

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7527294号
(P7527294)

(45)発行日 令和6年8月2日(2024.8.2)

(24)登録日 令和6年7月25日(2024.7.25)

(51)国際特許分類

F I

H O 4 L 41/06 (2022.01)

H O 4 L 41/06

請求項の数 15 (全26頁)

(21)出願番号	特願2021-535167(P2021-535167)	(73)特許権者	514136668
(86)(22)出願日	令和2年12月10日(2020.12.10)		パナソニック インテレクチュアル プロ
(86)国際出願番号	PCT/JP2020/046135		パティ コーポレーション オブ アメリカ
(87)国際公開番号	WO2021/153032		Panasonic Intellec
(87)国際公開日	令和3年8月5日(2021.8.5)		tual Property Corpo
審査請求日	令和5年10月2日(2023.10.2)		ration of America
(31)優先権主張番号	特願2020-14786(P2020-14786)		アメリカ合衆国 9 0 5 0 4 カリフォル
(32)優先日	令和2年1月31日(2020.1.31)		ニア州, トーランス, スイート 4 5 0
(33)優先権主張国・地域又は機関	日本国(JP)	(74)代理人	100109210
			弁理士 新居 広守
		(74)代理人	100137235
			弁理士 寺谷 英作
		(74)代理人	100131417
			弁理士 道坂 伸一

最終頁に続く

(54)【発明の名称】 異常検知方法、及び異常検知装置

(57)【特許請求の範囲】

【請求項1】

連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出し、

算出した前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出し、
抽出した前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する異常検知方法であって、

前記複数の特徴量の抽出では、それぞれが、前記複数の距離のうちの連続するL（Lは2以上の整数）個の距離からなる前記複数の特徴量を抽出し、

前記異常度情報の算出では、
抽出した前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、

N（NはL＋1以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該N個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する

異常検知方法。

【請求項2】

前記複数の距離の算出では、前記検知対象データとしてパケットのデータ部を利用して、前記複数の距離を算出する

請求項1に記載の異常検知方法。

【請求項 3】

前記複数の距離の算出では、前記複数の検知対象データ間のレーベンシュタイン距離を利用して、前記複数の距離を算出する

請求項 1 または請求項 2 に記載の異常検知方法。

【請求項 4】

前記複数の検知対象データ間のレーベンシュタイン距離は、前記検知対象データ列における P (P は 1 以上の整数) 個離れた検知対象データ間のそれぞれに対して算出され、前記データ部を構成するビット列の少なくとも一部を対象として、 Q (Q は、1 以上 16 以下の整数) ビットを単位として算出され、

前記複数の距離を算出する際に、前記検知対象データ列における P 個離れた検知対象データ間のそれぞれに対して距離を算出することで、前記複数の距離を算出する

請求項 2 に記載の異常検知方法。

【請求項 5】

前記複数の距離の算出では、前記複数の検知対象データのうち、同一種別の検知対象データ間に限定して、前記複数の距離を算出する

請求項 1 から請求項 4 のいずれか 1 項に記載の異常検知方法。

【請求項 6】

前記差異は、 L_p 距離である

請求項 1 から請求項 5 のいずれか 1 項に記載の異常検知方法。

【請求項 7】

前記補助情報の算出では、抽出した前記複数の特徴量毎に、予め記憶する複数のリファレンス特徴量のそれぞれとの L_p 距離を算出し、算出した L_p 距離のそれぞれを用いる K 近傍法を利用して、予め記憶する複数のリファレンス特徴量との K 近傍距離を前記補助情報として算出する

請求項 6 に記載の異常検知方法。

【請求項 8】

前記異常度情報の算出では、前記 1 以上の情報算出対象ウィンドウ毎に、前記 N 個の検知対象データから算出された複数の補助情報の全ての和を算出することで、前記異常度情報を算出する

請求項 1 から請求項 7 のいずれか 1 項に記載の異常検知方法。

【請求項 9】

さらに、

連続する複数のリファレンスデータからなる 1 以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出し、

算出した前記複数のリファレンス距離のそれぞれを用いて、前記 1 以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出し、

前記検知対象データ列の異常の度合いに係る異常度情報の算出では、さらに、抽出した前記リファレンス特徴量のそれぞれをも用いて、前記異常度情報を算出する

請求項 1 から請求項 8 のいずれか 1 項に記載の異常検知方法。

【請求項 10】

前記複数のリファレンス距離の算出では、前記複数のリファレンスデータ間のレーベンシュタイン距離を利用して、前記複数のリファレンス距離を算出する

請求項 9 に記載の異常検知方法。

【請求項 11】

前記複数のリファレンスデータ間のレーベンシュタイン距離は、前記 1 以上のリファレンスデータ列のそれぞれにおける P (P は 1 以上の整数) 個離れたリファレンスデータ間のそれぞれに対して算出され、前記リファレンスデータとしてのパケットのデータ部を構成するビット列の少なくとも一部を対象として、 Q (Q は、1 以上 16 以下の整数) ビットを単位として算出され、

前記複数のリファレンス距離の算出では、前記 1 以上のリファレンスデータ列のそれぞ

10

20

30

40

50

れにおけるP個離れたリファレンスデータ間のそれぞれに対してリファレンス距離を算出することで、前記複数のリファレンス距離をそれぞれ算出する

請求項10に記載の異常検知方法。

【請求項12】

さらに、

連続する複数のリファレンスデータからなる1以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出し、

算出した前記複数のリファレンス距離のそれぞれを用いて、前記1以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出し、

抽出した前記複数のリファレンス特徴量のそれぞれを、前記予め記憶する複数のリファレンス特徴量のそれぞれとして記憶する

請求項7に記載の異常検知方法。

【請求項13】

前記複数のリファレンス距離の算出では、前記複数のリファレンスデータのうち、同一種別のリファレンスデータ間に限定して、前記リファレンス距離のそれぞれを算出する

請求項9から請求項12のいずれか1項に記載の異常検知方法。

【請求項14】

連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出する検知対象データ間距離算出部と、

前記検知対象データ間距離算出部により算出された前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出する特徴量抽出部と、

前記特徴量抽出部により抽出された前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する情報算出部と、を備える異常検知装置であって、

前記特徴量抽出部は、それぞれが、前記複数の距離のうちの連続するL（Lは2以上の整数）個の距離からなる前記複数の特徴量を抽出し、

前記情報算出部は、

前記特徴量抽出部により抽出された前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、

N（NはL+1以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該N個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する

異常検知装置。

【請求項15】

さらに、

連続する複数のリファレンスデータからなる1以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出するリファレンスデータ間距離算出部と、

前記リファレンスデータ間距離算出部により算出された前記複数のリファレンス距離のそれぞれを用いて、前記1以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出するリファレンス特徴量抽出部と、を備え、

前記情報算出部は、さらに、前記リファレンス特徴量抽出部により抽出された前記リファレンス特徴量のそれぞれをも用いて、前記異常度情報を算出する

請求項14に記載の異常検知装置。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、パケット中のペイロード等のデータ列の異常を検知する異常検知方法、及び異常検知装置に関する。

【背景技術】

【0002】

従来、ネットワークシステム等で利用されるデータを対象として行う情報処理技術が知られている（例えば、非特許文献1、非特許文献2参照）。

【先行技術文献】

【非特許文献】

【0003】

【文献】Ye, N. (2000, June). A markov chain model of temporal behavior for anomaly detection. In Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop (Vol. 166, p. 169). West Point, NY.

