

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-186687
(P2003-186687A)

(43) 公開日 平成15年7月4日 (2003.7.4)

(51) Int.Cl.⁷

識別記号

F I

テ-マコ-ト* (参考)

G 0 6 F 11/00

G 0 6 F 9/06

6 6 0 N 5 B 0 7 6

審査請求 未請求 請求項の数13 OL (全 12 頁)

(21) 出願番号 特願2001-382592(P2001-382592)

(22) 出願日 平成13年12月17日 (2001. 12. 17)

特許法第30条第1項適用申請有り 平成13年10月31日～11月2日 社団法人情報処理学会主催の「コンピュータセキュリティシンポジウム2001」において文書をもって発表

(71) 出願人 593165487

学校法人金沢工業大学

石川県石川郡野々市町扇が丘7番1号

(72) 発明者 服部 進実

石川県石川郡野々市町扇が丘7番1号 学

校法人金沢工業大学内

(72) 発明者 千石 靖

石川県石川郡野々市町扇が丘7番1号 学

校法人金沢工業大学内

(74) 代理人 100105924

弁理士 森下 賢樹

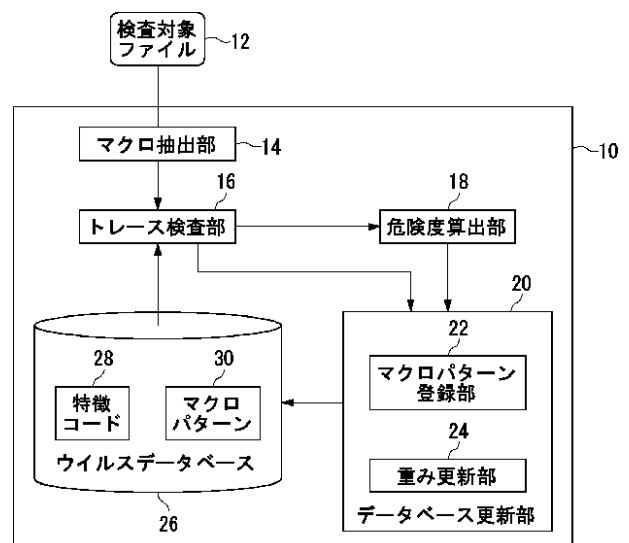
Fターム(参考) 5B076 FD08

(54) 【発明の名称】 ウイルス検出方法および装置

(57) 【要約】

【課題】 新種または変種のコンピュータウイルスを検出することは難しかった。

【解決手段】 マクロ抽出部14は、検査対象ファイル12からマクロ情報を抜き出し、ソースコードに再構築する。トレース検査部16は再構築されたマクロのソースコードを一行ずつトレースして、ウイルスの特徴コードが含まれていないかどうか検査する。危険度算出部18は、トレース検査部16により抽出された特徴コードの重みにもとづいて、マクロの危険度を算出し、マクロがウイルスであるかどうかを判定する。ウイルスが検出された場合、データベース更新部20の重み更新部24はトレース検査部16により抽出された特徴コードに関して、ウイルスデータベース26の特徴コードレコード28の重みを更新する。



【特許請求の範囲】

【請求項 1】 ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけてデータベースに登録する工程と、

検査対象ファイルをトレースして、前記データベースに登録された前記特徴コードを収集する工程と、前記収集された特徴コードの組み合わせにもとづいて、各特徴コードに関連づけられた前記重みを評価して、前記検査対象ファイルの危険度を算出する工程とを含むことを特徴とするウイルス検出方法。

【請求項 2】 前記収集された特徴コードに関して、前記データベースに格納された前記重みを更新する工程をさらに含むことを特徴とする請求項 1 に記載のウイルス検出方法。

【請求項 3】 前記重みの更新は、前記検査対象ファイルにウイルスが検出された場合になされることを特徴とする請求項 2 に記載のウイルス検出方法。

【請求項 4】 前記ウイルスの検出は、前記危険度にもとづいてなされることを特徴とする請求項 3 に記載のウイルス検出方法。

【請求項 5】 前記収集された特徴コードを階層的なレベルに分類した上でレベルによって前記重みの評価を異ならせて前記危険度を算出することを特徴とする請求項 1 に記載のウイルス検出方法。

【請求項 6】 ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけて格納したデータベースと、検査対象ファイルをトレースして、前記データベースに登録された前記特徴コードを収集し、収集した特徴コードの組み合わせを動作パターンとして特定する検査部と、前記収集した特徴コードの組み合わせと前記重みにもとづいて前記動作パターンの危険度を算出する危険度算出部とを含むことを特徴とするウイルス検査装置。

【請求項 7】 前記危険度に応じて、前記動作パターンを構成する前記特徴コードに関して、前記データベースに格納された前記重みを更新する更新部をさらに含むことを特徴とする請求項 6 に記載のウイルス検査装置。

【請求項 8】 前記データベースは、ウイルスの動作パターンを格納し、前記更新部は、前記特定された動作パターンを前記データベースに格納されたウイルスの動作パターンと比較して、その類似度に応じて、前記特定された動作パターンを構成する前記特徴コードの重みを更新することを特徴とする請求項 7 に記載のウイルス検査装置。

【請求項 9】 前記検査部は、前記収集した特徴コードを処理ルーチン単位とプリミティブなコマンド単位とに階層分けして分類し、前記危険度算出部は前記階層によって前記重みの評価を異ならせて前記危険度を算出することを特徴とする請求項 6 に記載のウイルス検査装置。

【請求項 10】 前記危険度算出部は、前記収集した特徴コードを命令と操作対象の種別により区別し、その命令と操作対象の種別の組み合わせによって前記重みの評価を異ならせて前記危険度を算出することを特徴とする請求項 6 または 9 に記載のウイルス検査装置。

【請求項 11】 ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけて登録したデータベースを参照して、検査対象ファイルから前記データベースに登録された前記特徴コードを収集する工程と、前記収集された特徴コードの組み合わせにもとづいて、各特徴コードに関連づけられた前記重みを評価して、前記検査対象ファイルの危険度を算出する工程とをコンピュータに実行させることを特徴とするコンピュータプログラム。

