

(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2019-21094
(P2019-21094A)

(43) 公開日 平成31年2月7日(2019.2.7)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 21/55 (2013.01)	G06F 21/55	
G06F 21/31 (2013.01)	G06F 21/55 320	
G06F 16/00 (2019.01)	G06F 21/31	
	G06F 17/30 350C	

審査請求 未請求 請求項の数 12 O L (全 26 頁)

(21) 出願番号 特願2017-139751 (P2017-139751)
(22) 出願日 平成29年7月19日 (2017.7.19)

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区丸の内一丁目6番6号
(74) 代理人 110001689
青稜特許業務法人
(72) 発明者 藤井 翔太
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
(72) 発明者 鬼頭 哲郎
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
(72) 発明者 藤井 康広
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

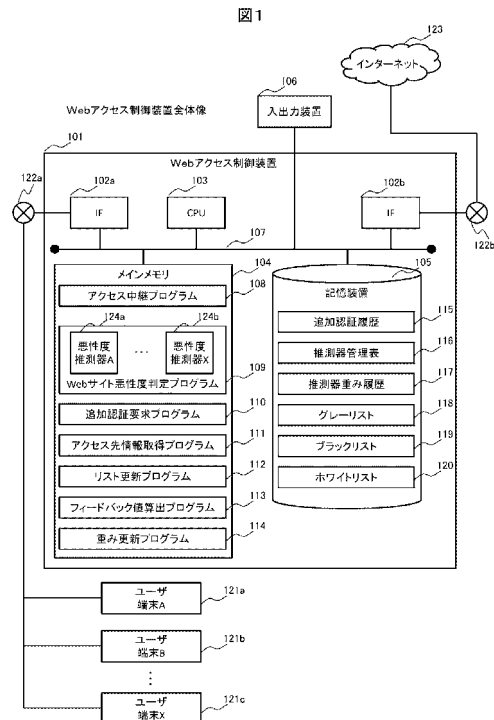
(54) 【発明の名称】 Webアクセス制御装置

(57) 【要約】

【課題】 Webアクセス制御装置において、Webサイトの悪性度を高精度で判定する。

【解決手段】 重みがそれぞれ割り当てられた複数の悪性度推測器 124 a ~ 124 b を用いてアクセス先Webサイトの悪性度を所定の閾値に基づいて判定するWebサイト悪性度判定プログラム 109 と、アクセス先Webサイトの悪性度が所定の閾値より高いと判定した場合にユーザ端末 121 a ~ 121 c に対してアクセス先Webサイトへのアクセスの可否を決定するための追加認証を要求する追加認証要求プログラム 110 と、追加認証の結果に基づいて悪性度推測器 124 a ~ 124 b にそれぞれ割り当てられた重みを更新する重み更新プログラム 114 を有する。

【選択図】 図 1



【特許請求の範囲】**【請求項 1】**

インターネットとユーザが操作するユーザ端末に接続されたWebアクセス制御装置であって、

重みがそれぞれ割り当てられた複数の悪性度推測器を用いて、Webサイト悪性度判定プログラムを実行することにより、アクセス先Webサイトの悪性度を所定の閾値に基づいて判定するWebサイト悪性度判定部と、

前記Webサイト悪性度判定部が前記アクセス先Webサイトの悪性度が前記所定の閾値より高いと判定した場合に、追加認証要求プログラムを実行することにより、前記ユーザ端末に対して前記アクセス先Webサイトへのアクセスの可否を決定するための追加認証を要求する追加認証要求部と、

前記追加認証要求部の前記追加認証の結果に基づいて、重み更新プログラムを実行することにより、前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みを更新する重み更新部と、

を有することを特徴とするWebアクセス制御装置。

【請求項 2】

フィードバック値算出プログラムを実行することにより、前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みにフィードバックする値を算出するフィードバック値算出部を更に有し、

前記重み更新部は、

前記フィードバック値算出部で算出された算出値に基づいて、前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みを更新することを特徴とする請求項 1 に記載のWebアクセス制御装置。

【請求項 3】

前記フィードバック値算出部は、

前記追加認証要求部の前記追加認証の結果に加え、前記追加認証の際に得られる付随情報を用いて、前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みへフィードバックする値を算出することを特徴とする請求項 2 に記載のWebアクセス制御装置。

【請求項 4】

前記フィードバック値算出部は、

前記付随情報として、前記追加認証に要した時間又は前記アクセス先Webサイトの情報表示の有無を用いることを特徴とする請求項 3 に記載のWebアクセス制御装置。

【請求項 5】

前記フィードバック値算出部は、

前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みの時系列の変化を用いて、前記悪性度推測器にそれぞれ割り当てられた前記重みへフィードバックする値を算出することを特徴とする請求項 2 に記載のWebアクセス制御装置。

【請求項 6】

前記追加認証要求部は、

前記Webサイト悪性度判定部が前記アクセス先Webサイトの悪性度が前記所定の閾値より高いと判定した場合に、前記アクセス先Webサイトが不審サイトと判定し、前記ユーザ端末に対して前記不審サイトに関する追加情報を提供することを特徴とする請求項 1 に記載のWebアクセス制御装置。

【請求項 7】

前記追加認証要求部は、

前記不審サイトに関する前記追加情報として、前記アクセス先Webサイトのサムネイル画像を用いることを特徴とする請求項 6 に記載のWebアクセス制御装置。

【請求項 8】

前記重み更新部は、

10

20

30

40

50

前記Webサイト悪性度判定プログラムの前記悪性度推測器の推測結果と、前記追加認証要求部の前記追加認証の認証結果とを比較し、この比較結果に基づいて前記悪性度推測器にそれぞれ割り当てられた前記重みを更新することを特徴とする請求項1に記載のWebアクセス制御装置。

【請求項9】

前記重み更新部は、

前記悪性度推測器の推測結果と、前記追加認証要求部の前記認証結果との比較結果の一致率が高い前記悪性度推測器の重みを大きくすることを特徴とする請求項8に記載のWebアクセス制御装置。

【請求項10】

前記Webサイト悪性度判定部の前記複数の悪性度推測器は、

第1の処理時間で前記悪性度を判定する第1のグループと、前記第1の処理時間より長い前記第2の処理時間で前記悪性度を判定する第2のグループとに階層的に分割され、

前記第1のグループに属する前記悪性度推測器で判定された悪性度に応じて、前記第2のグループに属する前記悪性度推測器で再度悪性度を判定するか否かを決定することを特徴とする請求項1に記載のWebアクセス制御装置。

【請求項11】

前記Webサイト悪性度判定部は、

前記複数の悪性度推測器における推測結果を予め定めた閾値と比較し、前記悪性度推測器ごとに前記アクセス先Webサイトが悪性か良性かを推定し、前記悪性の数の合計数と前記良性の数の合計数との比較結果に応じて、前記アクセス先Webサイトの前記悪性度を判定することを特徴とする請求項1に記載のWebアクセス制御装置。

【請求項12】

前記Webサイト悪性度判定部は、

前記複数の悪性度推測器の中から特定の悪性度推測器を予め決めておき、前記特定の悪性度推測器の推測結果を他の前記悪性度推測器の推測結果よりも優先させて、前記アクセス先Webサイトの前記悪性度を判定することを特徴とする請求項1に記載のWebアクセス制御装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、Webアクセス制御装置に関する。

【背景技術】

【0002】

近年、標的型攻撃に見られるように、攻撃が高度化しており、企業や国家にとって重大な脅威となっている。また、攻撃の高度化に伴い、外部からの侵入を完全に遮断することは困難になってきている。ここで、侵入された後の被害を抑制するためには、機密情報の窃取や感染拡大を目的とした外部への通信を遮断することが重要である。

【0003】

以上のような背景から、適切に外部への通信を制御し、不審な通信は遮断する技術が求められている。これに関連する技術として、例えば、特許文献1がある。

【0004】

特許文献1では、プロキシサーバ等のログから計算したWebサイトの悪性度に基づいて、悪性度が低いホワイトリスト、悪性度が高いブラックリスト及びいずれにも属さないグレーリストを自動的に生成する。グレーリストに含まれるWebサイトに接続する際に追加認証を要求し、追加認証に成功したときグレーリストをホワイトリストに振り分け、一定期間一度も成功しなかったときブラックリストに振り分ける。これにより、機械的にホワイトリストやブラックリストに振り分けられなかったWebサイトであっても、認証結果に応じていずれかのリストに振り分けることが可能である。

【先行技術文献】

10

20

30

40

50

【特許文献】

【0005】

【特許文献1】特開2015-170219号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかし、特許文献1では、Webサイトの悪性度を判定する機構の精度については考慮されていない。このため、悪性度判定機構の精度によっては、悪性度の高いWebサイトを安全と判断してしまい、アクセス時に被害を受けてしまう可能性がある。また、悪性度判定機構の精度によっては、悪性度の低いWebサイトを危険と判断してしまい、追加認証が頻発して利便性を低下してしまう可能性がある。

10

【0007】

本発明の目的は、Webアクセス制御装置において、Webサイトの悪性度を高精度で判定することにある。

【課題を解決するための手段】

【0008】

本発明の一態様のWebアクセス制御装置は、インターネットとユーザが操作するユーザ端末に接続され、重みがそれぞれ割り当てられた複数の悪性度推測器を用いて、Webサイト悪性度判定プログラムを実行することにより、アクセス先Webサイトの悪性度を所定の閾値に基づいて判定するWebサイト悪性度判定部と、前記Webサイト悪性度判定部が前記アクセス先Webサイトの悪性度が前記所定の閾値より高いと判定した場合に、追加認証要求プログラムを実行することにより、前記ユーザ端末に対して前記アクセス先Webサイトへのアクセスの可否を決定するための追加認証を要求する追加認証要求部と、前記追加認証要求部の前記追加認証の結果に基づいて、重み更新プログラムを実行することにより、前記Webサイト悪性度判定部の前記悪性度推測器にそれぞれ割り当てられた前記重みを更新する重み更新部と、を有することを特徴とする。