10

【文献】Otey, M. E., Ghoting, A., & Parthasarathy, S. (2006). Fast distributed outlier detection in mixed-attribute data sets. Data mining and knowledge discovery, 12(2-3), 203-228

【発明の概要】

【発明が解決しようとする課題】

【0004】

20

ネットワークシステム等で利用されるパケット列等のデータ列の異常を検知することが望まれる。

【0005】

そこで、本開示は、パケット列等のデータ列の異常を検知する異常検知方法、及び異常検知装置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】

【0006】

本開示の一態様に係る異常検知方法は、連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出し、算出した前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出し、抽出した前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する異常検知方法であって、前記複数の特徴量の抽出では、それぞれが、前記複数の距離のうちの連続する L （ L は2以上の整数）個の距離からなる前記複数の特徴量を抽出し、前記異常度情報の算出では、抽出した前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、 N （ N は $L+1$ 以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該 N 個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する。

30

【0007】

本開示の一態様に係る異常検知装置は、連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出する検知対象データ間距離算出部と、前記検知対象データ間距離算出部により算出された前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出する特徴量抽出部と、前記特徴量抽出部により抽出された前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する情報算出部と、を備える異常検知装置であって、前記特徴量抽出部は、それぞれが、前記複数の距離のうちの連続する L （ L は2以上の整数）個の距離からなる前記複数の特徴量を抽出し、前記情報算出部は、前記特徴量抽出部により抽出された前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、 N （ N は $L+1$ 以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該 N 個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する。

40

50

【発明の効果】

【0008】

本開示の一態様に係る異常検知方法、及び異常検知装置によると、パケット列等のデータ列の異常を検知することができる。

【図面の簡単な説明】

【0009】

【図1】図1は、実施の形態に係る異常検知装置の構成を示すブロック図である。

【図2】図2は、プロファイル情報の一例を示す模式図である。

【図3】図3は、Modbus/TCPプロトコルのパケットのデータ構造を示す模式図である。

【図4】図4は、BACnet/IPプロトコルのパケットのデータ構造を示す模式図である。

【図5】図5は、リファレンスパケットデータ列の一例を示す模式図である。

【図6】図6は、実施の形態に係るリファレンスパケットデータ間距離算出部がパケットデータを1バイトずつ切り出す様子の一例を示す模式図である。

【図7】図7は、実施の形態に係るリファレンスパケットデータ間距離算出部が2つの文字列間のレーベンシュタイン距離を算出する様子の一例を示す模式図である。

【図8】図8は、実施の形態に係るリファレンスパケットデータ間距離算出部が2つのバイト列間のレーベンシュタイン距離を算出する様子の一例を示す模式図である。

【図9】図9は、実施の形態に係るリファレンスパケットデータ間距離算出部がリファレンス距離を算出する様子の一例を示す模式図である。

【図10】図10は、実施の形態に係るリファレンス特徴量抽出部がリファレンス特徴量を抽出する様子の一例を示す模式図である。

【図11】図11は、実施の形態に係る検知対象パケットデータ間距離算出部が距離を算出する様子の一例を示す模式図である。

【図12】図12は、実施の形態に係る特徴量抽出部が検知対象パケットデータ列から、特徴量を抽出する様子の一例を示す模式図である。

【図13】図13は、実施の形態に係る補助情報算出部が検知対象パケット列の補助情報を算出する様子の一例を示す模式図である。

【図14】図14は、リファレンス特徴量抽出処理のフローチャートである。

【図15】図15は、異常検知処理のフローチャートである。

【図16】図16は、侵入検知システムが出力する検知ログの一例を示す図である。

【発明を実施するための形態】

【0010】

(本開示の一態様を得るに至った経緯)

ICS (Industrial Control System) におけるサイバー攻撃の事例が増加傾向にある。ICSでは、機器を制御する正常なフォーマットのコマンド列が深刻な攻撃となり得るため、悪意のある攻撃者によって生成された異常なパケット列であっても、パケット単体で見ると、正常なものとほとんど区別できない。

【0011】

従来、パケット列の異常を検知する方法として、ルールベース方式とアノマリ方式とが知られている。しかしながら、ルールベース方式には、ルールで記載されていないパケット列の異常を検知することが困難であるという問題があり、アノマリ方式には、極端なデータ量やパケット数の増加や、極端なバイト列の乱れがなければ、パケット列の異常を検知することが困難であるという問題がある。

【0012】

そこで、発明者は、これら問題を解決すべく、鋭意検討、実験を重ねた。そして、発明者は、機器を制御するコマンドを含む正常なパケット列が、一定のコンテキストを有していることに着目し、一定のコンテキストから逸脱するパケット列を、異常なパケット列として検知できることを見出した。その結果、発明者は、下記異常検知方法、及び異常検知

10

20

30

40

50

装置に想到した。

【0013】

本開示の一態様に係る異常検知方法は、連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出し、算出した前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出し、抽出した前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する異常検知方法であって、前記複数の特徴量の抽出では、それぞれが、前記複数の距離のうちの連続する L （ L は2以上の整数）個の距離からなる前記複数の特徴量を抽出し、前記異常度情報の算出では、抽出した前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、 N （ N は $L+1$ 以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該 N 個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する。

10

【0014】

上記異常検知方法によると、検知対象データ列のコンテキストの特徴を特徴量として抽出することができる。そして、抽出した特徴量を用いて、検知対象データ列の異常に係る情報を算出することができる。このように、上記異常検知方法によると、パケット列等のデータ列の異常を検知することができる。

【0015】

また、前記複数の距離の算出では、前記検知対象データとしてパケットのデータ部を利用して、前記複数の距離を算出するとしてもよい。

20

【0016】

また、前記複数の距離の算出では、前記複数の検知対象データ間のレーベンシュタイン距離を利用して、前記複数の距離を算出するとしてもよい。

【0017】

また、前記複数の検知対象データ間のレーベンシュタイン距離は、前記検知対象データ列における P （ P は1以上の整数）個離れた検知対象データ間のそれぞれに対して算出され、前記データ部を構成するビット列の少なくとも一部を対象として、 Q （ Q は、1以上16以下の整数）ビットを単位として算出され、前記複数の距離を算出する際に、前記検知対象データ列における P 個離れた検知対象データ間のそれぞれに対して距離を算出することで、前記複数の距離を算出するとしてもよい。

30

【0018】

また、前記複数の距離の算出では、前記複数の検知対象データのうち、同一種別の検知対象データ間に限定して、前記複数の距離を算出するとしてもよい。

【0019】

また、前記差異は、 L_p 距離であるとしてもよい。

【0020】

また、前記補助情報の算出では、抽出した前記複数の特徴量毎に、予め記憶する複数のリファレンス特徴量のそれぞれとの L_p 距離を算出し、算出した L_p 距離のそれぞれを用いる K 近傍法を利用して、予め記憶する複数のリファレンス特徴量との K 近傍距離を前記補助情報として算出するとしてもよい。

40

【0021】

また、前記異常度情報の算出では、前記1以上の情報算出対象ウィンドウ毎に、前記 N 個の検知対象データから算出された複数の補助情報の全ての和を算出することで、前記異常度情報を算出するとしてもよい。

【0022】

また、さらに、連続する複数のリファレンスデータからなる1以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出し、算出した前記複数のリファレンス距離のそれぞれを用いて、前記1以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出し、前記検知対象データ列の異常の度合いに係る異常度情報の算出では、さらに、抽出した前記リファレンス特

50

微量のそれぞれをも用いて、前記異常度情報を算出するとしてもよい。

【0023】

また、前記複数のリファレンス距離の算出では、前記複数のリファレンスデータ間のレーベンシュタイン距離を利用して、前記複数のリファレンス距離を算出するとしてもよい。

【0024】

また、前記複数のリファレンスデータ間のレーベンシュタイン距離は、前記1以上のリファレンスデータ列のそれぞれにおけるP（Pは1以上の整数）個離れたリファレンスデータ間のそれぞれに対して算出され、前記リファレンスデータとしてのパケットのデータ部を構成するビット列の少なくとも一部を対象として、Q（Qは、1以上16以下の整数）ビットを単位として算出され、前記複数のリファレンス距離の算出では、前記1以上のリファレンスデータ列のそれぞれにおけるP個離れたリファレンスデータ間のそれぞれに対してリファレンス距離を算出することで、前記複数のリファレンス距離をそれぞれ算出するとしてもよい。

10

【0025】

また、さらに、連続する複数のリファレンスデータからなる1以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出し、算出した前記複数のリファレンス距離のそれぞれを用いて、前記1以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出し、抽出した前記複数のリファレンス特徴量のそれぞれを、前記予め記憶する複数のリファレンス特徴量のそれぞれとして記憶するとしてもよい。

20

【0026】

また、前記複数のリファレンス距離の算出では、前記複数のリファレンスデータのうち、同一種別のリファレンスデータ間に限定して、前記リファレンス距離のそれぞれを算出するとしてもよい。

【0027】

本開示の一態様に係る異常検知装置は、連続する複数の検知対象データからなる検知対象データ列に対して、当該複数の検知対象データ間の複数の距離を算出する検知対象データ間距離算出部と、前記検知対象データ間距離算出部により算出された前記複数の距離を用いて、前記検知対象データ列の複数の特徴量を抽出する特徴量抽出部と、前記特徴量抽出部により抽出された前記複数の特徴量を用いて、前記検知対象データ列の異常の度合いに係る異常度情報を算出する情報算出部と、を備える異常検知装置であって、前記特徴量抽出部は、それぞれが、前記複数の距離のうちの連続するL（Lは2以上の整数）個の距離からなる前記複数の特徴量を抽出し、前記情報算出部は、前記特徴量抽出部により抽出された前記複数の特徴量毎に、当該特徴量に係る差異を用いて、前記異常度情報を算出するための補助情報を算出し、N（NはL+1以上の整数）個の前記検知対象データからなる1以上の情報算出対象ウィンドウ毎に、当該N個の検知対象データから算出された複数の補助情報を全て用いて、前記異常度情報を算出する。

