(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7632615号
(P7632615)

(45)発行日　令和7年2月19日(2025.2.19)

(24)登録日　令和7年2月10日(2025.2.10)

(51)国際特許分類

Ｈ０４Ｌ　43/02　(2022.01)		Ｈ０４Ｌ	43/02
Ｈ０４Ｌ　43/04　(2022.01)		Ｈ０４Ｌ	43/04

ＦＩ

請求項の数　10　（全17頁）

(21)出願番号　特願2023-527161(P2023-527161)
(86)(22)出願日　令和3年6月7日(2021.6.7)
(86)国際出願番号　PCT/JP2021/021581
(87)国際公開番号　WO2022/259317
(87)国際公開日　令和4年12月15日(2022.12.15)
審査請求日　令和5年9月22日(2023.9.22)

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　110002147
弁理士法人酒井国際特許事務所
(72)発明者　寺本 泰大
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　山田 真徳
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　山中 友貴
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　高橋 知克
最終頁に続く

(54)【発明の名称】　検出装置、検出方法及び検出プログラム

(57)【特許請求の範囲】

【請求項１】

パケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値を予め記憶する記憶部と、

正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換する変換部と、

前記第１の自然言語処理モデルを用いて変換された特徴量と、前記記憶部において記憶されたデータとを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に前記処理対象のパケットの異常の有無を判定する判定部と、

を有し、

前記判定部は、前記第１の自然言語処理モデルを用いて変換された特徴量のうち、正常ラベルを付与した特徴量を、攻撃を検出する第１の検出モデルの学習データとして出力することを特徴とする検出装置。

【請求項２】

パケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値を予め記憶する記憶部と、

正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換する変換部と、

前記第１の自然言語処理モデルを用いて変換された特徴量と、前記記憶部において記憶

されたデータとを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に前記処理対象のパケットの異常の有無を判定する判定部と、
を有し、
前記記憶部は、事前学習に用いた特徴量の全てまたは代表的な特徴量に、事前学習に用いた特徴量であることを示す事前学習ラベルを対応付けて記憶し、
前記判定部は、前記事前学習ラベルが付与された特徴量と、前記第１の自然言語処理モデルを用いて変換された前記処理対象のパケットの特徴量との類似度を基に、前記第１の自然言語処理モデルにおける適正な特徴量の変換の可否を判定することを特徴とする検出装置。

【請求項３】
前記判定部は、前記特徴量に前記ラベルが付与されていない前記処理対象のパケットに関する通知を出力し、該処理対象のパケットに対するラベルが入力された場合、前記記憶部に、該処理対象のパケットの特徴量と、入力された前記ラベルとを対応付けて記憶させることを特徴とする請求項１または２に記載の検出装置。

【請求項４】
前記判定部は、前記ラベルを付与した特徴量に対して、付与した前記ラベルが誤りであることが入力された場合には、前記記憶部が記憶する前記閾値を更新することを特徴とする請求項１～３のいずれか一つに記載の検出装置。

【請求項５】
前記判定部が、前記第１の自然言語処理モデルにおいて適正に特徴量が変換されていないと判定した場合、前記変換部に、新たな第２の自然言語処理モデルを設け、前記第２の自然言語処理モデルに前記処理対象のパケットの特徴量への変換を学習させる処理制御部をさらに有し、
前記判定部は、前記第２の自然言語処理モデルが変換した特徴量を、攻撃を検出する第２の検出モデルの学習データとして出力する請求項２に記載の検出装置。

【請求項６】
前記判定部は、前記第１の自然言語処理モデルと前記第２の自然言語処理モデルとのうち、変換後の特徴量が、学習データであるパケットの特徴量と最も類似する自然言語処理モデルを選択し、選択した自然言語処理モデルに対応する検出モデルを用いて攻撃の検出を実行させることを特徴とする請求項５に記載の検出装置。

【請求項７】
検出装置が実行する検出方法であって、
正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換する工程と、
前記第１の自然言語処理モデルを用いて変換された特徴量と、予め求められたパケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値とを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に前記処理対象のパケットの異常の有無を判定する工程と、
を含み、
前記判定する工程は、前記第１の自然言語処理モデルを用いて変換された特徴量のうち、正常ラベルを付与した特徴量を、攻撃を検出する第１の検出モデルの学習データとして出力することを特徴とする検出方法。

【請求項８】
検出装置が実行する検出方法であって、
正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換する工程と、
前記第１の自然言語処理モデルを用いて変換された特徴量と、予め求められたパケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値とを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与した

ラベルを基に前記処理対象のパケットの異常の有無を判定する工程と、
を有し、
前記予め求められたパケットの特徴量のうち、事前学習に用いた特徴量の全てまたは代表的な特徴量は、事前学習に用いた特徴量であることを示す事前学習ラベルが対応付けられており、
前記判定する工程は、前記事前学習ラベルが付与された特徴量と、前記第１の自然言語処理モデルを用いて変換された前記処理対象のパケットの特徴量との類似度を基に、前記第１の自然言語処理モデルにおける適正な特徴量の変換の可否を判定することを特徴とする検出方法。
【請求項９】
正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換するステップと、
前記第１の自然言語処理モデルを用いて変換された特徴量と、予め求められたパケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値とを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に前記処理対象のパケットの異常の有無を判定するステップと、
をコンピュータに実行させ、
前記判定するステップは、前記第１の自然言語処理モデルを用いて変換された特徴量のうち、正常ラベルを付与した特徴量を、攻撃を検出する第１の検出モデルの学習データとして出力する検出プログラム。
【請求項１０】
正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換するステップと、
前記第１の自然言語処理モデルを用いて変換された特徴量と、予め求められたパケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値とを基に、前記第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に前記処理対象のパケットの異常の有無を判定するステップと、
をコンピュータに実行させ、
前記予め求められたパケットの特徴量のうち、事前学習に用いた特徴量の全てまたは代表的な特徴量は、事前学習に用いた特徴量であることを示す事前学習ラベルが対応付けられており、
前記判定するステップは、前記事前学習ラベルが付与された特徴量と、前記第１の自然言語処理モデルを用いて変換された前記処理対象のパケットの特徴量との類似度を基に、前記第１の自然言語処理モデルにおける適正な特徴量の変換の可否を判定する検出プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、検出装置、検出方法及び検出プログラムに関する。
【背景技術】
【０００２】
ネットワーク内のトラフィックから不正なパケットを検出する技術として、シグネチャ型検出技術や、アノマリ型検出技術がある。
【０００３】
シグネチャ型検出技術は、事前に攻撃データから攻撃に含まれる特有のシグネチャを生成して、通信とシグネチャのマッチングにより攻撃を検出する。シグネチャ型検出技術は、既知の攻撃を良く検出出来る一方、攻撃の亜種・難読化や未知の攻撃への対応が難しい。また、シグネチャ型検出技術は、シグネチャ生成のために膨大な実攻撃データを収集する必要がある。
【０００４】