20

【発明の効果】

【0009】

本発明の一態様によれば、Webアクセス制御装置において、Webサイトの悪性度を高精度で判定することができる。

30

【図面の簡単な説明】

【0010】

【図1】実施例1のWebアクセス制御装置の構成例を示した図である。

【図2】認証情報管理表の一例を示す図である。

【図3】推測器管理表の一例を示す図である。

【図4】推測器重み履歴の一例を示す図である。

【図5】グレーリストの一例を示す図である。

【図6】ブラックリストの一例を示す図である。

【図7】ホワイトリストの一例を示す図である。

【図8】実施例1のWebアクセス制御装置の全体処理フローを示す図である。

40

【図9】Webサイト悪性度判定プログラムプログラムの処理フローを示す図である。

【図10】リスト更新プログラムの処理フローを示す図である。

【図11】フィードバック値算出の処理フローを示す図である。

【図12】重み更新の処理フローを示す図である。

【図13a】追加認証画面の一例を示す図である。

【図13b】追加認証画面の一例を示す図である。

【図14】実施例2の時系列変化を考慮した重み更新の処理フローを示す図である。

【図15】実施例3の推測器群構成を示す図であり、推測器群を階層的に並べた場合の処理例を示す図である。

【図16】実施例4の最終予測結果算出フローの一例を示す図であり、(a)は推測結果

50

と重みの積和をとる方法、(b)は多数決をとる方法、(c)は特定の推測器を重用する方法をそれぞれ示す。

【図17a】実施例5の低難易度の追加認証方式を用いた追加認証画面の一例を示す図である。

【図17b】実施例5の高難易度の追加認証方式を用いた追加認証画面の一例を示す図である。

【発明を実施するための形態】

【0011】

最初に、実施形態について説明する。

実施形態では、アクセス先Webサイトの悪性度を判定し、悪性サイトへの通信を遮断しつつ、ユーザによるフィードバックを基に判別精度を改善する。実施形態では、それぞれに重みが付与された悪性度を判定する推測器を複数用意してアクセス先Webサイトの悪性度を判定し、悪性度が一定の閾値より高かった場合にはユーザに追加認証を要求する。

10

【0012】

実施形態のWebアクセス制御装置は、ユーザが操作する端末と、インターネットとをネットワークを介して接続する装置であって、アクセス先Webサイトの悪性度をそれぞれが重み付けされた複数の機構を用いて推測するWebサイト悪性度判定プログラムと、悪性度の高いサイトへのアクセスを検出した際に、ユーザに対してその旨を伝え、ユーザの判断に基づいてアクセス可否を決定する追加認証要求プログラムを有する。

20

【0013】

また、ユーザの判断やその付随情報を基にWebサイト悪性度判定プログラムにフィードバックする内容を決定するフィードバック値算出プログラムと、Webサイトの悪性度判定プログラムを構成する各推測器の重みを前記フィードバック値算出プログラムが算出した値に基づいて更新する重み更新プログラムとを有する。

【0014】

実施形態によれば、Webサイトの悪性度を複数の推測器を用いて推測すること、及び推測結果の正否をユーザからのフィードバックによって判定することが可能である。この際、ユーザからのフィードバックを利用し、推測結果とユーザ判断の一致率が高い推測器の重みを大きくすることにより、全体の検出精度を自動的に改善する。これらの組み合わせによって、悪性サイトへの通信は遮断しつつ、不要な追加認証やブラックリストの手動での更新コストを抑制して利便性を向上する。

30

【0015】

ユーザは、追加認証を要求された際にアクセス先Webサイトへの接続可否を判断し、問題ないと判断した場合のみ追加認証を突破してWebサイトへアクセスする。これにより、ユーザが意図しない、あるいは追加認証の存在を想定しておらず追加認証を突破できないマルウェア等による悪性サイトへのアクセスを抑制できる。

【0016】

また、ユーザの判断とWebサイト悪性度判定プログラムが算出した結果を突合し、両者が一致したか否かによってWebサイト悪性度判定プログラムを構成する各推測器の重みを更新する。これにより、精度の高い推測器の判定を重視するようにし、全体としての悪性度判定精度が向上する。

40

【0017】

また、ユーザの判断を確認する際、付随的に得られる情報を用いて判断が正しいか否か推定し、より確からしい利用者判断の結果を悪性度判定機構に反映する。これにより、誤ったユーザの判断を反映することによる悪性度判定精度の低下を抑制できる。

【0018】

なお、以降では、サイバー攻撃に利用される悪質なWebサイトを悪性サイト、悪性サイトではないWebサイトを良性サイトと呼ぶ。

【実施例1】

50

【0019】

図1を参照して、実施例1のWebアクセス制御装置の構成について説明する。

図1に示すように、実施例1のWebアクセス制御装置101は、ユーザが操作するユーザ端末121a~121cと、インターネット123とをネットワーク122を介して接続される。

【0020】

Webアクセス制御装置101は、CPU(Central Processing Unit)103と、CPU103が処理を実行するために必要なデータを格納するためのメインメモリ104と、大量のデータを記憶する容量を持つハードディスクやフラッシュメモリなどの記憶装置105と、他装置と通信を行なうためのIF(インタフェース)102a、102bと、キーボード、ディスプレイなどの入出力を行うための入出力装置106と、これらの各装置を接続する通信路107と、を備えたコンピュータである。なお、通信路107は、例えば、バスやケーブルなどの情報伝達媒体である。

10

【0021】

CPU103は、メインメモリ104に格納されたアクセス中継プログラム108を実行することにより不審な通信の制御を行う。また、CPU103は、Webサイト悪性度判定プログラム109を実行することによりアクセス先Webサイトの悪性度の判定を行う。

【0022】

また、CPU103は、追加認証要求プログラム110を実行することによりユーザ端末121を操作するユーザの追加認証を行う。また、CPU103は、アクセス先情報取得プログラム111を実行することによりアクセス先Webサイトの情報取得を行う。また、CPU103は、リスト更新プログラム112を実行することによりグレーリスト118、ブラックリスト119、ホワイトリスト120の更新を行う。

20

【0023】

また、CPU103は、フィードバック値算出プログラム113を実行することによりWebサイト悪性度判定プログラム109を構成する各悪性度推測器124a~124bに割り当てられた重みへフィードバックする値の算出を行う。また、CPU103は、重み更新プログラム114を実行することによりフィードバック値算出プログラム113での算出値に基づいて推測器管理表116に記録されているWebサイト悪性度判定プログラム109を構成する各悪性度推測器124に割り当てられた重みの更新を行う。

30

【0024】

実施例1のWebアクセス制御装置101は、重みがそれぞれ割り当てられた複数の悪性度推測器124を用いて、Webサイト悪性度判定プログラム109を実行することにより、アクセス先Webサイトの悪性度を所定の閾値に基づいて判定するWebサイト悪性度判定部と、Webサイト悪性度判定部がアクセス先Webサイトの悪性度が所定の閾値より高いと判定した場合に、追加認証要求プログラム110を実行することにより、ユーザ端末121a~121cに対してアクセス先Webサイトへのアクセスの可否を決定するための追加認証を要求する追加認証要求部と、追加認証要求部の追加認証の結果に基づいて、重み更新プログラム114を実行することによりWebサイト悪性度判定部の悪性度推測器124にそれぞれ割り当てられた重みを更新する重み更新部を有する。

40

【0025】

さらに、フィードバック値算出プログラム113を実行することにより、Webサイト悪性度判定部の悪性度推測器124にそれぞれ割り当てられた重みにフィードバックする値を算出するフィードバック値算出部を有する。重み更新部は、フィードバック値算出部で算出された算出値に基づいて、Webサイト悪性度判定部の悪性度推測器124にそれぞれ割り当てられた重みを更新する。

【0026】

記憶装置105には、ユーザ端末121を操作するユーザの追加認証結果やその際の付随情報を示す追加認証履歴115、Webサイト悪性度判定プログラム109を構成する

50

悪性度推測器 1 2 4 の情報を管理する推測器管理表 1 1 6 が格納されている。

【 0 0 2 7 】

また、記憶装置 1 0 5 には、悪性度推測器 1 2 4 の時系列ごとの重みを管理する推測器重み履歴 1 1 7、不審な通信先の情報を示すグレーリスト 1 1 8、危険な通信先の情報を示すブラックリスト 1 1 9、安全な通信先の情報を示すホワイトリスト 1 2 0 が格納されている。

【 0 0 2 8 】

上記の各プログラムやデータは、あらかじめメモリ 1 0 4 又は記憶装置 1 0 5 に格納されていてもよいし、必要な時に、入出力装置 1 0 6 から又は I F 1 0 2 を介して他の装置から、インストール（ロード）されても良い。

10

【 0 0 2 9 】

次に、図 2 を参照して、追加認証履歴 1 1 5 の一例について説明する。

図 2 に示すように、追加認証履歴 1 1 5 は、I D 2 0 1 と、認証結果 2 0 2 と、付随情報 2 0 3 と、ユーザ識別情報 2 0 7 と、U R L 2 0 8 とを含んで構成される。また、付随情報は、例えば、認証に要した時間 2 0 4 と、正解数 2 0 5 と、サイト情報表示有無 2 0 6 とから構成される。

【 0 0 3 0 】

I D 2 0 1 は、認証情報を一意に識別できる情報を表す。認証結果 2 0 2 は、ユーザが追加認証に成功したか否かを表す。例えば、認証結果 2 0 2 が「成功」の場合は、ユーザが追加認証の突破に成功したことを、認証結果 2 0 2 が「失敗」の場合は、ユーザが追加

