(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6845819号
(P6845819)
(45)発行日　令和3年3月24日(2021.3.24)　(24)登録日　令和3年3月2日(2021.3.2)

(51)Int.Cl.　　F I
G06F 16/90 (2019.01)　G06F 16/90
G06F 11/34 (2006.01)　G06F 11/34　１５２
G06F 21/55 (2013.01)　G06F 21/55

請求項の数 13　(全 25 頁)

(21)出願番号　特願2018-30182(P2018-30182)
(22)出願日　平成30年2月22日(2018.2.22)
(65)公開番号　特開2019-144970(P2019-144970A)
(43)公開日　令和1年8月29日(2019.8.29)
審査請求日　令和2年2月18日(2020.2.18)

(73)特許権者　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　110001678
特許業務法人藤央特許事務所
(72)発明者　三村　和
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　對馬　雄次
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　池上　幸三
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

最終頁に続く

(54)【発明の名称】分析装置、分析方法、および分析プログラム

(57)【特許請求の範囲】
【請求項１】
プロセッサと、事象群の要因に対する結果を予測する予測モデル式を記憶する記憶デバイスと、を有する分析装置であって、
前記プロセッサは、
前記事象群の中の第１事象の要因に対する第１出現頻度を前記予測モデル式に与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を算出する予測誤差算出処理と、
前記事象群の中の第２事象の要因に対する第２出現頻度と、前記予測誤差算出処理によって算出された予測誤差と、の相関に基づいて、前記第１事象の要因の中から前記予測誤差の誤差要因を抽出する誤差要因抽出処理と、
を実行することを特徴とする分析装置。
【請求項２】
請求項１に記載の分析装置であって、
前記プロセッサは、
前記事象群の中の第３事象の要因に対する第３出現頻度と、前記第３出現頻度に対応する結果と、に基づいて、前記予測モデル式を作成する作成処理を実行し、
前記予測誤差算出処理では、前記プロセッサは、前記第１出現頻度を前記作成処理によって作成された予測モデル式に与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を算出する、

ことを特徴とする分析装置。
【請求項３】
請求項１に記載の分析装置であって、
前記事象群は、所定の時点以降に発生した事象の集合である、
ことを特徴とする分析装置。
【請求項４】
請求項１に記載の分析装置であって、
前記記憶デバイスは、前記第１事象の要因の重要度を示す重みを記憶しており、
前記プロセッサは、
前記第１事象の要因のうち前記誤差要因抽出処理によって抽出された誤差要因の重みを他の要因の重みよりも低くなるように設定する設定処理と、
前記事象群の中の第３事象の要因に対する第３出現頻度と、前記第３出現頻度に対応する結果と、前記設定処理によって設定された前記誤差要因の重みと、前記他の要因の重みと、に基づいて、前記予測モデル式を更新する更新処理と、
を実行することを特徴とする分析装置。
【請求項５】
請求項４に記載の分析装置であって、
前記予測誤差算出処理では、前記プロセッサは、前記更新処理による更新後の予測モデル式に前記第１出現頻度を与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を、算出する、
ことを特徴とする分析装置。
【請求項６】
請求項４に記載の分析装置であって、
前記プロセッサは、
前記更新処理による更新後の予測モデル式に、前記第２出現頻度を与えることにより、前記第２事象の第２予測値を算出する予測値算出処理と、
前記予測値算出処理によって算出された第２予測値を出力する出力処理と、
を実行することを特徴とする分析装置。
【請求項７】
請求項６に記載の分析装置であって、
前記プロセッサは、
前記第２事象の件数のうち前記第１予測値と第１結果との間に許容範囲外の誤差がある誤差件数に基づいて、前記設定処理および前記更新処理を試行するか否かを判断する判断処理を実行し、
前記プロセッサは、前記判断処理による判断結果に基づいて、前記設定処理および前記更新処理を試行する、
ことを特徴とする分析装置。
【請求項８】
請求項７に記載の分析装置であって、
前記プロセッサは、前記誤差件数がしきい値以上である場合、前記設定処理および前記更新処理を試行する、
ことを特徴とする分析装置。
【請求項９】
請求項７に記載の分析装置であって、
前記プロセッサは、前記誤差件数がしきい値以上でない場合、前記予測値算出処理および前記出力処理を試行する、
ことを特徴とする分析装置。
【請求項１０】
請求項６に記載の分析装置であって、
前記出力処理では、前記プロセッサは、前記予測モデル式の更新に用いられた要因を出

力する、
ことを特徴とする分析装置。
【請求項１１】
請求項６に記載の分析装置であって、
前記プロセッサは、
前記第３事象の要因に対する第３出現頻度と、前記第３出現頻度に対応する結果と、の相関を求め、当該相関と前記誤差要因とに基づいて、前記第３事象の要因の中から前記予測モデル式の精度を低下させる要因を抽出する要因抽出処理を実行し、
前記出力処理では、前記プロセッサは、前記要因抽出処理によって抽出された要因を出力する、
ことを特徴とする分析装置。
【請求項１２】
プロセッサと、事象群の要因に対する結果を予測する予測モデル式を記憶する記憶デバイスと、を有する分析装置による分析方法であって、
前記プロセッサは、
前記事象群の中の第１事象の要因に対する第１出現頻度を前記予測モデル式に与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を算出する予測誤差算出処理と、
前記事象群の中の第２事象の要因に対する第２出現頻度と、前記予測誤差算出処理によって算出された予測誤差と、の相関に基づいて、前記第１事象の要因の中から前記予測誤差の誤差要因を抽出する誤差要因抽出処理と、
を実行することを特徴とする分析方法。
【請求項１３】
事象群の要因に対する結果を予測する予測モデル式を記憶する記憶デバイスにアクセス可能なプロセッサに、
前記事象群の中の第１事象の要因に対する第１出現頻度を前記予測モデル式に与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を算出する予測誤差算出処理と、
前記事象群の中の第２事象の要因に対する第２出現頻度と、前記予測誤差算出処理によって算出された予測誤差と、の相関に基づいて、前記第１事象の要因の中から前記予測誤差の誤差要因を抽出する誤差要因抽出処理と、
を実行させることを特徴とする分析プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、データを分析する分析装置、分析方法、および分析プログラムに関する。
【背景技術】
【０００２】
サイバー空間では攻撃側が構造的に優位であり、その攻撃は日々高度化、増加、変化している。そのような中、攻撃対象は従来の金融サービス事業者やＩＴ（Ｉｎｆｏｒｍａｔｉｏｎ　Ｔｅｃｈｎｏｌｏｇｙ）サービス事業者からインフラ事業者へ拡大している。対策に必要な対策コストは右肩上がりだが、投資がそれに追いつかないのが現状である。セキュリティ専門家の人数も不足しており、将来に向けた人材確保が課題となっている。十分な数のセキュリティ専門家を確保できないために、情報システムや制御システムにおけるセキュリティインシデントの発生を監視するＳＯＣ（Ｓｅｃｕｒｉｔｙ　Ｏｐｅｒａｔｉｏｎ　Ｃｅｎｔｅｒ）の運用業務に支障を来たすことが懸念される。特に、社会インフラ事業者においては監視対象システム全体を監視する流れあり、これまでに比べて、ＳＯＣ１３０運用性能の大幅な向上が要求される。
【０００３】
ＳＯＣ１３０運用業務において最も工数を要するのは、ＦＷ（Ｆｉｒｅｗａｌｌ）／Ｉ

ＰＳ（Ｉｎｔｒｕｓｉｏｎ　Ｐｒｅｖｅｎｔｉｏｎ　Ｓｙｓｔｅｍ）などから通知されるセキュリティアラートの重要度を判断する作業（インシデントか誤検知かを人手で判断する作業）である。

【０００４】

従来、セキュリティアラートが発生した際には、ＳＯＣの専門家が監視対象システム内の各装置ログと外部脅威情報（ＵＲＬ（Ｕｎｉｆｏｒｍ　Ｒｅｓｏｕｒｃｅ　Ｌｏｃａｔｏｒ）やマルウェアの危険度評価など）を参照し、そのアラートの重要度を経験と勘に基づいて判断していた。増加し続けるサイバー攻撃や監視対象システムの大規模化に対して、将来に渡って持続可能なＳＯＣ運用を実現するには、上記セキュリティアラートの重要度判断を自動化、または支援することが必要である。

【０００５】

下記特許文献１の情報処理装置は、過去のアラートに関する通信情報の特徴量（ＩＰアドレス、ホスト名、検知ルール、一定時間内での同一アラート発生数、パケットペイロードのＮグラム出現頻度など）と、新たに発生したアラートに関する通信情報の特徴量の非類似度、すなわち距離を算出し、その距離と過去のアラートに対する判断結果から新たなアラートの重要度を決定する。

【０００６】

下記特許文献２の需要予測装置は、各種（来店者数、販売数量、電力消費量など）の需要量に関して、過去の需要量の予測値と実測値の誤差をとり、その誤差が異常値である場合には、それを目的変数として新たな説明変数を獲得して、その新たな説明変数を予測モデルに追加する。

【先行技術文献】

【特許文献】

【０００７】

【特許文献１】国際公開２０１６／２０８１５９号公報

【特許文献２】特開２０１７－１６６３２号公報

【発明の概要】

【発明が解決しようとする課題】

【０００８】

監視対象システムが大規模化すること、およびサイバー攻撃の手口が日々変化し増加していることに鑑みると、学習すべき各装置のログ項目や外部脅威情報の項目に関して、その特徴量の次元と値域は非常に多岐に渡り、かつ変化する。そのため、アラート重要度判断に影響を与えた要因の分析結果に多くのノイズを与え、その結果、正確に重要度を予測できなくなる問題がある。

【０００９】

また、一般に、特徴量の次元数が非常に多くなると算出される距離に差が出なくなる、すなわちすべてのアラートが類似に見えてしまう問題が知られている。したがって、特許文献１では、監視対象規模が大きくなる場合や、通信情報だけでなく多種多様なログを元にした特徴量を用いることで特徴量の次元数が非常に多くなる場合には、アラート重要度判断に対応することができない。また、特許文献２では、説明変数の増加に伴いノイズも増加することになり、逆に予測値の誤差が大きくなってしまう。

【００１０】

本発明は、予測誤差の誤差要因を特定することを目的とする。

【課題を解決するための手段】

【００１１】

本願において開示される発明の一側面となる分析装置は、プロセッサと、事象群の要因に対する結果を予測する予測モデル式を記憶する記憶デバイスと、を有する分析装置であって、前記プロセッサは、前記事象群の中の第１事象の要因に対する第１出現頻度を前記予測モデル式に与えることで得られる第１予測値と、前記第１出現頻度に対応する結果と、に基づいて、前記第１予測値の予測誤差を算出する予測誤差算出処理と、前記事象群の

中の第２事象の要因に対する第２出現頻度と、前記予測誤差算出処理によって算出された予測誤差と、の相関に基づいて、前記第１事象の要因の中から前記予測誤差の誤差要因を抽出する誤差要因抽出処理と、を実行することを特徴とする。

