュータは、必要ならば、クライエントコンピュータから送信されたデータを解凍し、解凍データには特徴ベクトルを含み、適当な位置に格納されたモデルに対して、特徴ベクトルを評価する。クライエントコンピュータが従来式のフィンガプリントまたはジェネリック・フィンガープリント等の他のデータを提供した場合、サーバーコンピュータは、任意的には、分類器からの結果の代わりに、より従来式の手段を介して得られた素性を優先させる。たとえば。クライエントコンピュータが関心のあるソフトウェア・アプリケーションのＳＨＡ－２５６の値を送信してこの値が安全なアプリケーションからなる既知のホワイトリストに含まれていたとき、サーバーコンピュータは、機械学習モデルの出力値にかかわらず、懸案のアプリケーションが安全なものと応答することができる。このアプローチの根拠となる前提は、直接的なホワイトリストまたはブラックリストより、機械学習モデルがより誤りを起こしやすいということにある（ただし、機械学習モデルは、これまで未知であったものを含めて、任意のファイルに対して適用可能であるのに対し、ホワイトリストおよびブラックリストは、そのエントリ数が限られているため、同様に制約があることに留意すべきである。）。サーバーコンピュータは、必要ならばクライエントコンピュータに実行させたい処理に関する情報を付加して、最終的な判断に関する応答をクライエントコンピュータに提供する。このトランザクション（変更処理）に関するトランザクション履歴は、クライエントコンピュータの識別子、タイムスタンプ、特徴ベクトル値、その他のフィンガプリント値、および最終的な素性および情報が得られた際の状況、中でも、サーバーコンピュータがクライエントコンピュータに実行させたい処理のタイプを含むものであるが、任意的に記録される。このトランザクション履歴は、学習コーパスに関する３つの特徴を含むものであるので、その後の新たな分類器の学習フェーズに利用することができる。第１に、このトランザクション履歴は、機械学習訓練アルゴリズムに対する入力値として与えられる特徴ベクトルを含む。第２に、このトランザクション履歴は、数多くの機械学習アルゴリズムが必要とする素性を含む。ただし、学習のためには、巡回フィードバックループを導入するリスクがあるので、従来式の機械学習に基づく素性ではなく、フィンガプリント値やジェ

ネリック・フィンガープリント値を用いて得られる素性を利用することが好ましいことに留意されたい。第３に、このデータにより生成された学習例は、当該分野における実際のユーザ事例に由来するものであり、通常のユーザが将来的に直面するものを如実に表現する傾向がある。

【００７１】

本発明に係る１つの実施形態において、クライエントコンピュータは、サーバーコンピュータから最終的な判断、およびその判断に付随する可能性のある処理指令を受信し、特化されたポリシーによる応答に基づいて処理を行う。１つの実施形態において、可能性のある応答は、限定するものではないが、次のものを含む。すなわち可能性のある応答には、アプリケーションが悪意のあるものと宣言し（システムから当該アプリケーションを取り除くか、ユーザがインストールするのを阻止する。）、任意的ではあるが、コピーをサーバーコンピュータに送信するか、または当該アプリケーションをシステム上に維持することを許容し、および／または当該アプリケーションを追加的に解析するために、クライエントコンピュータからサーバーコンピュータへ送信することを要求する。

【００７２】

上記最後の応答（オプション）は、たとえば当該アプリケーションを悪意のあるものである可能性があるが、その確からしさ（悪意ありと判断する確信度）が十分に高くなく、偽陽性判断するリスクが相当に高いとサーバーコンピュータにより判断された場合に意義あるものである（この場合、ファイルをサーバーコンピュータへ送信することにより、当該アプリケーションに対して追加的でより綿密な解析を行うことができるが、こうした解析は、あまりにも効果であるので、直面したすべてのファイルに対して実行することはできず、疑惑のある複数のファイルの一部を対象として実行することが好ましい。）。

【００７３】

本発明に係る１つの実施形態において、サーバーコンピュータは、数多くの防御手段（セーフガード手段）を所定位置に配置して、所与の安全なアプリケーションを誤って悪意のあるものと判断するリスクを低減することができる。防御手段は、限定するものではないが、以下のものが含まれる。第１に、より直接的な手段（ＳＨＡ－２５６等の従来式のフィンガープリントや、安全なソフトウェア・アプリケーションからなる既知のホワイトリストにあるものと一致すること）により、アプリケーションが安全であることが知られている場合、サーバーコンピュータは、機械学習分類器から提供された素性を無視することができる。第２に、機械学習分類器の使用を抑制することができる。たとえば、サーバーコンピュータは、この機械学習分類器による陽性判断される（悪意のあるものとの判断される）数を適当な数に制限することができる。さらに所与のアプリケーションに関する分類数を抑制することができる。たとえばＳＨＡ－２５６の関し、機械学習分類器を用いて、（Ｎが３等であるような適当な選択に対して）Ｎ倍を超えない数の陽性判断を行うことができる。この手法によれば、問題（mistake）があるとき、その問題による結果（damage）が確実に生じるようにすることができる（すなわち最も悪意のあるソフトウェアは、その怪しげな危機（fly-by-night danger）に起因して、低い確率（frequency）を有する傾向があるので、このように抑制することで、正検出率と誤検出率の好適なトレードオフを実現することができる。）。第３に、多少なりとも悪意のあるもの可能性がより高い特定の分類のファイルに対しては、機械学習分類器による陽性判断（悪意のあるものとの判断）を抑制することができる。たとえば、データサイズのより小さいファイルは、より大きいファイルに比較して、悪意のあるものである可能性が高いことが知られている。これは、悪意のある第三者が被害者のマシンに対し、より小さいファイルを送信することに成功する可能性が高いためである。当該技術分野において、デジタル署名されたファイルは、デジタル署名されていないファイルより悪意のあるものである可能性が低いということが知られている。同様の判断が他のファイル属性に対しても成り立つ。したがって、本発明に係る１つの実施形態において、とりわけデータサイズが所定の閾値以下であって、デジタル署名されていないソフトウェア・アプリケーションに関して、機械学習分類器による陽性判断（悪意のあるものとの判断）を任意的に抑制することができる。第４に、特

定の条件に関して、機械学習分類器による陽性判断（悪意のあるものとの判断）を記録することができる。

【００７４】

本発明に係る１つの実施形態において、コンピュータマシンが特定の脅威に感染する虞がある場合（たとえば過去においてこのタイプの脅威に遭遇したことがある場合や、特定の脅威に付随した地理的領域にある場合等）、このような場合に機械学習分類器を適用することができる。第５に、特定脅威事例をモデル化するために、機械学習分類器を構築することができる。たとえば、日常的に検出される１つの悪意のあるソフトウェア脅威としてコンフィッカー（Conficker）が当業者に知られている。コンフィッカーには数多くのバリエーションがあるが、バリエーションの中のそれぞれを同一系列の全体の一部として認識することを可能にする十分な共通性がある。したがって、本発明に係る１つの実施形態において、特定の脅威を標的として機械学習分類器を特別に学習させることができる。そのためには、安全なファイルおよびコーパス内の特徴ベクトルを一定に設定し、悪意のあるファイルおよび特定の脅威に関連する特徴ベクトルを含むようにすることができる。このアプローチの利点は、特定の脅威に対して微調整した機械学習分類器において、その脅威に対する偽陽性率（安全なアプリケーションを誤って悪意のあるものと認定する確率）を低く抑えることができ、エンドユーザがこれらのシステムが標的とする特定の脅威を知ることができる点にある。第６に、ポピュラリティ（popularity）が特定の閾値以下であるファイルに対して、機械学習分類器に関するアプリケーションを制限することができる。１つの実施形態において、システムにパラメータＮを導入して、Ｎ個未満のシステムが特定のアプリケーションを有するような場合にのみ、このアプリケーションが悪意のあるものと認定してもよい。第７に、懸案のシステムがある脅威に感染する可能性が閾値より多少なりとも大きい場合、いくつかの機械学習分類器に関するアプリケーションを制限することができる。脅威に感染する可能性の増大を示唆する指標は、限定するものではないが、そのシステムが最近感染した状況、システムが最近攻撃対象となった情報、脆弱なソフトウェア・アプリケーションがシステム上に存在するか否か、感染に関する共通の特徴ベクトルであるソフトウェア・アプリケーション（クライエントコンピュータを共有するピアツーピア・ファイル等）がシステム上に存在するか否か、およびオープンネットワークポートがシステム上に存在するか否か、が含まれる。