20

【 0 0 3 1 】

付随情報 2 0 3 は、ユーザが追加認証に臨んだ際に付随的に得られる情報を表す。ここでは、付随情報 2 0 3 は、認証に要した時間 2 0 4 と、正解数 2 0 5 と、サイト情報表示有無 2 0 6 とから構成されるものとして説明する。

【 0 0 3 2 】

認証に要した時間 2 0 4 は、ユーザに追加認証画面が提示されてから、認証が試行されるまでに要した時間を表す。例えば、追加認証画面が表示されてから認証の試行に 4 . 3 秒を要した場合、認証の成否に関わらず、「4 . 3」と記録される。また、あらかじめ定められた一定時間認証突破の試行がなかった場合、「t i m e o u t」と記録される。

30

【 0 0 3 3 】

正解数 2 0 5 は、追加認証に文字列での認証を用いた場合に、全体の文字数のうち、どれだけ正解したかの数と、その正解率とを表す。例えば、認証文字列が「a b c d」だった際、「a b v f」と入力があったならば、全 4 文字中前半の 2 文字が正解しているため、「2 (5 0 %)」と記録される。なお、この時間は、追加認証要求プログラム 1 1 0 において、追加認証画面を提示した時間とユーザからの追加認証文字列を受信した時間の差分によって算出される。

【 0 0 3 4 】

サイト情報表示有無 2 0 6 は、追加認証時にユーザが、アクセス先 W e b サイトが悪性か否か判断する材料として追加情報を要求したか否かを表す。例えば、「有」の場合は、ユーザが追加情報を要求したことを、「無」の場合には、ユーザが追加情報を要求しなかったことを表す。なお、ユーザの判断を補助するための追加情報としては、アクセス先 W e b サイトのサムネイル画像、w h o i s 情報、および D N S 情報等が挙げられる。

40

【 0 0 3 5 】

なお、追加認証履歴 1 1 5 の格納情報は、不審サイトへのアクセス試行があった際の追加認証時に取得され、格納されるものとする。また、今回は、付随情報として認証に要した時間 2 0 4 と、正解数 2 0 5 と、サイト情報表示有無 2 0 6 とを挙げたが、これらの情報以外に付随的に得られる情報を用いても良いものとする。

【 0 0 3 6 】

ユーザ識別情報 2 0 7 は、追加認証に臨んだユーザを一意に識別できる情報を表す。例

50

えば、ネットワーク環境において、「123」と識別子を付与されたユーザが認証に望んだ場合、ユーザを表す「123」が記録される。なお、今回はネットワークシステム上でユーザに付与された識別子を用いたが、ユーザを一意に識別できればよく、例えば、ユーザ端末のIP (Internet Protocol) アドレスや、ユーザのユーザ名を用いても良い。

【0037】

URL208は、追加認証の対象となった不審な接続先のURLを表す。例えば、「foo.com」の場合は、URLは、追加認証の対象となった不審な接続先であることを表す。なお、今回は、URLを用いたが、不審な接続先を識別できれば良く、例えば、IPアドレスや、FQDN (Fully Qualified Domain Name) を

10

【0038】

次に、図3を参照して、推測器管理表116の一例について説明する。

図3に示すように、推測器管理表116は、ID301と、正解率302と、重み303と、直近の悪性度判定結果304とを含んで構成される。

【0039】

ID301は、Webサイト悪性度判定プログラム109を構成する各悪性度推測器124を一意に識別できる情報を表す。正解率302は、これまでの各不審サイトに対する判定結果(悪性/良性)とユーザの追加認証結果(非突破/突破)の一致率を表す。例えば、「80%」の場合は、N回の不審サイトの悪性度判定機会のうち、0.8N回分の判定結果がユーザの追加認証結果と一致していたことを表す。正解率が高い推測器ほどWebアクセス制御装置を導入した組織での悪性サイトの判別精度が高いものとして、重みを大きく付与される。各悪性度推測器124に付与する重みを算出する際の具体的な処理手順については、図11を用いて後述する。

20

【0040】

重み303は、各悪性度推測器124に割り当てられた重みの合計値が1になるように正規化された値を表す。例えば、「0.2」の場合は、全悪性度推測器124a~124bのうち20%の判断権を当該悪性度推測器124が有することを表す。

【0041】

直近の悪性度判定結果304は、直近の悪性度を判断する機会において、判定対象の悪性度をどのように判定したかを表す。例えば、「75%」の場合は、当該悪性度推測器124は、判定対象が75%の確率で悪性であると判断したことを表す。なお、今回はアクセス先Webサイトの不審度を表す指標として、アクセス先Webサイトの悪性度を用いて説明したが、アクセス先Webサイトの良性度等、不審度を表すことのできる指標であれば、どのような情報を用いても良い。なお、推測器管理表116の各情報は、管理者が必要に応じて、入力または更新しても良い。

30

【0042】

次に、図4を参照して、推測器重み履歴117の一例について説明する。

図4に示すように、推測器重み履歴117は、時刻401と、重み402と、悪性度判定結果403と、認証結果との一致404とを含んで構成される。また、推測器重み履歴117は、Webサイト悪性度判定プログラム109を構成する悪性度推測器124毎に用意され、悪性度推測器124に関する情報を格納する。なお、図4では、図3のIDが0の悪性度推測器124に対応する悪性度重み履歴117を例示している。

40

【0043】

時刻401は、悪性度推測器124a~124bが不審サイトの悪性度を判定した時間を表す。なお、図4では、年/月/日 時間:分:秒.小数秒の表記を用いているが、Unix time等、時刻が判別できる情報であれば、どのような情報を用いても良い。

【0044】

重み402は、当該時刻における悪性度推測器124の重みを表す。例えば、「0.3」の場合は、当該時刻に悪性度推測器124a~124bへ0.3の重みが付与されてい

50

たことを表す。

【0045】

悪性度判定結果403は、当該時刻における悪性度判定対象のWebサイトに対して、悪性度推測器124が算出した悪性度を表す。例えば、「75%」の場合は、当該悪性度推測器124は、当該時刻に判定対象が75%の確率で悪性であると判断したことを表す。

【0046】

認証結果との一致404は、当該時刻における不審サイトに対する判定結果（悪性/良性）とユーザの追加認証結果（非突破/突破）が一致したか否かを表す。例えば、「一致」の場合は、両者が一致していたことを、「不一致」の場合は、両者が一致していなかったことを表す。

10

【0047】

次に、図5を参照して、グレーリスト118の一例について説明する。

図5に示すように、グレーリスト118は、ID501と、URL502と、認証成功ユーザ数503と、認証失敗ユーザ数504と、ユーザ毎の認証結果505を含んで構成される。ID501は、グレーリスト118を一意に識別できる情報を表す。

【0048】

URL502は、不審な接続先のURLを表す。例えば、「example.com」の場合は、URLは不審な接続先であることを表す。なお、今回は、URLを用いたが、不審な接続先を識別できれば良く、例えば、IPアドレスや、FQDNを用いても良い。

20

【0049】

認証成功ユーザ数503は、該接続先の追加認証を突破したユーザのユニーク数を表す。例えば、「2」の場合は、2人のユーザが該接続先の追加認証を突破したことを表す。なお、成功回数は、後述するユーザ毎の認証結果505を用いることによって算出できる。

【0050】

認証失敗ユーザ数504は、接続先の追加認証を突破しなかったユーザのユニーク数を表す。例えば、「4」の場合は、4人のユーザが該接続先の追加認証を突破しなかったことを表す。なお、失敗回数は、後述するユーザ毎の認証結果505を用いることによって算出できる。

30

【0051】

ユーザ毎の認証結果505は、接続先に対するユーザ毎の最新の認証結果を表す。例えば、「123:成功, 789:失敗」の場合は、識別子123のユーザが追加認証を突破したと識別子789のユーザが追加認証を突破しなかったことを表す。

【0052】

グレーリスト118は、Webサイト悪性度判定プログラム109によって悪性であると判定された接続先がリスト更新プログラム110によって登録される。Webサイト悪性度判定プログラム109と、リスト更新プログラム110との具体的な処理については、図9と、図10とを用いて後述する。なお、グレーリスト118の各情報は、管理者が必要に応じて、登録または更新しても良い。

40

【0053】

次に、図6を参照して、ブラックリスト119の一例について説明する。

図6に示すように、ブラックリスト119は、ID601と、URL602とを含んで構成される。ID601は、ブラックリスト119を一意に識別できる情報を表す。

【0054】

URL602は、悪性サイトのURLを表す。例えば、「black.com」の場合は、URLは悪性サイトであることを表す。なお、今回は、URLを用いたが、悪性サイトを識別できれば良く、例えば、IPアドレスや、FQDNを用いても良い。

【0055】

ブラックリスト119は、Webサイト悪性度判定プログラム109によって悪性であ

50

ると判定され、かつ一定数以上ユーザが認証を突破しなかった接続先がリスト更新プログラム110によって登録される。Webサイト悪性度判定プログラム109と、リスト更新プログラム110との具体的な処理については、図9と、図10とを用いて後述する。なお、グレーリスト119の各情報は、管理者が必要に応じて、登録または更新しても良い。

【0056】

次に、図7を参照して、ホワイトリスト120の一例について説明する。

図7に示すように、ホワイトリスト120は、ID701と、URL702を含んで構成される。ID701は、ホワイトリスト120を一意に識別できる情報を表す。URL702は、良性サイトのURLを表す。例えば、「white.com」の場合は、URLは良性サイトであることを表す。なお、今回は、URLを用いたが、悪性サイトを識別できれば良く、例えば、IPアドレスや、FQDNを用いても良い。