【請求項 12】 マクロウイルス特有の動作をとるために必要なコードと、そのコードがウイルスとして使用された場合の危険性を示す値を一つのレコードとしてデータベースに登録する工程と、マクロを一行ずつトレースし、前記登録したコードを収集し、コードの組み合わせを記録する工程と、収集したコードの組み合わせから、各重みを基にしてウイルスであるかどうかの判断値となる危険度を算出する工程と、危険度が基準となる値を超えたとき、そのマクロはウイルスであると判断する工程と、ウイルスを検出した場合、収集したコードを基に重みを増減させて前記データベースを更新する工程とを含むことを特徴とするウイルス検出方法。

【請求項 13】 前記更新する工程は、過去に検出したウイルスとパターン比較を行い、類似点が多ければ、収集したコードの重みを増加させることを特徴とする請求項 12 に記載のウイルス検出方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】この発明はコンピュータウイルスの検出技術、とくにコンピュータのプログラムファイルやデータファイルに感染するウイルスを検出する方法、装置およびシステムに関する。

【0002】

【従来技術】情報処理振興事業協会（IPA）が公表している国内のウイルス被害届出状況によると、1997年を境に急激に被害が増加しだしたことがわかる。1990年から1996年の間では、年間1,000件を越した年はわずか一年だけであったのに対し、1997年からは2,000件を下回ることは無くなり、2000年には11,109件と加速的に増加している。

【0003】このようにウイルス被害件数が急増してきた要因は、パソコンやネットワークが広く普及したことによる感染機会の増加や、個人や企業のウイルス対策意識が低いことその他、ウイルス対策の技術的な面にもその

要因があるといえる。なぜなら、1997年の被害の急増はマクロウイルスという新たな種のウイルスの出現によるものであり、近年突発的に被害が大きく増えたのも、VBS（ビジュアルベーシック（商標）スクリプト）ウイルスやWindows（商標）ウイルスなどの新種の出現によるものだからである。

【0004】

【発明が解決しようとする課題】このように、従来のウイルス対策は非常に新種ウイルスに対して弱いことを意味しており、これはウイルス検出方法に問題があるといえる。従来のウイルス検出方法は、過去に発見されたウイルスに固有のコードとのパターンマッチングに依存しており、新たなウイルスが出現するたびにウイルス定義ファイルを更新する手間があるため、ユーザの対応に遅れが生じる。

【0005】本発明はこうした状況に鑑みてなされたものであり、新種ウイルスまたは変種ウイルスを効果的に検出するウイルス検出技術を提供することを目的とする。

【0006】

【課題を解決するための手段】本発明のある態様はウイルス検出方法に関する。この方法は、ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけてデータベースに登録する工程と、検査対象ファイルをトレースして、前記データベースに登録された前記特徴コードを収集する工程と、前記収集された特徴コードの組み合わせにもとづいて、各特徴コードに関連づけられた前記重みを評価して、前記検査対象ファイルの危険度を算出する工程とを含む。ここで検査対象ファイルはプログラムファイルと、文書やマクロなどのデータファイルとを含む。

【0007】前記収集された特徴コードに関して、前記データベースに格納された前記重みを更新する工程をさらに含んでもよい。前記重みの更新は、前記検査対象ファイルにウイルスが検出された場合になされてもよい。前記検査対象ファイルにウイルスが含まれるかどうかの判定を外部から与えてもよい。また前記ウイルスの検出は、前記危険度にもとづいてなされてもよい。たとえば危険度が所定の基準値を超えた場合に、前記検査対象ファイルにウイルスが含まれると判定してもよい。

【0008】前記収集された特徴コードを階層的なレベルに分類した上でレベルによって前記重みの評価を異ならせて前記危険度を算出してもよい。階層的なレベルは、たとえばモジュール、サブルーチン、命令コード、およびオペランドといった処理コードの階層構造のレベルであり、モジュール、サブルーチンのような処理ルーチン単位レベルと、命令コード、オペランドのような下位のプリミティブなコマンドレベルとで重みの評価の仕方を異ならせ、その評価の組み合わせで前記危険度を算出してもよい。

【0009】本発明の別の態様はウイルス検査装置に関する。この装置は、ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけて格納したデータベースと、検査対象ファイルをトレースして、前記データベースに登録された前記特徴コードを収集し、収集した特徴コードの組み合わせを動作パターンとして特定する検査部と、前記収集した特徴コードの組み合わせと前記重みにもとづいて前記動作パターンの危険度を算出する危険度算出部とを含む。

【0010】前記危険度に応じて、前記動作パターンを構成する前記特徴コードに関して、前記データベースに格納された前記重みを更新する更新部をさらに含んでもよい。たとえば前記危険度が所定の基準値以上である場合に、前記動作パターンをウイルスと判定して、前記重みを更新するが、前記危険度が基準値に満たない場合には、前記重みを更新しないようにしてもよい。

【0011】前記データベースは、ウイルスの動作パターンを格納し、前記更新部は、前記特定された動作パターンを前記データベースに格納されたウイルスの動作パターンと比較して、その類似度に応じて、前記特定された動作パターンを構成する前記特徴コードの重みを更新してもよい。また前記特定された動作パターンの危険度が基準値以上である場合、この動作パターンを新たなウイルスの動作パターンとして前記データベースに格納してもよい。

【0012】前記検査部は、前記収集した特徴コードを処理ルーチン単位とプリミティブなコマンド単位とに階層分けして分類し、前記危険度算出部は前記階層によって前記重みの評価を異ならせて前記危険度を算出してもよい。また前記危険度算出部は、前記収集した特徴コードを命令と操作対象の種別により区別し、その命令と操作対象の種別の組み合わせによって前記重みの評価を異ならせて前記危険度を算出してもよい。たとえば、ファイルの自動オープンの命令のように、命令自体にリスクを伴うものや、システムフォルダ内のファイル、テンプレートファイル、実行形式のファイルなどに対するアクションのように、操作対象にリスクを伴うものがあり、命令と操作対象の種別の組み合わせによって重みの評価の仕方を変えてもよい。

【0013】本発明のさらに別の態様はコンピュータプログラムに関する。このプログラムは、ウイルス特有の動作に係る特徴コードに危険性を示す重みを関連づけて登録したデータベースを参照して、検査対象ファイルから前記データベースに登録された前記特徴コードを収集する工程と、前記収集された特徴コードの組み合わせにもとづいて、各特徴コードに関連づけられた前記重みを評価して、前記検査対象ファイルの危険度を算出する工程とをコンピュータに実行させる。

【0014】前記データベースはサーバに設けられ、ネットワークを介して前記データベースを参照してもよ

い。また当該プログラムがインストールされたユーザ端末と前記サーバを含むシステムが構成されてもよい。また当該プログラムはネットワーク上のユーザ端末を巡回してウイルス検査を行うモバイルエージェントとして構成されてもよい。