30

【0028】

上記異常検知装置によると、検知対象データ列のコンテキストの特徴を特徴量として抽出することができる。そして、抽出した特徴量を用いて、検知対象データ列の異常に係る情報を算出することができる。このように、上記異常検知装置によると、パケット列等のデータ列の異常を検知することができる。

40

【0029】

また、さらに、連続する複数のリファレンスデータからなる1以上のリファレンスデータ列のそれぞれに対して、当該複数のリファレンスデータの複数のリファレンス距離をそれぞれ算出するリファレンスデータ間距離算出部と、前記リファレンスデータ間距離算出部により算出された前記複数のリファレンス距離のそれぞれを用いて、前記1以上のリファレンスデータ列の複数のリファレンス特徴量をそれぞれ抽出するリファレンス特徴量抽出部と、を備え、前記情報算出部は、さらに、前記リファレンス特徴量抽出部により抽出された前記リファレンス特徴量のそれぞれをも用いて、前記異常度情報を算出するとして

50

もよい。

【0030】

以下、本開示の一態様に係る異常検知方法、及び異常検知装置の具体例について、図面を参照しながら説明する。ここで示す実施の形態は、いずれも本開示の一具体例を示すものである。従って、以下の実施の形態で示される数値、形状、構成要素、構成要素の配置及び接続形態、並びに、ステップ（工程）及びステップの順序等は、一例であって本開示を限定するものではない。以下の実施の形態における構成要素のうち、独立請求項に記載されていない構成要素については、任意に付加可能な構成要素である。また、各図は、模式図であり、必ずしも厳密に図示されたものではない。

【0031】

（実施の形態）

以下、実施の形態に係る異常検知装置について説明する。この異常検知装置は、検知対象パケット列の異常を検知する。

【0032】

[1. 構成]

図1は、実施の形態に係る異常検知装置1の構成を示すブロック図である。

【0033】

図1に示されるように、異常検知装置1は、第1プロファイル決定部11と、第1抽出部12と、リファレンスパケットデータ列群記憶部13と、リファレンスパケットデータ間距離算出部14と、リファレンス特徴量抽出部15と、リファレンス特徴量記憶部16と、第2プロファイル決定部21と、第2抽出部22と、検知対象パケットデータ列記憶部23と、検知対象パケットデータ間距離算出部24と、特徴量抽出部25と、特徴量記憶部26と、補助情報算出部27と、情報算出部28と、判定部29とを含んで構成される。

【0034】

異常検知装置1は、例えば、メモリと、メモリに記憶されたプログラムを実行するプロセッサとを含むコンピュータ装置によって実現される。この場合、異常検知装置1によって実現される各種機能は、異常検知装置1を構成するメモリに記憶されたプログラムを、異常検知装置1を構成するプロセッサが実行することにより実現される。

【0035】

第1プロファイル決定部11は、正常なパケット列からなるパケット列群、すなわち、異常なパケット列を含まないパケット列群であるリファレンスパケット列群10を取得する。そして、取得したリファレンスパケット列群10を構成する各リファレンスパケット列に含まれる各パケットに対して、その属性情報に基づいて、該当するプロファイルを決定する。属性情報は、送信元IP、宛先IP、送信元ポート、宛先ポート、プロトコル等であり、それらの組み合わせも含む。第1プロファイル決定部11は、例えば、プロファイル情報を記憶し、記憶するプロファイル情報に基づいて該当するプロファイルを決

【0036】

図2は、第1プロファイル決定部11が記憶するプロファイル情報の一例を示す模式図である。

【0037】

第1プロファイル決定部11は、例えば、図2に示されるプロファイル情報を記憶し、各パケットに対して、対象コマンドが一致する行のプロファイルIDで識別されるプロファイルを決

【0038】

図3は、Modbus / TCPプロトコルのパケットのデータ構造を示す模式図である。

【0039】

第1プロファイル決定部11は、例えば、パケットがModbus / TCPプロトコルである場合には、図3に示されるDataフィールドに格納されるビット列が、対象コマ

10

20

30

40

50

ンドに対応するビット列と一致するパケットを、対象コマンドが一致するパケットであると判定する。

【0040】

図4は、BACnet/IPプロトコルのパケットのデータ構造を示す模式図である。

【0041】

第1プロファイル決定部11は、例えば、パケットがBACnet/IPプロトコルである場合には、図4に示されるAPDU Typeフィールドに格納されるビット列、及びService Choiceフィールドに格納されるビット列が、対象コマンドに対応するビット列と一致するパケットを、対象コマンドが一致するパケットであると判定する。

【0042】

第1プロファイル決定部11は、例えば、プロファイルの決定対象とするパケットが、記憶するプロファイル情報に該当しない場合には、Deep Packet Inspection機能を備えるアプリケーションを実行することで、そのパケットのプロトコルを特定し、特定したプロトコルに基づいて、そのパケットに対してプロファイルを決定するとしてもよい。

【0043】

再び、図1に戻って、異常検知装置1の説明を続ける。

【0044】

第1抽出部12は、第1プロファイル決定部11によりプロファイルを決定された各パケットに対して、各パケットのペイロードフィールドに格納されるビット列をパケットデータとして抽出する。

【0045】

第1抽出部12は、例えば、抽出対象とするパケットが、Modbus/TCPプロトコルである場合には、図3に示されるTCP Payloadフィールドに格納されるビット列をパケットデータとして抽出する。

【0046】

第1抽出部12は、例えば、抽出対象とするパケットが、BACnet/IPプロトコルである場合には、図4に示されるUDP Payloadに格納されるビット列をパケットデータとして抽出する。

【0047】

第1抽出部12は、各パケットからパケットデータを抽出すると、第1プロファイル決定部11により取得されたリファレンスパケット列毎に、同一プロファイルのパケットデータからなるリファレンスパケットデータ列を出力する。

【0048】

図5は、第1抽出部12により出力されるリファレンスパケットデータ列の一例を示す模式図である。

【0049】

第1抽出部12は、例えば、図5に示される、「ID=1で識別される機器の電源をオンする」旨のwriteコマンドに対応するパケットデータと、「ID=2で識別される機器の電源をオンする」旨のwriteコマンドに対応するパケットデータと、「ID=3で識別される機器の電源をオンする」旨のwriteコマンドに対応するパケットデータと、「ID=4で識別される機器の電源をオンする」旨のwriteコマンドに対応するパケットデータと、「ID=1で識別される機器の温度設定を27.0℃に変更する」旨のwriteコマンドに対応するパケットデータと、「ID=1で識別される機器の風量設定を強に変更する」旨のwriteコマンドに対応するパケットデータと、「ID=2で識別される機器の温度設定を27.0℃に変更する」旨のwriteコマンドに対応するパケットデータと、「ID=2で識別される機器の風量設定を強に変更する」旨のwriteコマンドに対応するパケットデータとが順に並んで構成されるリファレンスパケットデータ列を出力する。

10

20

30

40

50

【0050】

再び、図1に戻って、異常検知装置1の説明を続ける。

【0051】

リファレンス packets データ列群記憶部13は、第1抽出部12から出力された1以上のリファレンス packets データ列からなるリファレンス packets データ列群を記憶する。

【0052】

リファレンス packets データ間距離算出部14は、リファレンス packets データ列群記憶部13に記憶されるリファレンス packets データ列のそれぞれに対して、リファレンス packets データ列を構成する複数のリファレンス packets データ間の複数のリファレンス距離を算出する。

10

【0053】

リファレンス packets データ間距離算出部14は、 packets データを、1バイトずつ切り出したバイト列として取り扱い、リファレンス packets データ間のリファレンス距離の算出を、リファレンス packets データ間におけるバイト列の距離を算出することで行う。

【0054】

図6は、リファレンス packets データ間距離算出部14が、 packets データを1バイトずつ切り出す様子の一例を示す模式図である。

【0055】

なお、ここでは、リファレンス packets データ間距離算出部14は、 packets データを、1バイトずつ切り出すとして説明するが、切り出す単位としては、必ずしも1バイトの例に限定される必要はない。切り出す単位は、例えば、1ビット以上、16ビット以下の任意の長さのビット列であってもよいし、他の任意の長さのビット列であってもよい。また、リファレンス packets データ間距離算出部14は、 packets データを、連続したビット単位で切り出す例に限定される必要はない。例えば、リファレンス packets データ間距離算出部14は、xビット切り出してyビットスキップするという処理を繰り返すことでビット列を切り出すとしてもよい。