10

【0057】

ホワイトリスト120は、Webサイト悪性度判定プログラム109によって悪性であると判定されたものの、一定数以上ユーザが認証を突破した接続先がリスト更新プログラム110によって登録される。Webサイト悪性度判定プログラム109と、リスト更新プログラム110との具体的な処理については、図9と、図10とを用いて後述する。なお、ホワイトリスト119の各情報は、管理者が必要に応じて、登録または更新しても良い。

【0058】

このように、実施例1のWebアクセス制御装置は、Webアクセス制御装置101のアクセス中継プログラム108が、端末121からの通信を中継する。次に、Webサイト悪性度判定プログラム109が、アクセス先Webサイトの悪性度を判定し、追加認証要求プログラム110がユーザ121に対して追加認証を要求する。

20

【0059】

次に、アクセス先情報取得プログラム111が、アクセス先Webサイトの情報を取得し、リスト更新プログラム112がグレーリスト118、ブラックリスト119、およびホワイトリスト120を更新する。次に、フィードバック値算出プログラム113が、Webサイト悪性度判定プログラム109を構成する各悪性度推測器124に割り当てられた重みへフィードバックする値を算出する。次に、重み更新プログラム114が、フィードバック値算出プログラム113での算出値に基づいてWebサイト悪性度判定プログラム109を構成する各悪性度推測器124に割り当てられた重みを更新する。

30

【0060】

以下、図8を参照して、Webアクセス制御装置101の全体処理フローについて説明する。

図8に示すように、アクセス中継プログラム108は、CPU103により実行され、ユーザ端末121からの通信をIF102a経由で中継する(ステップ801)。

【0061】

アクセス中継プログラム108は、ブラックリストを参照し、アクセス先Webサイトがブラックリストに該当した場合はステップ810aに進み、該当しなかった場合はステップ803に進む(ステップ802)。

40

【0062】

アクセス中継プログラム108は、ホワイトリストを参照し、アクセス先Webサイトがホワイトリストに該当した場合はステップ811aに進み、該当しなかった場合はステップ804に進む(ステップ803)。

【0063】

Webサイト悪性度判定プログラム109は、アクセス先Webサイトの悪性度を算出し、ステップ805に進む(ステップ804)。なお、Webサイト悪性度判定プログラム109の悪性度算出フローについては、図9を用いて後述する。

【0064】

50

アクセス中継プログラム108は、ステップ804で算出した悪性度をあらかじめ制定しておいた閾値と比較し、悪性度が閾値よりも低い場合はステップ811aに進み、高い場合はステップ806へ進む(ステップ805)。追加認証要求プログラム110は、ユーザに追加認証を要求し、ステップ807へ進む(ステップ806)。

【0065】

ここで、追加認証要求プログラム110は、ユーザへの追加認証に、CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart)のような人間と機械を分別するような方式を利用することが考えられる。

【0066】

これにより、マルウェアが機械的に悪性サイトへアクセスしようとした場合でも、該認証を突破することは困難であり、人間の悪性サイトへのアクセスに加え、マルウェアによる悪性サイトへのアクセスも抑制できる。なお、追加認証要求プログラム110の表示画面については、図13を用いて後述する。

【0067】

追加認証要求プログラム110は、ユーザがアクセスWeb先サイトの追加認証に際して追加情報を要求したか確認し、要求があった場合はステップ808へ、要求が無かった場合はステップ809へ進む(ステップ807)。

【0068】

アクセス先情報取得プログラム111は、アクセス先のサイト情報を取得し、取得した情報をユーザ端末118へ表示する(ステップ808)。

【0069】

追加認証要求プログラム110は、ユーザが追加認証に成功したか確認し、成功していた場合は811bへ、失敗していた場合は810bへ進む(ステップ809)。

【0070】

アクセス中継プログラム108は、ユーザの当該サイトへのアクセスを拒否し、処理を終了する(ステップ810a)。アクセス中継プログラム108は、ユーザの当該サイトへのアクセスを許可し、処理を終了する(ステップ811a)。

【0071】

アクセス中継プログラム108は、ユーザの当該サイトへのアクセスを拒否し、ステップ812へ進む(ステップ810b)。アクセス中継プログラム108は、ユーザの当該サイトへのアクセスを許可し、ステップ812へ進む(ステップ811b)。

【0072】

アクセス中継プログラム108は、ユーザからの追加認証文字列を受信し、追加認証履歴115に、該追加認証文字列による認証結果や認証時の付随情報等の追加認証の情報を記録し、ステップ813へ進む(ステップ812)。

【0073】

アクセス中継プログラム108は、ユーザの追加認証結果(認証を突破したか否か)をリスト更新プログラム112とフィードバック内容値プログラム113に通知し、各プログラムの処理であるステップ814と、ステップ815とに進む(ステップ813)。

【0074】

リスト更新プログラム112は、追加認証履歴115を参照し、ユーザの追加認証結果を取得後、該情報を用いてグレーリスト118、ブラックリスト119、およびホワイトリスト120を更新する(ステップ814)。なお、リスト更新プログラム112の各リスト更新フローについては、図10を用いて後述する。

【0075】

フィードバック値算出プログラム113は、追加認証履歴115を参照し、ユーザの追加認証結果や付随情報を取得後、該情報を用いてフィードバック値を算出し、ステップ815へ進む(ステップ815)。なお、フィードバック値算出プログラム113のフィードバック値算出フローについては、図11を用いて後述する。

10

20

30

40

50

【 0 0 7 6 】

重み更新プログラム 1 1 4 は、フィードバック値算出プログラム 1 1 3 が算出した値を基に推測器管理表 1 1 6 に記録されている各推測器の重みを更新し、処理を終了する（ステップ 8 1 6）。

【 0 0 7 7 】

なお、ステップ 8 0 6 において、アクセス先 Web サイトの悪性度が閾値以上の場合にユーザへ追加認証を要求しているが、ユーザの認証結果をキャッシュしておき、該ユーザが該アクセス先サイトに対する追加認証を 1 度でも突破していた場合、追加認証を要求せずアクセスを許可しても良い。これにより、ユーザの利便性が効率することが期待できる。

10

【 0 0 7 8 】

次に、図 9 を参照して、Web サイト悪性度判定プログラム 1 0 9 の処理フローの一例について説明する。

図 9 に示すように、CPU 1 0 3 により実行され、ユーザ端末 1 2 1 の通信先のうち、ブラックリストと、ホワイトリストとに記録されていないものをアクセス中継プログラム 1 0 8 より受信すると、処理を開始する（ステップ 9 0 1）。

【 0 0 7 9 】

なお、Web サイト悪性度判定プログラムは、それぞれが重みを持つ複数の悪性度推測器 1 2 4 から構成されており、各推測器の判断結果を統合して最終的な値を算出する。ここでの悪性度推測器 1 2 4 としては、URL の文字列から Web サイトの悪性度を推測するもの、Web サイトに関する外部情報（whois 情報や DNS 情報等）から Web サイトの悪性度を推測するもの、および Web サイトのコンテンツから Web サイトの悪性度を推測するものが考えられる。

20

【 0 0 8 0 】

Web サイト悪性度判定プログラム 1 0 9 は、アクセス中継プログラム 1 0 8 から受信したアクセス先 Web サイトを各悪性度推測器 1 2 4 へ入力し、ステップ 9 0 3 へ進む（ステップ 9 0 2）。

【 0 0 8 1 】

Web サイト悪性度判定プログラム 1 0 9 は、各悪性度推測器 1 2 4 の判定結果の受信を開始し、全ての推測器から結果を受信した後、ステップ 9 0 4 へ進む（ステップ 9 0 3）。

30

【 0 0 8 2 】

Web サイト悪性度判定プログラム 1 0 9 は、ステップ 9 0 3 で受信した各推測器の判定結果を推測器管理表 1 1 6 と推測器重み履歴 1 1 7 へ記録し、ステップ 9 0 5 へ進む（ステップ 9 0 4）。

【 0 0 8 3 】

Web サイト悪性度判定プログラム 1 0 9 は、ステップ 9 0 3 で受信した各推測器の推測結果を推測器管理表 1 1 6 を参照することによって得られる各推測器に付与された重みに掛け合わせ、その合計値を最終予測結果として算出し、処理を終了する（ステップ 9 0 5）。なお、上述の算出手順は、以下の計算式（数 1）のように示すことができる。

40

【 0 0 8 4 】

【 数 1 】

$$\sum_{i=0}^n P_i \times w_i \quad (n: \text{推測器の数、} w_i: \text{各推測器の重み、} P_i: \text{各推測器の推測結果})$$

【 0 0 8 5 】

なお、今回は上述の計算式を用いた最終予測結果の算出手順を例示したが、各推測器の予測結果に対して重みを反映させ、最終的に統合する方式であれば、どのような方式を用いても良い。また、悪性度推測器 1 2 4 の追加及び削除は可能であり、管理者が必要に応じて該悪性度推測器 1 2 4 の追加及び削除を行っても良い。

50

【0086】

次に、図10を参照して、リスト更新プログラム112の処理フローの一例について説明する。

図10に示すように、CPU103により実行され、ユーザ端末121の通信先のうち、Webサイト悪性度判定プログラム109によって、その悪性度が閾値以上と判定されたWebサイトに対するユーザの追加認証結果をアクセス中継プログラム108より受信すると、処理を開始する(ステップ1001)。

【0087】

リスト更新プログラム112は、グレーリスト118を参照し、アクセス先Webサイトが該グレーリストに該当した場合はステップ1004に進み、該当しなかった場合はステップ1003に進む(ステップ1002)。

10

【0088】

リスト更新プログラム112は、アクセス先Webサイトをグレーリスト118に登録し、ステップ1004に進む(ステップ1003)。

【0089】

リスト更新プログラム112は、アクセス中継プログラム108より受信したユーザの追加認証結果を参照し、該追加認証結果をグレーリスト118に記録し、グレーリスト118の認証成功ユーザ数、認証失敗ユーザ数、およびユーザ毎の認証結果を更新した後、ステップ1005に進む(ステップ1004)。