【0015】本発明のさらに別の態様もウイルス検出方法に関する。この方法は、マクロウイルス特有の動作をとるために必要なコードと、そのコードがウイルスとして使用された場合の危険性を示す値を一つのレコードとしてデータベースに登録する工程と、マクロを一行ずつトレースし、前記登録したコードを収集し、コードの組み合わせを記録する工程と、収集したコードの組み合わせから、各重みを基にしてウイルスであるかどうかの判断値となる危険度を算出する工程と、危険度が基準となる値を超えたとき、そのマクロはウイルスであると判断する工程と、ウイルスを検出した場合、収集したコードを基に重みを増減させて前記データベースを更新する工程とを含む。前記更新する工程は、過去に検出したウイルスとパターン比較を行い、類似点が多ければ、収集したコードの重みを増加させてもよい。

【0016】なお、以上の構成要素の任意の組み合わせ、本発明の表現を方法、装置、サーバ、システム、コンピュータプログラム、記録媒体などの中で変換したものもまた、本発明の態様として有効である。

【0017】

【発明の実施の形態】本発明では、新種ウイルスまたは変種ウイルスを効果的に検出することを目的とし、新種ウイルス検出に適すとされるヒューリスティック検査法の概念を取り入れたウイルス検出システムを提案、構築する。また、実施の形態では新種、変種ウイルスの多さと、この先感染機会が増えるであろうことからマクロウイルスを対象を絞っている。

【0018】まず検査対象としたマクロウイルスについて、マクロウイルスの特徴、感染メカニズム、およびマクロウイルス検出方法を述べる。

【0019】マクロウイルスは、文書ファイルなどのデータファイルに付加しているマクロ部分に感染するウイルスである。コンピュータで扱うファイルは大きく分けると「プログラムファイル」と「データファイル」の2種類があり、従来のウイルスはデータファイルには感染しないという定説があったがマクロウイルスの登場により覆されている。現在、マクロ機能を持つアプリケーションプログラムは多数存在するが、中でもシェアが大きく、マクロ機能に制限が少なく安全性が低いものはウイルス製作者の標的にされている。また、OLE (Object Linking and Embedding) オートメーション機能やDDE (Dynamic Data Exchange) 機能を使って、異なるアプリケーションプログラム間の感染が可能であり、これにより感染力の非常に強いウイルスを出現させる結果となっている。

【0020】マクロウイルスはその作成が非常に容易であるということも重要な点である。従来のウイルスと比較するとはるかに多くの新種、変種が存在するのはこのためである。従来のウイルスの作成には、アセンブラなどの低級言語によるプログラミングやOSの知識などが必要であったが、マクロウイルスの場合は、マクロ言語そのものかビジュアルベーシック (Visual Basic) (商標) についての知識をある程度持っていれば十分である。

【0021】マクロウイルス自体はデータファイルであるため、マクロを動作させるアプリケーションを介して感染や発病をする。以下にマクロウイルスが感染を広げていく流れを示す。

(1) 電子メールやダウンロードにより外部からコンピュータにウイルスが侵入する。

(2) ファイルを開くためにアプリケーションプログラムを起動する。

(3) アプリケーションプログラムがファイルを読み込みマクロを実行する。

(4) ウイルスマクロが実行されると、標準テンプレートファイルに自分自身のコピーを書込む。

(5) その後、そのアプリケーションプログラムに読み込まれたファイルには、ウイルスに感染したテンプレートが適応され、そのファイルのマクロもウイルスとなる。

【0022】先に述べたように、マクロウイルスには非常に多くの変種ウイルスが存在しており、従来の単にコードを比較するだけのウイルス検出方法では、全ての変種ウイルスを捕らえることは困難である。

【0023】そこで、本実施の形態ではマクロウイルスが感染を広げるために標準テンプレートに必ず自分自身のコピーを書込むことなど、マクロウイルスが特有の動作パターンを持つことに着目し、これらの動作パターンを捕らえることでマクロウイルスを検出する。また本実施の形態では、このマクロウイルス特有の動作パターンを捕らえるために、これから述べるヒューリスティック検査法の概念を取り入れている。

【0024】ヒューリスティック検査法というのは、検査対象のプログラムの中からコンピュータウイルス特有の動作パターンに必要なコードを収集し、そのプログラム中にどれだけそのコードが含まれるか、そのコードがウイルスである可能性はどれだけあるかを、ウイルス検出システム自身に試行錯誤させてウイルスであるかどうかを判断させる技術である。

【0025】この技術を使えば従来のコード比較によるウイルス検出とは違い、新種ウイルスが出現する度にウイルス定義ファイルを更新する手間が無くなり、対応の遅れをなくすることができる。また、ウイルスの特性だけを取り上げるので、新種や変種ウイルスの検出には最適である。

【0026】本実施の形態に係るウイルス検出システム

ではヒューリスティック検査法概念を取り入れ以下のようにしてウイルス特有の動作パターンを捕らえ、ウイルスであるかどうかの判定をしている。

(1) マクロウイルス特有の動作をとるために必要なコードと、そのコードがウイルスとして使用された場合の危険性を示す値(以下、重みという)などを1レコードとしてデータベースに登録する。

(2) マクロを一行ずつトレースし、登録したコードを収集する。このときコードがどのように組み合わせさせていたかを記録する。

(3) 収集したコードの組み合わせから、各重みを基にしてウイルスであるかどうかの判断値となる危険度を算出する。

(4) 危険度が基準となる値を超えたとき、そのマクロはウイルスであると判断する。

(5) ウイルスを検出した場合、このとき収集したコードや、過去にウイルスを検出した状態を基に重みを増減させてデータベースを更新する。

【0027】図1に示すように、本ウイルス検出システムは4つの処理ルーチンで構成されている。各ルーチンの働きとそれらの連携について述べる。

【0028】1. マクロ抜き出し処理ルーチン

このルーチンでは、マクロを含むファイルからマクロのみを抜きだす処理を行う。マクロをアプリケーションを介さずに読むにはOLE2の複合ファイルの仕組みや構造化記憶について理解する必要がある。図2に示すように、OLE2の複合ファイルはファイル内の構造がファイルシステムようになっており、ディレクトリに相当するストレージとファイルに相当するストリームで構成される。マクロや特にVBA(Visual Basic For Application)を使用しているファイルでは、それを扱うストリームを容易に特定することができるので、これを利用してマクロを再構築する。