【発明の効果】

【００１２】

本発明の代表的な実施の形態によれば、予測誤差の誤差要因を特定することができる。前述した以外の課題、構成及び効果は、以下の実施例の説明により明らかにされる。

【図面の簡単な説明】

【００１３】

【図１】図１は、監視システムのシステム構成例を示すブロック図である。

【図２】図２は、図１に示した各種コンピュータのハードウェア構成例を示すブロック図である。

【図３】図３は、アラート分析装置の機能的構成例を示すブロック図である。

【図４】図４は、アラート判断とログ統計集計の動作シーケンス例を示すシーケンス図である。

【図５】図５は、アラート判断集計テーブルの一例を示す説明図である。

【図６】図６は、ログ統計集計テーブルの一例を示す説明図である。

【図７】図７は、データ種別管理テーブルの一例を示す説明図である。

【図８】図８は、分析期間Ｔのアラート重要度予測の動作シーケンス例を示すシーケンス図である。

【図９】図９は、抽出要因テーブルの一例を示す説明図である。

【図１０】図１０は、誤差テーブルの一例を示す説明図である。

【図１１】図１１は、図８の点線枠の処理の繰り返し試行の終了条件を示す説明図である。

【図１２】図１２は、誤差要因テーブルの一例を示す説明図である。

【図１３】図１３は、更新後の抽出要因テーブルの一例を示す説明図である。

【図１４】図１４は、重要度予測値の出力画面表示例を示す説明図である。

【図１５】図１５は、要因抽出部による要因抽出処理手順例を示すフローチャートである。

【図１６】図１６は、重要度予測部による重要度予測処理手順例を示すフローチャートである。

【発明を実施するための形態】

【００１４】

＜システム構成例＞

図１は、監視システムのシステム構成例を示すブロック図である。監視システム１は、監視対象システム１００と、ＳＯＣ１３０と、を有する。監視対象システムとＳＯＣ１３０は、通信可能に接続される。

【００１５】

監視対象システム１００は、ＳＯＣ１３０に監視されるシステムである。監視対象システム１００は、第１ネットワーク１１０、１台以上のクライアント端末１１１、業務サーバ１１２、ネットワーク監視装置１１３、第１ＦＷ／ＩＰＳ１１４、およびプロキシサーバ１１６を有する。

【００１６】

第１ネットワーク１１０は、たとえば、バスであり、１台以上のクライアント端末１１１、業務サーバ１１２、ネットワーク監視装置１１３、第１ＦＷ／ＩＰＳ１１４、プロキシサーバ１１６、第２ＦＷ／ＩＰＳ１２３およびＳＯＣ１３０を通信可能に接続する。第１ＦＷ／ＩＰＳ１１４は、外部ネットワーク１１５に通信可能に接続される。外部ネットワーク１１５は、たとえば、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、ＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、インターネットである。

【００１７】

また、監視対象システム１００は、第２ネットワーク１２０、制御装置１２１、コントローラ１２２、および第２ＦＷ／ＩＰＳ１２３を有する。第２ネットワーク１２０は、たとえば、バスであり、第２ネットワーク１２０、制御装置１２１、コントローラ１２２、および第２ＦＷ／ＩＰＳ１２３を通信可能に接続する。

【００１８】

ＳＯＣ１３０は、アラート管理装置１３１と、ログ収集装置１３２と、アラート分析装置１３４と、第３ネットワーク１３５と、を有する。第３ネットワークは、たとえば、バスであり、アラート管理装置１３１、ログ収集装置１３２、アラート分析装置１３４、および外部脅威情報データベース１３３を通信可能に接続する。

【００１９】

アラート管理装置１３１は、事象の一例として、監視対象システム１００からウィルス検出、異常な挙動検出、未登録装置との接続検出といったアラートを取得して格納する。アラートは、たとえば、アラートの発生日時と、アラート対象（アラートの発生元）と、アラート対象の通信相手と、を含む情報である。ログ収集装置１３２は、監視対象システム１００からのログ（アラート除く）を取得して格納する。ログは、いつ、監視対象システム１００内のどのコンピュータ２００がどのようなデータをどの通信相手に送受信したかを示す履歴情報である。

【００２０】

アラート分析装置１３４は、アラート管理装置１３１で管理されているアラートとログ収集装置１３２で管理されているログと外部脅威情報データベース１３３に登録されている脅威情報を用いて、アラートを分析する。外部脅威情報データベース１３３は、たとえば、インターネット上で脅威情報を公開するデータベースである。脅威情報には、たとえば、マルウェア、プログラムの脆弱性、スパム、不正ＵＲＬがある。

【００２１】

＜コンピュータのハードウェア構成例＞

図２は、図１に示した各種コンピュータ（クライアント端末１１１、業務サーバ１１２、ネットワーク監視装置１１３、第１ＦＷ／ＩＰＳ１１４、プロキシサーバ１１６、制御装置１２１、コントローラ１２２、第２ＦＷ／ＩＰＳ１２３、アラート管理装置１３１、ログ収集装置１３２、アラート分析装置１３４）のハードウェア構成例を示すブロック図である。

【００２２】

コンピュータ２００は、プロセッサ２０１と、記憶デバイス２０２と、入力デバイス２０３と、出力デバイス２０４と、通信インターフェース（通信ＩＦ）２０５と、を有する。プロセッサ２０１、記憶デバイス２０２、入力デバイス２０３、出力デバイス２０４、および通信ＩＦ２０５は、バス２０６により接続される。プロセッサ２０１は、コンピュータ２００を制御する。

【００２３】

記憶デバイス２０２は、プロセッサ２０１の作業エリアとなる。また、記憶デバイス２０２は、各種プログラムやデータを記憶する非一時的なまたは一時的な記録媒体である。記憶デバイス２０２としては、たとえば、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）、フラッシュメモリがある。入力デバイス２０３は、データを入力する。入力デバイス２０３としては、たとえば、キーボード、マウス、タッチパネル、テンキー、スキャナがある。出力デバイス２０４は、データを出力する。出力デバイス２０４としては、たとえば、ディスプレイ、プリンタがある。通信ＩＦ２０５は、ネットワークと接続し、データを送受信する。

【００２４】

＜アラート分析装置１３４の機能的構成例＞

図３は、アラート分析装置１３４の機能的構成例を示すブロック図である。アラート分析装置１３４は、アラート判断集計部３０１と、ログ統計集計部３０２と、要因抽出部３

０３と、重要度予測部３０４と、誤差要因抽出部３０５と、表示部３０６と、を有する。これらは、具体的には、たとえば、図２に示した記憶デバイス２０２に記憶されたプログラムをプロセッサ２０１に実行させることにより実現される機能である。

【００２５】

アラート判断集計部３０１は、アラート管理装置１３１からアラートを取得してアラート判断集計テーブル５００を作成する。アラート判断情報とは、アラート（発生日時、アラート対象、通信相手）に、当該アラートの処理結果が追加された情報である。処理結果とは、当該アラートに対し、たとえば、「誤検知と判断」、「未対処の攻撃と判断」、「対処済みの攻撃と判断」、「未処理」のいずれかである。アラート判断集計テーブル５００の詳細については後述する。

【００２６】

ログ統計集計部３０２は、ログ収集装置１３２からログを取得してログ統計集計テーブル６００を作成する。ログ統計とは、収集したログ群のうち、アラート発生時のログに関する統計情報である。ログ統計には、たとえば、所定の分析期間内のキャッシュミス回数や、異常応答回数、アクセス回数、ＩＰアドレス危険度、ＵＲＬ危険度などがある。ログ統計収集テーブルの詳細については後述する。

【００２７】

要因抽出部３０３は、アラート判断集計テーブル５００内のアラート判断結果（学習データ）と、ログ統計集計テーブル６００内のログ統計（学習データ）とを用いて、アラート判断につながった要因を抽出する。アラート判断結果（学習データ）とは、所定の分析期間内のアラート判断情報のうち学習データとして選ばれたアラート判断情報である。ログ統計（学習データ）とは、所定の分析期間内のログ統計のうち学習データとして選ばれたログ統計である。要因とは、そのアラートが発生した原因を示す情報である。たとえば、『ある期間内でのプロキシサーバのキャッシュミス回数が１０～１５回』などがある。

【００２８】

要因抽出部３０３は、誤差要因抽出部３０５からの誤差要因結果を用いて、誤差要因に含まれる抽出要因の重みを低減することで、抽出要因結果を更新し、重要度予測部３０４に出力する。誤差要因とは、要因のうち、後述する予測モデル式から得られるアラートの重要度の予測値の誤差が発生する要因である。

【００２９】

重要度予測部３０４は、アラート判断結果（テストデータ）と要因抽出部３０３による要因抽出結果から、アラート重要度を予測する予測モデル式を作成する。アラート重要度とは、監視対象システム１００からのアラートがどの程度重要であるかを示す指標値である。本例では、たとえば、アラートの処理結果が「誤検知」（攻撃でないのに攻撃と判断）を示すアラート重要度Ｐ１と、アラートで特定される攻撃に対しアラートの処理結果が「対処済み」であることを示すアラート重要度Ｐ２（＞Ｐ１）と、アラートで特定される攻撃に対しアラートの処理結果が「未対処」であることを示すアラート重要度Ｐ３（＞Ｐ２）と、がある。

【００３０】

また、重要度予測部３０４は、後述するログ統計集計テーブル６００内のログ統計（テストデータ）を、作成した予測モデル式に与えることにより、アラート重要度の予測値（以下、重要度予測値）を算出し、アラート判断集計テーブル５００内のアラート判断情報（テストデータ）を用いて、重要度予測値の予測誤差を求める。重要度予測部３０４は、最終的に、要因抽出部３０３からの更新された抽出要因結果を用いて、予測モデル式を更新する。

【００３１】

また、重要度予測部３０４は、ログ統計（予測対象データ）を更新された予測モデル式に与えることにより、重要度予測値を算出する。ログ統計（予測対象データ）とは、ログ統計集計テーブル６００内で予測対象データとして選ばれたログ統計である。

【００３２】

誤差要因抽出部３０５は、重要度予測部３０４からの重要度予測値の予測誤差と、ログ統計（テストデータ）とを用いて、予測誤差につながった要因を抽出し、抽出した誤差要因結果を要因抽出部３０３に出力する。これにより、要因抽出部３０３は、抽出要因結果を更新することができる。

【００３３】

表示部３０６は、重要度予測部３０４からのログ統計（予測対象データ）についての予測結果をディスプレイに表示する。表示内容の詳細については後述する。

【００３４】

＜アラート判断とログ統計集計の動作シーケンス例＞

図４は、アラート判断とログ統計集計の動作シーケンス例を示すシーケンス図である。アラート判断集計部３０１が、たとえば、ユーザ操作により、処理済みのアラートの収集範囲を決定する（ステップＳ４０１）。処理済みのアラートとは、当該アラートのアラート判断集計部３０１による処理結果が「未処理」以外のアラートである。収集範囲とは、アラートを収集する期間であり、ここでは、アラート判断集計部３０１が、収集範囲を、過去のある時点から現在までの分析期間Ｔに決定したものとする。

【００３５】

アラート判断集計部３０１は、アラート管理装置１３１が収集したアラートをアラート管理装置１３１から受信する（ステップＳ４０２）。アラート判断集計部３０１は、受信したアラートのうち収集範囲内のアラートを用いて、アラート判断集計テーブル５００を作成する（ステップＳ４０３）。アラート判断集計部３０１は、アラート判断集計テーブル５００からのアラート判断情報をログ統計集計部３０２に送信する（ステップＳ４０４）。

【００３６】

ログ統計集計部３０２は、アラート判断集計部３０１からアラート判断情報を受信するとともに、ログ収集装置１３２からログを受信し（ステップＳ４０５）、外部脅威情報データベース１３３から外部脅威情報を受信する（ステップＳ４０６）。ログ統計集計部３０２は、受信したアラート判断情報、ログおよび外部脅威情報を用いて、アラート発生時のログ統計集計テーブル６００を作成する（ステップＳ４０７）。

【００３７】

また、アラート判断集計部３０１は、分析期間Ｔ内におけるアラート判断結果のデータ種別を決定する（ステップＳ４０８）。データ種別は、たとえば、学習、テスト、および予測対象の３種類である。学習は、予測モデル式の作成に用いられるデータ種別であり、テストは、作成された予測モデル式に与えて重要度予測値を算出するためのデータ種別であり、予測対象は、誤差要因が考慮されて更新された予測モデル式に与えて重要度予測値を算出するためのデータ種別である。