アノマリ型検出技術は、システム内のトラフィックから正常状態を学習して、正常状態の乖離を異常と検出する。アノマリ型検出技術は、シグネチャ型検出技術と比較すると正常な通信データの学習が必要なものの、未知の攻撃も含めて検出可能であり、攻撃データを事前に収集する必要もない。

【先行技術文献】

【特許文献】

【０００５】

【文献】特開２０２０－１０２６７１号公報

【非特許文献】

【０００６】

【文献】Diederik P. Kingma, Max Welling, “Auto-Encoding Variational Bayes”, ICLR 2014., ［online］, ［令和３年５月１４日検索］、インターネット＜ＵＲＬ：https://arxiv.org/pdf/1312.6114.pdf＞

【発明の概要】

【発明が解決しようとする課題】

【０００７】

しかしながら、機械学習を用いたアノマリ型検出技術は、学習に時間を要するため、学習が完了するまでの間、無防備状態のまま、運用と学習を並行する必要がある。このため、機械学習を用いたアノマリ型検出技術は、学習期間中に攻撃を受けるおそれや、攻撃パケットが混入していた場合には、この攻撃パケットを正常パケットとして誤認して学習するおそれがある。

【０００８】

シグネチャ型検出技術を併用する方法もあるが、シグネチャ生成のためには事前に様々な攻撃情報を収集する必要がある。したがって、アノマリ型検出技術において、システムの可用性を落とさずに、監視の空白期間を削減する、即時に攻撃を検出可能な学習モデルがあることが望ましい。

【０００９】

本発明は、上記に鑑みてなされたものであって、システムの可用性を保持しながら、監視の空白期間を削減することができる検出装置、検出方法及び検出プログラムを提供することを目的とする。

【課題を解決するための手段】

【００１０】

上述した課題を解決し、目的を達成するために、本発明の検出装置は、パケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値を予め記憶する記憶部と、正常通信のパケットを学習データとして事前に学習済みである第１の自然言語処理モデルを用いて、処理対象のパケットを特徴量に変換する変換部と、第１の自然言語処理モデルを用いて変換された特徴量と、記憶部において記憶されたデータとを基に、第１の自然言語処理モデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に処理対象のパケットの異常の有無を判定する判定部と、を有することを特徴とする。

【発明の効果】

【００１１】

本発明によれば、システムの可用性を落とさずに、監視の空白期間を削減することができる。

【図面の簡単な説明】

【００１２】

【図１】図１は、実施の形態に係る検出装置の構成の一例を示す図である。

【図２】図２は、ＶＡＥによる検出処理を説明する図である。

【図３】図３は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。

【図４】図４は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。

【図５】図５は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。

【図６】図６は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。
【図７】図７は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。
【図８】図８は、ＢＥＲＴモデルを用いた異常判定処理の処理手順を示すフローチャートである。
【図９】図９は、複数のＢＥＲＴモデルを用いた判定処理の処理手順を示すフローチャートである。
【図１０】図１０は、プログラムが実行されることにより、検出装置が実現されるコンピュータの一例を示す図である。
【発明を実施するための形態】
【００１３】
以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。
【００１４】
［実施の形態］
本発明の実施の形態について説明する。本実施の形態では、事前学習した自然言語処理モデルを用いてパケットから変換した特徴量と、過去のパケットの特徴量とを比較することで、パケットの異常の有無を判定する新たな判定方法を、アノマリ型検出モデルの学習期間中に使用する。この判定方法は、学習が不要であり、検出の処理時間が比較対象の過去のパケットの特徴量のデータ量に依存するのみであるため、即時に攻撃を検出可能である。このため、実施の形態では、この判定方法を、アノマリ型検出モデルの学習期間に用いることによって、システムの可用性を保持しながら、監視の空白期間を削減することができる。
【００１５】
［検出装置の構成］
続いて、図１を用いて、実施の形態に係る検出装置の構成及び処理について説明する。図１は、実施の形態に係る検出装置の構成の一例を示す図である。
【００１６】
実施の形態に係る検出装置１０は、例えば、ＲＯＭ（Read Only Memory）、ＲＡＭ（Random Access Memory）、ＣＰＵ（Central Processing Unit）等を含むコンピュータ等に所定のプログラムが読み込まれて、ＣＰＵが所定のプログラムを実行することで実現される。図２に示すように、検出装置１０は、収集部１１、エンコード部１２（変換部）、検出部１３、判定部１４、処理制御部１５及び特徴量データベース（ＤＢ）１６を有する。
【００１７】
収集部１５１は、処理対象のパケットを収集する。
【００１８】
エンコード部１５２は、自然言語処理モデル（例えば、ＢＥＲＴ（Bidirectional Encoder Representations from Transformers）モデル）を用いて、処理対象パケットを、特徴量である、１つの固定長ベクトルに変換する。ＢＥＲＴモデルは、正常通信のパケットを学習データとして事前に学習済みである。
【００１９】
ＢＥＲＴモデルは、１つのパケットを１つの文章とみなし、１つの固定長ベクトルへ変換する規則を学習したモデル、言い換えると、正常通信のパケットにおける内部のバイト列の順序等の頻出パターンを学習したモデルである。
【００２０】
例えば、ＢＥＲＴモデルは、文書内のある単語を周辺の単語から予測する、という補助タスクを解くことで、パケットの特徴を反映した良い中間表現、すなわち、固定長ベクトルを習得する。また、ＢＥＲＴモデルは、パケット内のある位置のバイトを、周辺のバイトから予測することで、パケットの特徴を反映したベクトル表現を獲得する。エンコード

部１５２は、ＢＥＲＴモデルを用いて、各パケットを、各パケットの特徴を反映した固定長ベクトルに変換する。

【００２１】

エンコード部１５２は、一つのＢＥＲＴモデル（第１の自然言語処理モデル）を保持するほか、後述するように、このＢＥＲＴモデルが未対応であるプロトコルのパケットを処理する場合には、新たなＢＥＲＴモデル（第２の自然言語処理モデル）を用いてもよい。

【００２２】

検出部１３は、検出モデルを用いて、エンコード部１５２によって変換された固定長ベクトルを基に、パケットの異常の有無を検出することで攻撃を検出する。検出装置１０では、判定部１４（後述）によって正常通信であると判定されたパケットの固定長ベクトルのパターンを学習させる。

【００２３】

検出部１５３では、例えば、ＶＡＥ（Variational Auto Encoder）、ＡＥ（Auto Encoder）、ＬｏＦ（Local Outlier Factor）等の教師なし学習による検出モデルを検出モデルとして用いる。以降、検出モデルとして、ＶＡＥを用いる場合を例に説明する。