【0029】ストリームに格納されているマクロの情報は、命令や、変数名を文字列として格納するデータ群や、それをどのように配置するかを表すデータ群で構成されている。このままでは、次のマクロのトレース処理を効率よく行えないため、一度マクロをトレースしやすい形態に再構築する。図3は、マクロを構成する文字列データ群を格納するストリームの一部である。これを規則に従い再構築しテキストに書き出したのが図4である。

【0030】2. マクロトレース処理ルーチン

マクロを一行ずつトレースしながら、データベースに登録した特徴コードを検査していく。ここでいう特徴コードとは、ウイルスの自己伝染、発病、潜伏機能を実行するために使われるであろうコードであり、データベースにはこの特徴コードとそれに対応した重みを記録しておく。レコードの詳細については後述の学習手法で説明する。

【0031】特徴コードが発見された場合は、図5に示す4つのレベルに分類してそれぞれの重みを危険度算出ルーチンに送る。すなわちモジュールレベル、サブルーチンレベル、命令コードレベル、および引数(以下、オペランドともいう)レベルの4つの階層化されたレベルである。コードがどのレベルであるかの判断はマクロを再構築する際に得た情報を利用する。モジュールレベルとサブルーチンレベルでは、その名前によってウイルス活動のトリガーとしての働きをし、命令コードレベルと引数レベルでは、その組み合わせによって大きくウイルスである危険度が変わる。

【0032】3. 危険度算出ルーチン

危険度はそのままウイルスらしさを表し、基準値以上のものをウイルスであると判断する。危険度は、先のトレース処理(以下、トレース検査ともいう)で収集した重みを基に以下のような流れで算出する。

(1) まず、モジュールレベルの重みが得られるのでこれを記憶する。

(2) 次にサブルーチンレベルの重みが得られるので、ここでウイルス活動のトリガーになり得ないかをモジュールレベルの重みと合わせて算出し、その値をAとする。

(3) 命令コードレベルの重みを得たときに、引数レベルの重みが同時に得られていれば、その組み合わせからウイルスの行動に対する値Bを算出する。引数レベルの重みが無い場合は、命令コードレベルの重みをBとする。

(4) マクロをトレース中にサブルーチンの終了を検出した時点で、AとBの組み合わせから危険度を算出する。全てのサブルーチンについて危険度を算出し、一番高い危険度をそのマクロの危険度とする。

【0033】4. データベース更新ルーチン

このルーチンでは、データベースに登録した重みを増減させる。重みは、ウイルスであるかどうかを判定するための基になる重要な値であり、この重みを巧く増減させることで、ウイルス検出精度の向上が期待できる。

【0034】重みを増減させるタイミングは、ウイルスを検出した時とし、そのコードがウイルスに使われた頻度、時間間隔を考慮して増減させるが、重みの減少させすぎによる検出ミスは避ける必要がある。その対策としては重みの下限を設定し、重みを増加させる量よりも、減少させる量を少なくしている。

【0035】また、後述の学習手法で説明するが、過去にウイルスを検出した時のマクロのパターンもデータベースに登録するため、マクロトレース処理ルーチンから送られたコードも全てデータベースに登録する。

【0036】次に学習手法について詳細に説明する。本ウイルス検出システムでは、マクロウイルスの本質的な動作パターンを突き詰めていくことが必要不可欠であり、そのために過去にウイルスを検出した経験を反映さ

せる方法をとっている。また、ここでいう学習とは重みを適切な値に近づけるもので、データベースの自動更新がこれにあたる。

【0037】まずデータベースの構成を説明する。データベースは2つのテーブルからなり、1つは、特徴コードと重みの対を格納し、1つは過去に検出したウイルスマクロをパターン化したものを格納している。マクロのパターン化は、特徴コードに通し番号を付け、それを集めたものである。レコードの構成は次のようになっている。

【0038】特徴コード用テーブルのレコード

- (1) 通し番号
- (2) 基本値
- (3) 付加値
- (4) 特徴コードの文字列
- (5) 出現頻度

【0039】マクロパターン化用テーブルのレコード

- (1) タイムスタンプ
- (2) 通し番号列の文字列
- (3) 出現頻度

【0040】これまで述べてきた重みは基本値のことをいっており、付加値は過去の経験を反映させる重みであり、ここで述べる学習処理で使用する。通し番号列は、特徴コード用テーブルの通し番号を集めて文字列にしたものである。

【0041】次に図6を参照しながら学習手順を示す。

- (1) 基本値による危険度を前述のように算出する。このときマクロのパターン化も同時に行う。
- (2) 次に、過去に検出したウイルスマクロとパターン比較を行う。類似点が多ければ、過去に同じような動作をするウイルスを検出していることになる。また、まったく同じ動作をしている部分があればそれは、ウイルスの本質的な動作であるとみなすことができ、そのコードの付加値を増加させる。
- (3) すでにコンピュータ内にウイルスが広まっている場合も考慮し、全てが同じパターンである場合は付加値を増加させない。これは、特定のウイルスの影響を受けて重みが不適切に偏ることを防ぐためである。
- (4) 危険度に付加値を与え、新たな危険度を算出する。
- (5) 出現頻度の一番多いものを基準に、その差が一定値以上ついたコードの付加値を減少させる。

【0042】図7は、実施の形態に係るウイルス検出装置10の構成図である。この構成は、ハードウェア的には、任意のコンピュータのCPU、メモリ、その他のLSIで実現でき、ソフトウェア的にはメモリにロードされたウイルス検出機能のあるプログラムなどによって実現されるが、ここではそれらの連携によって実現される機能ブロックを描いている。したがって、これらの機能ブロックがハードウェアのみ、ソフトウェアのみ、また

はそれらの組み合わせによっていろいろな形で実現できることは、当業者には理解されるところである。

【0043】ウイルス検出装置10のマクロ抽出部14、トレース検査部16、危険度算出部18、およびデータベース更新部20は、ソフトウェア処理としては、それぞれ前述のウイルス検出システムにおけるマクロ抜き出し処理ルーチン、マクロトレース処理ルーチン、危険度算出ルーチン、およびデータベース更新ルーチンを実行するものである。