【００３８】

＜アラート判断集計テーブル５００＞

図５は、アラート判断集計テーブル５００の一例を示す説明図である。アラート判断集計テーブル５００は、アラート判断情報を収集するテーブルであり、アラート判断集計部３０１により作成され（ステップＳ４０３）、アラート分析装置１３４の記憶デバイス２０２に記憶される。アラート判断集計テーブル５００は、アラート識別子５０１と、発生日時５０２と、アラート対象５０３と、通信相手５０４と、処理結果５０５と、重要度換算値５０６と、をフィールドとして有する。

【００３９】

アラート識別子５０１は、アラートを一意に特定する識別情報である。発生日時５０２は、アラートが発生した日付時刻である。アラート対象５０３は、アラートの発生元である。通信相手５０４は、アラート対象５０３が送信したデータの宛先またはアラート対象５０３にデータを送信した送信元である。アラート識別子５０１、発生日時５０２、アラート対象５０３、および通信相手５０４が、アラートを構成する。

【００４０】

処理結果５０５は、上述したように、当該アラートに対し、たとえば、「誤検知と判断」、「未対処の攻撃と判断」、「対処済みの攻撃と判断」、「未処理」のいずれかである。処理結果５０５は、アラート分析装置１３４が、ユーザ操作により入力された情報である（未入力の場合は「未処理」となる）。アラート識別子５０１、発生日時５０２、アラート対象５０３、通信相手５０４、および処理結果５０５がアラート判断結果を構成する。

【００４１】

重要度換算値５０６は、処理結果５０５を数値化した値である。重要度換算値５０６は、たとえば、０．０以上１．０以下の値の範囲をとる。本例では、処理結果５０５が「誤検知」の場合、重要度換算値５０６は「０．０」、処理結果５０５が「対処済み」の場合、重要度換算値５０６は「０．５」、処理結果５０５が「未対処の攻撃」の場合、重要度換算値５０６は「１．０」である。重要度換算値５０６が高いほど、危険性が高いことを示す。

【００４２】

＜ログ統計集計テーブル６００＞

図６は、ログ統計集計テーブル６００の一例を示す説明図である。ログ統計集計テーブル６００は、ログ統計を収集するテーブルであり、ログ統計集計部３０２により作成され（ステップＳ４０７）、アラート分析装置１３４の記憶デバイス２０２に記憶される。ログ統計集計テーブル６００は、アラート識別子５０１と、集計日時６０２と、プロキシサーバログ６０３と、業務サーバログ６０４と、外部脅威情報６０５と、をフィールドとして有する。

【００４３】

プロキシサーバログ６０３、業務サーバログ６０４、および外部脅威情報６０５以外にも監視対象システム１００内の他のコンピュータ（クライアント端末１１１やＦＷ／ＩＰＳ１１４，１２３、ネットワーク監視装置１１３など）についてのログがあってもよいが、図６では省略する。集計日時６０２は、アラート識別子５０１で特定されるアラートの発生日時５０２から所定時間遡った時刻から一定時間間隔で発生日時５０２までログ収集装置１３２がログを集計した日付時刻である。

【００４４】

本例では、所定時間を１時間とし、一定時間間隔を１０分とする。集計日時６０２は、一定時間間隔の終了時刻を示す。たとえば、集計日時６０２が「１０／１０　１２：５７」のエントリは、１０／１０の１２：４８から１２：５７までの１０分間で集計されたログの統計（ログ統計）を示す。

【００４５】

たとえば、アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるアラートの発生日時５０２は「１０／１０　１３：５７」であるため（図５参照）、アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるアラートの集計日時６０２は、発生日時５０２である「１０／１０　１３：５７」から1時間遡った「１０／１０　１２：５７」と、「１０／１０　１３：５７」から１０分刻みの「１０／１０　１３：０７」、「１０／１０　１３：１７」、「１０／１０　１３：２７」、「１０／１０　１３：３７」、「１０／１０　１３：４７」、および「１０／１０　１３：５７」（発生日時５０２）となる。このようにして、アラート発生時のログ統計の集計タイミングが設定される。

【００４６】

プロキシサーバログ６０３は、サブフィールドとして、キャッシュミス回数６３１と異常応答回数６３２とを有する。キャッシュミス回数６３１は、集計日時６０２においてプロキシサーバ１１６がキャッシュミスした回数である。異常応答回数６３２は、集計日時６０２においてプロキシサーバ１１６が異常応答を受信した回数である。なお、プロキシサーバログ６０３のサブフィールドは、キャッシュミス回数６３１や異常応答回数６３２以外（たとえば、通信バイト数）であってもよいが、図６では省略する。

【００４７】

業務サーバログ６０４は、サブフィールドとして、異常応答回数６４１とアクセス回数６４２とを有する。異常応答回数６４１は、集計日時６０２において業務サーバ１１２が異常応答を受信した回数である。アクセス回数６４２は、集計日時６０２で特定される一定時間間隔の集計期間において業務サーバ１１２が他のコンピュータ２００にアクセスされた回数である。なお、業務サーバログ６０４のサブフィールドは、異常応答回数６４１やアクセス回数６４２以外（たとえば、認証失敗回数）であってもよいが、図６では省略する。

【００４８】

外部脅威情報６０５は、サブフィールドとして、ＩＰアドレス危険度６５１とＵＲＬ危険度６５２とを有する。ＩＰアドレス危険度６５１は、集計日時６０２におけるアラート対象５０３の通信相手５０４がＩＰアドレスで特定された場合に、外部脅威情報データベース１３３において当該ＩＰアドレスの危険度を段階的に示した指標値である。本例では、０～５の６段階とし、５が最も危険度が高いことを示す。

【００４９】

ＵＲＬ危険度６５２は、集計日時６０２におけるアラート対象５０３の通信相手５０４がＵＲＬで特定された場合に、外部脅威情報データベース１３３において当該ＵＲＬの危険度を段階的に示した指標値である。本例では、０～５の６段階とし、５が最も危険度が高いことを示す。なお、外部脅威情報６０５のサブフィールドは、ＩＰアドレス危険度６５１やＵＲＬ危険度６５２以外であってもよいが、図６では省略する。

【００５０】

＜データ種別管理テーブル７００＞

図７は、データ種別管理テーブル７００の一例を示す説明図である。データ種別管理テーブル７００は、アラートごとにデータ種別を規定するテーブルであり、アラート判断集計部３０１により作成され（ステップＳ４０８）、アラート分析装置１３４の記憶デバイス２０２に記憶される。データ種別管理テーブル７００は、アラート識別子５０１と、分析期間（Ｔ－２）７０２と、分析期間（Ｔ－１）７０３と、分析期間（Ｔ）７０４と、をフィールドとして有する。

【００５１】

分析期間（Ｔ－２）７０２は、分析期間Ｔの２つ前にステップＳ４０１で決定された分析期間Ｔ－２における、アラートのデータ種別である。分析期間（Ｔ－１）７０３は、分析期間Ｔの１つ前にステップＳ４０１で決定された分析期間Ｔ－１における、アラートのデータ種別である。分析期間（Ｔ）７０４は、ステップＳ４０１で決定された最新の分析期間Ｔにおける、アラートのデータ種別である。

【００５２】

アラート判断集計部３０１は、分析期間Ｔ内で発生したアラートについて、ランダムにデータ種別を決定し、データ種別管理テーブル７００に格納する。この場合、「学習」と「テスト」のデータ種別の比率があらかじめ設定されていてもよい。アラート判断集計部３０１は、分析期間Ｔ以降のアラート、すなわち、処理結果５０５が「未処理」のアラートのデータ種別を「予測対象」に決定する。

【００５３】

あらたな分析期間Ｔおよびデータ種別がステップＳ４０１、Ｓ４０８で決定される都度、分析期間（Ｔ－２）７０２、分析期間（Ｔ－１）７０３、および分析期間（Ｔ）７０４は更新され、当該決定前の最古の分析期間Ｔ－２のデータ種別は消去される。なお、分析期間Ｔ－３以前のフィールドもあってもよいが、図７では省略する。

【００５４】

なお、データ種別が「学習」であるアラート識別子５０１で特定されるアラートのアラート判断情報が、アラート判断結果（学習データ）であり、データ種別が「テスト」であるアラート識別子５０１で特定されるアラートのアラート判断情報が、アラート判断結果（テストデータ）である。

【００５５】

また、データ種別が「学習」であるアラート識別子５０１で特定されるログ統計（図６のエントリ）が、ログ統計（学習データ）であり、データ種別が「テスト」であるアラート識別子５０１で特定されるログ統計（図６のエントリ）が、ログ統計（テストデータ）であり、データ種別が「予測対象データ」であるアラート識別子５０１で特定されるログ統計（図６のエントリ）が、ログ統計（予測対象データ）である。

【００５６】

＜分析期間Ｔのアラート重要度予測の動作シーケンス＞

図８は、分析期間Ｔのアラート重要度予測の動作シーケンス例を示すシーケンス図である。アラート判断集計部３０１は、アラート判断結果（学習データ）を要因抽出部３０３に出力する（ステップＳ８０１）。また、ログ統計集計部３０２は、ログ統計（学習データ）を要因抽出部３０３および重要度予測部３０４に出力する。

【００５７】

要因抽出部３０３は、アラート判断集計テーブル５００内のアラート判断結果（学習データ）と、ログ統計集計テーブル６００内のログ統計（学習データ）とを用いて、抽出要因テーブル９００を作成し、アラート判断につながった要因を抽出する（ステップＳ８０３）。ここで、要因抽出部３０３による要因抽出について具体的に説明する。

【００５８】

図９は、抽出要因テーブル９００の一例を示す説明図である。抽出要因テーブル９００は、要因項目９０１と、値域９０２と、第１相関度９０３と、重み９０４と、をフィールドとして有する。要因項目９０１は、抽出対象となる要因であり、ログ統計集計テーブル６００のプロキシサーバログ６０３や業務サーバログ６０４、外部脅威情報６０５の各サブフィールドを示す。値域９０２は、要因項目９０１の値が取り得る範囲である。たとえば、要因項目９０１が「プロキシサーバ　キャッシュミス回数」の値域９０２が「３～４」となっている場合、プロキシサーバ１１６のキャッシュミス回数６３１が３～４回である場合の第１相関度９０３が求められる。

【００５９】

第１相関度９０３は、アラート判断における要因項目９０１の値域９０２と重要度換算値５０６との相関を示す情報である。第１相関度９０３は、たとえば、ログ統計（学習データ）における値域９０２の出現回数をログ統計（学習データ）の集計回数で除算した値域９０２の出現頻度ｐ１（発生確率）と、重要度換算値５０６（ここでは、重要度換算値ｑとする）と、の相関係数Ｒ１である。具体的には、たとえば、相関係数Ｒ１は、出現頻度ｐ１の標準偏差σｐ１と、重要度換算値ｑの標準偏差σｑと、出現頻度ｐ１および重要度換算値ｑの共分散Ｓｐ１ｑと、により、下記式（１）で求められる。

【００６０】

Ｒ１＝Ｓｐ１ｑ／（σｐ１×σｑ）・・・（１）

【００６１】

ここで、出現頻度ｐ１について詳細に説明する。図７に示したように、分析期間Ｔにおいてデータ種別が「学習」であるアラート識別子５０１は、「Ａｌｅｒｔ＿００５」，「Ａｌｅｒｔ＿００７」，「Ａｌｅｒｔ＿００８」，「Ａｌｅｒｔ＿０１０」，および「Ａｌｅｒｔ＿０１１」である。要因抽出部３０３は、これらのアラート識別子５０１ごとに、出現頻度ｐ１および重要度換算値ｑを求める。アラート識別子５０１が「Ａｌｅｒｔ＿００５」で、かつ、要因項目９０１が「プロキシサーバ　キャッシュミス回数」を例に挙げる。