【００７５】

しかしながら、実際には悪意のあるマルウェアが、偽陽性率を低減しようとする安全策（セイフティネット）を取ったために安全なものと過誤認識されるため、一般には、偽陽性率を低減すする試みにより、真陽性率をも低減することになることに留意されたい。当該技術分野においては、こうしたトレードオフ（二律相反）が存在し、特定のアプリケーションにもよるが、図らずも、そのトレードオフが望ましいか否か判断することがあるものと理解されている。たとえば、真陽性率がほんの少しだけ低減する一方、偽陽性判断のリスクが劇的に低減する場合、上記トレードオフが望ましいものである。択一的には、偽陽性判断に要するコストがきわめて高い場合、すなわち安全なアプリケーションを遮断することにより、財政的なビジネス上の損失を招く可能性がきわめて高い場合、真陽性率が実質的に低減するとしても、偽陽性判断コストを実質的に低減させるより従来的な立場を取る方が望ましい。他方、誤った検出（すなわち偽陽性検出）に対するコストがきわめて高い場合、たとえばシステムが非常に高いセキュリティを要求する場合、システムに侵入する脅威のリスクをきわめて小さく抑制するために、高い偽陽性率が許容されない場合もある。

【実施例２】

【００７６】

この実施例は、本発明に係る特定の具体例について示し、その手順に沿ったステップまたは手続きについて説明する。この実施例は、本発明に係る説明を明確にするものであって、本発明を限定するものと理解すべきではない。たとえば本発明に係る上記説明は、数多くの変形例および変更例にも適用することができる。不明確な説明を回避するために、

これらの変形例および変更例については以下において説明しない。

【００７７】

まず始めに、ユーザのコンピュータマシン上で実行される１つのエージェントソフトウェアを考慮されたい。この実施例において、エージェントソフトウェアは、マイクロソフト・ウィンドウズのファイルシステムのミニフィルタ・ドライバ（filesystem mini-filter driver）を含み、これは新しい（実行可能な）ファイルをファイルシステムへ書き込もうとしていることを検出することができるものである。同様に、新しい実行可能なファイルをファイルシステムへ書き込もうとしていることを検出することができるその他のソフトウェアを用いることができる。ファイルをファイルシステムへ書き込んだこと、または書き込もうとしていることを検出した後、ソフトウェアエージェントは２つの値を計算する。第１に、ファイル上の「従来式の」フィンガプリント値（たとえばＳＨＡ－２５６）を計算する。第２に、ファイルから機械学習特徴ベクトル値を計算する。特徴ベクトル値は、このシステム上のファイルに関連する数多くの属性を有し、限定するものではないが、アプリケーションが参照するダイナミック・リンク・ライブラリ（ＤＬＬ）、バイナリコンテンツの特定の位置に関する値、ファイル領域の個数（および読み出し可能、書き込み可能、または実行可能等の各領域が有する属性）、シンボルの個数、バイナリコンテンツのテータサイズ、バイナリコンテンツに対するデジタル署名の有無等が含まれる。これらすべての属性は、ファイルのバイナリコンテンツから容易に計算することかできる。さらに、その他の文脈上の情報要素が特徴ベクトルに含まれており、限定するものではないが、ファイルシステムのタイムスタンプ、ファイル名の素性（同一ファイルが異なるシステム上で異なるファイル名を有し、その属性が所与のシステムにおけるファイルの具体例に固有のものとなる場合があることに留意されたい。）、システムにインストールされたその他のソフトウェア・アプリケーションに関する情報（たとえばシステムが脆弱なソフトウェア、または一般にシステム感染を引き起こすソフトウェアに関する情報）、および最近のシステムの感染履歴（たとえば最近３０分の間にユーザのコンピュータが感染したか否か）等が特徴ベクトルに含まれている。これらの属性は、適当に符号化されるとともに、（伝送容量を小さくするために）圧縮される。

【００７８】

クライエントコンピュータは、フィンガプリントおよび特徴ベクトルをサーバーコンピュータに送信する。これら２つの値に加えて、クライエントコンピュータは、（同一のクライエントコンピュータからのその他のトランザクションとリンクさせやすくするために）識別子を有していてもよい。

【００７９】

次にサーバーコンピュータは、任意のブラックリストおよびホワイトリスト内のファイルを検索する。このときサーバーコンピュータは、この検索を実行するために、たとえば従来式のフィンガプリントを用いてもよい。この検索により、決定的な素性（特性）が明らかとなった場合（たとえば、このファイルが悪意のあるものか、安全なものであるか決定的であることが知られている場合）、この素性がクライエントコンピュータに送信される。サーバーコンピュータは、任意的ではあるが、この段階でファイルについて追加的な情報（たとえば当該ファイルを保有するユーザの数等）を検索し、フィンガプリント値、基本的な特徴ベクトル、追加的情報、クエリのタイムスタンプ、ユーザの識別子、ブラックリストおよびホワイトリストのリストごとの素性を記録する。記録フォーマットは、サーバーコンピュータのトランザクション履歴であってもよい。

【００８０】

サーバーコンピュータがブラックリストおよびホワイトリストのどちらにもファイルを見出さなかった場合には、以下のステップを実行する。第１に、サーバーコンピュータは、任意的に、その他の計算可能な属性を用いて、クライエントコンピュータから提供される特徴ベクトルを増補（拡張）する。これらの属性は、限定するものではないが、ファイルがユーザコンピュータに出現する頻度、およびファイルがサーバーコンピュータで初めて確認された時を示すサーバーコンピュータ側のタイムスタンプを含むものであってもよ

い。

【００８１】

そしてサーバーコンピュータは、機械学習分類器（サポートベクターマシン、決定木、ニューラルネットワーク等）を用いて増補された特徴ベクトルを評価する。クライエントコンピュータは、素性（悪意のあるもの／安全なもの）および任意的な確信度を受信し、将来の解析のためのトランザクション（変更処理）が記録される。

【００８２】

サーバーコンピュータは、周期的に、過去のすべての履歴を調べて、既知のブラックリスト／ホワイトリストにフィンガプリント値が記録されたファイルに関するすべての特徴ベクトルを呼び出すことができる。サーバーコンピュータは、フィンガプリント値に対応する特徴ベクトルに関する学習コーパスを形成する（たとえばホワイトリスト上のアイテムは学習コーパスの「安全な」サブセットであり、ブラックリスト上のアイテムは学習コーパスの「悪意のある」サブセットである。）。

【００８３】

このコーパスに基づいて、機械学習分類器（サポートベクターマシン、決定木、ニューラルネットワーク等）を学習させることができる。システムを起動、すなわち「ジャンプスタート（jumpstart）」させる方法がいくつかあることに留意されたい。我々は、データ収集フェーズから開始することができる（複数のタイプのサイレント検出機能（silent detection capability）を想定されたい。）。