【0044】ウイルスデータベース26は、特徴コードレコード28とマクロパターンレコード30を格納している。特徴コードレコード28は、ウイルス特有の特徴コードに危険性を示す重みを関連づけたものであり、前述の特徴コード用テーブルのレコードである。マクロパターンレコード30は、特徴コードの組み合わせで構成されるウイルスのパターンを示すものであり、前述のマクロパターン化用テーブルのレコードである。

【0045】特徴コードは図5のように4つのレベルに分けられて重みづけされている。ファイルの自動オープンのようにその動作自体がリスクを伴うようなモジュールやサブルーチンの場合、モジュール名またはサブルーチン名から危険度を把握できる。またシステムフォルダ内のファイルの操作、テンプレートファイルの操作、実行形式のファイルのオープンなど操作対象の種別から危険度を把握できるものもある。したがって命令と操作対象を別々に登録しておいて、その組み合わせについて重みづけをすることが必要である。また特徴コードを階層的にレベル分けして、モジュールやサブルーチンといった大きな処理単位で危険性を判断するとともに、命令コードやオペランドといったよりプリミティブな単位で危険性を判断することが必要である。たとえばファイルの自動オープンを使用せずに、ダイアログを表示して、ユーザにダイアログのボタンをクリックさせることで実質的にファイルの自動オープンを行うことも可能である。したがってモジュール名やサブルーチン名だけに頼った危険性の判断だけでは不十分であり、より下位の命令コードやオペランドのレベルでの危険性の判断が要求される。

【0046】図8を参照して、ウイルス検出装置10によるウイルス検出手順の大まかな流れを説明する。マクロ抽出部14は、検査対象ファイル12を読込む(S10)。次にマクロ抽出部14は、検査対象ファイルからマクロ情報を抜き出し、マクロ情報をソースコードの状態に再構築する(S12)。トレース検査部16は再構築されたマクロのソースコードを一行ずつトレースして、ウイルスの特徴コードが含まれていないかどうか検査する(S14)。このときトレース検査部16は、特徴コードの組み合わせから構成される動作パターンを特定する。

【0047】危険度算出部18は、トレース検査部16

により抽出された特徴コードの重みにもとづいて、マクロの危険度を算出し、マクロがウイルスであるかどうかを判定する(S16)。ウイルス判定により、ウイルスが検出された場合(S18のY)、データベース更新部20は、トレース検査部16により特定されたマクロの動作パターンを構成する特徴コードに関して、ウイルスデータベース26の特徴コードレコード28の重みを更新する(S20)。ウイルスが検出されない場合(S18のN)、ウイルスデータベース26の更新は行わない。

【0048】図9はマクロの再構築処理S12の詳細な手順を示すフローチャートである。既に図3に示したように、マクロ情報は検査対象の文書ファイルのバイナリデータ中に散らばっており、このままではトレースが難しい。マクロ抽出部14は検査対象ファイルからマクロ情報を抽出して、図4のようなソースコードの形に組み上げる再構築の処理を行う。まず検査対象ファイルが構造化記憶を利用したものであるかどうかを調べる(S30)。構造化記憶を利用していない場合(S30のN)、ウイルスの検査を行わずに終了する。構造化記憶を利用している場合(S30のY)、検査対象ファイルにマクロが含まれるかどうかを調べる(S32)。マクロが含まれない場合(S32のN)、ウイルスの検査を行わずに終了する。マクロが含まれる場合(S32のY)、マクロ情報を組み上げ、ソースコードの状態にして書き出す(S34)。

【0049】図10はマクロのトレース検査処理S14の詳細な手順を示すフローチャートである。トレース検査部16は、再構築されたマクロのソースコードを一行ずつトレースしながら、ウイルスデータベース26を参照して、登録された特徴コードレコード28に一致するものがあるかどうか調べ、一致した特徴コードとその重みを収集する(S40)。またマクロのサブルーチンごとに、そのサブルーチンを構成する特徴コードの種類を動作パターンとして記憶する(S42)。

【0050】図11はウイルス判定処理S16の詳細な手順を示すフローチャートである。危険度算出部18は、トレース検査部16により抽出された特徴コードの重みにもとづいて、マクロのサブルーチンごとに危険度を算出する。トレース検査部16は、特徴コードを図5に示した4つのレベルに分類している。危険度算出部18は、モジュールレベルの重みMとサブルーチンレベルの重みSとを組み合わせ、ウイルス活動のトリガーになりうる危険度VTを算出する(S50)。たとえば、 $VT = \max(M, S)$ とする。次にサブルーチンの一つ一つの命令について、命令コードレベルの重みIとオペランドレベルの重みOとを組み合わせ、ウイルス活動に対する危険度VAを算出する(S52)。たとえば $VA = I \times O$ とする。

【0051】次にサブルーチンごとに、評価した重みの

階層レベルが異なる2種類の危険度VT、VAを組み合わせ、サブルーチンの危険度を算出する(S54)。たとえば、VTの値とサブルーチン内の最も大きいVAの値の和をそのサブルーチンの危険度とする。マクロのすべてのサブルーチンについてサブルーチンの危険度が算出されると、それらのサブルーチンの危険度の内、もっとも高い危険度をマクロの危険度とする(S56)。マクロの危険度が所定の基準値を越える場合(S58のY)、そのマクロはウイルスであると判定し(S60)、そうでない場合(S58のN)、そのマクロをウイルスとは判定しない。

【0052】図12はデータベース更新処理S20の詳細な手順を示すフローチャートである。データベース更新部20は、マクロパターン登録部22と重み更新部24を含む。マクロがウイルスと判定された場合に、マクロパターン登録部22はトレース検査部16により特定されたマクロの動作パターンを新たなマクロパターンレコード30としてウイルスデータベース26に登録し、重み更新部24はその動作パターンを構成する特徴コードに関して、特徴コードレコード28の重みを更新する。

【0053】学習アルゴリズムの基本方針は、多種のウイルスに共通する動作パターンはウイルスの本質的な動作とみなし、その動作パターンを構成する特徴コードの重みを増加させることである。たとえば、ウイルスA、B、Cの動作が以下であるとする。