【００６２】

図６に示したように、アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるプロキシサーバログ６０３のキャッシュミス回数６３１は、「３」（１０／１０　１２：５７）、「４」（１０／１０　１３：０７）、…、「４」（１０／１０　１３：５７）である。なお、集計日時６０２が「１０／１０　１３：１７」、「１０／１０　１３：２７」、「１０／１０　１３：３７」、および「１０／１０　１３：４７」のキャッシュミス回数６３１を「３」および「４」以外の値とする。

【００６３】
アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるプロキシサーバログ６０３のキャッシュミス回数６３１における値域９０２「３～４」の出現回数は３回である。また、アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるプロキシサーバログ６０３のキャッシュミス回数６３１の集計回数は、「１０／１０　１２：５７」、「１０／１０　１３：０７」、「１０／１０　１３：１７」、「１０／１０　１３：２７」、「１０／１０　１３：３７」、「１０／１０　１３：４７」、および「１０／１０　１３：５７」の７回である。したがって、アラート識別子５０１が「Ａｌｅｒｔ＿００５」であるプロキシサーバログ６０３のキャッシュミス回数６３１における値域９０２「３～４」の出現頻度ｐ１は、３／７である。また、アラート識別子５０１が「Ａｌｅｒｔ＿００５」である重要度換算値ｑは、「０」である（図５を参照）。

【００６４】
要因抽出部３０３は、データ種別が「学習」であるアラート識別子５０１ごとに、出現頻度ｐ１と重要度換算値ｑとの組み合わせを求め、各出現頻度ｐ１から出現頻度ｐ１の標準偏差σｐ１を求め、各重要度換算値ｑから重要度換算値ｑの標準偏差σｑを求め、さらに、共分散Ｓｐ１ｑを求める。そして、要因抽出部３０３は、上記式（１）により、データ種別が「学習」であるアラート識別子５０１についての要因項目９０１「プロキシサーバ　キャッシュミス回数」の値域９０２「３～４」の相関係数Ｒ１（＝－０．５４）を算出する。

【００６５】
相関係数Ｒ１が正の相関の場合（Ｒ１＞０）、アラート判断が正しく、相関係数Ｒ１が高いほど要因項目９０１による危険度が高いことを示す。逆に、相関係数Ｒ１が負の相関の場合（Ｒ１＜０）、アラート判断が間違っている、すなわち、誤検知であり、相関係数Ｒ１が低いほど、要因項目９０１による危険度が低く、誤検知が多発していることを示す。このように、要因項目９０１と値域９０２との組み合わせごとに第１相関度が求められるため、要因項目９０１と値域９０２とのどの組み合わせにアラート判断につながった要因があるかを統計的に抽出することができる。

【００６６】
重み９０４は、要因項目９０１と値域９０２の組み合わせの重要度を示す。重み９０４は、０．０以上１．０以下の範囲を取り、初期値を１．０とする。第１相関度が正の相関係数Ｒ１になると、要因抽出部３０３は、対応する重み９０４を低下させる。重み９０４は上述した出現頻度ｐ１と乗算して、予測モデル式の作成に用いられる。したがって、重み９０４が低下すると、その出現頻度ｐ１、すなわち、要因項目９０１および値域９０２の組み合わせの影響度が低下して、予測モデル式が更新される。

【００６７】
図８に戻り、要因抽出部３０３は、抽出要因結果（抽出要因テーブル９００から得られた出現頻度ｐ１および重要度換算値ｑ）を重要度予測部３０４に出力する（ステップＳ８０４）。重要度予測部３０４は、アラート判断結果（学習データ）、ログ統計（学習データ）、および抽出要因結果を用いて、予測モデル式を作成する（ステップＳ８０５）。ここで、目的変数Ｙを重要度換算値５０６、説明変数ＸｎをＰ（要因項目９０１＋値域９０２）とする。ただし、Ｐ（Ｚ）は、事象Ｚの発生確率（出現頻度ｐ１）を表す。説明変数Ｘｉは、抽出要因テーブル９００を参照することで、
Ｘ１＝Ｐ（プロキシサーバ　キャッシュミス回数［３～４］）
Ｘ２＝Ｐ（プロキシサーバ　キャッシュミス回数［１０～１５］）
・・・
などとする。このとき、予測モデル式の一例として、下記式（２）のような重回帰式が作成される。ｎは、要因項目９０１および値域９０２との組み合わせの総数、すなわち、事象Ｚの総数である。

【００６８】

【数１】

$$Y = b0 + b1 \times X1 + b2 \times X2 + \cdots + bn \times Xn \quad \cdots(2)$$

【００６９】

ここで、ログ統計（学習データ）のエントリｋ（たとえば、Ａｌｅｒｔ＿００５に関する目的変数Ｙと説明変数Ｘｎの組み合わせ）に対する目的変数Ｙの値をｙ＿ｋ（重要度換算値ｑ＝（０．０）、説明変数Ｘ１の値をｘ１＿ｋ（出現頻度ｐ１＝３／７）、説明変数Ｘ２の値をｘ２＿ｋ、・・・、説明変数Ｘｎの値をｘｎ＿ｋとすれば、上記式（２）の各係数ｂ０、ｂ１、ｂ２、・・・、ｂｎは、一例として下記式（３）のような行列式によって求めることができる。式（３）中、ｉは、ログ統計（学習データ）のエントリ１～ｋの任意のエントリを示す。

【００７０】

【数２】

$$\begin{pmatrix} b0 \\ b1 \\ b2 \\ \vdots \\ bn \end{pmatrix} = \begin{pmatrix} n & \Sigma x1\_i & \Sigma x2\_i & \cdots & \Sigma xn\_i \\ \Sigma x1\_i & \Sigma x1\_i^2 & \Sigma x2\_i \times x1\_i & \cdots & \Sigma xn\_i \times x1\_i \\ \Sigma x2\_i & \Sigma x1\_i \times x2\_i & \Sigma x2\_i^2 & \cdots & \Sigma xn\_i \times x2\_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma xn\_i & \Sigma x1\_i \times xn\_i & \Sigma x2\_i \times xn\_i & \cdots & \Sigma xn\_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma y\_i \\ \Sigma y\_i \times x1\_i \\ \Sigma y\_i \times x2\_i \\ \vdots \\ \Sigma y\_i \times xn\_i \end{pmatrix} \quad \cdots(3)$$

【００７１】

上記式（３）による予測モデル式の作成方法は一例であり、一般的に知られている正則化や決定木、アンサンブル学習、ニューラルネットワーク、ベイジアンネットワークなどの手法を用いて導出してもよい。

【００７２】

アラート判断集計部３０１は、アラート判断結果（テストデータ）を重要度予測部３０４に出力し（ステップＳ８０６）、ログ統計集計部３０２は、ログ統計（テストデータ）を重要度予測部３０４に出力する（ステップＳ８０７）。

【００７３】

重要度予測部３０４は、テストデータに対してアラート重要度を予測する（ステップＳ８０８）。具体的には、たとえば、重要度予測部３０４は、ログ統計（テストデータ）の説明変数ｘ１＿ｋ、ｘ２＿ｋ、…ｘｎ＿ｋを予測モデル式に与えることにより、重要度予測値ｙ＿ｋを算出する。

【００７４】

つぎに、重要度予測部３０４は、誤差テーブル１０００を作成して、ステップＳ８０８で算出した重要度予測値と、アラート判断結果（テストデータ）に含まれる重要度換算値５０６と、の予測誤差を、アラート識別子５０１ごとに算出する（ステップＳ８０９）。ここで、誤差テーブル１０００について説明する。

【００７５】

図１０は、誤差テーブル１０００の一例を示す説明図である。誤差テーブル１０００は、アラート識別子５０１と、重要度換算値５０６と、重要度予測値１００１と、予測誤差１００２とを、フィールドとして有する。重要度予測値１００１は、そのアラート識別子５０１について予測モデル式から算出された予測値である。重要度予測値１００１は、重要度換算値５０６と同様、たとえば、０．０以上１．０以下の値の範囲をとる。本例では、重要度予測値１００１が０．０以上０．３未満であれば「誤検知」、０．３以上０．７未満であれば「対処済み」、０．７以上１以下であれば「未対処の攻撃」であることを示

す。重要度予測値１００１が高いほど、危険性が高いことを示す。
【００７６】
予測誤差１００２は、重要度予測値１００１の誤差を示す値である。具体的には、たとえば、予測誤差１００２は、重要度換算値５０６と重要度予測値１００１との差分を丸めた値である。差分が許容範囲内であれば、予測が当たっていることを示し、重要度予測部３０４は、予測誤差１００２を「０」に設定する。たとえば、アラート識別子５０１が「Ａｌｅｒｔ＿００６」のエントリでは、差分が「０．１３」であり、許容範囲内とする。この場合、予測誤差１００２は「０」に設定される。
【００７７】
一方、差分が許容範囲外であれば、予測が外れていることを示し、重要度予測部３０４は、予測誤差１００２を「１」に設定する。たとえば、アラート識別子５０１が「Ａｌｅｒｔ＿００９」のエントリでは、差分が「０．４６」であり、許容範囲外とする。この場合、予測誤差１００２は「１」に設定される。
【００７８】
図８に戻り、重要度予測部３０４は、図８の点線枠の処理の繰り返し終了確認を実行する（ステップＳ８１０）。図８の点線枠の処理の繰り返しは、予測モデル式の更新（再作成）を示す。図８の点線枠の処理の繰り返し試行の終了条件について具体的に説明する。
【００７９】
図１１は、図８の点線枠の処理の繰り返し試行の終了条件を示す説明図である。図１１は、横軸を図８の点線枠の処理の繰り返し試行回数１１０１、縦軸を誤差件数１１０２とするグラフ１１０３を示す。誤差件数１１０２は、誤差テーブル１０００の予測誤差１００２の値が「１」の個数である。繰り返し試行回数１１０１が増加するにしたがって、予測モデル式が更新されるため、誤差件数が減少傾向になる。繰り返し試行回数１１０１がＮ回目でしきい値１１０４を下回った場合、図８の点線枠の処理の繰り返しが終了する。
【００８０】
図８に戻り、点線枠の処理の繰り返しが終了していない場合、重要度予測部３０４は、誤差結果を誤差要因抽出部３０５に出力する（ステップＳ８１１）。誤差結果とは、アラート識別子５０１ごとの予測誤差１００２である。また、ログ統計集計部３０２は、ログ統計（テストデータ）を誤差要因抽出部３０５に出力する（ステップＳ８１２）。
【００８１】
誤差要因抽出部３０５は、誤差結果（予測誤差１００２）とログ統計（テストデータ）とを用いて誤差要因テーブル１２００を作成し、誤差につながった要因である誤差要因を抽出する（ステップＳ８１３）。ここで、誤差要因抽出部３０５による誤差要因抽出について具体的に説明する。
【００８２】
図１２は、誤差要因テーブル１２００の一例を示す説明図である。誤差要因テーブル１２００は、抽出要因テーブル９００と同様に作成される。誤差要因テーブル１２００は、要因項目９０１と、値域９０２と、第２相関度１２０３と、をフィールドとして有する。要因項目９０１の値は、抽出要因テーブル９００と同じである。値域９０２は、要因項目９０１の値が取り得る範囲であるが、誤差要因テーブル１２００の場合、ログ統計（テストデータ）のエントリにより設定される。
【００８３】
第２相関度１２０３は、アラート判断における要因項目９０１の値域９０２と予測誤差１００２との相関を示す情報である。第２相関度１２０３は、たとえば、ログ統計（テストデータ）における値域９０２の出現回数をログ統計（テストデータ）の集計回数で除算した値域９０２の出現頻度ｐ２（発生確率）と、予測誤差１００２（ここでは、予測誤差ｅとする）と、の相関係数Ｒ２である。具体的には、たとえば、相関係数Ｒ２は、出現頻度ｐ２の標準偏差σｐ２と、予測誤差ｅの標準偏差σｅと、出現頻度ｐ２および予測誤差ｅの共分散Ｓｐ２ｅと、により、下記式（４）で求められる。
【００８４】