【００２４】

例えば、ＶＡＥでは、正常通信パケットの確率密度を学習後、確率密度の低い通信を異常として検出するため、正常通信パケットのみが分かればよく、全ての悪性データを学習せずとも異常検出が可能である。ＶＡＥは、正常通信のパケットに対応する固定長ベクトルを学習データとして、異常度の学習を行う。

【００２５】

検出部１３は、一つのＶＡＥ（第１の検出モデル）を保持する。このほか、エンコード部１２が、未対応プロトコルのパケットの処理のために新たなＢＥＲＴモデルを用いる場合、検出部１３は、この新たなＢＥＲＴモデルが変換した固定長ベクトルを処理するために、新たなＶＡＥ（第２の検出モデル）を用いてもよい。

【００２６】

特徴量ＤＢ１６は、パケットの特徴量、各パケットの特徴量に付与されたラベル、及び、判定に使用する閾値を予め記憶する。特徴量は、ＢＥＲＴモデルによって変換されたパケットの固定長ベクトルである。

【００２７】

ラベルは、正常ラベル、異常ラベル、事前学習ラベルがある。正常ラベルは、正常通信のパケットを変換した固定長ベクトルであることを示すラベルである。異常ラベルは、異常通信のパケットを変換した固定長ベクトルであることを示すラベルである。事前学習ラベルは、事前学習に用いた正常通信のパケットを変換した固定長ベクトルであることを示すラベルである。特徴量ＤＢ１６は、事前学習に用いた固定長ベクトルの全てまたは代表的な固定長ベクトルに、事前学習ラベルを対応付けて記憶する。閾値は、判定部１４における、処理対象のパケットの固定長ベクトルと、特徴量ＤＢ１６に記憶される固定長ベクトルとに対する類似判定に使用される。

【００２８】

判定部１４は、エンコード部１２のＢＥＲＴモデルを用いて変換された固定長ベクトルと、特徴量ＤＢ１６において記憶されたデータとを基に、ＢＥＲＴモデルを用いて変換された特徴量にラベルを付与し、付与したラベルを基に処理対象のパケットの異常の有無を判定する。

【００２９】

判定部１４は、ＢＥＲＴモデルを用いて変換された固定長ベクトルに対し、閾値以上の類似度を有する固定長ベクトルを特徴量ＤＢ１６から検索する。そして、判定部１４は、検索した固定長ベクトルに付与されたラベルを、ＢＥＲＴモデルによって変換された固定長ベクトルに付与する。判定部１４は、ＢＥＲＴモデルを用いて変換された固定長ベクトルのうち、正常ラベルを付与した固定長ベクトルを、検出部１３のＶＡＥの学習データとして出力する。

【００３０】
判定部１４は、事前学習ラベルが付与された固定長ベクトルと、エンコード部１２のＢＥＲＴモデルを用いて変換された処理対象のパケットの固定長ベクトルとの類似度を基に、このＢＥＲＴモデルにおける適正な固定長ベクトルの変換の可否を判定する。
【００３１】
処理制御部１５は、判定部１４が、ＢＥＲＴモデルにおいて適正に特徴量が変換されていないと判定した場合、エンコード部１２に、新たなＢＥＲＴモデルを設け、この新たなＢＥＲＴモデルに、処理対象のパケットの固定長ベクトルへの変換を学習させる。
【００３２】
［ＶＡＥの処理］
まず、検出部１３のＶＡＥの学習処理及び検出処理について説明する。図２は、ＶＡＥによる検出処理を説明する図である。
【００３３】
図２に示すように、収集部１１は、学習対象或いは評価対象のパケットとして、可変長のパケットを複数取得する（図２の（１））。
【００３４】
エンコード部１２は、各パケットのバイト列の順序等の頻出パターンを学習したＢＥＲＴモデルを用いて、各パケットを、各パケットの特徴を反映した固定長ベクトルにそれぞれ変換する（図２の（２），（３））。
【００３５】
検出部１３は、ＶＡＥを用いて、エンコード部１２によって変換された固定長ベクトルに対する異常度（異常パケットの発生頻度）を取得する。学習段階では、ＶＡＥは、正常通信のパケットに対応する固定長ベクトルに基に、異常度（例えば、アノマリスコア）の学習を行う（図２の（４））。この場合、ＶＡＥが取得する異常度が最小化されるように、ＶＡＥの検出モデルのパラメータが調整される。
【００３６】
評価段階では、検出部１３は、ＶＡＥを用いて、エンコード部１２によって変換された固定長ベクトルに対する異常度（例えば、アノマリスコア）を取得する（図２の（５））。そして、検出部１３は、アノマリスコアが所定の閾値を超えている場合に、評価対象のパケットの異常を検出し、アラートを発生する。閾値は、例えば、要求されている検出精度や、検出装置１０のリソース等に応じて、設定される。
【００３７】
［ＢＥＲＴモデルを用いた異常判定処理］
次に、ＢＥＲＴモデルを用いた異常判定処理を説明する。図３～図５は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。
【００３８】
検出部１３のＶＡＥでは、通信内容の分析のために、本番環境での異常度を算出するための学習が必要である（図３の（１）及び図４の（Ｂ））。しかしながら、ＶＡＥでは、学習に時間を要するため、学習期間中に攻撃を受けるおそれや、攻撃パケットを正常状態として誤認して学習するおそれがある。
【００３９】
ここで、適切にチューニングされた検出装置では、エンコード部１２のＢＥＲＴモデルがパケットのプロトコル構造を分析し（図３の（２））、検出部１３のＶＡＥがその環境における通信のパラメータを分析する。このＢＥＲＴモデルは、正常通信のパケットの内部のバイト列の順序等の頻出パターンを学習させることで、事前に作成可能である。
【００４０】
ここで、ＢＥＲＴモデルのエンコードによって出力された固定長ベクトルを用いることで、パケット同士の近さを比較することが可能である。これを基に、ＢＥＲＴモデルが変換した処理対象のパケットの固定長ベクトルと、予め求めておいた正常通信または異常通信のパケットの固定長ベクトルとのいずれに類似するかを求めることで、処理対象のパケ

ットが正常または異常であるかを判別できる。そこで、本実施の形態では、ＶＡＥの学習が完了するまで、ＢＥＲＴモデルを用いた異常パケット判定処理を攻撃検出に利用する（図３の（３））。

【００４１】

具体的には、異常パケット判定処理として、判定部１４が、ＢＥＲＴモデルが変換した処理対象のパケットの固定長ベクトルと、予め求めておいた正常通信または異常通信のパケットの固定長ベクトルとの類似度を用いて、処理対象のパケットの異常の有無を判定する（図４の（Ａ））。例えば、判定部１４は、類似度として、ｃｏｓ類似度やＬ２ノルムを用いる。