【0054】(A)ウイルスAの動作：

(A-1)ファイルオープンをトリガーとする、(A-2)レジスタにXを書込む、(A-3)標準テンプレートファイルを書き換える。

【0055】(B)ウイルスBの動作：

(B-1)ファイルオープンをトリガーとする、(B-2)システムファイルを削除する、(B-3)標準テンプレートファイルを書き換える。

【0056】(C)ウイルスCの動作：

(C-1)アプリケーションの起動をトリガーとする、(C-2)ファイルYを改ざんする、(C-3)標準テンプレートファイルを書き換える。

【0057】このとき、「標準テンプレートファイルを書き換える」という動作パターンは3種のウイルスA、B、Cに共通しており、「ファイルオープンをトリガーとする」という動作パターンは2種のウイルスA、Bに共通している。そこで「標準テンプレートファイルを書き換える」という動作パターンを構成する特徴コードについては重みの増加量を大きくとり、「ファイルオープンをトリガーとする」という動作パターンを構成する特徴コードについては重みの増加量を小さくとる。

【0058】重みは基本値と付加値に分かれている。基本値は特徴コードに対して予想される危険度を示す固定の値であり、付加値は初期値をゼロとして、学習アルゴ

リズムにより更新される値である。ウイルスに頻繁に使用される特徴コードについては付加値が増加し、ほとんどウイルスに使用されない特徴コードについては付加値が減少する。特徴コードの危険度を示す重みは基本値と付加値の和で与えられるため、過去に発見されたウイルスの動作パターンによって重みが更新されることになる。

【0059】具体的な学習手順を説明する。発見されたウイルスの動作パターンについて、その動作パターンを構成する特徴コードの種類ごとに出現回数をカウントする(550)。たとえばウイルスの動作パターンが3種の特徴コードa、b、cから構成される場合、3種の特徴コードa、b、cのそれぞれの出現回数のカウント数を1だけインクリメントする。ウイルスのすべての動作パターンについて、その動作パターンに含まれる特徴コードの出願回数をこのようにカウントする。次に、特徴コードの付加値をこの出現回数のカウント数に比例して増加させる(552)。たとえば出現回数に所定の増加係数をかけた値を新たな付加値とする。これにより、多種のウイルスに共通する動作パターンの特徴コードほど付加値が大きくなる。

【0060】出現回数のカウント数が特定数以上である場合(554のY)、カウント数が極端に大きい特徴コードの付加値を減少させる(556)。たとえば特徴コード間の出現回数の差に比例して付加値を減少させる。すなわち出現回数の差に所定の減少係数をかけた値を付加値から差し引く。付加値の減少は、特定のウイルスの影響を受けて重みが偏るのを防ぐために行われる。

【0061】以上述べたように、実施の形態では検査対象にマクロウイルスを取り上げ、ヒューリスティック検査法の概念を取り入れた新種ウイルスに有効な検出方法を提案した。マクロウイルスに特有のコードをデータベース化し、その危険性を学習するシステムであるため、単純なコードのパターンマッチングによるウイルス検出の方法とは違って、既存ウイルスの一部を変えた変形型のウイルスにも柔軟に対応でき、またこれまでになかった新種のウイルスであっても検出することが可能となる。

【0062】以上、本発明を実施の形態をもとに説明した。実施の形態は例示であり、それらの各構成要素や各処理プロセスの組合せにいろいろな変形例が可能なこと、またそうした変形例も本発明の範囲にあることは当業者に理解されるところである。以下そのような変形例を説明する。

【0063】実施の形態では、新種、変種のウイルスが多く出現するマクロウイルスを例にあげてウイルス検出手法を説明したが、マクロウイルスに限定する趣旨ではなく、本発明はプログラムコードに感染するウイルス一般にも適用可能である。

【0064】実施の形態では、ウイルス特有のコードを

収集したデータベースが検査対象のコンピュータ内に設けられ、このようなデータベースがサーバに設けられ、ネットワークを介して必要なデータを検索したり、ダウンロードする構成でもよい。また本実施の形態のウイルス検出プログラムを有するモバイルエージェントがユーザのコンピュータに配信されて、分散型のウイルス検出が行われてもよい。このようなエージェントを用いる構成においても、ウイルスコードのデータベースはユーザのコンピュータ内に設けられてもサーバに設けられてもよい。またエージェントプログラムのデータ領域にそのようなデータベースの少なくとも一部が含まれていてもよい。

【0065】またウイルス検出システムは、コンピュータのハードディスク等をスキャンしてウイルスの検査、除去を行うウイルス検疫ソフトウェアのような独立したアプリケーションとして提供されてもよい。その場合、CD-ROMなどの記録媒体から読みとられたデータやネットワークを介してダウンロードされたデータがハードディスクに書込まれるときに、オンデマンドでデータをスキャンしてもよい。また、ウイルス検出システムは、マクロ機能をもつアプリケーションに組み込まれる形態で提供されてもよい。この場合、アプリケーションがマクロファイルを開く際、本ウイルス検出システムによりマクロの危険度を測定して、マクロファイルを開く前に実行する前に危険度をユーザに通知して警告を与えてもよい。

【0066】上記の学習手順では、ウイルスが検出された場合に重みの更新が行われたが、ウイルスかどうかの判定は、外部から与えられてもよい。たとえばユーザがウイルスかどうかを判断して、判定結果をウイルス検出システムに与えてもよい。またデータベース更新部は、外部から提供されるウイルス定義ファイルを用いて、危険度算出部によるウイルス判定の成否を評価してもよい。