【００８４】

文脈上の確からしさ（悪意ありと判断する確信度）に関する実施形態

本発明に係る１つの実施形態によれば、クライエントコンピュータおよびサーバーコンピュータの構成部品は、以下のように機能する。クライエントコンピュータが、あるソフトウェア・アプリケーションに遭遇（アクセス）し、それが悪意のあるものか、安全なものかを分類しようとするとき、そのアプリケーションに関する両方の情報を収集し、システムに関する文脈データ情報とともに用いて、マルウェアの従来式の検出を実施する。収集されるデータ情報として、限定するものではないが、システムに対する最近の感染履歴、クライエントコンピュータの地理的位置情報、インターネットプロトコルすなわちクライエントコンピュータのＩＰアドレス、および同一のクライエントコンピュータ上で複数回行われるトランザクション（変更処理）をリンクさせるために用いられるクライエントコンピュータ識別子が含まれる。

【００８５】

感染イベントに晒されたカスタムエージェント（custom agent）または第三者エージェント（third-party agent）を用いて、感染履歴を収集することができる。クライエントコンピュータは、ソフトウェア・アプリケーションに関する従来式のデータ情報および文脈上のデータ情報の両方を送信する。データ情報は、未処理の状態で伝送してもよいし、またはネットワークにおける効率的な伝送を可能にするように符号化してもよい。符号化機構の選択は、本発明に係る主な態様に対して直交し、当該技術分野においてデータを符号化するための数多くの技術が存在する。サーバーコンピュータは、クライエントコンピュータからデータを受信し、そのアプリケーションが悪意のあるものか否かを判定する。ホワイトリストまたはブラックリストにあるシグネチャ等の従来式の手段を用いて、アプリケーションが悪意のあるものか、または安全なものであると判断された場合、クライエントコンピュータから送信される追加的な文脈を参照することなく、最終的な判断がなされる。送信されたデータに基づいて、そのアプリケーションが疑念を抱かせるものであるが、完全に悪意のあるものであると警告するほどの疑念を抱かせるものではない場合、文脈上の情報が考慮される。１つの実施形態において、アプリケーションが疑念を抱かせるものであって、そのコンピュータマシンが１つまたはそれ以上の感染を受けていた場合には、そのアプリケーションは悪意のあるものと、サーバーコンピュータは判断することができる。一旦、サーバーコンピュータがその勧告を通知すると、この情報はクライエントコンピュータに返信され、その後クライエントコンピュータは、その勧告に応じたポリシ

ーを適用することができる。１つの実施形態では、アプリケーションは悪意のあるものと、サーバーコンピュータが判定すると、クライエントコンピュータは、そのシステムから当該アプリケーションを削除するか、システムへのインストールを妨害することができる。異なる実施形態では、クライエントコンピュータは、セキュリティ上、より危険な状態にある場合には、当該アプリケーションを阻止することができる。たとえばマシンが銀行取引アプリケーション等の高いセキュリティを要するソフトウェアを現在実行中である場合、（セキュリティ侵害はより直接的に金銭的損失をもたらすため、）安全面において、より慎重に扱うべき状態にある。この場合、クライエントコンピュータは、銀行取引アプリケーションの実行が完了するまで、疑念を抱かせるソフトウェアの実行を阻止することができる。

【００８６】

本発明の別の実施形態によれば、クライエント側の構成部品は、文脈上の確からしさ（悪意ありと判断する確信度）を生成することに関連する情報を収集する。１つの実施形態では、クライエント側の構成部品は、単に、クライエントコンピュータの識別子をサーバーコンピュータへ送信する。クライエントコンピュータは、任意的に、以下の１つまたはそれ以上の情報要素をサーバーコンピュータへ送信する。すなわち、その情報要素には、最近の感染、ならびにそれらの感染に関するタイムスタンプおよびウィルス識別子、クライエントコンピュータが最近アクセスしたウェブサイトに関する情報、システム上で実行しているアプリケーションに関する情報、システムにインストールされたアプリケーションに関する情報、システム上で開いた（アクセス可能な）ネットホークポートに関する情報、クライエントコンピュータの地理的位置、クライエントコンピュータのインタネットプロトコルまたはＩＰアドレスが含まれる。１つの実施形態において、このクライエント側の構成部品は、バックグラウンドで、常に実行し、情報を収集し、そして周期的間隔でサーバーコンピュータへ送信することができ、または懸案のアプリケーションに遭遇したときにサーバーコンピュータへ送信することができる。さまざまな実施形態において、このクライエント側の構成部品は、必要があるタイミングで情報を収集することができる。さらに別の実施形態では、このクライエント側の構成部品は、バックグラウンドで収集した情報と、必要なタイミングで収集した情報とを組み合わせてもよい。

【００８７】

本発明の別の実施形態によれば、サーバー側の構成部品は、そのマシンに関する文脈上の情報に加え、アプリケーションに関する情報を解析し、その情報を用いて、当該アプリケーションが悪意のあるものか否かについて判断する。１つの実施形態において、入力された文脈により、アプリケーションが悪意のあるものである可能性が以前より増大するとき、サーバーコンピュータは、疑わしいアプリケーションを悪意のあるアプリケーションに更新することを選択してもよい。別の実施形態では、最近の感染がシステム上で確認されていた場合に、その疑わしいアプリケーションを悪意のあるアプリケーションとみなすようにしてもよい。上記説明した実施形態においては、サーバーコンピュータ上で当該判断を行うものとしたが、ロジック（一連の処理）そのものは、クライエントコンピュータが行ってもよいし、あるいはクライエントコンピュータおよびサーバーコンピュータが協働して行ってもよい。１つの実施形態では、サーバーコンピュータがクライエントコンピュータから提供されたクライエントコンピュータ識別子を参照し、その識別子を用いて、サーバーコンピュータとクライエントコンピュータとの間の履歴を検索することができる。この情報を用いて、文脈を判断に加えることができる。たとえばクライエントコンピュータが最近にサーバーコンピュータと通信して、クライエントコンピュータが問い合わせた懸案のアプリケーションが悪意のあるものであった場合、サーバーコンピュータは、アプリケーションが悪意のあるものである可能性が以前より増大した状況として処理することができる。別の実施形態では、サーバーコンピュータは、複数のクライエントコンピュータから収集した文脈上の情報を利用することができる。この場合サーバーコンピュータは、限定するものではないが、特定のアプリケーションに対して問い合わせされた頻度や、この問い合わせに関連した他のクライエントコンピュータからの文脈情報を含む情報を