【００４２】

ＶＡＥを用いた検出処理では、検出の処理時間は一定であるものの、ある程度データを収集した上で学習しないと、誤検出に繋がる場合があり、また、過検出補正はモデル全体の再学習が必要である。

【００４３】

これに対し、ＢＥＲＴモデルを用いた異常判定処理では、検出の処理時間はデータ量に依存するものの、過去のパケットとの比較のみでパケットの異常の有無を判定することが可能であるため、学習が不要である。また、ＢＥＲＴモデルを用いた異常判定処理では、過検出補正は過去のパケットのラベル付けを行うのみで足り、ラベル付けや利用データの工夫によって検出以外の様々な用途に利用可能である。

【００４４】

続いて、図５を参照して、ＢＥＲＴモデルを用いた異常判定処理の適用について説明する。図５に示すように、特徴量ＤＢ１６には、ＢＥＲＴモデルの事前学習で用いた正常通信のパケットの特徴量（固定長ベクトル）、及び、本番学習で用いた正常通信のパケットの特徴量、異常通信のパケットの特徴量と、各パケットに対応付けられたラベルと、判定に使用するための閾値とが記憶される（矢印Ｙ１）。

【００４５】

検出部１３のＶＡＥの学習中、判定部１４は、事前学習したエンコード部１２のＢＥＲＴモデルがパケットから変換した固定長ベクトルと（矢印Ｙ２）、特徴量ＤＢ１６が記憶するデータ（矢印Ｙ３）とを用いて、入力されたパケットのラベル付けを行う。

【００４６】

具体的には、判定部１４は、ＢＥＲＴモデルを用いて変換された固定長ベクトルと、特徴量ＤＢ１６の固定長ベクトルとの間の類似度を算出する（図５の（１））。判定部１４は、閾値以上の類似度を有する固定長ベクトルを特徴量ＤＢ１６から抽出する。そして、判定部１４は、抽出した固定長ベクトルに付与されたラベルを、処理対象のパケットの固定長ベクトルに付与する。判定部１４は、付与したラベルを基に処理対象のパケットの異常の有無を判定する。

【００４７】

判定部１４は、付与したラベルが異常である場合は、処理対象のパケットが異常通信のパケットと類似すると判定し（矢印Ｙ４）、異常アラートを発生させる（図５の（２））。そして、判定部１４は、付与したラベルが正常である場合は、処理対象のパケットが正常通信のパケットと類似すると判定し（矢印Ｙ５）、ＶＡＥの学習データに利用する（図５の（３））。

【００４８】

また、判定部１４は、算出した類似度が閾値を下回る場合、この処理対象のパケットが未知のデータであるとし、管理者に確認要求を通知する。管理者は、この通知を参照すると、処理対象のパケットの正常性を判断し（図５の（４））、判断結果を検出装置１０に入力する。この判断結果は、判定部１４によって、特徴量ＤＢ１６にフィードバックされ、次回の判定に使用されることで、過検出を抑制する。言い換えると、判定部１４は、特徴量にラベルが付与されていない処理対象のパケットに関する通知を出力し、該処理対象のパケットに対するラベルが入力された場合、特徴量ＤＢ１６に、該処理対象のパケット

の特徴量と、入力されたラベルとを対応付けて記憶させる。

【００４９】

また、判定部１４は、事前学習ラベルが付された特徴量との類似度を算出し、算出した類似度が閾値を下回る場合には、ＢＥＲＴモデルが未対応であるプロトコルのパケットとして管理者への通知を行い（矢印Ｙ６）、正常度判定を行わない。または、判定部は、7-tuple判定を利用して、判定を行ってもよい。

【００５０】

判定部１４における判定例を説明する。例えば、パケット１については、管理者によって正常と確認される。判定部１４は、パケットの１の特徴量に正常ラベルを対応付けて特徴量ＤＢ１６に保存する。パケット２については、判定部１４は、パケット１との類似度を算出し、類似度が閾値を下回るため、管理者に確認要求を通知する。パケット２については、管理者によって正常と確認されたため、判定部１４は、パケットの２の特徴量に正常ラベルを対応付けて特徴量ＤＢ１６に保存する。

【００５１】

判定部１４は、パケット３については、パケット１，２との類似度を算出し、いずれのパケットについても類似度が閾値を下回るため、管理者に確認要求を通知する。パケット３については、管理者によって異常と確認されたため、判定部１４は、パケットの３の特徴量に異常ラベルを対応付けて特徴量ＤＢ１６に保存する。

【００５２】

判定部１４は、パケット４については、パケット１～３との類似度を算出し、パケット２と最も類似するため、正常通信と判定する。なお、パケット４については、特徴量ＤＢ１６への特徴量等の保存は行わない。

【００５３】

判定部１４は、パケット５については、パケット１～３との類似度を算出し、パケット３と最も類似するため、異常通信と判定し、アラートを出力する。なお、初期の過検出抑制のために、判定部１４は、最初の数分の間は、処理パケットを正常データとして処理し、各特徴量に正常ラベルを付与して特徴量ＤＢ１６に保管してもよい。また、新たな通信パターンが発生するたびに特徴量ＤＢ１６のレコードが増え、パフォーマンスが落ちることから、検出装置１０は、検出部１３のＶＡＥの学習が完了した場合には、検出方法を、ＶＡＥを介する方法（図３の（Ｂ））に切り替える。

【００５４】

［複数のＢＥＲＴモデルを用いた判定処理］

また、検出装置１０では、複数のＢＥＲＴモデル及びＶＡＥを用いて判定を行ってもよい。図６及び図７は、ＢＥＲＴモデルを用いた異常判定処理を説明する図である。

【００５５】

検出装置１０では、第１ＶＡＥ１３－１の学習期間中、事前学習済みの第１ＢＥＲＴモデル１２－１（第１の自然言語処理モデル）が正常通信であると判定したパケットについては、第１ＢＥＲＴモデル１２－１において対応済みであるプロトコルのパケットとして（矢印Ｙ１１）、この第１ＢＥＲＴモデル１２－１に対応する第１ＶＡＥ１３－１に出力する。

【００５６】

そして、事前学習に用いたパケットに含まれていないプロトコル（独自プロトコルやマイナープロトコル）は、事前学習データの代表的な特徴量と比較することで検出が可能である。判定部１４は、事前学習ラベルが付された特徴量と処理対象のパケットの特徴量との類似度を算出し、算出した特徴量が、閾値を下回る場合には、処理対象のパケットが、第１ＢＥＲＴモデル１２－１において未対応であるプロトコルの通信であると判定する（図６の（１））。