20

【0056】

再び図1に戻って、異常検知装置1の説明を続ける。

【0057】

リファレンス packets データ間距離算出部14は、 packets データ間のレーベンシュタイン距離を利用してリファレンス距離を算出する。

30

【0058】

レーベンシュタイン距離は、2つの文字列又はバイト列間に定めることができる距離である。レーベンシュタイン距離は、1文字又は1バイトの挿入、削除、置換によって、一方の文字列又はバイト列を他方の文字列又はバイト列に変形するのに必要な最小回数として定義される。

【0059】

図7は、リファレンス packets データ間距離算出部14が、2つの文字列間（ここでは、一例として、“ELEPHANT”と“RELEVANT”との文字列間）のレーベンシュタイン距離を算出する様子の一例を示す模式図である。

40

【0060】

図7に示されるように、“ELEPHANT”を“RELEVANT”に変形するのに必要な、挿入、削除、置換の最小回数は3である。このため、リファレンス packets データ間距離算出部14は、“ELEPHANT”と“RELEVANT”とのレーベンシュタイン距離を「3」と算出する。

【0061】

図8は、リファレンス packets データ間距離算出部14が、2つのバイト列間のレーベンシュタイン距離を算出する様子の一例を示す模式図である。

【0062】

図8に示されるように、一方のバイト列を他方のバイト列に変形するのに必要な、挿入

50

、削除、置換の最小回数は3である。このため、リファレンス packets データ間距離算出部 14 は、図 8 に図示されるバイト列間のレーベンシュタイン距離を「3」と算出する。

【0063】

リファレンス packets データ間距離算出部 14 は、リファレンス packets データ列を構成する複数のリファレンス packets データ間の複数のリファレンス距離を、 P (P は 1 以上の整数) 個離れたリファレンス packets 間のそれぞれに対して算出する。

【0064】

図 9 は、 P が 1 である場合において、リファレンス packets データ間距離算出部 14 が、リファレンス packets データ列から複数のリファレンス距離を算出する様子の一例を示す模式図である。ここでは、 P は 1 であるとして説明するが、必ずしも P が 1 である場合

10

【0065】

図 9 に示されるように、リファレンス packets データ間距離算出部 14 は、対象とするリファレンス packets データ列が 9 個の連続するリファレンス packets データにより構成される場合には、対象とするリファレンス packets データ列から、1 離れたリファレンス packets データ間、すなわち、互いに隣接するリファレンス packets データ列間の 8 個のリファレンス距離を算出する。

【0066】

なお、リファレンス packets データ間距離算出部 14 は、packets データ間のレーベンシュタイン距離として、packets データを構成するビット列の少なくとも一部を対象としてレーベンシュタイン距離を算出すればよく、必ずしも、packets データを構成するビット列の全てを対象とする場合に限定されない。

20

【0067】

再び図 1 に戻って、異常検知装置 1 の説明を続ける。

【0068】

リファレンス特徴量抽出部 15 は、リファレンス packets データ列群記憶部 13 に記憶されるリファレンス packets データ列のそれぞれに対して、リファレンス packets データ間距離算出部 14 により算出された複数のリファレンス距離のそれぞれを用いて、リファレンス特徴量をそれぞれ抽出する。より具体的には、リファレンス特徴量抽出部 15 は、リファレンス packets データ列群記憶部 13 に記憶されるリファレンス packets データ列のそれぞれに対して、連続する $L + 1$ (L は補助情報算出対象ウィンドウの長さであって、2 以上の整数) 個のリファレンス packets データからなる 1 以上の補助情報算出対象ウィンドウ毎に、その補助情報算出対象ウィンドウに属するリファレンス packets データについて算出された L 個のリファレンス距離からなるリファレンス距離列を算出し、算出したリファレンス距離列のそれぞれを、リファレンス特徴量のそれぞれとして抽出する。

30

【0069】

図 10 は、リファレンス特徴量抽出部 15 が、1 のリファレンス packets データ列から、リファレンス特徴量を抽出する様子の一例を示す模式図である。図 10 は、 L が 2 であり、リファレンス packets データ間距離算出部 14 が、1 のリファレンス packets データ列に対して、図 9 に示される複数のリファレンス距離を算出した場合の模式図となっている。

40

【0070】

図 10 に示されるように、第 1 の補助情報算出対象ウィンドウに属するリファレンス packets データについて算出されたリファレンス距離は、「1」、「1」である。このため、リファレンス特徴量抽出部 15 は、第 1 の補助情報算出対象ウィンドウにおける特徴量を $(1, 1)$ という 2 次元のベクトル A_1 として抽出する。また、第 2 の補助情報算出対象ウィンドウに属するリファレンス packets データについて算出されたリファレンス距離は、「1」、「1」である。このため、リファレンス特徴量抽出部 15 は、第 2 の補助情報算出対象ウィンドウにおける特徴量を $(1, 1)$ という 2 次元のベクトル B_1 として抽出する。また、第 3 の補助情報算出対象ウィンドウに属するリファレンス packets データ

50

について算出されたリファレンス距離は、「1」、「3」である。このため、リファレンス特徴量抽出部15は、第3の補助情報算出対象ウィンドウにおける特徴量を(1, 3)という2次元のベクトルC1として抽出する。また、第4の補助情報算出対象ウィンドウに属するリファレンスパケットデータについて算出されたリファレンス距離は、「3」、「7」である。このため、リファレンス特徴量抽出部15は、第4の補助情報算出対象ウィンドウにおける特徴量を(3, 7)という2次元のベクトルD1として抽出する。そして、リファレンス特徴量抽出部15は、算出したベクトルA1とベクトルB1とベクトルC1とベクトルD1とを、リファレンス特徴量として抽出する。

【0071】

ここで、ベクトルA1とベクトルB1とは、同一であるため、どちらか一方、例えばベクトルA1のみをリファレンス特徴量として抽出するとしてもよい。

【0072】

同様にして、リファレンス特徴量抽出部15は、第5の補助情報算出対象ウィンドウにおけるベクトルE1(7, 2)、第6の補助情報算出対象ウィンドウにおけるベクトルF1(2, 1)、第7のウィンドウにおけるベクトルG1(1, 7)をリファレンス特徴量として抽出する。

【0073】

再び図1に戻って、異常検知装置1の説明を続ける。

【0074】

リファレンス特徴量記憶部16は、リファレンス特徴量抽出部15によって抽出されたリファレンス特徴量を記憶する。

【0075】

第2プロファイル決定部21は、異常の検知対象となる検知対象パケット列20を取得する。そして、取得した検知対象パケット列20に含まれる各パケットに対して、その属性情報に基づいて、該当するプロファイルを決定する。属性情報は、送信元IP、宛先IP、送信元ポート、宛先ポート、プロトコル等であり、それらの組み合わせも含む。第2プロファイル決定部21は、第1プロファイル決定部11が行うプロファイルの決定と同様のアルゴリズムでプロファイルの決定を行う。

【0076】

第2抽出部22は、第2プロファイル決定部21によりプロファイルを決定された複数のパケットに対して、各パケットのペイロードフィールドに格納されるビット列をパケットデータとして抽出する。第2抽出部22は、第1抽出部12が行うパケットデータの抽出と同様のアルゴリズムでパケットデータの抽出を行う。

【0077】

第2抽出部22は、パケットデータを抽出すると、同一プロファイルのパケットデータからなる検知対象パケットデータ列を出力する。

【0078】

検知対象パケットデータ列記憶部23は、第2抽出部22から出力された検知対象パケットデータ列を記憶する。

【0079】

検知対象パケットデータ間距離算出部24は、検知対象パケットデータ列記憶部23に記憶される検知対象パケットデータ列に対して、検知対象パケットデータ列を構成する複数の検知対象パケットデータ間の複数の距離を算出する。

【0080】

検知対象パケットデータ間距離算出部24は、リファレンスパケットデータ間距離算出部14が行うリファレンス距離の算出に対して、リファレンスパケットを検知対象パケットに読み替え、リファレンスパケットデータを検知対象パケットデータに読み替え、リファレンスパケットデータ列を検知対象パケットデータ列に読み替え、リファレンス距離を距離に読み替えたアルゴリズムと同様のアルゴリズムで、距離の算出を行う。