利用することができる。

【００８８】

本発明に係る別の実施形態によれば、アプリケーションが悪意のあるものか、または安全なものかについて判断の支援するために保有された文脈データ（コンテキストデータ）を収集するために、ある方法が、クライエントコンピュータシステム上で実行される。この方法は以下のステップを有し、各ステップは任意的なものである。第１に、サーバーコンピュータは、同一のシステムからの複数のトランザクション（変更処理）を関連付けるために用いることができるクライエントコンピュータ識別子を収集する。１つの実施形態では、この識別子は、グローバル一意識別子（ＧＵＩＤ：Globally Unique Identifier）であってもよい。別の実施形態では、識別子は、クライエントコンピュータが初期化された時にサーバーコンピュータにより構成され、クライエントコンピュータに送信されるものであってもよい。その後、クライエントコンピュータは、このデータを不揮発性記憶媒体に記憶する。第２に、サーバーコンピュータは、悪意のある脅威がシステムに侵入した時に関する情報とともに、カスタムエージェント（custom agent）または第三者エージェント（third-party agent）を用いて、システム上で特定された悪意のある脅威を記録する。アンチマルウェア技術に関連して、こうした脅威はウィルスＩＤ、ジェネリックシグネチャ、ＳＨＡ－２５６、またはこれらの組み合わせにより特定することができる。通常、ウィルスＩＤは、脅威に関するほとんどのジェネリックラベル（ジェネリックシグネチャ）を提供し、唯一の脅威を特定する特定ラベル（特定シグネチャ）を提供する。ジェネリック・フィンガープリント値は、両極にある間の特異性のレベルを示唆するものである。第３に、サーバーコンピュータは、ユーザのクライエントコンピュータがアクセスしたすべてのウェブサイトを記録する。第４に、サーバーコンピュータは、ユーザが特定範囲の期間内にイントールしたすべてのソフトウェア・アプリケーションを記録する。第５に、サーバーコンピュータは、懸案のアプリケーション（すなわち、我々の興味を引く素性を有するアプリケーション）が導入された時に実行していたすべてのアプリケーションを記録する。第６に、サーバーコンピュータは、クライエントコンピュータのインターネットプロトコル・アドレス（ＩＰアドレス）を取得する。第７に、サーバーコンピュータは、クライエントコンピュータのネットブロック（netblock）に関する情報を取得する。第８に、サーバーコンピュータは、クライエントコンピュータの地理的情報に関する情報を取得する。第９に、サーバーコンピュータは、クライエントコンピュータのシステム上で用いられている言語に関する情報を取得する。第１０に、サーバーコンピュータは、システム上で開いた（アクセス可能な）ネットホークポートに関する情報を取得する。第１１に、サーバーコンピュータは、システム上で実行中のアプリケーションに関する情報を取得する。第１２に、サーバーコンピュータは、懸案のアプリケーションがシステムに送信された状況に関する情報を取得する。この情報は、限定するものではないが、ウェブブラウザ等を介して取得されたソフトウェア・アプリケーション、ならびにウェブサイト、ＣＤ－ＲＯＭ、またはＵＳＢドライブ等、ファイルが導入されたソース位置が含まれる。第１３に、サーバーコンピュータは、懸案のアプリケーションが取得しようとしている権限、たとえばサーバーコンピュータは、管理者権限の下で実行するか否かについての情報を取得する。第１４に、サーバーコンピュータは、懸案のアプリケーションに関する問い合わせがあった時に、ユーザが現在閲覧しているウェブサイトに関する情報を取得する。第１５に、サーバーコンピュータは、アプリケーションの現在の状況、たとえばアプリケーションがシステム上で実行中であるか否か、または休止中であるか否かについての情報を取得する。これらのすべての情報要素が必ずしも強制的なものではなく、冗長でさえある可能性があることに留意されたい。このリストは、本発明の異なる態様を明瞭にするために含まれるものである。たとえばクライエントコンピュータが識別子とともに、懸案のアプリケーションに関するデータをサーバーコンピュータに送信した場合、サーバーコンピュータは、クライエントコンピュータの過去のトランザクション（変更処理）および懸案のアプリケーションに対するリクエストに関する情報を用いて、文脈上の情報を策定（生成）することができる。とりわけサーバーコンピュータは、クライエントコンピュータが

懸案のアプリケーションについて悪意のあるものと問い合わせした（クエリを行った）時に、過去に問い合わせしたアプリケーションを特定し、特定可能な場合には、対応するアプリケーションの脅威を特定することかできる。この情報から、クライエントコンピュータの感染履歴を構築することができる。同様に、サーバーコンピュータは、クライエントコンピュータのインターネットプロトコル・アドレス（ＩＰアドレス）を取得することができるため、クライエントコンピュータの地理的位置に関する情報を取得することができるが、サーバーコンピュータと通信するために、クライエントコンピュータが利用するネットワークプロトコルの一部として含まれる情報を利用することにより、サーバーコンピュータと通信することができる。特に、利用されるプロトコルが通信制御プロトコル／インターネットプロトコル（ＴＣＰ／ＩＰ）である場合、インターネットプロトコル・アドレスは、自動的に包含されるものである。

【００８９】

本発明に係る別の実施形態によれば、懸案のアプリケーションに関連するメタデータとともに文脈情報を用いて、当該アプリケーションが悪意のあるものか、安全なものかについての最終的に判断する方法が提供される。この方法は、以下のステップを有する。第１に、懸案のアプリケーションについて従来式の評価を行う。アプリケーションが悪意のあるものか、安全なものかについて最終的に決定された場合、ある提案とともに、この情報をクライエントコンピュータに送信することができる。アプリケーションの素性（特性）が不明である場合、アプリケーションに関する収集データおよび送信された文脈情報を解析する。１つの実施形態において、収集データおよび送信された文脈情報を機械学習システムの特徴ベクトルとして利用することができる場合、機械学習分類器の結果を利用することができる。こうしたコーパス内の具体例をラベル付けするためには、実行ファイル（executables）について、従来的な手法または手作業による解析を実施する必要がある場合がある。しかしながら、このプロセスは、処理を「ジャンプスタート（jumpstart）」させる手法として提案される。当業者に広く知られた数多くの手法で、学習コーパス構築のための具体例をラベル付けすることができる。あるファイルに関連して、十分な個数の特徴ベクトルがラベル付けされると、本願の実施形態に係る機械学習について上記説明したように、機械学習分類器に学習させることができる。その結果、新しい（分類されていない）事案（アプリケーション）に適応できる「モデル」が得られる。

【００９０】

別の実施形態によれば、アプリケーションが真偽不明であり、マシンが最新の感染履歴を有することを、収集データが示唆するものである場合、アプリケーションを悪意のあるものとみなすことができる。別の実施形態において、マシンのセキュリティ状態が危険に晒されていることを、文脈情報が示唆するものである場合、より攻撃的な（厳格な）検出機能を適用することができる。

【００９１】

これらの検出機能は、限定するものではないが、悪意のあるアプリケーションのジェネリック・フィンガープリント（これは脅威となる変形されたアプリケーションを捉えることができるが、偽陽性の可能性もより高くなるものである。）、ジェネリックな特徴に基づいて脅威を検出できる攻撃的な（厳格な）機械学習分類器、および悪意のある可能性が高いが、未だ精査されていないソフトウェアサンプルのフィンガプリントを含むものであってもよい。マシンのセキュリティ状態が危険に晒されていることを示す文脈情報には、限定するものではないが、以下のものを含んでいてもよい。すなわち、この文脈情報には、システム上における最近の感染、セキュリティ上の危険に晒されていたことが確認されたウェブサイトにアクセスしたこと（こうしたサイトを特定するリストおよびこうしたサイトを特定する技術が開示された発明に直交する場合）、およびピアツーピア・ファイル共有クライエントコンピュータ等の危険を含むと考えられているソフトウェア・アプリケーションをイントールしたことを含む。さらに、マシンが危険に晒されているという潜在的なリスクを有するか否かについて判断する上で、いくつかの文脈情報は有用である。こうした文脈情報は、限定するものではないが、以下のものを含む。すなわち文脈情報は、