【００５７】

ここで、未対応であるプロトコルの通信の学習のために、第１ＢＥＲＴモデル１２－１全体を再学習すると、ＢＥＲＴモデルが生成する特徴量そのものが変化してしまい、第１

ＶＡＥ１３－１も全て再学習する必要がある。
【００５８】
このため、検出装置１０では、第１ＢＥＲＴモデル１２－１が対応していないプロトコルのみ、新規に、第２ＢＥＲＴモデル１２－２（第２の自然言語処理モデル）と、この第２ＢＥＲＴモデル１２－２に対応する第２ＶＡＥ１３－２（第２の検出モデル）とを設けて、検出に利用する（図６の（２））。検出装置１０は、第２ＢＥＲＴモデル１２－２に、処理対象のパケットの特徴量への変換を学習させる。そして、判定部１４は、第２ＢＥＲＴモデル１２－２が変換した特徴量を、第２ＶＡＥ１３－２の学習データとして出力する。このように、検出装置１０によれば、第１ＢＥＲＴモデル１２－１が未対応であるプロトコルの通信パケットが入力された場合、第１ＶＡＥ１３－１の再学習を行わずとも、適切に処理可能である。
【００５９】
そして、監視中においては、図７に示すように、判定部１４は、第１ＢＥＲＴモデル１２－１と第２ＢＥＲＴモデル１２－２とのうち、処理対象のパケットの変換後の特徴量が、学習データの特徴量と最も類似するＢＥＲＴモデルを選択し、選択したＢＥＲＴモデルに対応するＶＡＥモデルを用いて攻撃の検出を実行させる。
【００６０】
例えば、判定部１４は、処理対象のパケットについて、第１ＢＥＲＴモデル１２－１が変換した特徴量と、特徴量ＤＢ１６が保持する事前学習に用いた正常通信のパケットの特徴量との類似度を算出する。そして、判定部１４は、処理対象のパケットについて、第２ＢＥＲＴモデル１２－２が変換した特徴量と、特徴量ＤＢ１６－２が保持する、第２ＢＥＲＴモデル１２－２が学習した本番環境において学習されたパケットの特徴量との類似度を算出する。
【００６１】
続いて、判定部１４は、算出した類似度を比較し、類似度が高い方のＢＥＲＴモデルに対応するＶＡＥを選択して、検出を実行させる。このように、検出装置１０は、検出に適したＢＥＲＴモデル及びＶＡＥを選択して検出を行うことで、検出精度を上げることができる。なお、検出装置１０は、ＶＡＥの再学習のタイミングに合わせて、第１ＢＥＲＴモデル１２－１と第２ＢＥＲＴモデル１２－２も１つのモデルに圧縮してもよい。
【００６２】
［ＢＥＲＴモデルを用いた異常判定処理の処理手順］
次に、ＶＡＥの学習中における、ＢＥＲＴモデルを用いた異常判定処理の処理手順について説明する。図８は、ＢＥＲＴモデルを用いた異常判定処理の処理手順を示すフローチャートである。
【００６３】
図８に示すように、処理対象のパケットが入力されると（ステップＳ１１）、エンコード部１２は、ＢＥＲＴモデルを用いて、特徴量である固定長ベクトルに変換する（ステップＳ１２）。
【００６４】
判定部１４は、ＢＥＲＴモデルによって変換された処理対象のパケットの特徴量と、閾値以上の類似度を有する特徴量及びそのラベルを特徴量ＤＢ１６から検索する（ステップＳ１３）。判定部１４は、検索した特徴量のラベルを、処理対象のパケットの特徴量に付与する。続いて、判定部１４は、処理対象のパケットの特徴量に付与したラベルが正常ラベルであるか否かを判定する（ステップＳ１４）。
【００６５】
ラベルが正常ラベルである場合（ステップＳ１４：Ｙｅｓ）、判定部１４は、この処理対象のパケットの特徴量を、検出部１３のＶＡＥの学習データとして出力し（ステップＳ１５）、検出部１３のＶＡＥの学習を進めさせる。
【００６６】
これに対し、ラベルが正常ラベルではない場合（ステップＳ１４：Ｎｏ）、判定部１４

は、処理対象のパケットの特徴量に付与したラベルが異常であるか否かを判定する（ステップＳ１６）。ラベルが異常ラベルである場合（ステップＳ１６：Ｙｅｓ）、判定部１４は、処理対象のパケットが異常である旨を示すアラートを発する（ステップＳ１７）。

【００６７】

また、ラベルが異常ラベルでない場合（ステップＳ１６：Ｎｏ）、判定部１４は、処理対象のパケットの特徴量に付与したラベルが事前学習ラベルであるか否かを判定する（ステップＳ１８）。

【００６８】

ラベルが事前学習ラベルである場合（ステップＳ１８：Ｙｅｓ）、処理対象のパケットが未知のデータである旨を示すアラートを発する（ステップＳ１９）。判定部１４は、ステップＳ１７，Ｓ１９のアラートに対して、管理者から、このパケットのラベルを含むフィードバックを取得すると（ステップＳ２０）、特徴量ＤＢ１６に、この処理対象のパケットの特徴量に対応させて、管理者の判断したラベルを保存し、類似判定の際に使用する閾値を更新する（ステップＳ２１）。

【００６９】

そして、判定部１４は、管理者によって判断されたラベルが正常ラベルである場合には（ステップＳ２２：Ｙｅｓ）、この処理対象のパケットの特徴量を、検出部１３のＶＡＥの学習データとして出力し（ステップＳ２３）、検出部１３のＶＡＥの学習を進めさせる。管理者によって判断されたラベルが正常ラベルでない場合には（ステップＳ２２：Ｎｏ）、この処理対象のパケットに対する異常判定処理を終了する。

【００７０】

そして、ラベルが事前学習ラベルではない場合（ステップＳ１８：Ｎｏ）、処理対象のパケットがＢＥＲＴモデルの未対応プロトコルのパケットである旨を示すアラートを発する（ステップＳ２４）。この場合、検出装置１０は、エンコード部１２に、新たなＢＥＲＴモデルを設け（ステップＳ２５）、この新たなＢＥＲＴモデルに、処理対象のパケットの固定長ベクトルへの変換を学習させる。そして、検出装置１０は、検出部１３の新たなＶＡＥに、新たなＢＥＲＴモデルが変換した固定長ベクトルを学習データとして学習させる。

【００７１】

［複数のＢＥＲＴモデルを用いた異常判定処理の処理手順］

図９は、複数のＢＥＲＴモデルを用いた判定処理の処理手順を示すフローチャートである。

【００７２】

処理対象のパケットが入力されると（ステップＳ３１）、処理対象のパケットについて、第１ＢＥＲＴモデル１２－１及び第２ＢＥＲＴモデル１２－２は、特徴量の変換を行う（ステップＳ３２）。

【００７３】

判定部１４は、第１ＢＥＲＴモデル１２－１が変換した特徴量と、特徴量ＤＢ１６が保持する事前学習に用いた正常通信のパケットの特徴量との類似度、及び、第２ＢＥＲＴモデル１２－２が変換した特徴量と、特徴量ＤＢ１６－２が保持する特徴量との類似度を算出する。そして、判定部１４は、算出した類似度を比較し、類似度が高いＢＥＲＴモデルを判定する（ステップＳ３３）。