20

【0090】

リスト更新プログラム112は、アクセス中継プログラム108より受信したユーザの追加認証結果を参照し、認証に成功していた場合はステップ1006に進み、認証に失敗していた場合はステップ1009に進む(ステップ1005)。

【0091】

リスト更新プログラム112は、グレーリスト118を参照し、アクセス先Webサイトの累計認証成功ユーザ数が一定数を越えていた場合はステップ1007に進み、越えていなかった場合は処理を終了する(ステップ1006)。

【0092】

リスト更新プログラム112は、アクセス先Webサイトをホワイトリスト120に登録し、ステップ1008に進む(ステップ1007)。

30

【0093】

リスト更新プログラム112は、アクセス先Webサイトに対応する行をグレーリスト118から削除し、処理を終了する(ステップ1008)。

【0094】

リスト更新プログラム112は、グレーリスト118を参照し、アクセス先Webサイトの累計認証失敗ユーザ数が一定数を越えていた場合はステップ1010に進み、越えていなかった場合は処理を終了する(ステップ1009)。

【0095】

リスト更新プログラム112は、アクセス先Webサイトをブラックリスト119に登録し、ステップ1011に進む(ステップ1010)。

40

【0096】

リスト更新プログラム112は、アクセス先Webサイトに対応する行をグレーリスト118から削除し、処理を終了する(ステップ1011)。

【0097】

次に、図11を参照して、フィードバック値算出プログラム113の処理フローについて説明する。

図11に示すように、CPU103により実行され、ユーザ端末121の通信先のうち、Webサイト悪性度判定プログラム109によって、その悪性度が閾値以上と判定されたアクセス先Webサイトに対するユーザの追加認証結果をアクセス中継プログラム108より受信すると、処理を開始する(ステップ1101)。

50

【0098】

なお、今回は認証結果202と付随情報203として認証に要した時間204、正解数205、およびサイト情報表示有無206を用いた場合の一例を説明するが、該情報以外を用いることや同様の情報を用いた場合でも異なる処理フローで最終的な値を算出することも可能である。

【0099】

フィードバック値算出プログラム113は、値の初期値1.0を設定し、ステップ1103に進む(ステップ1102)。

【0100】

フィードバック値算出プログラム113は、追加認証履歴115の最新エントリから認証情報を取得し、ステップ1104に進む(ステップ1103)。

10

【0101】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、認証に成功していた場合はステップ1109に進み、認証に失敗していた場合はステップ1105に進む(ステップ1104)。

【0102】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、認証がタイムアウトしていた場合はステップ1112に進み、していなかった場合はステップ1106に進む(ステップ1105)。

【0103】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、サイト情報の表示があった場合はステップ1112に進み、なかった場合はステップ1107に進む(ステップ1106)。

20

【0104】

フィードバック値算出プログラム113は、値を0.5倍し、ステップ1108に進む(ステップ1107)。

【0105】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、値を正解数に応じて減少し、ステップ1112に進む(ステップ1108)。

【0106】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、値を認証に要した時間に応じて減少し、ステップ1110に進む(ステップ1109)。

30

【0107】

フィードバック値算出プログラム113は、ステップ1103で取得した認証情報を参照し、サイト情報の表示があった場合はステップ1112に進み、なかった場合はステップ1111に進む(ステップ1110)。

【0108】

フィードバック値算出プログラム113は、値を0.5倍し、ステップ1112に進む(ステップ1111)。

40

【0109】

フィードバック値算出プログラム113は、ここまでのステップで算出した最終的な値を返却し、処理を終了する(ステップ1112)。

【0110】

本処理により、ユーザの認証結果の確からしさを数値化し、確からしさに応じてWebサイト悪性度判定プログラム109を構成する各推測器124へフィードバックを行うことが可能となる。

【0111】

なお、今回付随情報203として認証に要した時間204、正解数205、およびサイト情報表示有無206を用いた理由は、以下の通りである。

50

【0112】

認証に要した時間204は、主に人間によるアクセスとマルウェアによるアクセスを分別するために用いる。不審サイトへアクセスした際に要求する追加認証は、通常のWebアクセス時には発生せず、人間は追加認証が出た際に人間は柔軟に対応できるのに対し、追加認証を想定していないマルウェアは、これを突破できずタイムアウトになる。

【0113】

あるいは、近年の攻撃の高度化に伴い、追加認証を突破する機能を持つマルウェアも報告されているが、この場合は、プログラムによって人間では達成困難な入力速度で認証突破を試みると推察される。

【0114】

以上のように、マルウェアが認証に要する時間には、タイムアウト、あるいはきわめて短いといった特徴があり、人間による認証と分類に有用であることから、該情報を利用した。

10

【0115】

また、認証結果が成功だった場合、フィードバック値算出プログラム113は、ステップ1109において、認証に要した時間に応じてフィードバックする値を減算している。一方で、認証結果が失敗の場合は、同減算は行っていない。

【0116】

これは、アクセス先Webサイトが不審な場合でもリスクを鑑みずにアクセスしてしまうユーザを考慮したものである。セキュリティレベルの低いユーザは、アクセス先Webサイトが不審であると警告を受けたとしても、そのリスクを鑑みずにアクセスしてしまうことが考えられる。

20

【0117】

一方で、追加認証を突破せずにアクセスを中止した場合は、アクセス先Webサイトのリスクを考慮した上のものであると推察される。上述の追加認証結果の性質から、認証成功の信頼性は、認証失敗の信頼性よりも高いと推察される。

【0118】

以上の理由から、認証突破の場合は、認証を突破しなかった場合にはないフィードバックする値の減算を行っており、これによって、該追加認証の信頼性を反映している。

【0119】

正解数205は、主に認証失敗時、該失敗がヒューマンエラーによるものか否かを分別するために用いる。例えば、認証に失敗した際、意図したものであれば、何も入力しない、あるいは適当な文字を入力することにより、認証突破のために要求された文字列から大きく外れることが推察される。

30

【0120】

一方で、意図しないもの（ヒューマンエラー）は、突破しようとしたものの外れてしまった入力であることから、認証突破のために要求された文字列から大きくは外れないことが推察される。以上のように、認証の失敗がヒューマンエラーによるか否かは、認証突破のために要求された文字列に対する入力の正解数に現れることから、該情報を利用した。

【0121】

サイト情報表示有無206は、主にユーザの認証結果の確からしさを検証するために用いる。Webアクセス制御装置101は、前述の通り、ユーザが追加認証を要求された際にユーザの要求に応じてアクセス先Webサイトの情報を追加で取得・表示する機能を有する。

40

【0122】

該情報を要求したユーザは、追加情報を要求し、アクセス先Webサイトの性質を多角的に判断しようとするセキュリティ意識の高いユーザであることが推察されることに加え、その際の判断は追加情報を見た上のものであることから、該追加情報を見ていない判断よりも信頼性が高いと推察される。以上のように、サイト情報表示有無206は、ユーザの認証結果の信頼性を検証するのに有用であることから、該情報を利用した。

50

【0123】

次に、図12を参照して、重み更新プログラム114の処理フローについて説明する。

図12に示すように、CPU103により実行され、フィードバック値算出プログラム113よりWebサイト悪性度判定プログラムを構成する各悪性度推測器124に反映するフィードバック値を受信すると、処理を開始する(ステップ1201)。

【0124】

重み更新プログラム114は、悪性度推測器124のIDを意味する変数*i*に初期値0を設定し、ステップ1203に進む(ステップ1202)。

【0125】

重み更新プログラム114は、追加認証履歴115と推測器管理表116を参照し、IDが変数*i*の推測器の推測結果と認証結果を取得した後、ステップ1204に進む(ステップ1203)。

10

【0126】

重み更新プログラム114は、ステップ1103で取得したIDが変数*i*の推測器の推測結果と認証結果を比較し、両者が一致していた場合はステップ1205に進み、一致していなかった場合はステップ1206に進む(ステップ1204)。

【0127】

重み更新プログラム114は、IDが変数*i*の推測器の重みをフィードバック値算出プログラム113より受信した値に応じて加算し、推測器管理表116に記録されている該推測器の重みを加算後の値に更新した後、ステップ1207へ進む(ステップ1205)

20

【0128】

重み更新プログラム114は、IDが変数*i*の推測器の重みをフィードバック値算出プログラム113より受信した値に応じて減算し、推測器管理表116に記録されている該推測器の重みを減算後の値に更新した後、ステップ1207へ進む(ステップ1206)

【0129】

重み更新プログラム114は、変数*i*に1加算し、ステップ1203に進む(ステップ1207)。

【0130】

重み更新プログラム114は、変数*i*と推測器管理表116を参照することによって得ることのできる推測器数を比較し、変数*i*が推測器数に満たない場合はステップ1203に戻り、推測器数を上回っていた場合はステップ1209へ進む(ステップ1208)。

30

【0131】

重み更新プログラム114は、推測器の重みの合計が1になるように、推測器管理表116に記録されている各悪性度推測器124の重みを正規化し、処理を終了する(ステップ1209)。

【0132】

図13a、図13bを参照して、追加認証要求プログラム110によって表示される追加認証画面の一例について説明する。

40

図13aに示すように、アクセス先Webサイト悪性度判定プログラムによって閾値以上の悪性度を示したWebサイトへアクセスしようとした際、追加認証画面を表示し、画面内の追加認証を突破した場合のみ、Webサイトへのアクセスを許可する。追加認証画面は、アクセス先Webサイトが不審であるという警告文1301と、追加認証用の文字列表示欄1302と、追加認証文字列入力フォーム1303と、追加認証文字列送信ボタン1304と、追加情報要求ボタン1305とから構成される。