既知のセキュリティ脆弱性を有するソフトウェア・アプリケーションの有無、およびシステムに脅威となるソフトウェア・アプリケーションをダウンロードさせようとする加害者による情報ルートとして用いられるウェブブラウザ等のソフトウェア・アプリケーションの有無が含まれる。別の実施形態において、銀行取引アプリケーション等のセキュリティ脆弱性アプリケーションがシステム上で実行中であることを文脈データが示す場合、遠隔的に疑いのあるものとみなされるとき、懸案のアプリケーションを一時的に停止するように提案することができる。これは、こうした状況下において、潜在的に危険に晒されたときのコストを考慮したとき、偽陽性のリスクを容認できることが前提となる。別の実施形態において、クライエントコンピュータが特定の地理的領域からアクセスしようとしていること、または特定の地理的領域において処理実行していることを文脈情報が示す場合、当該特定の地理的領域に関する脅威に関連する検出機能を適用することができる。たとえば、銀行トロイ（Bancos Trojan、銀行コンピュータシステムにおけるトロイの木馬）は、（とりわけブラジルの銀行アカウントに関する情報の窃取を目的とする）ブラジルにおけるユーザを標的にした脅威を与える既知のマルウェアである。保護の対象とされるコンピュータシステムがブラジルにある場合、銀行を特定するためのより厳格な技術を適用することができる。この技術は、たとえば専ら銀行を特定するための機械学習分類器であってもよい。関連する実施形態においては、ユーザが特定のウェブサイトにアクセスしたことを文脈情報が示す場合、そのウェブサイトに関連する脅威を特定するための攻撃的な技術を利用することができる。上記実施例において、ユーザが銀行トロイの標的リストと一致する銀行のウェブサイトにアクセスした場合には、銀行トロイの検出機能を適用することができる。同様のラインの中で、ユーザがフェイスブック等のサイトにアクセスした場合には、クーブフェイス・ワーム（Koobface worm）等の脅威に関する検出機能を適用することができる。

【実施例３】

【００９２】

この実施例は、１つの可能性のある実施形態を俯瞰することにより、本発明を明確にすることを支援するように、本発明に係る１つの態様を説明するものである。すなわち、この実施例は、本発明の範囲を制限するものと解すべきではない。

【００９３】

（本発明の一部である）エージェントソフトウェアは、クライエントコンピュータ（ラップトップコンピュータまたはデスクトップ・パーソナルコンピュータ）のシステム上で実行されるものである。このソフトウェアは、セキュリティ関連イベントの有無をモニタするものである。たとえばエージェントソフトウェアは、ファイルアクセスをモニタするマイクロソフト・ウィンドウズのミニフィルタ・ドライバ（mini-filter driver）を実行するものであってもよい。これは新しいファイルがファイルシステムに作成されるときは常に、エージェントソフトウェアは、そのファイルを解析し、従来式の技術を用いて（たとえばブラックリストに含まれるか否かにより）悪意のあるものか否かを確認するものである。このプロセスは、任意の場所にあるホストコンピュータによる遠隔操作サービス（たとえば「クラウド型」サービス）に対する問い合わせ（クエリ）により実行することができる。

【００９４】

他方、このような問い合わせ（クエリ）を受信したとき、アプリケーションが悪意のあるものか否かについて判断するためのいくつかの方法を適用することができる。これらの方法は、ブラックリストを用いたアプローチと同様、ヒューリスティックな（発見的な）アプローチを含むものであってもよい。あるファイルが最終的に悪意のあるものと（その他の証拠を必要とすることなく）判断された場合、その結果をクライエントコンピュータに返信してもよい（また将来のプロセスのためにトランザクション（変更処理）を記録してもよい。）。

【００９５】

あるファイルが最終的に悪意のあるものでないと判断されたが、（悪意のあるものであ

る確率が７０％という経験則（ヒューリスティック）に基づいて）依然として疑わしい場合、追加的な文脈情報が精査される。たとえば、このファイルを含むシステムが、クライエントコンピュータ共有ピアツーピア・ファイルを最近インストールし、最近日中に最終的に悪意のある３つのファイルが存在すると判断された場合、疑わしいものとして処理したにも拘わらず、この新たなファイルを最終的には悪意のあるものとして認識（ラベル付け）してもよい。

【００９６】

主たるアイデアは、システム上で起こった最近の感染を利用して（梃として）、基準の底上げを支援することである。この場合の判断基準ルールは、きわめて単純である（悪意のある３つのファイルに最近感染したこと、およびクライエントコンピュータ共有ピアツーピア・ファイルを最近インストールしたことを判断基準ルールとする。）。ただし、より洗練された判断基準ルールを適用してもよい。さらに機械学習技術を用いて、判断基準ルール（または判断基準ルールを効率的に符号化するモデル）を設定してもよい。

【００９７】

複合的な実施形態

本発明に係る複合的な実施形態によれば、２つまたはそれ以上の上記説明した実施形態を、クライエントコンピュータ（クライエントアプリケーション）およびサーバーコンピュータ（サーバーアプリケーション）のいずれか一方または両方で一体として、または独立して実施する。換言すると、ａ）ジェネリックシグネチャ、ｂ）文脈上の確からしさ（悪意ありと判断する確信度）、ｃ）機械学習によるモデルのうちの２つまたはそれ以上を適用して、ソフトウェア・アプリケーションが悪意のあるものか否かを判断する。この実施形態によれば、クライエントコンピュータは、（ｉ）ソフトウェア・アプリケーションから特徴ベクトルを抽出するステップと、（ii）ソフトウェア・アプリケーションに関するメタテータを抽出し、ソフトウェア・アプリケーションがインストールされる可能性のあるシステムに関する文脈情報を収集するステップと、（iii）ソフトウェア・アプリケーションのジェネリック・フィンガープリント値を計算するステップのうちの２つまたはそれ以上のステップを実行し、得られたデータに関する情報をサーバーコンピュータに送信してもよい。サーバーコンピュータは、上記情報を処理した後、真偽判断または関連情報をクライエントコンピュータに返信し、クライエントコンピュータは、サーバーコンピュータから受信した情報に基づいて、ソフトウェア・アプリケーションに対する処理を行ってもよい。

【００９８】

これに応じて、サーバーコンピュータは、（ｉ）ソフトウェア・アプリケーションから得た特徴ベクトル、（ii）ソフトウェア・アプリケーションに関するメタデータ、およびソフトウェア・アプリケーションがインストールされる可能性のあるシステムに関する文脈情報、ならびに（iii）ソフトウェア・アプリケーションのジェネリック・フィンガープリント値のうちの２つまたはそれ以上を、クライエントコンピュータから受信するものであってもよい。クライエントコンピュータから特徴ベクトル情報を受信した場合、サーバーコンピュータは、機械学習による分類アルゴリズムを特徴ベクトルに適用する。ソフトウェア・アプリケーションに関するメタデータ、およびクライエントシステムに関する文脈情報を受信した場合、サーバーコンピュータは、このデータ情報を精査する。ソフトウェア・アプリケーションのジェネリックシグネチャを受信した場合、サーバーコンピュータは、そのジェネリックシグネチャが悪意のものあるものとみなすべきか否かを判断する。サーバーコンピュータは、１つまたはそれ以上の上記評価に基づいて、ソフトウェア・アプリケーションが悪意のものあるものとみなすべきか否かについて判断を行い、ソフトウェア・アプリケーションが悪意のものあるものとみなすべきか否かについての判断に関する情報を、クライエントコンピュータに送信する。