【００７４】

類似度が高いＢＥＲＴモデルが第１ＢＥＲＴモデル１２－１である場合には（ステップＳ３３：第１ＢＥＲＴモデル１２－１）、判定部１４は、第１ＢＥＲＴモデル１２－１が変換した特徴量を第１ＶＡＥ１３－１に入力し、第１ＶＡＥ１３－１を用いて検出を実行させる（ステップＳ３４）。類似度が高いＢＥＲＴモデルが第２ＢＥＲＴモデル１２－２である場合には（ステップＳ３３：第２ＢＥＲＴモデル１２－２）、第２ＢＥＲＴモデル１２－２が変換した特徴量を第２ＶＡＥ１３－２に入力し、第２ＶＡＥ１３－２を用いて検出を実行させる（ステップＳ３５）。

【００７５】
［実施の形態の効果］
実施の形態に係る検出装置１０では、事前学習したＢＥＲＴモデルを用いてパケットから変換した特徴量と、過去のパケットの特徴量とを比較することで、パケットの異常の有無を判定する新たな判定方法を適用する。この判定方法は、監視時における学習が不要であり、検出の処理時間が比較対象の過去のパケットの特徴量のデータ量に依存するのみであるため、即時に攻撃を検出可能である。このため、検出装置１０では、この判定方法を、アノマリ型検出モデルの学習期間に用いることによって、システムの可用性を保持しながら、監視の空白期間を削減することができる。
【００７６】
なお、ＢＥＲＴモデルは、事前学習後に別環境のデータを変換する際も正しく認識するか認証済みであり、ＢＥＲＴモデルを用いた異常の有無の判定も精度よく実行可能である。
【００７７】
［実施の形態のシステム構成について］
検出装置１０の各構成要素は機能概念的なものであり、必ずしも物理的に図示のように構成されていることを要しない。すなわち、検出装置１０の機能の分散及び統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散または統合して構成することができる。
【００７８】
また、検出装置１０においておこなわれる各処理は、全部または任意の一部が、ＣＰＵ、ＧＰＵ（Graphics Processing Unit）、及び、ＣＰＵ、ＧＰＵにより解析実行されるプログラムにて実現されてもよい。また、検出装置１０においておこなわれる各処理は、ワイヤードロジックによるハードウェアとして実現されてもよい。
【００７９】
また、実施の形態において説明した各処理のうち、自動的におこなわれるものとして説明した処理の全部または一部を手動的に行うこともできる。もしくは、手動的におこなわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上述及び図示の処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて適宜変更することができる。
【００８０】
［プログラム］
図１２は、プログラムが実行されることにより、検出装置１０が実現されるコンピュータの一例を示す図である。コンピュータ１０００は、例えば、メモリ１０１０、ＣＰＵ１０２０を有する。また、コンピュータ１０００は、ハードディスクドライブインタフェース１０３０、ディスクドライブインタフェース１０４０、シリアルポートインタフェース１０５０、ビデオアダプタ１０６０、ネットワークインタフェース１０７０を有する。これらの各部は、バス１０８０によって接続される。
【００８１】
メモリ１０１０は、ＲＯＭ１０１１及びＲＡＭ１０１２を含む。ＲＯＭ１０１１は、例えば、ＢＩＯＳ（Basic Input Output System）等のブートプログラムを記憶する。ハードディスクドライブインタフェース１０３０は、ハードディスクドライブ１０９０に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１１００に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ１１００に挿入される。シリアルポートインタフェース１０５０は、例えばマウス１１１０、キーボード１１２０に接続される。ビデオアダプタ１０６０は、例えばディスプレイ１１３０に接続される。
【００８２】
ハードディスクドライブ１０９０は、例えば、ＯＳ（Operating System）１０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３、プログラムデータ１０９４を記憶する。すなわち、検出装置１０の各処理を規定するプログラムは、コン

ピュータ１０００により実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０９０に記憶される。例えば、検出装置１０及における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０９０に記憶される。なお、ハードディスクドライブ１０９０は、ＳＳＤ（Solid State Drive）により代替されてもよい。

【００８３】

また、上述した実施の形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０９０に記憶される。そして、ＣＰＵ１０２０が、メモリ１０１０やハードディスクドライブ１０９０に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてＲＡＭ１０１２に読み出して実行する。

【００８４】

なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０９０に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１１００等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

【００８５】

以上、本発明者によってなされた発明を適用した実施の形態について説明したが、本実施の形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施の形態に基づいて当業者等によりなされる他の実施の形態、実施例及び運用技術等は全て本発明の範疇に含まれる。

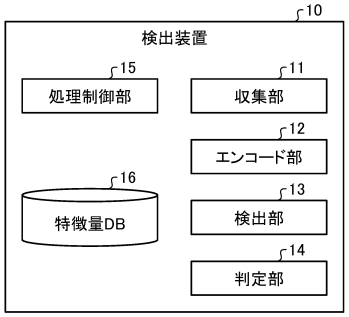
【符号の説明】

【００８６】

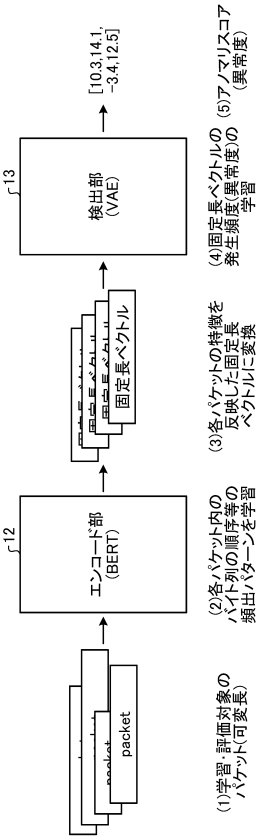
１０　検出装置
１１　収集部
１２　エンコード部
１３　検出部
１４　判定部
１５　処理制御部
１６　特徴量ＤＢ