【0081】

10

20

30

40

50

図 1 1 は、P が 1 である場合において、検知対象パケットデータ間距離算出部 2 4 が、検知対象パケットデータ列から複数の距離を算出する様子の一例を示す模式図である。ここでは、P は 1 であるとして説明するが、必ずしも P が 1 である場合に限定されない。

【0082】

図 1 1 に示されるように、検知対象パケットデータ間距離算出部 2 4 は、対象とする検知対象パケットデータ列が 9 個の連続する検知対象パケットデータにより構成される場合には、対象とする検知対象パケットデータ列から、1 離れた検知対象パケットデータ間、すなわち、互いに隣接する検知対象パケットデータ列間の 8 個の距離を算出する。

【0083】

なお、検知対象パケットデータ間距離算出部 2 4 は、パケットデータ間のレーベンシュタイン距離として、パケットデータを構成するビット列の少なくとも一部を対象としてレーベンシュタイン距離を算出すればよく、必ずしも、パケットデータを構成するビット列の全てを対象とする場合に限定されない。

【0084】

再び図 1 に戻って、異常検知装置 1 の説明を続ける。

【0085】

特徴量抽出部 2 5 は、検知対象パケットデータ列記憶部 2 3 に記憶される検知対象パケットデータ列に対して、検知対象パケットデータ間距離算出部 2 4 により算出された複数の距離を用いて、特徴量を抽出する。より具体的には、特徴量抽出部 2 5 は、検知対象パケットデータ列記憶部 2 3 に記憶される検知対象パケットデータ列に対して、連続する L + 1 個の検知対象パケットデータからなる 1 以上の補助情報算出対象ウィンドウ毎に、その補助情報算出対象ウィンドウに属する検知対象パケットデータについて算出された L 個の距離からなる距離列を算出し、算出した距離列のそれぞれを、特徴量として抽出する。

【0086】

図 1 2 は、特徴量抽出部 2 5 が、検知対象パケットデータ列から、特徴量を抽出する様子の一例を示す模式図である。図 1 2 は、L が 2 であり、検知対象パケットデータ間距離算出部 2 4 が、検知対象パケットデータ列に対して、図 1 1 に示される複数の距離を算出した場合の模式図となっている。

【0087】

図 1 2 に示されるように、第 1 の補助情報算出対象ウィンドウに属する検知対象パケットデータについて算出された距離は、「1」、「5」である。このため、特徴量抽出部 2 5 は、第 1 の補助情報算出対象ウィンドウにおける特徴量を (1, 5) という 2 次元のベクトル A 2 として抽出する。また、第 2 のウィンドウに属する検知対象パケットデータについて算出された距離は、「5」、「1」である。このため、特徴量抽出部 2 5 は、第 2 の補助情報算出対象ウィンドウにおける特徴量を (5, 1) という 2 次元のベクトル B 2 として抽出する。また、第 3 の補助情報算出対象ウィンドウに属する検知対象パケットデータについて算出された距離は、「1」、「1」である。このため、特徴量抽出部 2 5 は、第 3 の補助情報算出対象ウィンドウにおける特徴量を (1, 1) という 2 次元のベクトル C 2 として抽出する。また、第 4 の補助情報算出対象ウィンドウに属する検知対象パケットデータについて算出された距離は、「1」、「3」である。このため、特徴量抽出部 2 5 は、第 4 の補助情報算出対象ウィンドウにおける特徴量を (1, 3) という 2 次元のベクトル D 2 として抽出する。そして、特徴量抽出部 2 5 は、算出したベクトル A 2 とベクトル B 2 とベクトル C 2 とベクトル D 2 とを、特徴量として抽出する。

【0088】

同様にして、特徴量抽出部 2 5 は、第 5 の補助情報算出対象ウィンドウにおけるベクトル E 2 (3, 7)、第 6 の補助情報算出対象ウィンドウにおけるベクトル F 2 (7, 6)、第 7 のウィンドウにおけるベクトル G 2 (6, 5) を特徴量として抽出する。

【0089】

再び図 1 に戻って、異常検知装置 1 の説明を続ける。

【0090】

10

20

30

40

50

特徴量記憶部 26 は、特徴量抽出部 25 によって抽出された特徴量を記憶する。

【0091】

補助情報算出部 27 は、特徴量記憶部 26 に記憶される特徴量のそれぞれについて、当該特徴量と、リファレンス特徴量記憶部 16 に記憶される特徴量のそれぞれとの間の L_p 距離を算出し、算出した L_p 距離のそれぞれを用いる K 近傍法を利用して、リファレンス特徴量記憶部 16 に記憶されるリファレンス特徴量との K 近傍距離を、補助情報として算出する。

【0092】

L_p 距離は 2 つの同次元の実数値からなるベクトル間に定義される距離であり、2 つのベクトルの差の p -ノルムを計算することで求められる。

【0093】

ベクトル x の p -ノルム ($p > 0$) の求め方を示す算出式を、下記 (式 1)、(式 2) に示す。

【0094】

$p \geq 1$ のとき

【数 1】

$$\|x\|_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{\frac{1}{p}} \quad (\text{式 1})$$

【0095】

$0 < p < 1$ のとき

【数 2】

$$\|x\|_p = |x_1|^p + |x_2|^p + \dots + |x_n|^p \quad (\text{式 2})$$

【0096】

(式 1)、(式 2) に示されるように、ベクトル x の p -ノルム ($p > 0$) は、 p が 1 以上の場合と、 p が 1 未満の場合とで算出式が異なる。

【0097】

(式 1) に示されるように、 p が 1 以上の場合には、ベクトル x の p -ノルムは、 x の各成分の絶対値を p 乗した値の和の p 乗根により求められる。また、(式 2) に示されるように、 p が 1 未満の場合には、ベクトル x の p -ノルムは、 x の各成分の絶対値を p 乗した値の和により求められる。

【0098】

同一の次元を持つ 2 つのベクトル x 、 y 間の L_p 距離の求め方を示す算出式を、下記 (式 3) に示す。

【数 3】

$$L_p(x, y) = \|x - y\|_p \quad (\text{式 3})$$

【0099】

(式 3) に示されるように、同一の次元を持つ 2 つのベクトル x 、 y 間の L_p 距離は、ベクトル (x 、 y) の p -ノルムにより求められる。

【0100】

図 13 は、補助情報算出部 27 が、特徴量記憶部 26 に記憶される特徴量のそれぞれについて、当該特徴量と、リファレンス特徴量記憶部 16 に記憶される特徴量のそれぞれとの間の L_p 距離を算出し、算出した L_p 距離のそれぞれを用いる K 近傍法を利用して、リファレンス特徴量記憶部 16 に記憶される特徴量との K 近傍距離を、補助情報として算出する様子の一例を示す模式図である。

【0101】

図13は、補助情報算出対象ウインドウの長さ L が2であり、 p が1である場合において、特徴量記憶部26に距離列として記憶される特徴量が、 $(1, 5)$ 、 $(5, 1)$ 、 $(1, 1)$ 、 $(1, 3)$ 、 $(3, 7)$ 、 $(7, 6)$ 、 $(6, 5)$ であり、リファレンス特徴量記憶部16にリファレンス距離列として記憶されるリファレンス特徴量が、 $(1, 1)$ 、 $(1, 3)$ 、 $(1, 7)$ 、 $(2, 1)$ 、 $(3, 7)$ 、 $(7, 2)$ であるときの模式図となっている。

【0102】

図13に示されるように、補助情報算出部27は、距離列 $(1, 5)$ と距離列 $(1, 1)$ との間の $L1$ 距離を4と算出し、距離列 $(1, 5)$ と距離列 $(1, 3)$ との間の $L1$ 距離を2と算出し、距離列 $(1, 5)$ と距離列 $(1, 7)$ との間の $L1$ 距離を2と算出し、距離列 $(1, 5)$ と距離列 $(2, 1)$ との間の $L1$ 距離を5と算出し、距離列 $(1, 5)$ と距離列 $(3, 7)$ との間の $L1$ 距離を4と算出し、距離列 $(1, 5)$ と距離列 $(7, 2)$ との間の $L1$ 距離を9と算出する。

【0103】

そして、補助情報算出部27は、 $K=1$ となる K 近傍法を適用して、距離列 $(1, 5)$ である特徴量と、リファレンス特徴量記憶部16に記憶されるリファレンス特徴量との K 近傍距離「2」を、補助情報として算出する。すなわち、補助情報算出部27は、距離列 $(1, 5)$ である特徴量の補助情報を「2」と算出する。