【００９９】

当業者が上記説明を読んだ後においては、本発明に係る数多くの変形例および変更例が明らかとなるところ、図示され、記載された特定の各実施形態は説明を目的として記載さ

れたものであり、限定しようとする意図はまったくないことを理解されたい。

【０１００】

図１５は、上記説明した１つまたはそれ以上の処理手続を実行する例示的なコンピュータシステムに関するブロック図である。図１５を参照すると、コンピュータシステムは、例示的なクライエントコンピュータシステムまたはサーバーコンピュータシステムを備えたものである。コンピュータシステムは、通信メカニズムまたは情報を通信するためのバスと、情報を処理するためのバスに接続されたプロセッサとを有する。プロセッサは、マイクロプロセッサを含むが、ペンティアム（Pentium）、パワーＰＣ（PowerPC）、アルファ等のマイクロプロセッサに限定されるものではない。このシステムは、（メインメモリと呼ばれる）ランダム・アクセス・メモリ（ＲＡＭ）や、バスに接続され、プロセッサで実行された情報および指令を記憶するためのその他の動的記憶デバイスを備える。メインメモリは、プロセッサが実行している間または指令を出している間、一時的な変数またはその他の中間的な情報を記憶するために用いられる。

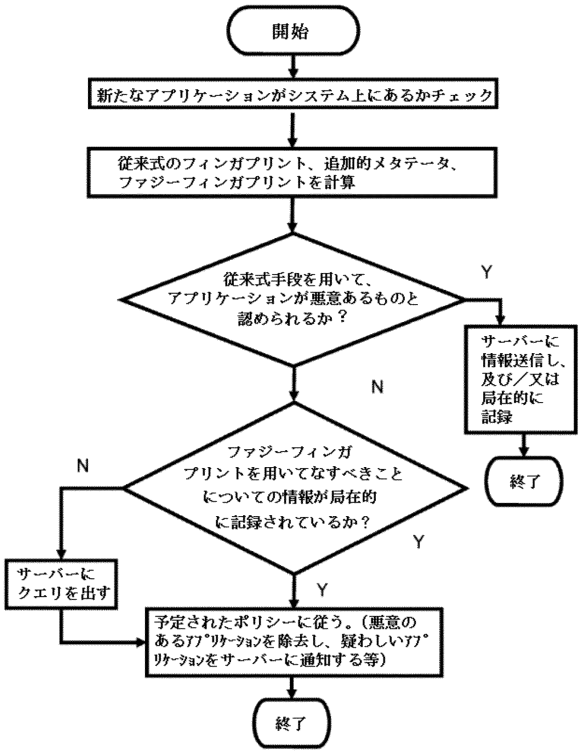
【０１０１】

コンピュータシステムは、リード・オンリー・メモリ（ＲＯＭ）、および／または磁気ディスクもしくは光ディスクや対応するディスクドライブ等、バスに接続され、プロセッサのための情報または指令を記憶するための静的記憶デバイスを備える。データ記憶デバイスは、情報または指令を記憶するために、バスに接続されている。コンピュータシステムは、コンピュータのユーザに情報を表示するためにバスに接続されたブラウン管（ＣＲＴ）や液晶ディスプレイ（ＬＣＤ）等のディスプレイデバイスをさらに備えている。英数字キーおよびその他の複数のキーを含む英数字入力デバイスが、同様にバスに接続され、情報および命令選択をプロセッサに送信することができる。追加的なユーザ入力デバイスとして、マウス、トラックボール、トラックパッド、スタイラス、カーソル矢印キー等のカーソル制御デバイスがバスに接続され、方向情報、命令選択をプロセッサに送信し、ディスプレイ上のカーソル移動を制御することができる。バスに接続され得る別のデバイスは、指令、データ、またはその他の情報を、紙やフィルム、同様のタイプの媒体等の媒体上に印刷するためのハードコピーデバイスである。さらに、スピーカおよび／またはマイクロフォン等の録音再生デバイスを任意的にバスに接続して、コンピュータシステムとの音響インターフェイスを実現することができる。バスに接続することができる別のデバイスは、電話や携帯パームトップデバイスと通信するための優先／無線の通信機能デバイスである。

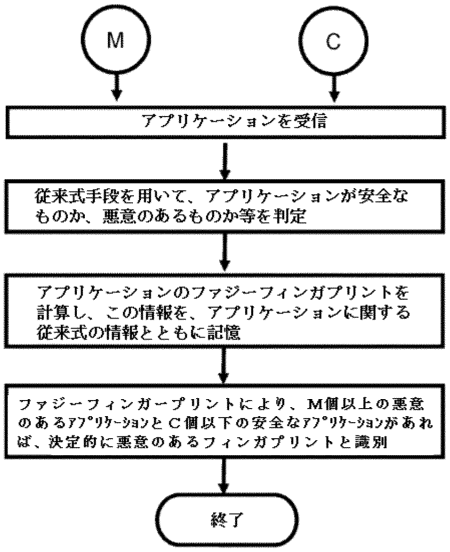
【０１０２】

本発明において、このシステムの任意のまたはすべての構成部品および付随するハードウェアを採用してもよい。しかし、コンピュータシステムの他の構成は、いくつかの、またはすべてのデバイスを有するものであってもよい。

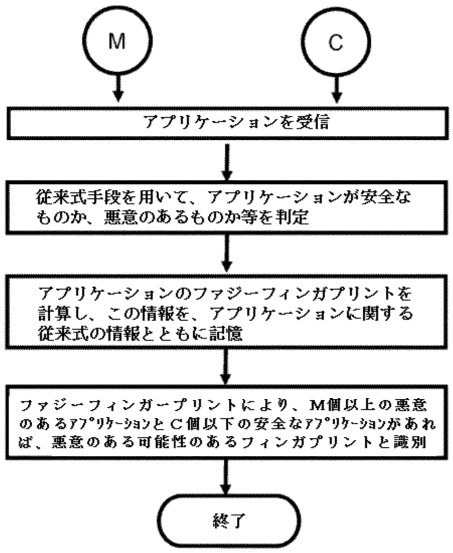
【図１】


開始
新たなアプリケーションがシステム上にあるかチェック
従来式のフィンガプリント、追加的メタデータ、ファジーフィンガプリントを計算
従来式手段を用いて、アプリケーションが悪意あるものと認められるか？
Y
サーバーに情報送信し、及び／又は局在的に記録
終了
N
ファジーフィンガプリントを用いてなすべきことについての情報が局在的に記録されているか？
N
Y
サーバーにクエリを出す
Y
予定されたポリシーに従う。（悪意のあるｱﾌﾟﾘｹｰｼｮﾝを除去し、疑わしいｱﾌﾟﾘｹｰｼｮﾝをサーバーに通知する等）
終了

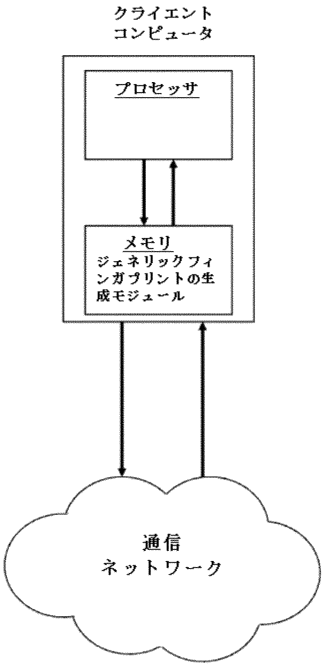
【図２】


M
C
アプリケーションを受信
従来式手段を用いて、アプリケーションが安全なものか、悪意のあるものか等を判定
アプリケーションのファジーフィンガプリントを計算し、この情報を、アプリケーションに関する従来式の情報とともに記憶
ファジーフィンガープリントにより、Ｍ個以上の悪意のあるｱﾌﾟﾘｹｰｼｮﾝとＣ個以下の安全なｱﾌﾟﾘｹｰｼｮﾝがあれば、決定的に悪意のあるフィンガプリントと識別
終了