【図面】

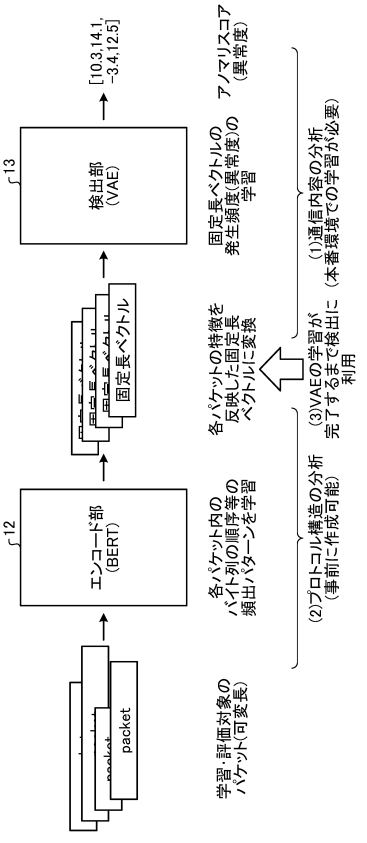
【図１】


10
検出装置
15
処理制御部
11
収集部
12
エンコード部
16
特徴量DB
13
検出部
14
判定部

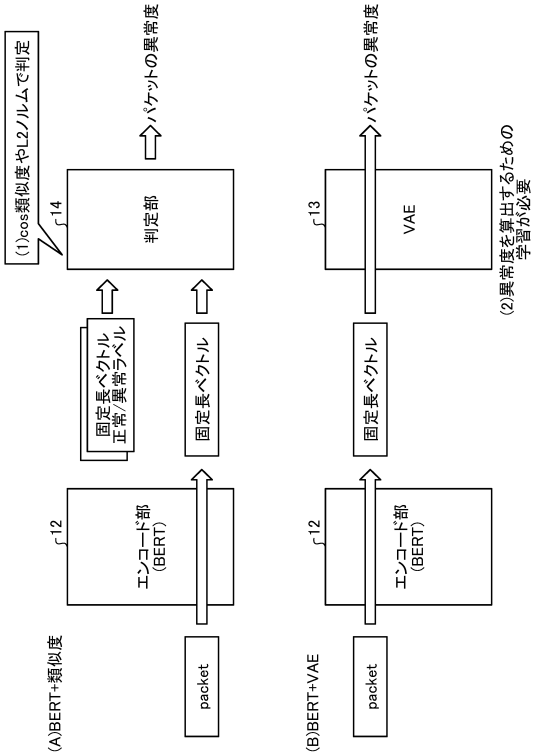
【図２】


[10.3,14.1,
-3.4,12.5]
13
検出部
(VAE)
(5)アノマリスコア
(異常度)
(4)固定長ベクトルの
発生頻度(異常度)の
学習
(3)各パケットの特徴を
反映した固定長
ベクトルに変換
固定長ベクトル
12
エンコード部
(BERT)
(2)各パケット内の
バイト列の順序等の
頻出パターンを学習
packet
(1)学習・評価対象の
パケット(可変長)

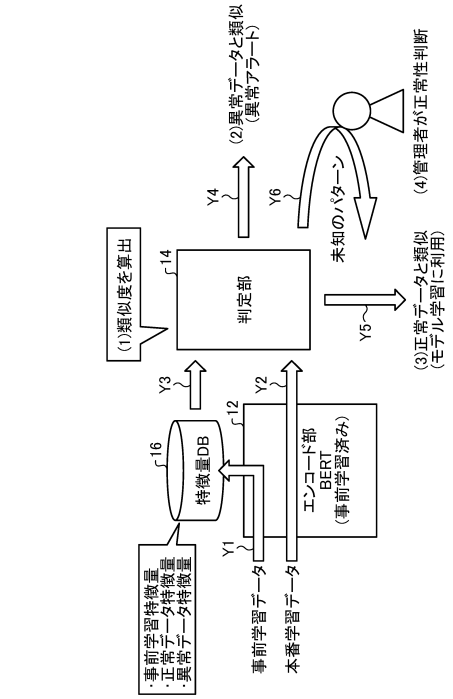
【図３】


[10.3,14.1,
-3.4,12.5]
13
検出部
(VAE)
アノマリスコア
(異常度)
固定長ベクトルの
発生頻度(異常度)の
学習
(1)通信内容の分析
(本番環境での学習が必要)
各パケットの特徴を
反映した固定長
ベクトルに変換
(3)VAEの学習が
完了するまで検出に
利用
固定長ベクトル
12
エンコード部
(BERT)
各パケット内の
バイト列の順序等の
頻出パターンを学習
(2)プロトコル構造の分析
(事前に作成可能)
packet
学習・評価対象の
パケット(可変長)

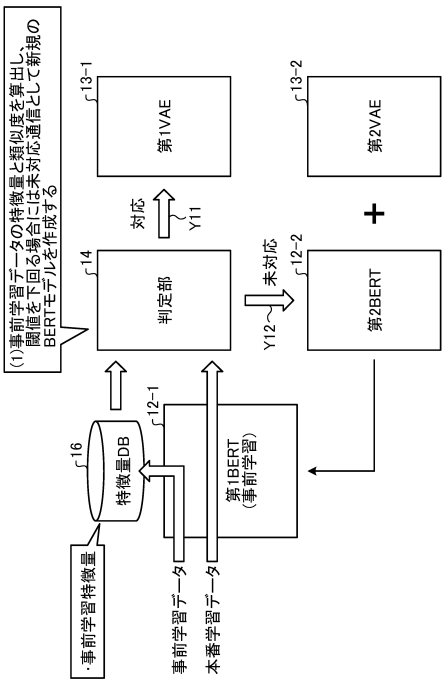
【図４】


(1)cos類似度やL2ノルムで判定
14
判定部
パケットの異常度
固定長ベクトル
正常/異常ラベル
固定長ベクトル
12
エンコード部
(BERT)
packet
(A)BERT+類似度
13
VAE
パケットの異常度
(2)異常度を算出するための
学習が必要
固定長ベクトル
12
エンコード部
(BERT)
packet
(B)BERT+VAE

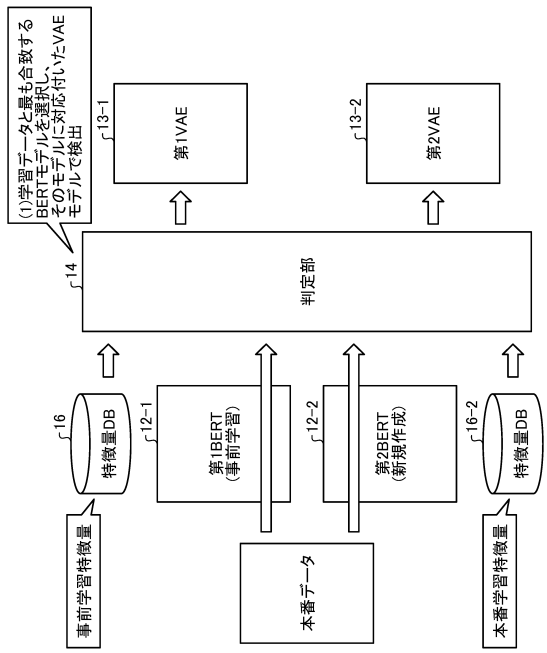
【図５】


(1)類似度を算出
判定部
14
Y4
(2)異常データと類似
(異常アラート)
Y6
未知のパターン
(4)管理者が正常性判断
Y5
(3)正常データと類似
(モデル学習に利用)
Y3
Y2
12
エンコード部
BERT
(事前学習済み)
16
特徴量DB
Y1
事前学習データ
本番学習データ
・事前学習特徴量
・正常データ特徴量
・異常データ特徴量