【0104】

同様に、補助情報算出部27は、距離列 $(5, 1)$ である特徴量の補助情報を「3」と算出し、距離列 $(1, 1)$ である特徴量の補助情報を「0」と算出し、距離列 $(1, 3)$ である特徴量の補助情報を「0」と算出し、距離列 $(3, 7)$ である特徴量の補助情報を「0」と算出し、距離列 $(7, 6)$ である特徴量の補助情報を「4」と算出し、距離列 $(6, 5)$ である特徴量の補助情報を「4」と算出する。

【0105】

再び図1に戻って、異常検知装置1の説明を続ける。

【0106】

情報算出部28は、 N (N は、情報算出ウインドウの長さであって、 $L+1$ 以上の整数)個の検知対象パケットデータからなる1以上の情報算出対象ウインドウ毎に、当該 N 個の検知対象パケットデータから算出された W ($W=N-L+1$)個の補助情報を全て用いて、検知対象パケット列20の異常の度合いに係る異常度情報を算出する。より具体的には、情報算出部28は、 W 個の連続する補助情報の全ての和を算出することで、異常度情報を算出する。

【0107】

図13に示される模式図の例において、 N が5の場合、 W は、 $W=5-2+1=4$ となる。このため、情報算出部28は、4つの補助情報の和を異常度情報として算出する。すなわち、情報算出部28は、 $2+3+0+0=5$ となる異常度情報と、 $3+0+0+0=3$ となる異常度情報と、 $0+0+0+4=4$ となる異常度情報と、 $0+0+4+4=8$ となる異常度情報とを算出する。

【0108】

再び図1に戻って、異常検知装置1の説明を続ける。

【0109】

判定部29は、情報算出部28により算出された、検知対象パケット列20の異常度情報に基づいて、検知対象パケット列20が異常であるか否かを判定する。そして、判定部29は、判定結果を外部に出力する。判定部29は、例えば、閾値を記憶し、検知対象パケット列20の異常度情報のうちのいずれか1つが閾値以上である場合に、検知対象パケット列20が異常であると判定し、検知対象パケット列20の異常度情報のうちのいずれもが閾値未満である場合に、検知対象パケット列20が異常でないと判定してもよい。また、判定部29は、第1プロファイル決定部11及び第2プロファイル決定部21の決定対象となるプロファイル毎に、互いに独立した値となる閾値を記憶し、検知対象パケット

10

20

30

40

50

列 2 0 から抽出された検知対象パケット列 2 0 のプロファイルに応じて、そのプロファイルに対応する閾値を用いて上記判定を行うとしてもよい。

【0 1 1 0】

〔2．動作〕

上記構成の異常検知装置 1 は、リファレンス特徴量抽出処理と、異常検知処理とを行う。以下、これらの処理について、図面を用いて順に説明する。

【0 1 1 1】

まず、リファレンス特徴量抽出処理について説明する。

【0 1 1 2】

リファレンス特徴量抽出処理は、リファレンスパケット列群 1 0 から、リファレンス特徴量を抽出する処理である。リファレンス特徴量抽出処理は、例えば、異常検知装置 1 を利用するユーザが、異常検知装置 1 に対して、リファレンス特徴量抽出処理を開始させる旨の操作を行うことで開始される。

10

【0 1 1 3】

図 1 4 は、リファレンス特徴量抽出処理のフローチャートである。

【0 1 1 4】

リファレンス特徴量抽出処理が開始されると、第 1 プロファイル決定部 1 1 は、リファレンスパケット列群 1 0 を取得する（ステップ S 5）。

【0 1 1 5】

第 1 プロファイル決定部 1 1 は、リファレンスパケット列群 1 0 を取得すると、リファレンスパケット列群 1 0 に含まれるリファレンスパケット列の中から、未選択のリファレンスパケット列を 1 つ選択する（ステップ S 1 0）。ここで、未選択のリファレンスパケット列とは、ステップ S 1 0 の処理～ステップ S 5 5 : Y e s の処理（後述）で形成されるループ処理において、過去のステップ S 1 0 の処理で選択されたことのないリファレンスパケット列のことをいう。

20

【0 1 1 6】

第 1 プロファイル決定部 1 1 は、未選択のリファレンスパケット列を 1 つ選択すると、選択したリファレンスパケット列に含まれるパケットの中から、未選択のパケットを 1 つ選択する（ステップ S 1 5）。ここで、未選択のパケットを選択するとは、ステップ S 1 5 の処理～ステップ S 3 5 : Y e s の処理（後述）で形成されるループ処理において、過去のステップ S 1 5 の処理で選択されたことのないパケットのことをいう。

30

【0 1 1 7】

第 1 プロファイル決定部 1 1 は、1 のパケットを選択すると、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定することができるか否かを調べる（ステップ S 2 0）。

【0 1 1 8】

ステップ S 2 0 の処理において、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定することができる場合に（ステップ S 2 0 : Y e s）、第 1 プロファイル決定部 1 1 は、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決

40

【0 1 1 9】

ステップ S 2 0 の処理において、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決

【0 1 2 0】

第 1 プロファイル決定部 1 1 は、選択したパケットのプロファイルを決

50

を調べる（ステップ S 3 5）。

【0121】

ステップ S 3 5 の処理において、未選択の packets がある場合に（ステップ S 3 5 : Yes）、第 1 リファレンス特徴量抽出処理は、再びステップ S 1 5 の処理に進む。

【0122】

ステップ S 3 5 の処理において、未選択の packets がない場合に（ステップ S 3 5 : No）、第 1 抽出部 1 2 は、第 1 プロファイル決定部 1 1 によりプロファイルを決定された各 packets に対して、各 packets のペイロードフィールドに格納されるビット列を packet データとして抽出する。そして第 1 抽出部 1 2 は、同一プロファイルの packet データからなるリファレンス packet データ列を抽出する（ステップ S 4 0）。そして、リファレンス packet データ列群記憶部 1 3 は、第 1 抽出部 1 2 によって抽出されたリファレンス packet データ列を記憶する。

【0123】

リファレンス packet データ列が抽出されると、リファレンス packet データ間距離算出部 1 4 は、リファレンス packet データ列を構成する複数のリファレンス packet データ間の複数のリファレンス距離を算出する（ステップ S 4 5）。

【0124】

複数のリファレンス距離が算出されると、リファレンス特徴量抽出部 1 5 は、算出された複数のリファレンス距離から、補助情報算出対象ウィンドウ毎に、リファレンス距離列をリファレンス特徴量として抽出する（ステップ S 5 0）。そして、リファレンス特徴量記憶部 1 6 は、リファレンス特徴量抽出部 1 5 によってリファレンス特徴量として抽出されたリファレンス距離列を記憶する。

【0125】

リファレンス距離列が抽出されると、第 1 プロファイル決定部 1 1 は、取得したリファレンス packet 列群 1 0 の中に、未選択のリファレンス packet 列があるか否かを調べる（ステップ S 5 5）。

【0126】

ステップ S 5 5 の処理において、未選択のリファレンス packet 列がある場合に（ステップ S 5 5 : Yes）、リファレンス特徴量抽出処理は、再びステップ S 1 0 の処理に進む。

【0127】

ステップ S 5 5 の処理において、未選択のリファレンス packet 列がない場合に（ステップ S 5 5 : No）、リファレンス特徴量抽出処理は、その処理を終了する。

【0128】

次に、異常検知処理について説明する。

【0129】

異常検知処理は、検知対象 packet 列 2 0 の異常を検知する処理である。異常検知処理は、例えば、異常検知装置 1 を利用するユーザが、異常検知装置 1 に対して、異常検知処理を開始する旨の操作を行うことで開始される。

【0130】

図 1 5 は、異常検知処理のフローチャートである。

【0131】

異常検知処理が開始されると、第 2 プロファイル決定部 2 1 は、検知対象 packet 列 2 0 を取得する（ステップ S 1 0 5）。

【0132】

第 2 プロファイル決定部 2 1 は、検知対象 packet 列 2 0 を取得すると、取得した検知対象 packet 列 2 0 に含まれる packet の中から、未選択の packet を 1 つ選択する（ステップ S 1 1 5）。ここで、未選択の packet を選択するとは、ステップ S 1 1 5 の処理～ステップ S 1 3 5 : Yes の処理（後述）で形成されるループ処理において、過去のステップ S 1 1 5 の処理で選択されたことのない packet のことをいう。

【0133】

第2プロファイル決定部21は、1のパケットを選択すると、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定することができるか否かを調べる（ステップS120）。