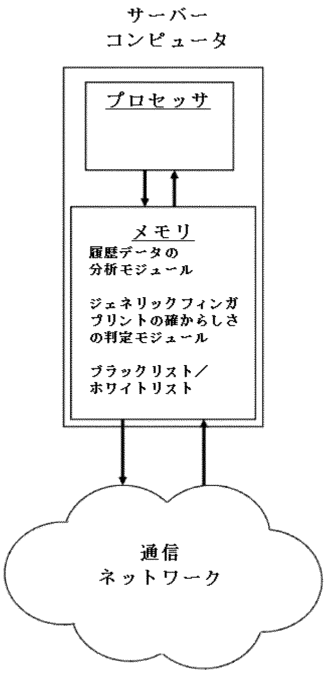
【図３】


M
C
アプリケーションを受信
従来式手段を用いて、アプリケーションが安全なものか、悪意のあるものか等を判定
アプリケーションのファジーフィンガプリントを計算し、この情報を、アプリケーションに関する従来式の情報とともに記憶
ファジーフィンガープリントにより、Ｍ個以上の悪意のあるｱﾌﾟﾘｹｰｼｮﾝとＣ個以下の安全なｱﾌﾟﾘｹｰｼｮﾝがあれば、悪意のある可能性のあるフィンガプリントと識別
終了

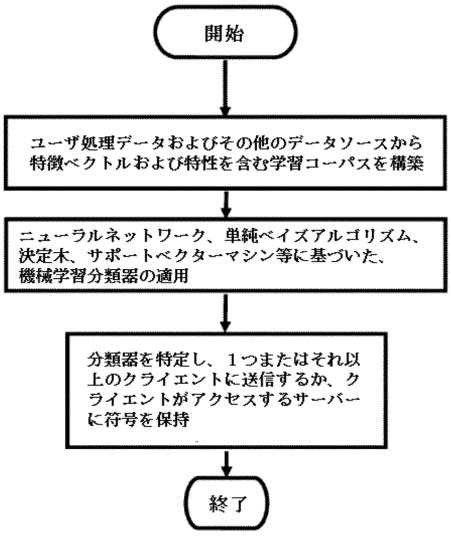
【図４】


クライエント
コンピュータ
プロセッサ
メモリ
ジェネリックフィンガプリントの生成モジュール
通信
ネットワーク

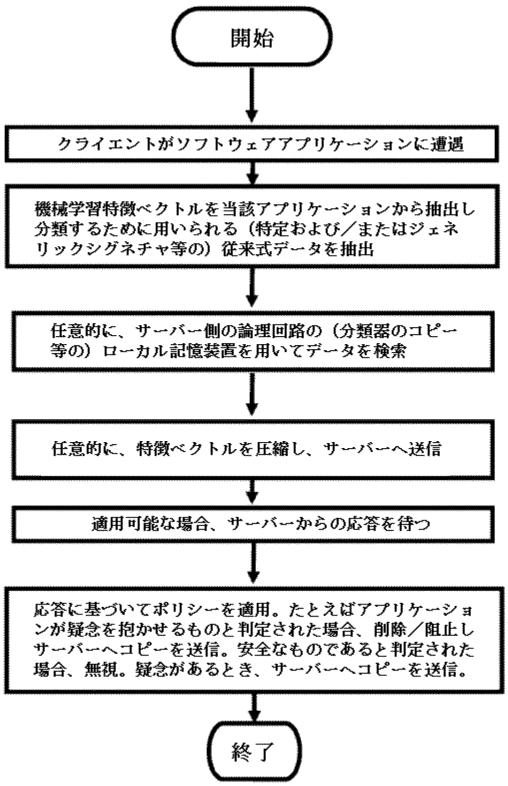
【図５】


サーバー
コンピュータ
プロセッサ
メモリ
履歴データの
分析モジュール
ジェネリックフィンガ
プリントの確からしさ
の判定モジュール
ブラックリスト／
ホワイトリスト
通信
ネットワーク

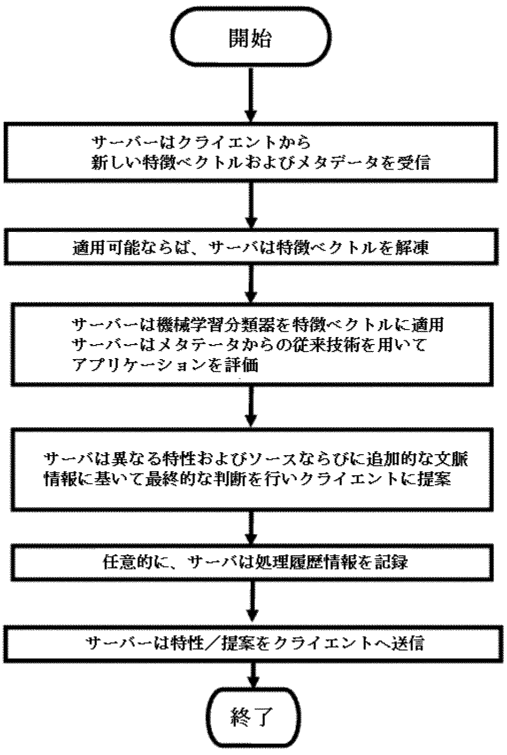
【図６】


開始
ユーザ処理データおよびその他のデータソースから
特徴ベクトルおよび特性を含む学習コーパスを構築
ニューラルネットワーク、単純ベイズアルゴリズム、
決定木、サポートベクターマシン等に基づいた、
機械学習分類器の適用
分類器を特定し、１つまたはそれ以
上のクライエントに送信するか、ク
ライエントがアクセスするサーバー
に符号を保持
終了

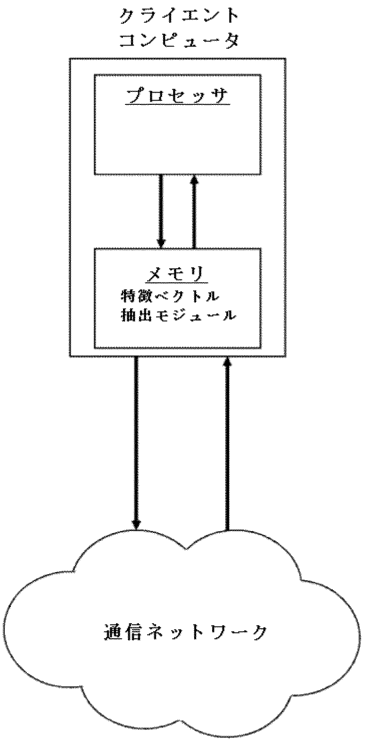
【図７】


開始
クライエントがソフトウェアアプリケーションに遭遇
機械学習特徴ベクトルを当該アプリケーションから抽出し
分類するために用いられる（特定および／またはジェネ
リックシグネチャ等の）従来式データを抽出
任意的に、サーバー側の論理回路の（分類器のコピー
等の）ローカル記憶装置を用いてデータを検索
任意的に、特徴ベクトルを圧縮し、サーバーへ送信
適用可能な場合、サーバーからの応答を待つ
応答に基づいてポリシーを適用。たとえばアプリケーショ
ンが疑念を抱かせるものと判定された場合、削除／阻止し
サーバーへコピーを送信。安全なものであると判定された
場合、無視。疑念があるとき、サーバーへコピーを送信。
終了