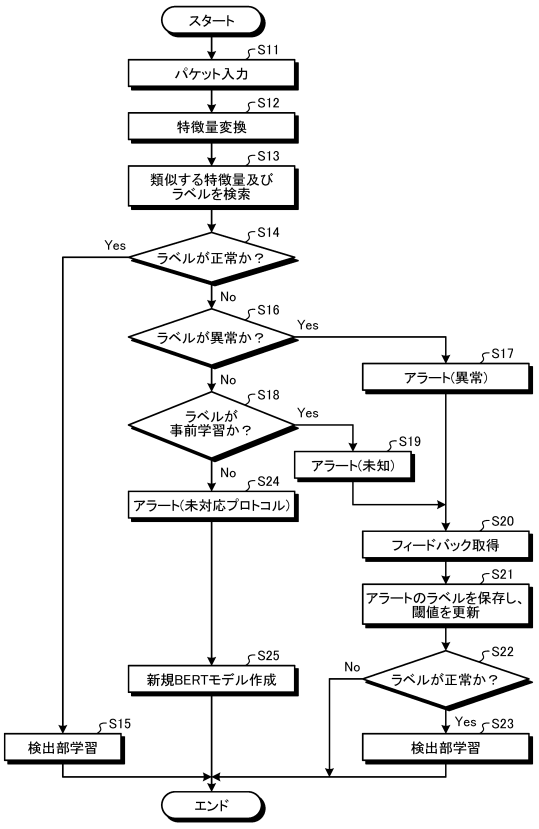
【図６】


(1)事前学習データの特徴量と類似度を算出し、
閾値を下回る場合には未対応通信として新規の
BERTモデルを作成する
13-1
第1VAE
13-2
第2VAE
対応
Y11
+
14
判定部
未対応
Y12
12-2
第2BERT
12-1
第1BERT
(事前学習)
16
特徴量DB
事前学習データ
本番学習データ
・事前学習特徴量

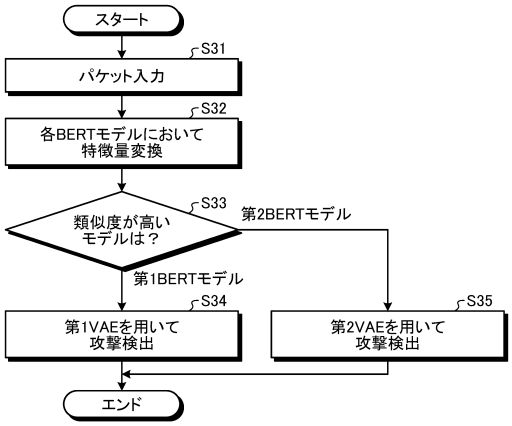
【図７】


(1)学習データと最も合致する
BERTモデルを選択し、
そのモデルに対応付いたVAE
モデルで検出
13-1
第1VAE
13-2
第2VAE
14
判定部
16
特徴量DB
12-1
第1BERT
(事前学習)
12-2
第2BERT
(新規作成)
16-2
特徴量DB
事前学習特徴量
本番データ
本番学習特徴量

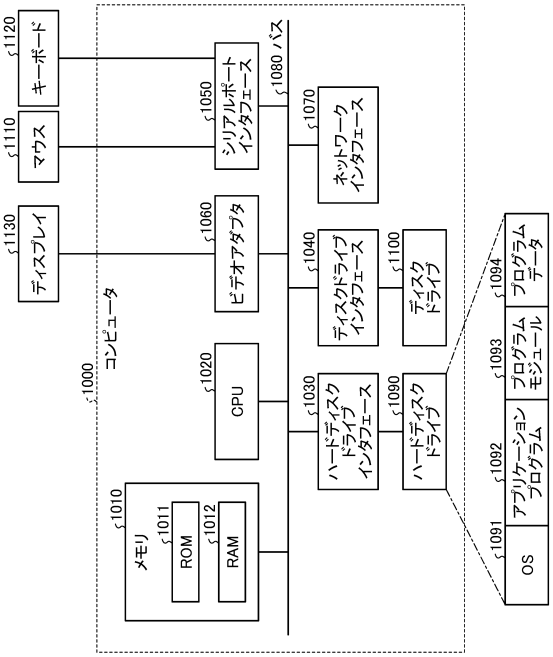
【図８】


スタート
S11
パケット入力
S12
特徴量変換
S13
類似する特徴量及び
ラベルを検索
S14
ラベルが正常か？
Yes
No
S16
ラベルが異常か？
Yes
No
S17
アラート(異常)
S18
ラベルが
事前学習か？
Yes
No
S19
アラート(未知)
S24
アラート(未対応プロトコル)
S20
フィードバック取得
S21
アラートのラベルを保存し、
閾値を更新
S22
ラベルが正常か？
No
Yes
S23
検出部学習
S25
新規BERTモデル作成
S15
検出部学習
エンド

【図９】


スタート
S31
パケット入力
S32
各BERTモデルにおいて
特徴量変換
S33
類似度が高い
モデルは？
第2BERTモデル
第1BERTモデル
S34
第1VAEを用いて
攻撃検出
S35
第2VAEを用いて
攻撃検出
エンド

【図１０】


1000 コンピュータ
1120 キーボード
1110 マウス
1130 ディスプレイ
1050 シリアルポート インタフェース
1080 バス
1070 ネットワーク インタフェース
1060 ビデオアダプタ
1040 ディスクドライブ インタフェース
1100 ディスク ドライブ
1020 CPU
1030 ハードディスク ドライブ インタフェース
1090 ハードディスク ドライブ
1010 メモリ
1011 ROM
1012 RAM
1094 プログラム データ
1093 プログラム モジュール
1092 アプリケーション プログラム
1091 OS

フロントページの続き

東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内
(72)発明者 永井 智大
東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内
(72)発明者 小山 高明
東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内

審査官 安井 雅史

(56)参考文献 中国特許出願公開第１１２４４６３９９（ＣＮ，Ａ）
米国特許第１０９９０７６７（ＵＳ，Ｂ１）
特表２０１７－５１９２８２（ＪＰ，Ａ）
中国特許出願公開第１１１１８１９３９（ＣＮ，Ａ）
GOODMAN, Eric L et al.，Packet2Vec: Utilizing Word2Vec for Feature Extraction in Packet Data，[オンライン]，2020年04月，[検索日 2024.08.06], インターネット 〈URL：https://arxiv.org/abs/2004.14477〉

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｈ０４Ｌ １２／００－１２／６６
Ｈ０４Ｌ ４１／００－６９／４０