Ｒ２＝Ｓｐ２ｅ／（σｐ２×σｅ）・・・（４）

【００８５】

なお、出現頻度ｐ２の求め方は、用いるアラート識別子５０１が、分析期間Ｔにおいてデータ種別が「テスト」であるアラート識別子５０１であること以外は、出現頻度ｐ１と同じである。相関係数Ｒ２が正の相関の場合（Ｒ２＞０）、相関係数Ｒ２が高いほど、その要因項目９０１は予測誤差１００２を生む要因であることを示す。逆に、相関係数Ｒ２が負の相関の場合（Ｒ２＜０）、相関係数Ｒ２が低いほど、その要因項目９０１は予測誤差１００２を生む要因ではないことを示す。このように、要因項目９０１と値域９０２との組み合わせごとに第２相関度１２０３が求められるため、要因項目９０１と値域９０２とのどの組み合わせに予測誤差１００２につながった要因があるかを統計的に抽出することができる。

【００８６】

図８に戻り、誤差要因抽出部３０５は、誤差要因結果を要因抽出部３０３に出力する（ステップＳ８１４）。誤差要因結果とは、第２相関度１２０３（相関係数Ｒ２）が正の相関（Ｒ２＞０）である要因項目９０１および値域９０２との組み合わせである。図１２の例では、誤差要因結果は、エントリ１２１１～１２１５における要因項目９０１および値域９０２との組み合わせである。

【００８７】

要因抽出部３０３は、誤差要因抽出部３０５からの誤差要因結果を参照して、誤差要因に含まれる抽出要因の重み９０４を減らす（ステップＳ８１５）。具体的には、たとえば、要因抽出部３０３は、誤差要因結果に該当する要因項目９０１および値域９０２の組み合わせが存在するエントリを抽出要因テーブル９００から特定する。そして、要因抽出部３０３は、特定したエントリのうち第１相関度９０３が正の相関（Ｒ１＞０）のエントリの重み９０４を低減させて更新する。

【００８８】

図１３は、更新後の抽出要因テーブル９００の一例を示す説明図である。誤差要因結果に該当する要因項目９０１および値域９０２との組み合わせが、上述したエントリ１２１１～１２１５の場合、要因抽出部３０３は、エントリ１２１１～１２１５のうち、要因項目９０１および値域９０２が一致するエントリ１３０１～１３０３を特定する。そして、要因抽出部３０３は、特定したエントリ１３０１～１３０３のうち第１相関度９０３が正の相関（Ｒ１＞０）のエントリ１３０１，１３０２の重み９０４を低減させて更新する。

【００８９】

図１３は、特定したエントリ１３０１，１３０２の重み９０４が「１．０」から「０．５」に低減された例である。低減量は、一例として「０．５」としたが、０より大きく１以下の範囲であれば、ユーザが任意に設定可能である。重み９０４を０よりも大きく１．０よりも小さい値に低減させることで、予測誤差に影響を与えている要因（＝誤差要因）の重要度予測精度の悪化を抑制することができる。さらに、重み９０４を０にすることで、予測誤差に影響を与えている要因（＝誤差要因）を取り除き、予測精度の悪化をより効果的に抑制することができる。

【００９０】

図８に戻り、要因抽出部３０３は、更新した抽出要因結果を重要度予測部３０４に出力する（ステップＳ８１６）。具体的には、たとえば、要因抽出部３０３は、更新後の抽出要因テーブル９００を重要度予測部３０４に参照可能にする。重要度予測部３０４は、更新後の抽出要因テーブル９００を参照して、アラート判断結果（学習データ）、ログ統計（学習データ）、および更新後の抽出要因結果を用いて、ステップＳ８０５と同様の処理により、予測モデル式を再作成（更新）する（ステップＳ８１７）。そして、ステップＳ８０８に戻る。

【００９１】

具体的には、たとえば、重要度予測部３０４は、説明変数Ｘｎの重みをＷｎとした場合、一例として、下記式（５）により説明変数Ｘ´ｎに変換する。

【００９２】
【数３】

$$\acute{X}n = \begin{cases} Wn \times Xn & (0 \leq Xn < Wn) \\ Wn^2 & (Wn \leq Xn) \end{cases} \quad \cdots(5)$$

【００９３】
重要度予測部３０４は、予測モデル式を再作成する場合、説明変数ＸｎをＸ´ｎに置き換えて、上記式（２）の係数ｂ０、ｂ１、ｂ２、・・・、ｂｎを再計算することになる。
【００９４】
一方、ステップＳ８１０の繰り返しの終了条件の確認において、繰り返しの終了条件を満たした場合、点線枠の繰り返し試行が終了する。この場合、ログ統計集計部３０２は、ログ統計（予測対象データ）を重要度予測部３０４に出力する（ステップＳ８１８）。この場合、重要度予測部３０４は、ログ統計（予測対象データ）を更新された予測モデル式に与えることにより、重要度予測値１００１を算出する（ステップＳ８２０）。ログ統計（予測対象データ）とは、ログ統計集計テーブル６００内で予測対象データとして選ばれたログ統計である。このあと、重要度予測部３０４は、予測結果を表示部３０６に出力する（ステップＳ８２１）。重要度予測値１００１の出力画面表示例について説明する。
【００９５】
＜重要度予測値１００１の出力画面表示例＞
図１４は、重要度予測値１００１の出力画面表示例を示す説明図である。出力画面１４００は、アラート通知タブ１４０１を有する。アラート通知タブ１４０１は、アラートリスト１４０２と、予測モデル式の作成に用いた要因１４０３と、予測モデル式を悪化させる要因（予測モデル式の予測精度を低下させる要因）１４０４と、を表示する。これらは、表示部３０６が、予測結果を用いて生成する。
【００９６】
すなわち、重要度予測部３０４からの予測結果には、ステップＳ８２０で算出した重要度予測値１００１のほか、当該重要度予測値１００１を求めるために予測モデル式に与えられたログ統計（予測対象データ）に関連するアラート判断情報（図５）が含まれる。
【００９７】
たとえば、ログ統計（予測対象データ）のアラート識別子５０１が「Ａｌｅｒｔ＿０１３」，「Ａｌｅｒｔ＿０１４」であれば、アラート判断集計テーブル５００のアラート識別子５０１が「Ａｌｅｒｔ＿０１３」，「Ａｌｅｒｔ＿０１４」のエントリにおける発生日時５０２、アラート対象５０３および通信相手５０４が、予測結果に含まれるアラート判断情報となる。表示部３０６は、このアラート判断情報とステップＳ８２０で算出した重要度予測値１００１とをアラート識別子５０１で関連付けて、アラートリスト１４０２として出力画面１４００に表示する。
【００９８】
また、重要度予測値１００１からの予測結果には、予測モデル式の作成に用いた要因１４０３である要因項目９０１および値域９０２との組み合わせ（図１３の重み９０４が「１．０」のエントリ）が含まれてもよい。表示部３０６は、この図１３の重み９０４が「１．０」のエントリを、予測モデル式の作成に用いた要因１４０３として出力画面１４００に表示する。
【００９９】
また、重要度予測値１００１からの予測結果には、予測モデル式を悪化させる要因１４０４である要因項目９０１および値域９０２との組み合わせ（図１３の重み９０４が「１．０」でないエントリ）が含まれてもよい。表示部３０６は、この図１３の重み９０４が「１．０」でないエントリを、予測モデル式を悪化させる要因１４０４として出力画面１４００に表示する。
【０１００】

なお、本実施例では、表示部３０６が予測結果を表示することとしたが、アラート分析装置１３４は、他のコンピュータに予測結果を送信してもよい。この場合、予測結果の宛先のコンピュータが予測結果を表示してもよい。

【０１０１】

＜要因抽出処理＞

図１５は、要因抽出部３０３による要因抽出処理手順例を示すフローチャートである。要因抽出部３０３は、アラート判断集計部３０１から、学習データに分類されたアラート判断結果を取得する（ステップＳ１５０１）。ステップＳ１５０１は図８のステップＳ８０１に対応する。要因抽出部３０３は、ログ統計集計部３０２から、学習データに分類されたアラートのログ統計を取得する（ステップＳ１５０２）。ステップＳ１５０２は図８のステップＳ８０２に対応する。要因抽出部３０３は、アラート判断結果とログ統計とを分析し、アラート判断に繋がったログ統計の要因を抽出する（ステップＳ１５０３）。ステップＳ１５０３は図８のステップＳ８０３に対応する。

【０１０２】

要因抽出部３０３は、重要度予測部３０４に、抽出要因結果を渡す（ステップＳ１５０４）。ステップＳ１５０４は図８のステップＳ８０４に対応する。要因抽出部３０３は、誤差要因抽出部３０５から、誤差要因結果を取得する（ステップＳ１５０５）。ステップＳ１５０５は図８のステップＳ８１４に対応する。要因抽出部３０３は、現在の抽出要因テーブル９００に対して、正相関の誤差要因に含まれる項目の重み９０４を減らす（ステップＳ１５０６）。ステップＳ１５０６は図８のステップＳ８１５に対応する。

【０１０３】

要因抽出部３０３は、重要度予測部３０４に、更新した抽出要因結果を渡す（ステップＳ１５０７）。ステップＳ１５０７は図８のステップＳ８０１に対応する（ステップＳ８１６）。要因抽出部３０３は、繰り返し試行が終了したか否かを判断する（ステップＳ１５０８）。ステップＳ１５０８は図８のステップＳ８１０に対応する。終了していない場合（ステップＳ１５０８：Ｎｏ）、ステップＳ１５０５に戻る。終了した場合（ステップＳ１５０８：Ｙｅｓ）、要因抽出部３０３は要因抽出処理を終了する。

【０１０４】

＜重要度予測処理＞

図１６は、重要度予測部３０４による重要度予測処理手順例を示すフローチャートである。重要度予測部３０４は、要因抽出部３０３から、抽出要因結果を取得する（ステップＳ１６０１）。ステップＳ１６０１は図８のステップＳ８０４に対応する。重要度予測部３０４は、抽出要因を用いて予測モデル式を作成する（ステップＳ１６０２）。ステップＳ１６０２は図８のステップＳ８０５に対応する。

【０１０５】

重要度予測部３０４は、アラート判断集計部３０１から、テストデータに分類されたアラート判断結果を取得する（ステップＳ１６０３）。ステップＳ１６０３は図８のステップＳ８０６に対応する。重要度予測部３０４は、ログ統計集計部３０２から、テストデータに分類されたアラートのログ統計を取得する（ステップＳ１６０４）。ステップＳ１６０４は図８のステップＳ８０７に対応する。

【０１０６】

重要度予測部３０４は、要因抽出部３０３から、テストデータに対して、アラート重要度を予測する（ステップＳ１６０５）。ステップＳ１６０５は図８のステップＳ８０８に対応する。重要度予測部３０４は、予測値と実際の判断結果とを比較して、誤差を出す（ステップＳ１６０６）。ステップＳ１６０６は図８のステップＳ８０９に対応する。