【0133】

ユーザは、該Webサイトが悪性でないと判断した場合のみ追加認証用の文字列表示欄1302に示された文字列を追加認証文字列入力フォーム1303に正しく入力し、追加認証文字列送信ボタン1304を押下することにより、アクセスを続行できる。また、ユ

50

ーザがアクセス可否の判断に窮した際には、追加情報要求ボタン1305を押下することにより、該判断のための追加情報を取得・表示できる。

【0134】

追加情報表示後の画面の一例を図13bに示す。

追加情報表示後の追加認証画面は、Webサイト情報表示欄1306と、Webサイトサムネイル表示欄1307とから構成される。Webサイト情報表示欄1306には、whois情報やDNS情報といったアクセス先Webサイトに関する公開情報が表示される。Webサイトサムネイル表示欄1307には、実際に該アクセス先Webサイトにアクセスした際のサムネイルが画像で表示される。ユーザはこれらの情報を基に、アクセス先Webサイトへのアクセス可否を判断する。

10

なお、実施例1の一部を変更して以下のように実施しても良い。

【0135】

アクセス先Webサイトの悪性度の最終予測結果が閾値以下であっても、ある1つ以上の推測器が高い悪性度を算出した場合は、追加認証を要求しても良い。これにより、汎用性の低さから重みが少なくなりがちな特定の悪性サイトを検出することに特化したような推測器であっても、Webサイト悪性度判定プログラム110に組み込むことが可能になり検出漏れを抑制できる。

【0136】

また、悪性度の値によって追加認証の難易度を上下させても良い。例えば、悪性度が相対的に低い場合はボタンのクリックのみ、高い場合はCAPTCHA認証を行うことが考

20

【0137】

また、検証後のユーザの追加認証結果を正解データ(Ground Truth)として扱い、データが与えられる毎に逐次的に学習し、モデルを構築するオンライン学習の教師データに利用しても良い。これにより、各推測器に付与する重みの最適化だけでなく、推測器そのものの最適化も可能となる。

【実施例2】

【0138】

次に、図14を参照して、実施例2のWebアクセス制御装置について説明する。

30

実施例2は、実施例1に係るWebアクセス制御装置101を含み、さらに悪性度推測器124に付与する重みを更新する際に、該重みの時系列変化に着目し、より適した形で重みを付与する機能を有するWebアクセス制御装置である。

【0139】

実施例2では、推測器の重みの時系列変化に着目し、重みの変化にこれまでと異なる傾向が見られた場合、重みの更新を保留する。例えば、潜在的に精度の高い推測器の重みは、単調増加を続け、上昇値が飽和した値へ収束することが推察されるが、ユーザの認証結果があるべきものと異なっていた場合、単調増加、あるいは一定値に収束していた重みが減少方向に向かう。この変化点を検出し、その場合重みの更新を一旦保留することにより、その重みの変動が正しいものなのか、あるいはユーザの認証ミス等によるノイズかを見極め、不適切な重みの上下を抑制でき、Webサイト悪性度の判定精度の低下を抑制できる。

40

【0140】

図14は、本実施例における重み更新フローの例である。

実施例1と同一の構成要素には同一の符号を付すことによってその説明を省略し、以下では、実施例1と異なる点を中心に説明する。

【0141】

重み更新プログラム114は、図12に示したフローに加え、重み更新時にその時系列変化に着目した処理フローを有する。具体的には、ステップ1204において、該推測器の推測結果と認証結果が一致したか否かを検証し、重みを加算/減算するフローに分岐し

50

た際に、重みの時系列変化を基に実際に重みを加算/減算するか判断するステップ1210とステップ1211を有する。以降では、ステップ1210とステップ1211の処理内容について説明する。

【0142】

重み更新プログラム114は、IDが変数*i*の推測器の重みの時系列変化を該推測器に対応する推測器時系列重み管理表117を参照し取得する。また、仮に重みを加算した場合、重みの変化の傾向がこれまでと異なるか検証し、これまでの変化と異なる傾向が見られた場合、重み更新を保留するために、ステップ1207へ進む。重みの変化にこれまでと異なる傾向が見られなかった場合は、ステップ1205へ進む(ステップ1210)。

【0143】

重み更新プログラム114は、IDが変数*i*の推測器の重みの時系列変化を該推測器に対応する推測器時系列重み管理表117を参照し取得する。また、仮に重みを減算した場合、重みの変化の傾向がこれまでと異なるか検証し、これまでの変化と異なる傾向が見られた場合、重み更新を保留するために、ステップ1207へ進む。重みの変化にこれまでと異なる傾向が見られなかった場合は、ステップ1206へ進む(ステップ1211)。

【実施例3】

【0144】

図15を参照して、実施例3のWebアクセス制御装置について説明する。

実施例3は、実施例1に係るWebアクセス制御装置101を含み、さらにWebサイト悪性度判定プログラム109において、複数の悪性度推測器124を階層的に並べることによって、悪性度算出時間を効率化する機能を有する。

【0145】

実施例3では、複数の悪性度推測器124を処理時間の比較短い表層解析部分(第1のグループ)と、処理時間を要する分より高精度でWebサイトの悪性度を算出可能な詳細解析部分(第2のグループ)に階層的に分割する。これにより、明らかに良性、あるいは明らかに悪性なWebサイトに対しては、処理時間の短い表層解析部分の複数の悪性度推測器124のみでWebサイトの性質を判定できるため、判定に要する処理時間を短縮できる。また、表層解析部分ではWebサイトの性質を判断することが難しい場合は、詳細解析部分を用いて判断を行うため、判定精度が著しく低下することも無い。

【0146】

図15は、実施例3における推測器群構成方式の一例である。なお、図15は、実施例1におけるステップ804でのWebサイト悪性度判定プログラム109において、複数の悪性度推測器124を階層的に並べた場合の処理例である。Webサイト悪性度判定プログラム109以外の処理フローは実施例1と同じであるためその説明は省略し、以下では、実施例1と異なる点である階層的に複数の悪性度推測器124を並べた場合の推測結果導出方法を中心に説明する。

【0147】

Webサイト悪性度判定プログラム109は、図9に示したフローに加え、複数の悪性度推測器124を表層解析部分(悪性度推測器A、悪性度推測器B)と詳細解析部分(悪性度推測器C、悪性度推測器D、悪性度推測器E)に分割することによる追加の悪性度計算フローを有する。具体的には、ステップ902において、全ての複数の悪性度推測器124にアクセス先Webサイトを入力するのではなく、表層解析部分(悪性度推測器A、悪性度推測器B)にのみ入力する。その後、表層解析部分(悪性度推測器A、悪性度推測器B)から推測結果(推測結果A、推測結果B)を受信し、その推測結果(途中推測結果)が明らかに良性である、あるいは明らかに悪性である場合は、推測結果を最終予測結果としてステップ904に進む。

【0148】

ここでの明らかに悪性とは、例えば、悪性度が90%以上、明らかに良性とは、例えば悪性度が10%以下のような状態を指す。なお、具体的な数値は、この限りではない。表層解析部分(悪性度推測器A、悪性度推測器B)で明らかに良性、あるいは明らかに悪性

10

20

30

40

50

だと判断できなかった場合は、詳細解析部分（悪性度推測器 C、悪性度推測器 D、悪性度推測器 E）へと処理を進める。この際、詳細解析部分を構成する複数の悪性度推測器 C、悪性度推測器 D、悪性度推測器 E にアクセス先 Web サイトを入力する。その後、推測結果（推測結果 C、推測結果 D、推測結果 E）を受信し、該推測結果を最終予測結果とする。

【0149】

なお、図 15 では、複数の悪性度推測器 124 を 2 段に分割したが、3 段以上に分割しても良い。実施例 3 では、以上の手法によって、悪性度算出時間を効率化することを可能とする。

【実施例 4】

10

【0150】

図 16 を参照して、実施例 4 の Web アクセス制御装置について説明する。

実施例 4 は、実施例 1 に係る Web アクセス制御装置 101 を含む。さらに Web サイト悪性度判定プログラム 109 において、複数の悪性度推測器 124 から受領した悪性度を基に最終予測結果を算出する際の手法に着目し、より適した形で悪性度を算出することによって Web サイトの悪性度判定精度を向上する機能を有する。

【0151】

実施例 4 は、実施例 1 で示した最終予測結果算出方法（a）の他に、複数の悪性度推測器 124 の推測結果から多数決をとる方法（b）、および実施例 1 の算出方法に加えて特定の場合に単体の推測器が算出した予測結果を重用する方法（c）を有する。

20

【0152】

多数決をとる方法（b）では、悪性度推測器 A、悪性度推測器 B、悪性度推測器 C において、予測結果と重みを掛け算した後に合計をとる方法（a）における計算処理が不要になるため処理時間を短縮できる。

【0153】

特定の場合に単体の推測器が算出した予測結果を重用する方法（c）では、特定の性質を有する Web サイトの性質判断に特化した悪性度推測器 C を有効活用し全体としての推測精度を向上できる。例えば、フィッシングサイトやマルウェアダウンロードサイトのような特定の悪性サイトに特化した悪性度推測器 C は、よりアクセスの多い良性サイトに対しては推測精度が低く、普段の運用の中で、悪性度推測器 C に与えられる重みは小さくなることが考えられる。

30

【0154】

この場合には、悪性度推測器 C の得意とするフィッシングサイトやマルウェアダウンロードサイトへのアクセスが発生した場合でも悪性度推測器 C の重みが小さく、最終予測結果にほとんど反映されない。このため、最終予測結果が適切に算出できない可能性がある。そこで、上述した特定の悪性サイトに特化した悪性度推測器 C がアクセス先の Web サイトに対し、高い悪性度を示した場合、全体としての予測結果が良性であったとしても、悪性の恐れがあるとして利用者に追加認証を要求する。これにより、悪性サイトの見落としを抑制し、前述したような全体としての推測精度向上が可能になる。