【0134】

ステップS120の処理において、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定することができる場合に（ステップS120：Yes）、第2プロファイル決定部21は、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定する（ステップS130）。

【0135】

ステップS120の処理において、記憶するプロファイル情報を用いて、選択したパケットのプロファイルを決定できない場合に（ステップS120：No）、第2プロファイル決定部21は、Deep Packet Inspection機能を備えるアプリケーションを実行することで、選択したパケットのプロトコルを特定する（ステップS125）。そして、第2プロファイル決定部21は、特定したプロトコルに基づいて、選択したパケットのプロファイルを決定する（ステップS130）。

【0136】

第2プロファイル決定部21は、選択したパケットのプロファイルを決定すると、取得した検知対象パケット列20に含まれるパケットの中に、未選択のパケットがあるか否かを調べる（ステップS135）。

【0137】

ステップS135の処理において、未選択のパケットがある場合に（ステップS135：Yes）、第1異常検知処理は、再びステップS115の処理に進む。

【0138】

ステップS135の処理において、未選択のパケットがない場合に（ステップS135：No）、第2抽出部22は、第2プロファイル決定部21によりプロファイルを決定された各パケットに対して、各パケットのペイロードフィールドに格納されるビット列をパケットデータとして抽出する。そして第2抽出部22は、同一プロファイルのパケットデータからなる検知対象パケットデータ列を抽出する（ステップS140）。そして、検知対象パケットデータ列記憶部23は、第2抽出部22によって抽出された検知対象パケットデータ列を記憶する。

【0139】

検知対象パケットデータ列が抽出されると、検知対象パケットデータ間距離算出部24は、検知対象パケットデータ列を構成する複数の検知対象パケットデータ間の複数の距離を算出する（ステップS145）。

【0140】

複数の距離が算出されると、特徴量抽出部25は、算出された複数の距離から、補助情報算出対象ウィンドウ毎に、距離列を特徴量として抽出する（ステップS150）。そして、特徴量記憶部26は、特徴量抽出部25によって特徴量として抽出された距離列を記憶する。

【0141】

距離列が抽出されると、補助情報算出部27は、抽出された距離列のそれぞれについて、リファレンス特徴量記憶部16にリファレンス特徴量として記憶されるリファレンス距離列のそれぞれとの間のLp距離を算出し（ステップS155）、算出したLp距離を用いるK近傍法を利用して、K近傍距離を補助情報として算出する（ステップS160）。

【0142】

補助情報が算出されると、情報算出部28は、情報算出対象ウィンドウ毎に、連続するW個の補助情報の和を、異常度情報として算出する（ステップS165）。

【0143】

各異常度情報が算出されると、判定部29は、算出されたそれぞれの異常度情報が、記

10

20

30

40

50

憶する閾値以上であるか否かをそれぞれ調べる（ステップ S 1 7 0）。

【0144】

ステップ S 1 7 0 の処理において、算出されたそれぞれの異常度情報のうちいずれか 1 つが、記憶する閾値以上である場合に（ステップ S 1 7 0：Y e s）、判定部 2 9 は、検知対象パケット列 2 0 が異常であると判定し（ステップ S 1 7 5）、検知対象パケット列 2 0 が異常である旨の信号を外部に出力する。

【0145】

ステップ S 1 7 0 の処理において、算出されたそれぞれの異常度情報が、いずれも記憶する閾値以上でない場合に（ステップ S 1 7 0：N o）、判定部 2 9 は、検知対象パケット列 2 0 が異常でないとして判定し（ステップ S 1 8 0）、検知対象パケット列 2 0 が異常でない旨の信号を外部に出力する。

10

【0146】

ステップ S 1 7 5 の処理が終了した場合、及びステップ S 1 8 0 の処理が終了した場合に異常検知処理は、その処理を終了する。

【0147】

〔3. 考察〕

上記構成の異常検知装置 1 は、リファレンス特徴量抽出処理を実行することで、正常なパケット列からなるリファレンスパケット列群 1 0 から、正常なパケット列のコンテキストの特徴を、リファレンス特徴量として抽出して記憶する。そして、異常検知装置 1 は、異常検知処理を実行することで、検知対象パケット列 2 0 から、検知対象パケット列 2 0 のコンテキストの特徴を特徴量として抽出し、抽出した特徴量が記憶するリファレンス特徴量から逸脱している度合いを示す異常度情報を算出し、算出した異常度情報に基づいて、検知対象パケット列 2 0 が異常であるか否かを判定する。このように、異常検知装置 1 によると、検知対象パケット列 2 0 の異常を検知することができる。

20

【0148】

また、上記補助情報を用いて異常度情報を算出することにより、例えば、以下の効果が得られる。

【0149】

すなわち、（A）検知対象パケット列 2 0 がリファレンスパケット列群 1 0 と同様の遷移を行う場合には、異常度情報のスコアが低くなり、（B）検知対象パケット列 2 0 で、確率は低いが確実に起こる遷移が起こった場合にも、異常度情報のスコアが低くなり、（C）検知対象パケット列 2 0 で、リファレンスパケット列群 1 0 で起こったことがない遷移が起こった場合に、その異常性を L p 距離で定量化することができ、異常度情報のスコアが高くなる、という効果が得られる。

30

【0150】

（補足）

以上のように、本出願において開示する技術の例示として、実施の形態について説明した。しかしながら本開示による技術は、これらに限定されず、適宜、変更、置き換え、付加、省略等を行った実施の形態にも適用可能である。

【0151】

以下に、本開示における変形例の一例について列記する。

【0152】

（1）実施の形態において、異常検知装置 1 は、判定部 2 9 が、検知対象パケット列 2 0 が異常であるか否かを判定し、その判定結果を外部に出力する構成の例であった。これに対して、他の構成の例として、異常検知装置 1 は、情報算出部 2 8 が、検知対象パケット列 2 0 の異常の度合いを示す異常度を算出し、算出した異常度を外部に出力する構成の例も考えられる。この場合、異常検知装置 1 は、判定部 2 9 を含まない構成であっても構わない。

40

【0153】

（2）実施の形態において、異常検知装置 1 は、リファレンス特徴量抽出処理を実行す

50

ることで、リファレンス特徴量を抽出し、抽出したリファレンス特徴量を、リファレンス特徴量抽出部 15 で記憶する構成の例であった。これに対して、他の構成の例として、異常検知装置 1 は、外部装置等で抽出されたリファレンス特徴量を外部から取得し、取得したリファレンス特徴量を、予めリファレンス特徴量記憶部 16 で記憶しておく構成の例も考えられる。この場合、異常検知装置 1 は、リファレンス特徴量抽出処理を行う必要はなく、第 1 プロファイル決定部 11 と、第 1 抽出部 12 と、リファレンスパケットデータ列群記憶部 13 と、リファレンスパケットデータ間距離算出部 14 と、リファレンス特徴量抽出部 15 とを含まない構成であっても構わない。また、他の構成の例として、異常検知装置 1 は、リファレンス特徴量記憶部 16 が予めリファレンス特徴量を記憶している状態で製造される構成の例も考えられる。この場合も、異常検知装置 1 は、リファレンス特徴量抽出処理を行う必要はなく、第 1 プロファイル決定部 11 と、第 1 抽出部 12 と、リファレンスパケットデータ列群記憶部 13 と、リファレンスパケットデータ間距離算出部 14 と、リファレンス特徴量抽出部 15 とを含まない構成であっても構わない。

10

【0154】

(3) 実施の形態において、異常検知装置 1 が検知の対象とするのはパケット列であったが、必ずしもパケット列である必要はなく、データの列で構成されるものであればよい。例えば図 16 に示される侵入検知システムの検知ログのデータ列の中で、Description カラムに現れる文字列の系列を検知の対象とするデータ列とみなして本開示に係る異常検知方法を適用しても構わない。また、syslog のメッセージテキストを検知の対象のデータ列とみなして本開示に係る異常検知方法を適用しても構わない。また、制御機器や持つ制御に関するログの情報から適切なカラムを抽出し、そのカラムのデータを検知対象のデータ列とみなして本開示に係る異常検知方法を適用しても構わない。

20

【0155】

(4) 情報算出部 28 は、補助情報算出部 27 が算出する、連続する W 個の補助情報の和を異常度情報として算出するとして説明した。しかしながら、異常度情報を算出する関数は、必ずしも連続する W 個の補助情報の和である必要はなく、補助情報算出部 27 が算出する補助情報の連続する W 次元の値を入力とし、異常度情報を返す関数であれば良い。