【０１０７】

重要度予測部３０４は、誤差件数がしきい値以下であるか否かを判断する（ステップＳ１６０７）。ステップＳ１６０７は図８のステップＳ８１０に対応する。誤差件数がしきい値以下でない場合（ステップＳ１６０７：Ｎｏ）、重要度予測部３０４は、誤差要因抽出部３０５に予測値の誤差結果を渡す（ステップＳ１６０８）。ステップＳ１６０１は図

８のステップＳ８１１に対応する。

【０１０８】

重要度予測部３０４は、要因抽出部３０３から、更新された抽出要因結果を取得する（ステップＳ１６０９）。ステップＳ１６０９は図８のステップＳ８１６に対応する。重要度予測部３０４は、要因抽出部３０３から、更新された抽出要因を用いて予測モデル式を再作成して（ステップＳ１６１０）、ステップＳ１６０５に戻る。ステップＳ１６１０は図８のステップＳ８１７に対応する。一方、ステップＳ１６０７において、誤差件数がしきい値以下でない場合（ステップＳ１６０７：Ｎｏ）、重要度予測部３０４は、図８の点線枠で示した繰り返し試行を終了する。

【０１０９】

なお、上述した説明では、監視対象システムへの攻撃に対するアラートについて説明したが、アラート以外の事象にも適用可能である。たとえば、電力需要予測に適用した場合、たとえば、目的変数Ｙを１時間当たりの電力需要、説明変数Ｘｎを過去数時間分の電力需要の変動、各地点の気象データ（天気、気温、湿度、風向、風速、気圧、日照など）、各地点の人口流動統計、カレンダー情報（曜日、祝日など）、太陽光発電量など、とすることにより、アラート分析装置１３４は、前日までの１時間毎の電力需要と、各時間での説明変数Ｘｎのデータを元に学習を行って予測モデル式を作成し、テストデータを与えて予測誤差を求めることで予測モデル式を再作成して最適化することにより、翌日の１時間毎の電力需要を予測することができる。

【０１１０】

また、売上予測に適用した場合、たとえば、目的変数Ｙを１週間での店舗売上金額、説明変数Ｘｎを商品分類（生鮮品、惣菜、一般食品、日用品、衣料品など）ごとの売り場面積、商品分類ごとの顧客滞留時間、商品分類ごとの広告掲載数、顧客データ（来店者数、性別、年代、職業、住所など）などとすることにより、過去における各店舗の１週間毎の売上金額と各週での説明変数のデータを元に学習を行って予測モデル式を作成し、テストデータを与えて予測誤差を求めることで予測モデル式を再作成して最適化することにより、翌週の店舗売上金額を予測することができる。

【０１１１】

（１）このように、本実施例のアラート分析装置１３４は、事象群の中の第１事象（データ種別：テストのアラート）の要因に対する第１出現頻度（ログ統計（テストデータ）の説明変数ｘ１＿ｋ、ｘ２＿ｋ、…ｘｎ＿ｋ）を予測モデル式に与えることで得られる第１予測値（重要度予測値ｙ＿ｋ）と、第１出現頻度に対応する結果（重要度換算値５０６）と、に基づいて、第１予測値の予測誤差を算出する予測誤差算出処理（Ｓ８０９）と、事象群の中の第２事象（データ種別：予想対象のアラート）の要因に対する第２出現頻度（出現頻度ｐ２）と、予測誤差算出処理によって算出された予測誤差と、の相関（第２相関度１２０３）に基づいて、第１事象の要因の中から予測誤差の誤差要因（エントリ１２１１～１２１５の要因項目９０１および値域９０２）を抽出する誤差要因抽出処理（Ｓ８１３）と、を実行する。これにより、予測誤差の誤差要因を特定することができる。したがって、ユーザは、特定された誤差要因を考慮して、事象が発生しないように対策を取ることができる。

【０１１２】

（２）また、上記（１）のアラート分析装置１３４は、事象群の中の第３事象（データ種別：学習のアラート）の要因に対する第３出現頻度（出現頻度ｐ１）と、第３出現頻度に対応する結果（重要度換算値ｑ）と、に基づいて、前記予測モデル式を作成する作成処理を実行する。このように、学習データを用いて予測モデル式を事前に作成することにより、予測モデル式を学習することができる。

【０１１３】

（３）また、上記（１）のアラート分析装置１３４において、前記事象群は、所定の時点以降に発生した事象の集合である。このように、所定の時点以降に発生した事象を用いることにより、換言すれば、当該事象以前の過去の事象を用いないことにより、監視対象

システム１００への攻撃が変化して既に事象の特性が変わっている場合にも、過去の判断要因に引きずられることなく、誤差要因を特定することができる。

【０１１４】

（４）また、上記（１）のアラート分析装置１３４において、記憶デバイス２０２は、第１事象の要因の重要度を示す重み９０４を記憶しており、アラート分析装置１３４は、第１事象の要因のうち誤差要因抽出処理によって抽出された誤差要因（エントリ１２１１～１２１５の要因項目９０１および値域９０２）の重み９０４を他の要因の重み９０４よりも低くなるように設定する設定処理（ステップＳ８１６）と、事象群の中の第３事象の要因に対する第３出現頻度（出現頻度ｐ１）と、第３出現頻度に対応する結果（重要度換算値ｑ）と、設定処理によって設定された誤差要因の重み９０４と、他の要因の重み９０４と、に基づいて、予測モデル式を更新する更新処理（ステップＳ８１７）と、を実行する。このように、誤差要因による影響が低くなるように予測モデル式を更新することにより、予測値の予測精度の向上を図ることができる。

【０１１５】

（５）また、上記（４）のアラート分析装置１３４は、予測誤差算出処理では、更新処理による更新後の予測モデル式に第１出現頻度を与えることで得られる第１予測値と、第１出現頻度に対応する結果と、に基づいて、第１予測値の予測誤差を、算出する。このように、更新された予測モデル式を用いて予測誤差を再算出することにより、予測誤差を小さくすることができ、誤差要因の絞り込みの効率化を図ることができる。

【０１１６】

（６）また、上記（４）のアラート分析装置１３４は、更新処理による更新後の予測モデル式に、第２出現頻度（出現頻度ｐ２）を与えることにより、第２事象の第２予測値を算出する予測値算出処理（ステップＳ８１９）と、予測値算出処理によって算出された第２予測値を出力する出力処理（ステップＳ８２１）と、を実行する。このように、更新された予測モデル式に、予測対象データを与えることにより、事象の予測値を算出することにより、当該予測値の予測精度の向上を図ることができる。

【０１１７】

（７）また、上記（６）のアラート分析装置１３４は、第２事象の件数のうち第１予測値と第１結果との間に許容範囲外の誤差がある誤差件数に基づいて、設定処理（ステップＳ８１６）および更新処理（ステップＳ８１７）を試行するか否かを判断する判断処理（ステップＳ８１０）を実行し、判断処理による判断結果に基づいて、設定処理（ステップＳ８１６）および更新処理（ステップＳ８１７）を試行する。このように、データ種別がテストである第２事象のうち、重要度予測値１００３と重要度換算値５０６との間に許容範囲外の誤差がある誤差件数により、予測モデル式の更新処理の試行を判断するため、予測モデル式の更新頻度を調整することができる。

【０１１８】

（８）また、上記（７）のアラート分析装置１３４は、誤差件数がしきい値以上である場合、設定処理（ステップＳ８１６）および更新処理（ステップＳ８１７）を試行する。このように、誤差件数がしきい値以上の場合、予測モデル式の更新処理を試行するため、誤差件数がしきい値未満となるまで、予測モデル式の更新処理が繰り返されることになり、予測モデル式から算出される予測値の高精度化を図ることができる。

【０１１９】

（９）また、上記（７）のアラート分析装置１３４は、誤差件数がしきい値以上でない場合、予測値算出処理（ステップＳ８１９）および出力処理（ステップＳ８２１）を試行する。このように、誤差件数がしきい値以上でない場合、予測値の算出を実行するため、誤差件数がしきい値以上では予測値は算出されない。したがって、予測モデル式から算出される予測値の精度低下を抑制することができる。

【０１２０】

（１０）また、上記（６）のアラート分析装置１３４は、予測モデル式の更新に用いられた要因を出力する。このように、予測モデル式の更新に用いられた要因を出力すること

により、どの要因が予測モデル式の更新に寄与したかを把握することができる。

【０１２１】

（１１）また、上記（６）のアラート分析装置１３４は、第３事象（データ種別：学習のアラート）の要因に対する第３出現頻度（出現頻度ｐ１）と、第３出現頻度に対応する結果（重要度換算値ｑ）と、の相関（第１相関度９０３）を求め、当該相関（第１相関度９０３）と誤差要因（エントリ１２１１～１２１５の要因項目９０１および値域９０２）とに基づいて、第３事象の要因の中から予測モデル式の精度を低下させる要因（エントリ１３０１，１３０２の要因項目９０１および値域９０２）を抽出する要因抽出処理（ステップＳ８０３、Ｓ８１５）を実行し、出力処理（ステップＳ８２１）では、要因抽出処理によって抽出された要因（エントリ１３０１，１３０２の要因項目９０１および値域９０２）を出力する。このように、予測モデル式の精度を低下させる要因（エントリ１３０１，１３０２の要因項目９０１および値域９０２）を抽出することにより、どの要因が予測モデル式の精度に悪影響を与えたかを把握することができる。

【０１２２】

以上説明したように、本実施例によれば、アラート重要度の予測に用いる特徴量の次元（要因項目の種類）、または値域またはその両方が多岐に渡っても、アラート重要度の予測精度の低下を抑制することができる。また、アラート重要度の予測に用いる特徴量の次元（要因項目の種類）、または値域またはその両方が多岐に渡っても、予測誤差を与える要因である誤差要因を取り除くことができ、監視対象システムの大規模化し、またサイバー攻撃の手口が日々変化し増加しても、アラート重要度の予測精度を向上することができる。その結果、将来に渡って持続可能なＳＯＣ１３０の運用の実現に貢献することができる。

【０１２３】

なお、本発明は前述した実施例に限定されるものではなく、添付した特許請求の範囲の趣旨内における様々な変形例及び同等の構成が含まれる。例えば、前述した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに本発明は限定されない。また、ある実施例の構成の一部を他の実施例の構成に置き換えてもよい。また、ある実施例の構成に他の実施例の構成を加えてもよい。また、各実施例の構成の一部について、他の構成の追加、削除、または置換をしてもよい。

【０１２４】

また、前述した各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等により、ハードウェアで実現してもよく、プロセッサがそれぞれの機能を実現するプログラムを解釈し実行することにより、ソフトウェアで実現してもよい。

【０１２５】

各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリ、ハードディスク、ＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）等の記憶装置、又は、ＩＣ（Ｉｎｔｅｇｒａｔｅｄ　Ｃｉｒｃｕｉｔ）カード、ＳＤカード、ＤＶＤ（Ｄｉｇｉｔａｌ　Ｖｅｒｓａｔｉｌｅ　Ｄｉｓｃ）の記録媒体に格納することができる。

【０１２６】

また、制御線や情報線は説明上必要と考えられるものを示しており、実装上必要な全ての制御線や情報線を示しているとは限らない。実際には、ほとんど全ての構成が相互に接続されていると考えてよい。