【0155】

40

次に、図 16 を参照して、実施例 4 における最終予測結果算出フローの一例について説明する。

なお、図 16 は、実施例 1 におけるおよびステップ 805 での Web サイト悪性度判定プログラム 109 において、アクセス先 Web サイトの性質が悪性か良性か判定する際の各手法における一例である。

【0156】

また、左から順に、実施例 1 の推測結果と重みの積和をとる方法（a）、多数決をとる方法（b）、および特定の悪性度推測器を重用する方法（c）を図示している。Web サイト悪性度判定プログラム以外の処理フローは実施例 1 と同一のためその説明は省略する。以下では、実施例 1 と異なる方法である多数決をとる方法（b）と、特定の悪性度推測

50

器を重用する方法(c)を中心に説明する。

【0157】

まず、多数決をとる方法(b)について説明する。

多数決をとる方法は、悪性度推測器A、悪性度推測器B、悪性度推測器Cにおける推測結果を事前に制定していた閾値と比較し、悪性度推測器A、悪性度推測器B、悪性度推測器Cごとにアクセス先Webサイトの性質が悪性か良性かを推定する。その後、悪性度推測器A、悪性度推測器B、悪性度推測器Cから集計した性質の合計数を比較し、該合計数が最も多いものを最終予測結果として算出する。(b)の例では、3つの悪性度推測器A、悪性度推測器B、悪性度推測器Cのうち、2つが良性、1つが良性と推測しており、良性の方が多いため、最終予測結果を良性としている。

10

【0158】

次に、特定の推測器を重用する方法(c)について説明する。

特定の推測器を重用する方法では、まず重用する悪性度推測器Cが高い悪性度を示していないか確認する。このとき、悪性度推測器Cが高い値を出していた場合は、他の悪性度推測器A、悪性度推測器Bが算出した結果に関わらず、Webサイトは悪性の可能性があるとして利用者に追加認証を要求する。悪性度推測器Cが高い値を出していなかった場合は、実施例1と同様の方法で最終予測結果を算出し、該値に応じてアクセス制御を実施する。

【0159】

(c)の例では、3つの悪性度推測器A、悪性度推測器B、悪性度推測器Cのうち、悪性度推測器Cが特定の悪性サイトに特化したものである。悪性度推測器Cが高い悪性度を出していることから、アクセス先Webサイトを悪性と判定している。なお、実施例1の計算方式だとアクセス先Webサイトは良性と判定される。また、重用する特定の悪性度推測器Cは、管理者が事前に指定しているものとし、本方法では、図3の推測器管理表116に各悪性度推測器について重用するか否かを表す追加項目を持つ。

20

【0160】

実施例4では、以上の手法によって、処理時間の短縮や推測精度の向上が可能となる。

【実施例5】

【0161】

図17a及び図17bを参照して、実施例5のWebアクセス制御装置について説明する。

30

実施例5は、実施例1に係るWebアクセス制御装置101を含む。さらに追加認証要求プログラム110において、状況に応じて追加認証方法を変更することによって、利用者の利便性を向上するとともに、悪性サイトへの誤アクセスをより高い確度で抑制する機能を有する。実施例5では、以下の3つの観点で追加認証の難易度を変更する。

【0162】

1つ目は、悪性度に応じた追加認証の難易度変更である。具体的には、悪性サイトへのアクセスの可能性があるとして利用者に追加認証を要求する際、Webサイトの悪性度を非常に高いと判断した際には、高難易度の追加認証を要求する。一方、Webサイトの悪性度が閾値よりは高いものの閾値に近く悪性と判断されるものの中では相対的に低いと破断した際には、低難易度の追加認証を要求する。これにより、より悪性度の高い可能性があるWebサイトに誤アクセスする可能性を抑制し、悪性度が相対的には低いサイトにはアクセス性を向上して利便性を向上する。

40

【0163】

本方法では、悪性度が何%以上であれば非常に高いか、悪性度が何%以下であれば悪性度が相対的に低いかを表す追加の表を図1の記憶装置105の部分を持つ。また、各値は、事前に管理者が制定するものとする。

【0164】

2つ目は、グレーリストにおける情報源の信頼度に応じた追加認証の難易度変更である。グレーリストには、手動で悪性サイトらしいWebサイトを追加できるが、この際の情

50

報源としては、信頼性の高い組織が公開しているものから、そうでないものまでが含まれる。

【0165】

そこで、該情報源の信頼度に応じて、追加認証の難易度を変更する。信頼性の高い情報源から得られたWebサイトへのアクセスに対する追加認証には、低難易度のものを利用する。一方、そうでない情報源から得られたWebサイトへのアクセスに対する追加認証には、高難易度のものを利用する。これにより、利用者の利便性を向上するとともに、悪性サイトへの誤アクセスをより高い確度で抑制する。

【0166】

本方法では、図5のグレーリスト118に格納された各URLについて、情報源の信頼性を、例えば、高、中及び低で表す追加項目を持つ。また、各値は、事前に管理者が制定するものとし、情報源の信頼性を表す方法は、上記の限りではない。

10

【0167】

3つ目は、特定の悪性サイトに特化した悪性度推測器が高い悪性度を示していた場合における追加認証の高難易度化である。特定の悪性サイトに特化した悪性度推測器が高い悪性度を示していた場合は、アクセスするべきでない特定の性質を持った悪性サイトである可能性が高まる。特に、悪性度推測器が高い悪性度を示し、かつ全体の推定結果としては良性サイトよりの判定であった場合は、他の悪性度推測器では検知できないような良性サイトを模した悪性サイトである可能性がある。このため、Webサイトへのアクセスはより遮断する必要がある。

20

【0168】

以上の理由から、利用者に要求する追加認証を高難易度のものにし、これにより、悪性サイトへの誤アクセスをより高い確度で抑制する。また、重用する特定の悪性度推測器は、管理者が事前に指定しているものとし、本方法では、図3の推測器管理表116に各悪性度推測器について重用するか否かを表す追加項目を持つ。

【0169】

実施例5においては、どの認証方式を利用するかを表す追加の表を図1の記憶装置105の部分を持つ。どの認証方式を利用するかは、事前に管理者が制定するものとする。また、実施例5では、難易度の異なる追加認証方式を実施例1に示したものの他に2つ有する。

30

【0170】

図17aは、低難易度の追加認証方式を用いた追加認証画面の一例を示す図である。

実施例1のように文字列を画面から読み取り入力するといったステップは無く、画面に表示された接続ボタン1701を押下するだけでWebサイトへのアクセスを続行できる。

【0171】

図17bは、高難易度の追加認証方式を用いた追加認証画面の一例を示す図である。

追加認証用の計算式表示欄1702を備え、計算式表示欄1702から計算式を読み取り、計算式の計算結果を追加認証文字列入力フォーム1703に正しく入力し、追加認証文字列送信ボタン1704を押下することにより、アクセスを続行できる。

40

【0172】

実施例5では、以上の方法によって、利用者の利便性を向上するとともに、悪性サイトへのアクセスをより高い確度で抑制することが可能となる。

【符号の説明】

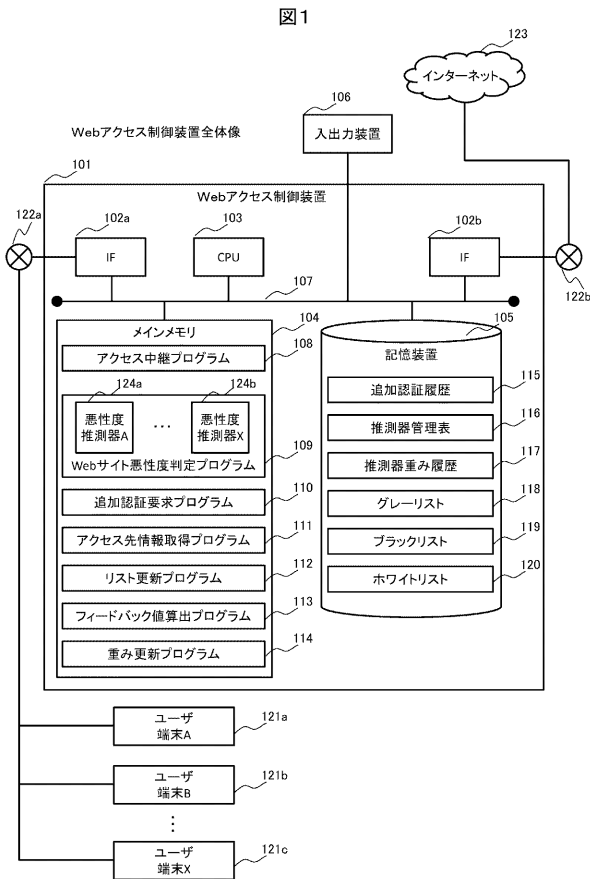
【0173】

- 101：Webアクセス制御装置
- 108：アクセス中継プログラム
- 109：Webサイト悪性度判定プログラム
- 110：追加認証要求プログラム
- 111：アクセス先情報取得プログラム

50

- 1 1 2 : リスト更新プログラム
- 1 1 3 : フィードバック値算出プログラム
- 1 1 4 : 重み更新プログラム
- 1 1 5 : 追加認証履歴
- 1 1 6 : 推測器管理表
- 1 1 7 : 推測器重み履歴
- 1 1 8 : グレーリスト
- 1 1 9 : ブラックリスト
- 1 2 0 : ホワイトリスト
- 1 2 1 : ユーザ端末
- 1 2 3 : インターネット
- 1 2 4 : 悪性度推測器

【 図 1 】



【 図 2 】

追加認証履歴115

図2

ID	認証結果	付随情報 204			ユーザ識別情報 207	URL 208
		認証に要した時間 (s) 202	正解数 203	サイト情報表示有無 206		
1	成功	4.3	4 (100%)	無	123	foo.com
2	失敗	7.2	2 (50%)	有	456	bar.net
3	失敗	timeout	0 (0%)	無	789	baz.jp
...						