【0067】

【発明の効果】本発明によれば、新種または変種のウイルスを効果的に検出することができる。

【図面の簡単な説明】

【図1】 実施の形態に係るウイルス検出システムの全体構成図である。

【図2】 ウイルス検査対象ファイルの一例である複合ファイルのデータ構造を説明する図である。

【図3】 図2の複合ファイルのストリームに格納されているマクロ情報を示す図である。

【図4】 図3のマクロ情報を再構築してテキストに書き出した状態を示す図である。

【図5】 ウイルスの特徴コードに対応づける重みのレベル分けを説明する図である。

【図6】 図1のデータベースの学習手順を説明する図である。

【図7】 実施の形態に係るウイルス検出装置の構成図である。

【図8】 ウイルス検出装置におけるウイルス検出手順を示すフローチャートである。

【図9】 図8のマクロの再構築処理の詳細な手順を示すフローチャートである。

【図10】 図8のマクロのトレース検査処理の詳細な手順を示すフローチャートである。

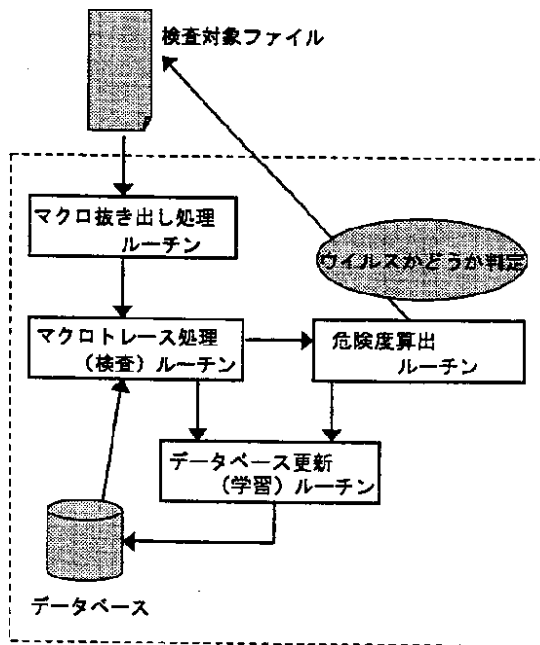
【図11】 図8のウイルス判定処理の詳細な手順を示すフローチャートである。

【図12】 図8のデータベース更新処理の詳細な手順を示すフローチャートである。

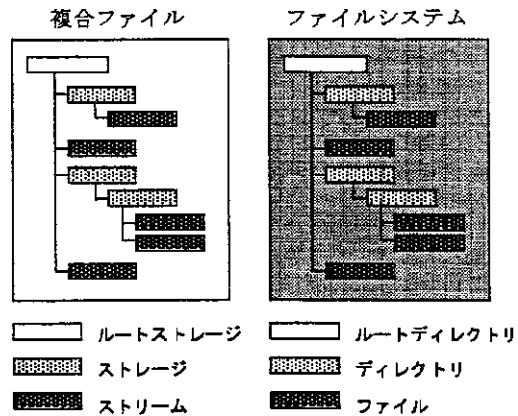
【符号の説明】

- 10 ウイルス検出装置、 12 検査対象ファイル、
- 14 マクロ抽出部、 16 トレース検査部、 1
- 8 危険度算出部、 20 データベース更新部、 2
- 2 マクロパターン登録部、 24 重み更新部、 2
- 6 ウイルスデータベース、 28 特徴コードレコード、
- 30 マクロパターンレコード。

【図1】



【図2】



【図3】

```

10 00 08 04 53 61 60 65 70 6C 65 6E 09 2A .\...SampleIn*.
04 04 40 41 49 4E 88 2C 10 00 09 04 66 69 ...MAIN-...fi
40 61 63 72 6F 31 54 10 00 09 04 67 6C 6F leMacroIT...glo
61 63 72 6F FC A3 10 00 09 04 40 61 63 72 bMacro.J...Macr
69 6C 65 55 78 10 00 09 04 65 72 72 43 61 oFileUv...errCa
68 74 73 60 10 00 09 00 57 6F 72 64 42 61 ights'...WordBa
63 F1 92 10 00 0F 00 46 69 6C 65 53 75 6D sic'...FileSum
72 79 49 6E 6E 6F 58 EE 10 00 06 00 55 70 maryInfoY...Up
74 65 D7 FC 10 00 03 04 64 6C 67 2D 84 10 date*....dig-..
00 44 69 61 6C 6F 67 52 65 63 6F 72 64 92 ...DialogRecord4
00 09 00 43 75 72 58 61 6C 75 65 73 C8 A4 ...CurValues*.
09 00 44 69 72 55 63 74 6F 72 79 6D 47 10 ...DirectorynC.
00 46 69 6C 65 4E 61 60 65 6A C3 10 00 05 ...FileName*.
43 61 73 65 36 0A 10 00 06 80 80 00 FF 03 .UCase$.
52 69 67 68 74 24 79 31 10 00 0E 80 80 00 ..Right$yl.
00 00 40 61 63 72 6F 46 69 6C 65 4E 61 6D ...MacroFileNam
9C 4A 10 00 0A 80 80 00 FF 03 00 00 40 61 e$.
6F 4E 61 60 65 24 8D 01 10 00 09 00 00 69 croName$.
53 61 76 65 41 73 CE EE 10 00 07 04 63 65 leSaveAs*.

```

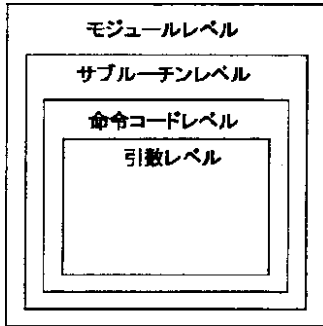
【図4】

```

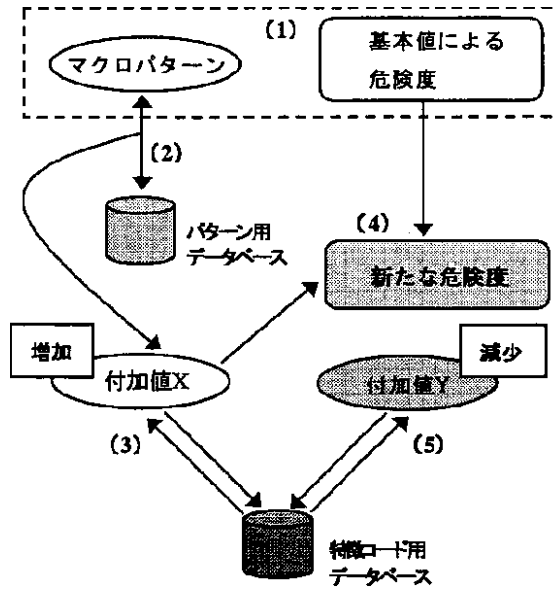
Sub MAIN
Dim fileMacro
Dim MacroFile
On Error GoTo -1: On Error GoTo errCaught
WordBasic.FileSummaryInfo Update:=1
Dim Word: dig = WordBasic.DialogRecord.FileSummaryInfo(False)
WordBasic.CurValues.FileSummaryInfo dig
fileMacro = dig.Directory + "\ + dig.FileName + ".autoOpen"
MacroFile = UCase(WordBasic.Right$(WordBasic.MacroFileName$))
If MacroFile = "sample.001" Then
MsgBox "sample.doc"
Else
MsgBox "test.doc"
End If
SetString
GoTo bye
errCaught:
bye:
On Error GoTo -1: On Error GoTo 0
End Sub
Sub SetString
Dim i