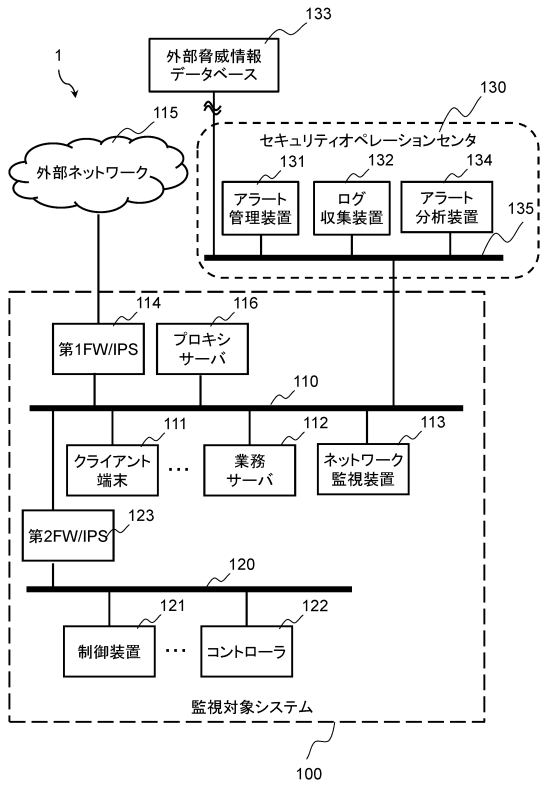
【符号の説明】

【０１２７】

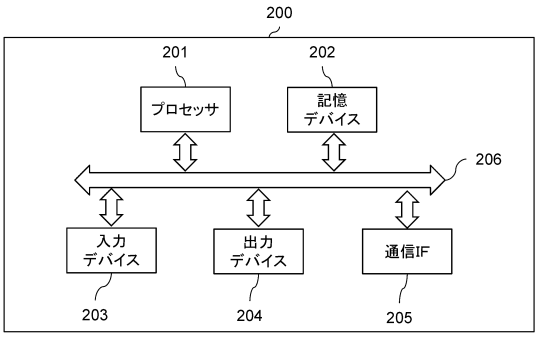
１００　監視対象システム
１３１　アラート管理装置
１３２　ログ収集装置
１３３　外部脅威情報データベース

１３４　アラート分析装置
３０１　アラート判断集計部
３０２　ログ統計集計部
３０３　要因抽出部
３０４　重要度予測部
３０５　誤差要因抽出部
３０６　表示部
５００　アラート判断集計テーブル
６００　ログ統計集計テーブル
７００　データ種別管理テーブル
９００　抽出要因テーブル
１０００　誤差テーブル
１２００　誤差要因テーブル

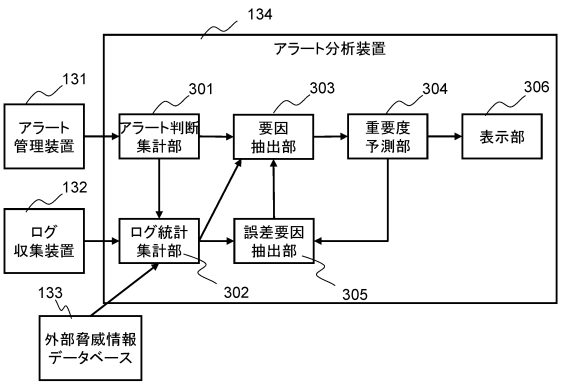
【図１】



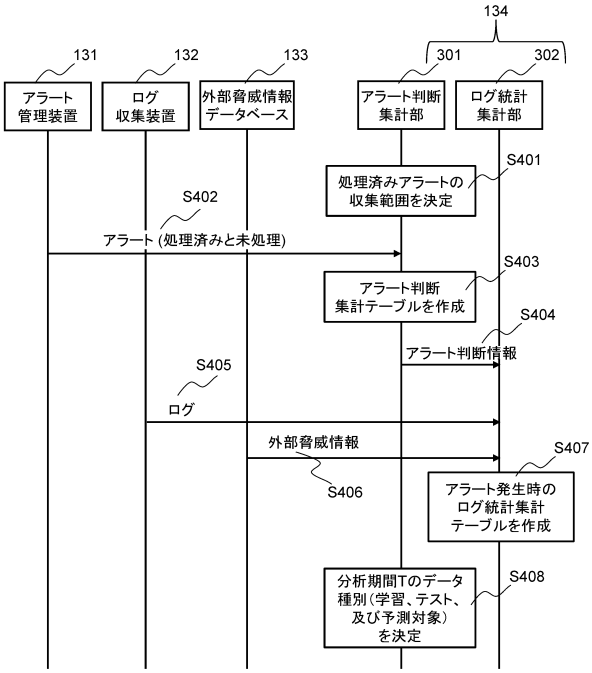
【図２】



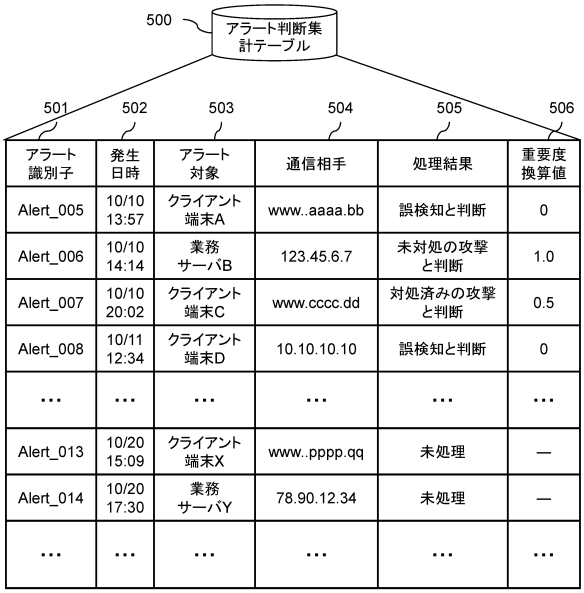
【図３】



【図４】

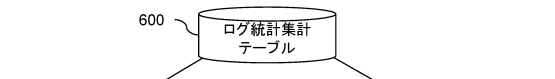


【図５】



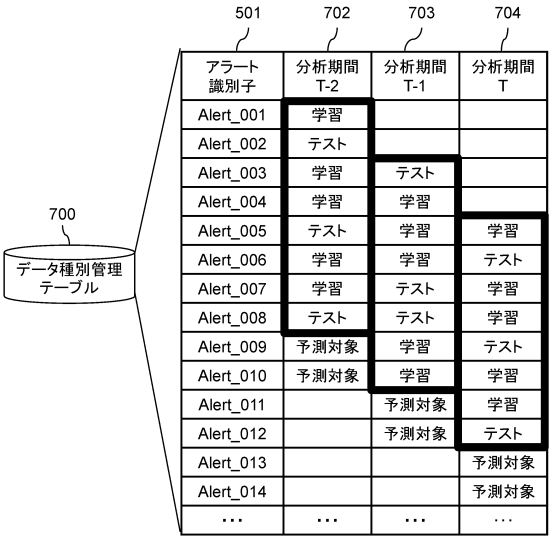
アラート識別子	発生日時	アラート対象	通信相手	処理結果	重要度換算値
Alert_005	10/10 13:57	クライアント端末A	www..aaaa.bb	誤検知と判断	0
Alert_006	10/10 14:14	業務サーバB	123.45.6.7	未対処の攻撃と判断	1.0
Alert_007	10/10 20:02	クライアント端末C	www.cccc.dd	対処済みの攻撃と判断	0.5
Alert_008	10/11 12:34	クライアント端末D	10.10.10.10	誤検知と判断	0
…	…	…	…	…	…
Alert_013	10/20 15:09	クライアント端末X	www..pppp.qq	未処理	—
Alert_014	10/20 17:30	業務サーバY	78.90.12.34	未処理	—
…	…	…	…	…	…

【図６】



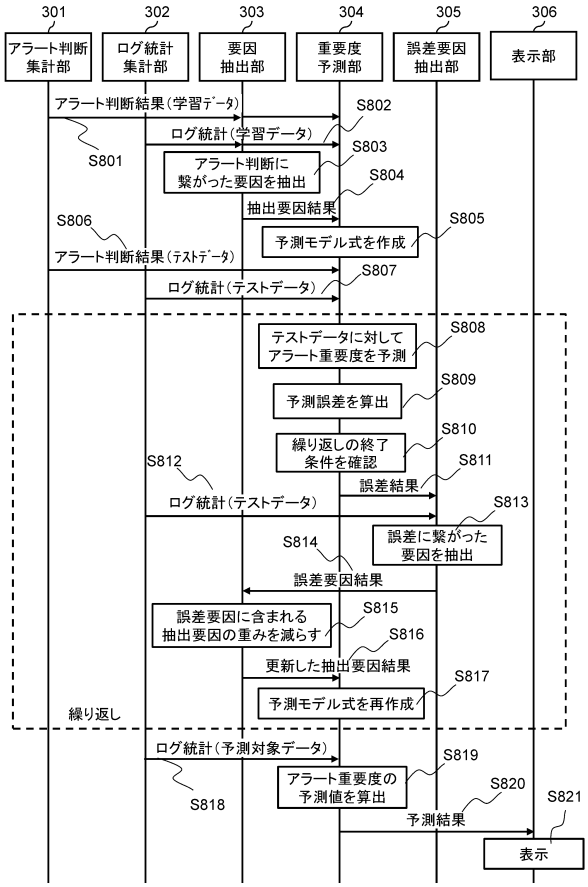
アラート識別子	集計日時	プロキシサーバログ		業務サーバログ		外部脅威情報	
		キャッシュミス回数	異常応答回数	異常応答回数	アクセス回数	IPアドレス危険度	URL危険度
Alert_005	10/10 12:57	3	0	0	4	0	0
Alert_005	10/10 13:07	4	1	1	1	0	0
…	…	…	…	…	…	…	…
Alert_005	10/10 13:57	4	0	0	7	0	0
Alert_006	10/10 13:14	0	0	300	350	4	0
Alert_006	10/10 13:24	3	0	100	120	4	0
…	…	…	…	…	…	…	…

【図７】

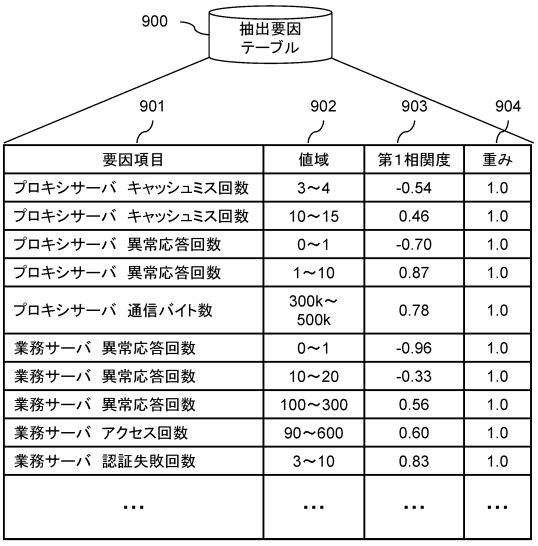


アラート識別子	分析期間T-2	分析期間T-1	分析期間T
Alert_001	学習		
Alert_002	テスト		
Alert_003	学習	テスト	
Alert_004	学習	学習	
Alert_005	テスト	学習	学習
Alert_006	学習	学習	テスト
Alert_007	学習	テスト	学習
Alert_008	テスト	テスト	学習
Alert_009	予測対象	学習	テスト
Alert_010	予測対象	学習	学習
Alert_011		予測対象	学習
Alert_012		予測対象	テスト
Alert_013			予測対象
Alert_014			予測対象
…	…	…	…