【 図 3 】

推測器管理表116

図3

ID 301	正解率 302	重み 303	直近の悪性度判定結果 304
0	80%	0.4	75%
1	90%	0.5	82%
2	30%	0.05	19%
...			

【図4】

図4

推測器重み履歴117

時刻	重み	悪性度判定結果	認証結果との一致
2016/12/01 21:04:42.20	0.3	12%	一致
2016/12/01 21:04:46.30	0.4	82%	一致
2016/12/01 21:04:51.40	0.5	64%	不一致
2016/12/01 21:06:03.50	0.4	75%	不一致
...			

【図5】

図5

グレーリスト118

ID	URL	認証成功ユーザ数	認証失敗ユーザ数	ユーザ毎の認証結果
0	example.com	2	0	123:成功, 456:成功
1	malware.com	0	4	123:失敗, 789:失敗...
2	abc.com	4	1	456:成功, 789:成功...
3	zyz.org	1	1	123:成功, 789:失敗
...				

【図7】

図7

ホワイトリスト120

ID	URL
0	white.com
1	safe.com
2	ok.com
...	

【図6】

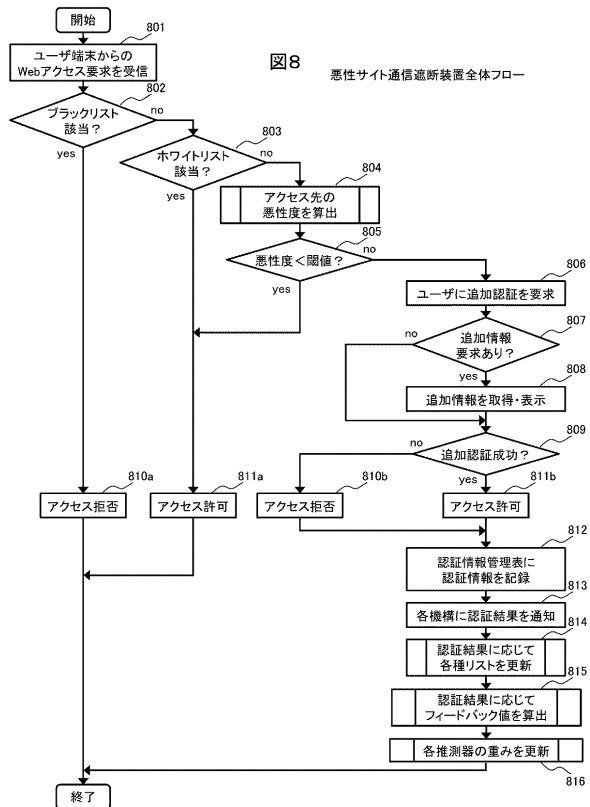
図6

ブラックリスト119

ID	URL
0	black.com
1	unsafe.com
2	ng.com
...	

【図8】

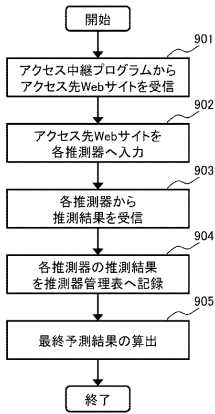
図8 悪性サイト通信遮断装置全体フロー



【 図 9 】

図9

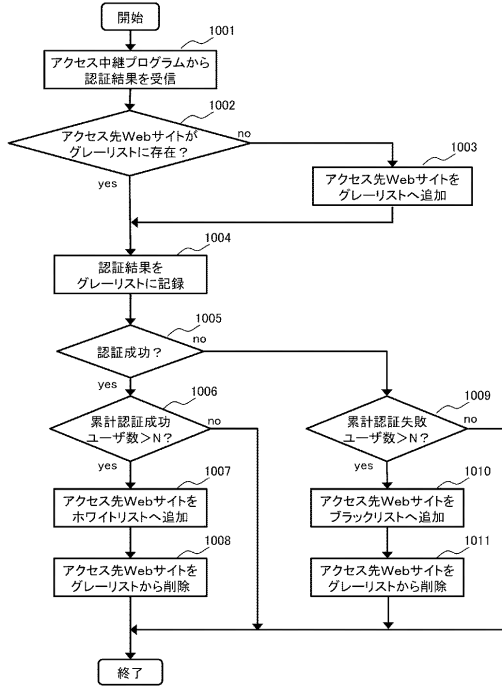
Webサイト悪性度判定フロー



【 図 10 】

図10

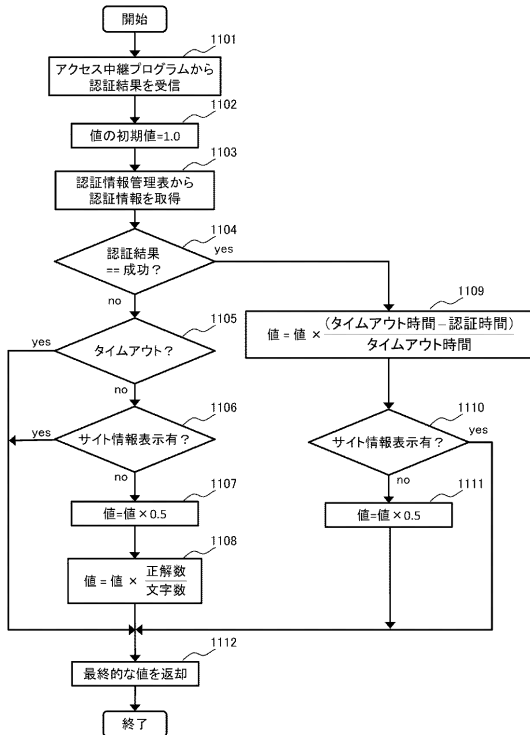
リスト更新フロー



【 図 11 】

図11

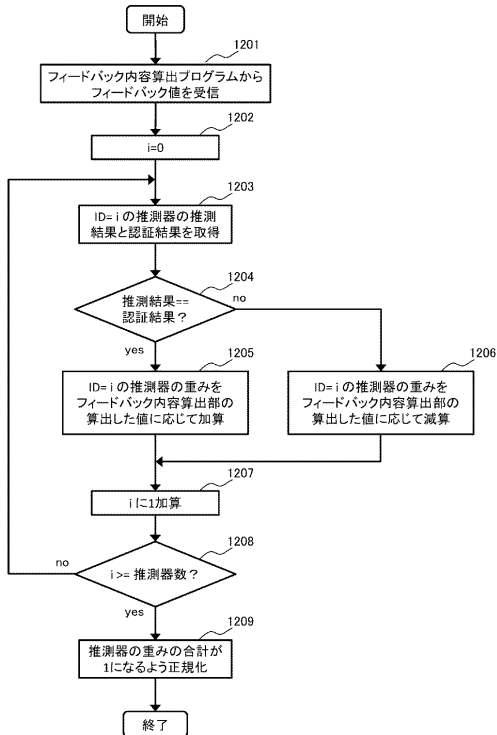
フィードバック内容算出フロー



【 図 12 】

図12

重み更新フロー



【 図 1 3 a 】

図13a

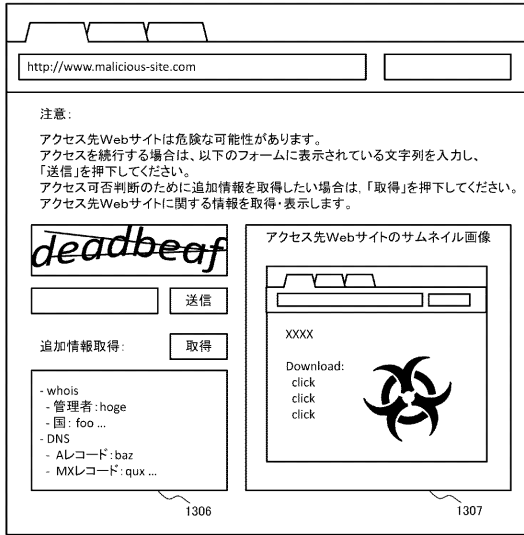
追加認証画面の一例(追加情報取得前)



【 図 1 3 b 】

図13b

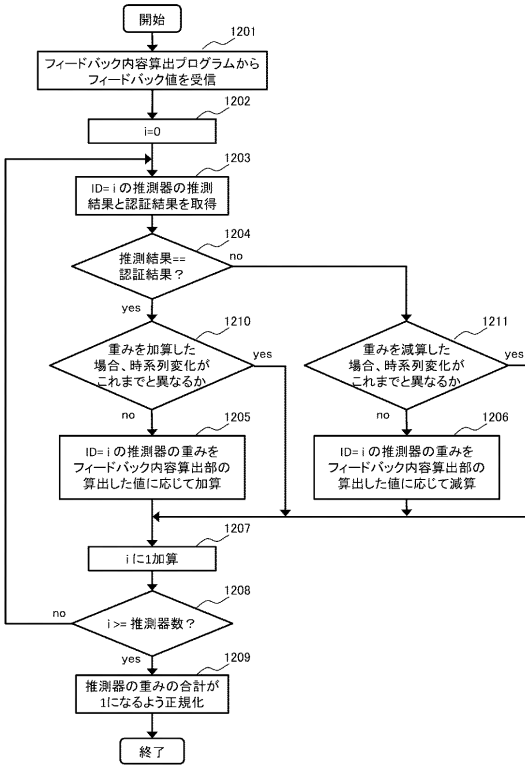
追加認証画面の一例(追加情報取得後)



【 図 1 4 】

図14

重み更新フロー(時系列変化考慮)



【 図 1 5 】

図15

推測器を階層的に並べた場合の悪性度判定例