```

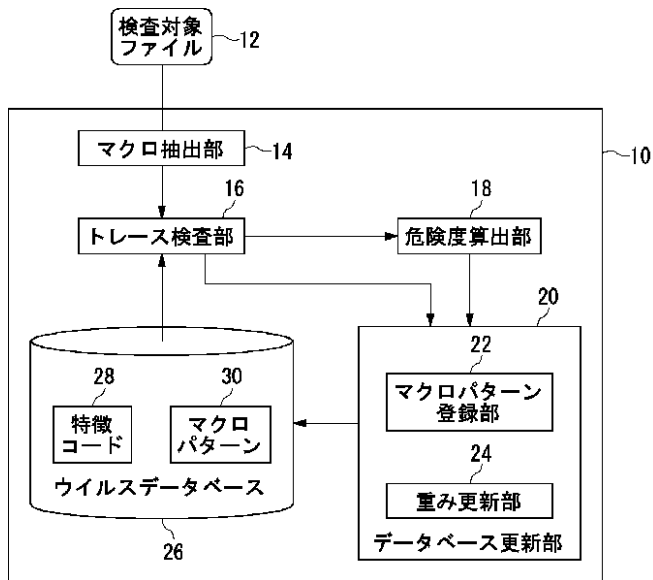
【図5】



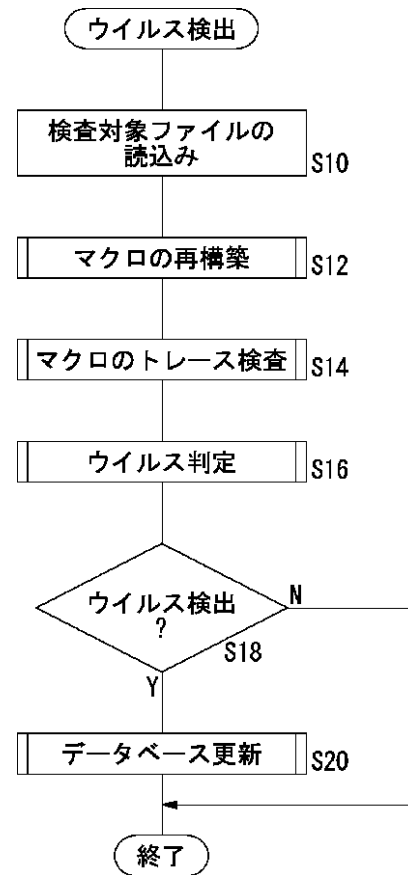
【図6】



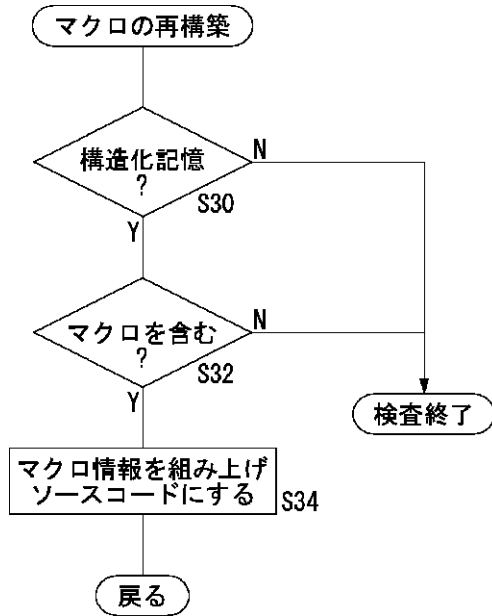
【図7】



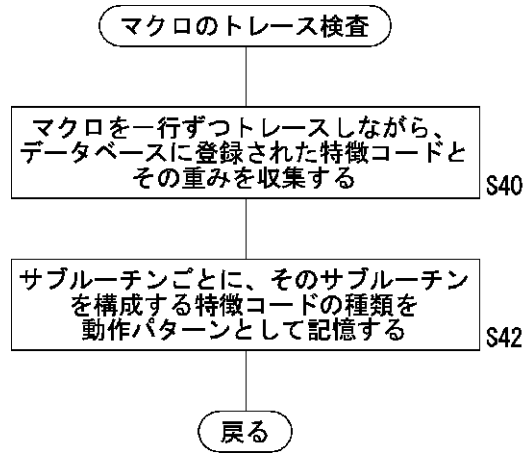
【図8】



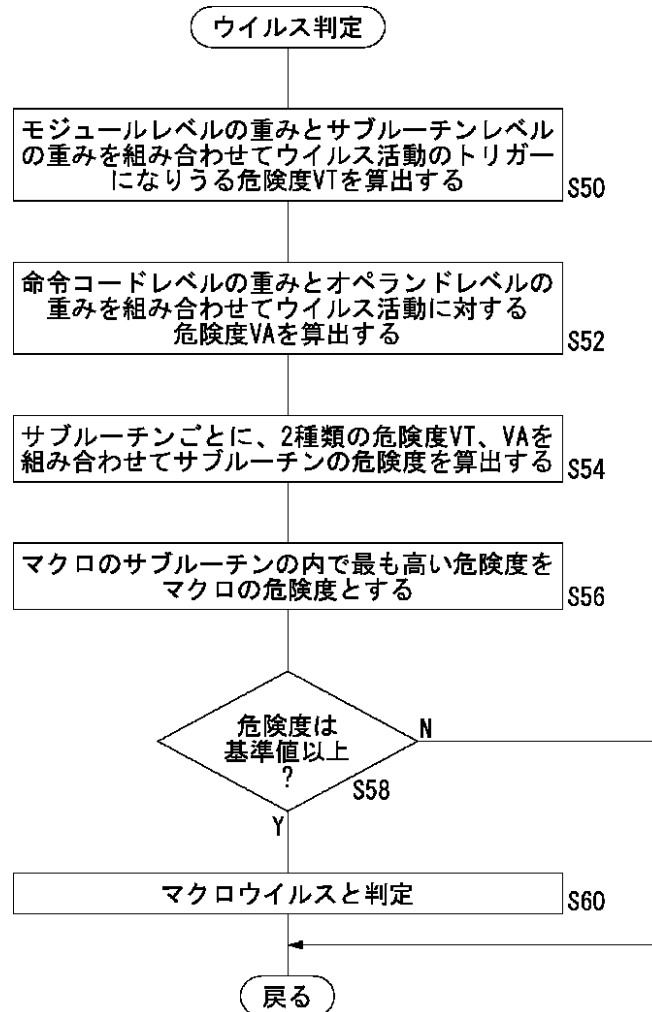
【図9】



【図10】



【図11】



【図12】