【図８】

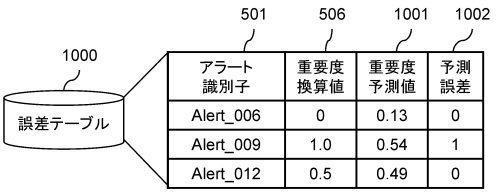


【図９】

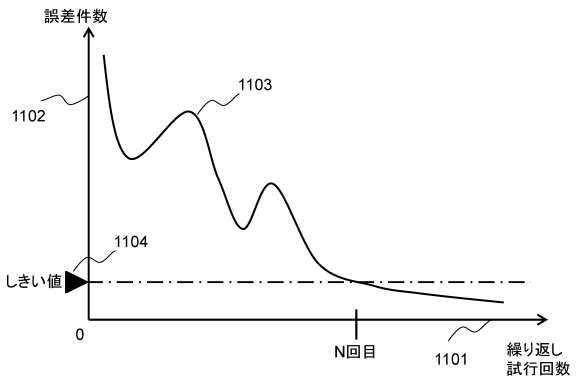


要因項目	値域	第1相関度	重み
プロキシサーバ キャッシュミス回数	3～4	-0.54	1.0
プロキシサーバ キャッシュミス回数	10～15	0.46	1.0
プロキシサーバ 異常応答回数	0～1	-0.70	1.0
プロキシサーバ 異常応答回数	1～10	0.87	1.0
プロキシサーバ 通信バイト数	300k～500k	0.78	1.0
業務サーバ 異常応答回数	0～1	-0.96	1.0
業務サーバ 異常応答回数	10～20	-0.33	1.0
業務サーバ 異常応答回数	100～300	0.56	1.0
業務サーバ アクセス回数	90～600	0.60	1.0
業務サーバ 認証失敗回数	3～10	0.83	1.0
…	…	…	…

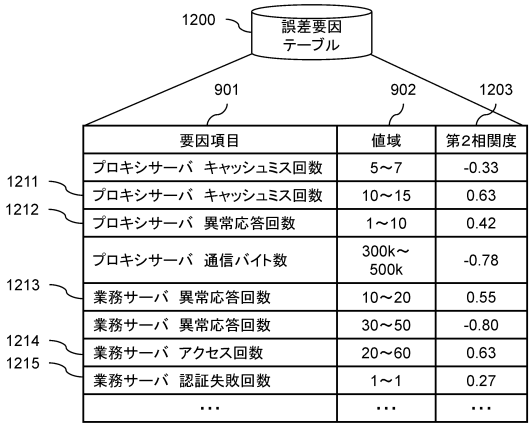
【図１０】



【図１１】

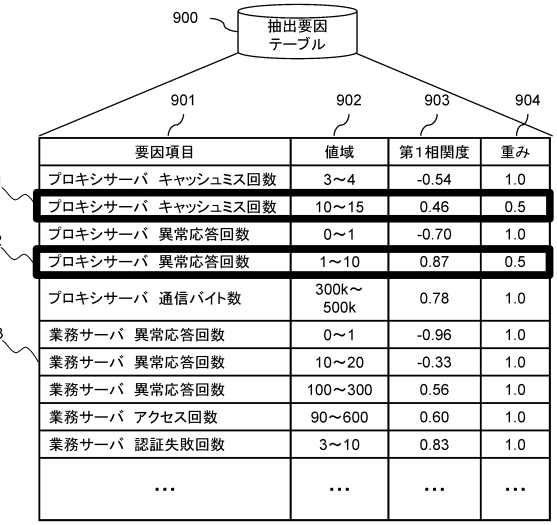


【図１２】



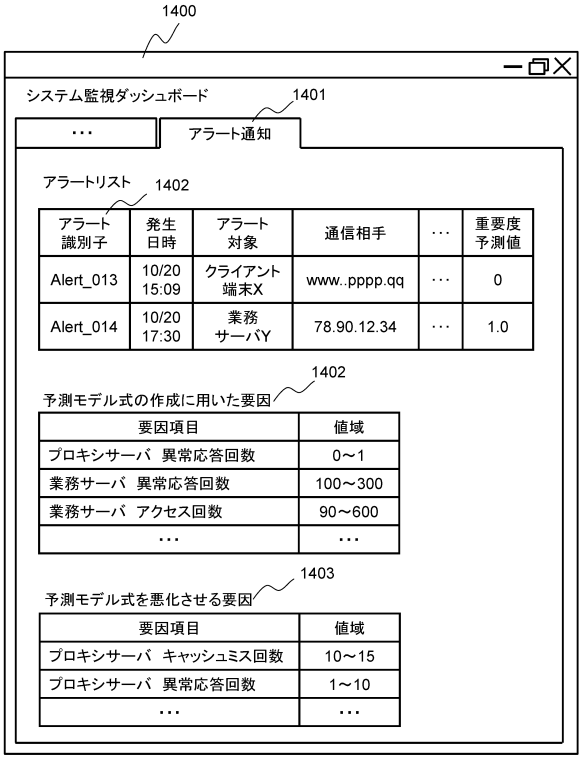
要因項目	値域	第2相関度
プロキシサーバ キャッシュミス回数	5～7	-0.33
プロキシサーバ キャッシュミス回数	10～15	0.63
プロキシサーバ 異常応答回数	1～10	0.42
プロキシサーバ 通信バイト数	300k～500k	-0.78
業務サーバ 異常応答回数	10～20	0.55
業務サーバ 異常応答回数	30～50	-0.80
業務サーバ アクセス回数	20～60	0.63
業務サーバ 認証失敗回数	1～1	0.27
…	…	…

【図１３】

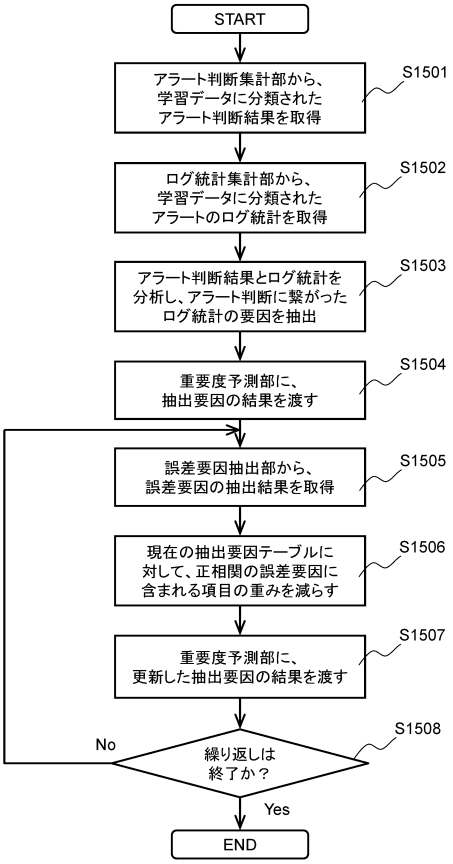


要因項目	値域	第1相関度	重み
プロキシサーバ キャッシュミス回数	3～4	-0.54	1.0
プロキシサーバ キャッシュミス回数	10～15	0.46	0.5
プロキシサーバ 異常応答回数	0～1	-0.70	1.0
プロキシサーバ 異常応答回数	1～10	0.87	0.5
プロキシサーバ 通信バイト数	300k～500k	0.78	1.0
業務サーバ 異常応答回数	0～1	-0.96	1.0
業務サーバ 異常応答回数	10～20	-0.33	1.0
業務サーバ 異常応答回数	100～300	0.56	1.0
業務サーバ アクセス回数	90～600	0.60	1.0
業務サーバ 認証失敗回数	3～10	0.83	1.0
…	…	…	…

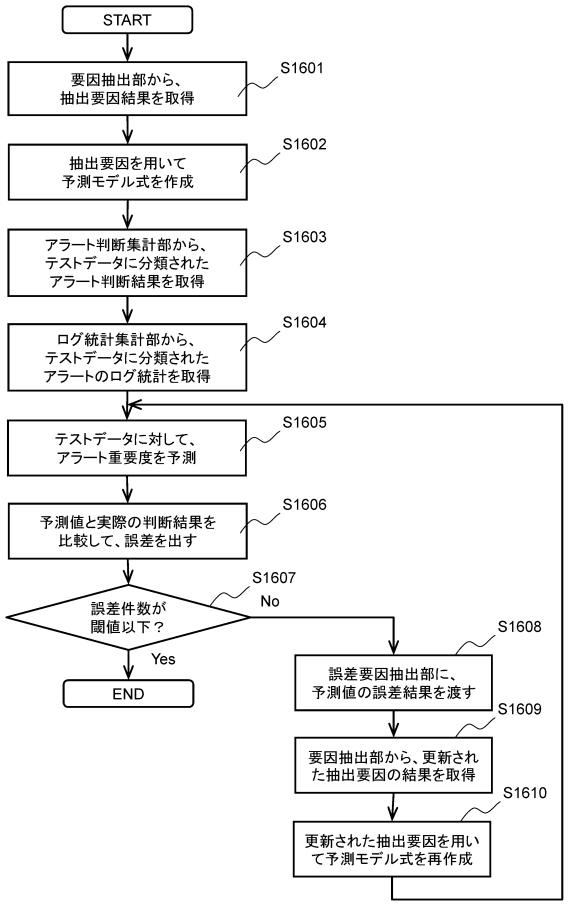
【図１４】


1400
システム監視ダッシュボード
1401
…
アラート通知
アラートリスト
1402
アラート識別子
発生日時
アラート対象
通信相手
…
重要度予測値
Alert_013
10/20 15:09
クライアント端末X
www..pppp.qq
…
0
Alert_014
10/20 17:30
業務サーバY
78.90.12.34
…
1.0
1402
予測モデル式の作成に用いた要因
要因項目
値域
プロキシサーバ 異常応答回数
0～1
業務サーバ 異常応答回数
100～300
業務サーバ アクセス回数
90～600
…
…
1403
予測モデル式を悪化させる要因
要因項目
値域
プロキシサーバ キャッシュミス回数
10～15
プロキシサーバ 異常応答回数
1～10
…
…

【図１５】


START
アラート判断集計部から、学習データに分類されたアラート判断結果を取得
S1501
ログ統計集計部から、学習データに分類されたアラートのログ統計を取得
S1502
アラート判断結果とログ統計を分析し、アラート判断に繋がったログ統計の要因を抽出
S1503
重要度予測部に、抽出要因の結果を渡す
S1504
誤差要因抽出部から、誤差要因の抽出結果を取得
S1505
現在の抽出要因テーブルに対して、正相関の誤差要因に含まれる項目の重みを減らす
S1506
重要度予測部に、更新した抽出要因の結果を渡す
S1507
繰り返しは終了か？
S1508
No
Yes
END

【図１６】


START
要因抽出部から、抽出要因結果を取得
S1601
抽出要因を用いて予測モデル式を作成
S1602
アラート判断集計部から、テストデータに分類されたアラート判断結果を取得
S1603
ログ統計集計部から、テストデータに分類されたアラートのログ統計を取得
S1604
テストデータに対して、アラート重要度を予測
S1605
予測値と実際の判断結果を比較して、誤差を出す
S1606
誤差件数が閾値以下？
S1607
No
Yes
END
誤差要因抽出部に、予測値の誤差結果を渡す
S1608
要因抽出部から、更新された抽出要因の結果を取得
S1609
更新された抽出要因を用いて予測モデル式を再作成
S1610

フロントページの続き

審査官　松尾　真人

(56)参考文献　国際公開第２０１５／００４７４２（WO，Ａ１）
国際公開第２０１６／０６３４４６（WO，Ａ１）
特開２００９－００３５６１（ＪＰ，Ａ）
特開２０１７－０９０９４７（ＪＰ，Ａ）
特開２０１３－０７３４８９（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)

Ｇ０６Ｆ　１６／００－１６／９５８
Ｇ０６Ｆ　１１／０７
１１／２８－１１／３６
Ｇ０６Ｆ　２１／１２－２１／１６
２１／５０－２１／５７
Ｇ０６Ｆ　１７／００－１７／１８
Ｇ０６Ｑ　１０／００－９９／００
Ｇ１６Ｚ　９９／００