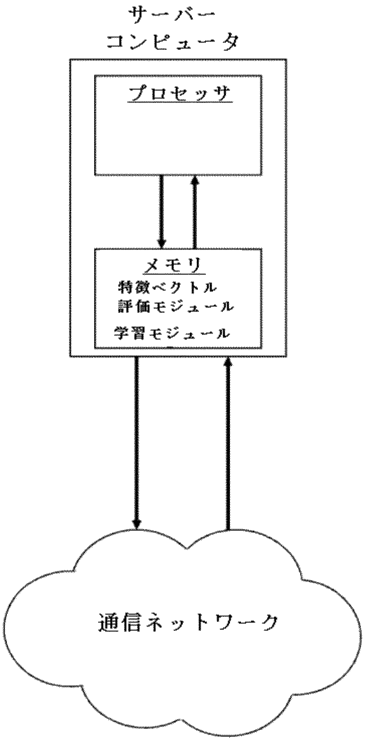
【図８】


開始
サーバーはクライエントから
新しい特徴ベクトルおよびメタデータを受信
適用可能ならば、サーバは特徴ベクトルを解凍
サーバーは機械学習分類器を特徴ベクトルに適用
サーバーはメタテータからの従来技術を用いて
アプリケーションを評価
サーバは異なる特性およびソースならびに追加的な文脈
情報に基いて最終的な判断を行いクライエントに提案
任意的に、サーバは処理履歴情報を記録
サーバーは特性／提案をクライエントへ送信
終了

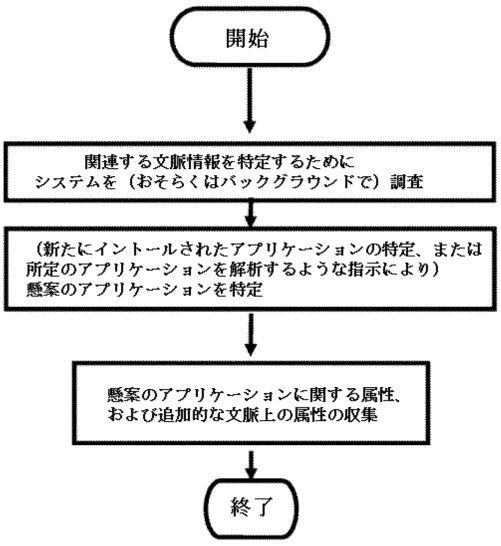
【図９】



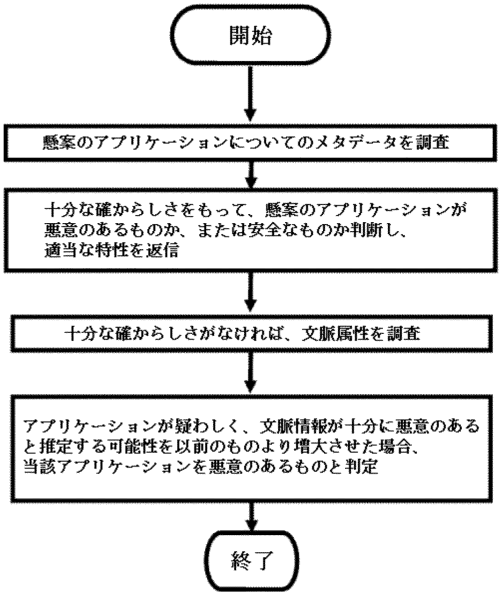
【図１０】



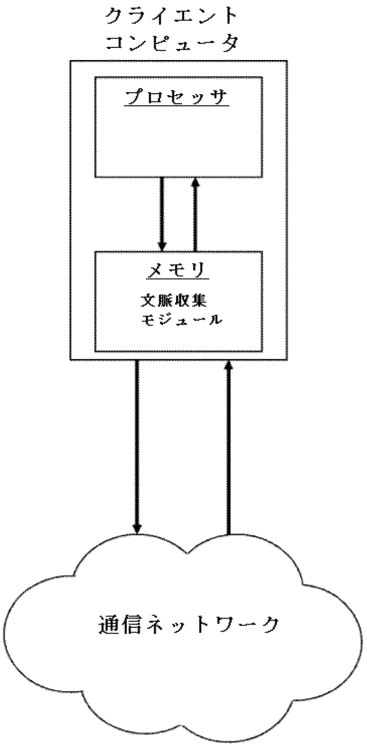
【図１１】



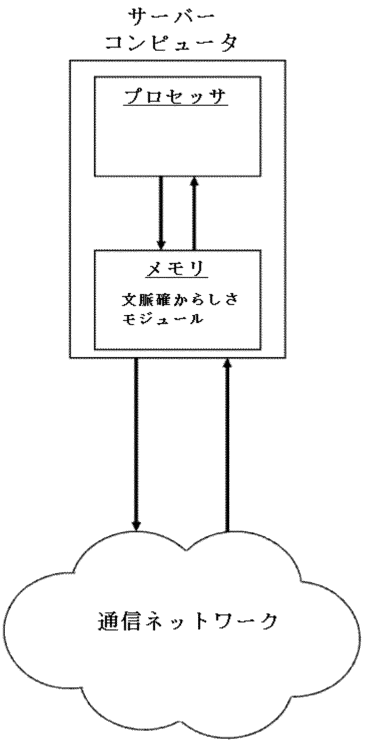
【図１２】



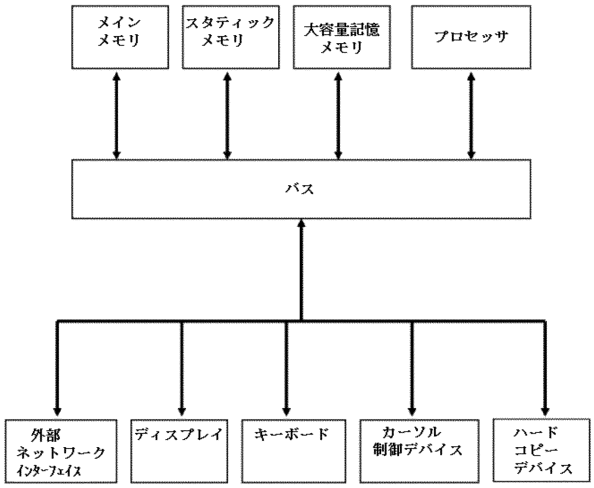
【図１３】


クライエント
コンピュータ
プロセッサ
メモリ
文脈収集
モジュール
通信ネットワーク

【図１４】


サーバー
コンピュータ
プロセッサ
メモリ
文脈確からしさ
モジュール
通信ネットワーク

【図１５】


メイン
メモリ
スタティック
メモリ
大容量記憶
メモリ
プロセッサ
バス
外部
ネットワーク
インターフェイス
ディスプレイ
キーボード
カーソル
制御デバイス
ハード
コピー
デバイス

フロントページの続き

(72)発明者　オリバー・フリードリックス
アメリカ合衆国２１０４６メリーランド州コロンビア、パタクセント・ウッズ・ドライブ９７７０番
(72)発明者　アルフレッド・ヒューガー
アメリカ合衆国２１０４６メリーランド州コロンビア、パタクセント・ウッズ・ドライブ９７７０番
(72)発明者　アダム・ジェイ・オドネル
アメリカ合衆国２１０４６メリーランド州コロンビア、パタクセント・ウッズ・ドライブ９７７０番

審査官　宮司　卓佳

(56)参考文献　米国特許出願公開第２０１０／０１６９９７２（ＵＳ，Ａ１）
米国特許出願公開第２００９／０３００７６１（ＵＳ，Ａ１）
米国特許出願公開第２００７／００１６９５３（ＵＳ，Ａ１）
米国特許出願公開第２０１０／０１６２３９５（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／５６