【0156】

(5) 実施の形態において、異常検知装置 1 における各構成要素は、IC (Integrated Circuit)、LSI (Large Scale Integration) 等の半導体装置により個別に 1 チップ化されてもよいし、一部又は全部を含むように 1 チップ化されてもよい。また、集積回路化の手法は LSI に限るものではなく、専用回路又は汎用プロセッサで実現してもよい。LSI 製造後に、プログラムすることが可能な FPGA (Field Programmable Gate Array) や、LSI 内部の回路セルの接続や設定を再構成可能なリコンフィギュラブル・プロセッサを利用してもよい。更には、半導体技術の進歩又は派生する別技術により LSI に置き換わる集積回路化の技術が登場すれば、その技術を用いて機能ブロックの集積化を行ってもよい。バイオ技術の適用等が可能性としてあり得る。

30

【産業上の利用可能性】

【0157】

本開示は、データ列を利用するシステムの異常検知に広く利用可能である。

40

【符号の説明】

【0158】

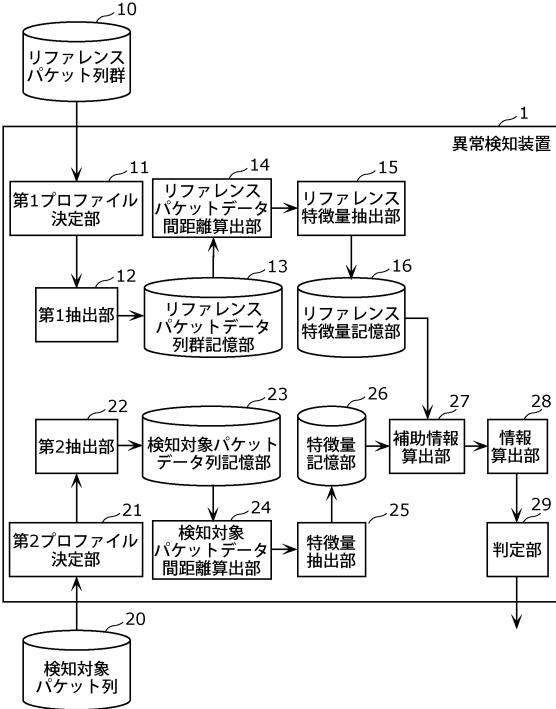
- 1 異常検知装置
- 10 リファレンスパケット列群
- 11 第 1 プロファイル決定部
- 12 第 1 抽出部
- 13 リファレンスパケットデータ列群記憶部
- 14 リファレンスパケットデータ間距離算出部
- 15 リファレンス特徴量抽出部

50

- 16 リファレンス特徴量記憶部
- 20 検知対象パケット列
- 21 第2プロファイル決定部
- 22 第2抽出部
- 23 検知対象パケットデータ列記憶部
- 24 検知対象パケットデータ間距離算出部
- 25 特徴量抽出部
- 26 特徴量記憶部
- 27 補助情報算出部
- 28 情報算出部
- 29 判定部

【図面】

【図 1】



【図 2】

プロファイル ID	送信元 IP	対象プロトコル	対象コマンド	...
1	192.168.0.1	BACnet/IP	write コマンド	...
2	192.168.0.1	BACnet/IP	read コマンド	...
3	192.168.0.2	BACnet/IP	状態通知コマンド	...
4	192.168.0.3	BACnet/IP	状態通知コマンド	...

10

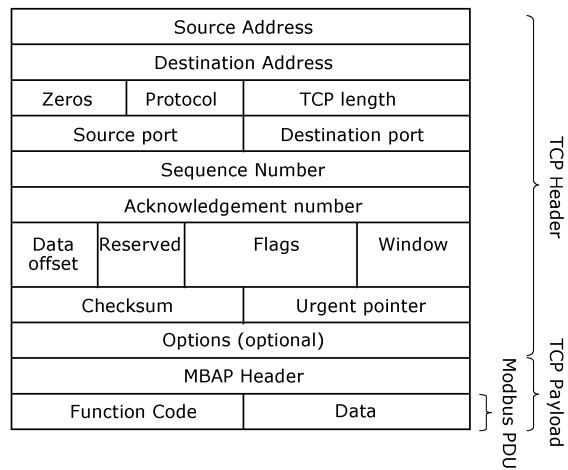
20

30

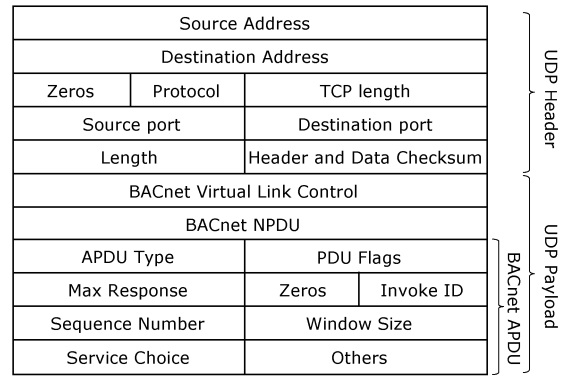
40

50

【図 3】



【図 4】

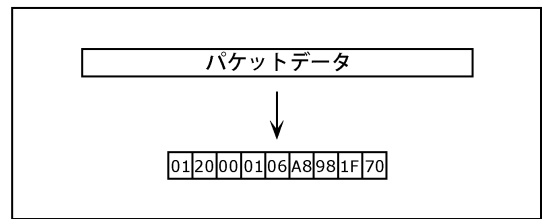


10

【図 5】



【図 6】



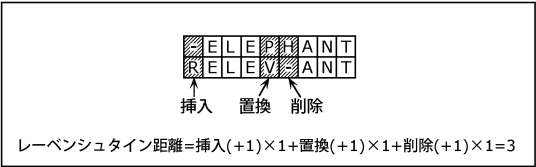
20

30

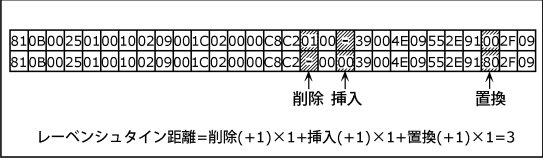
40

50

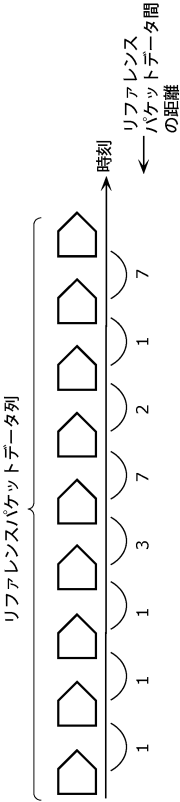
【図 7】



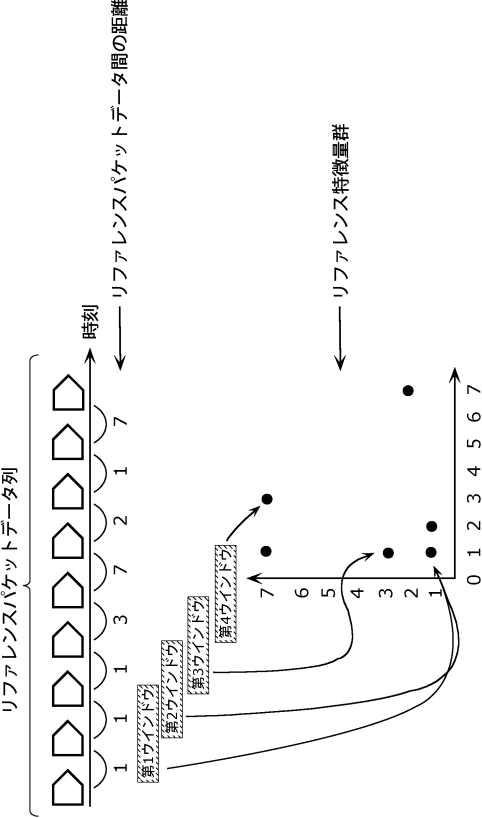
【図 8】



【図 9】



【図 10】



10

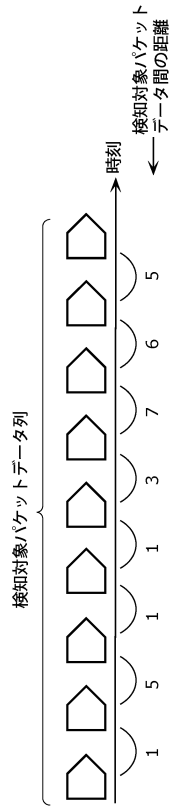
20

30

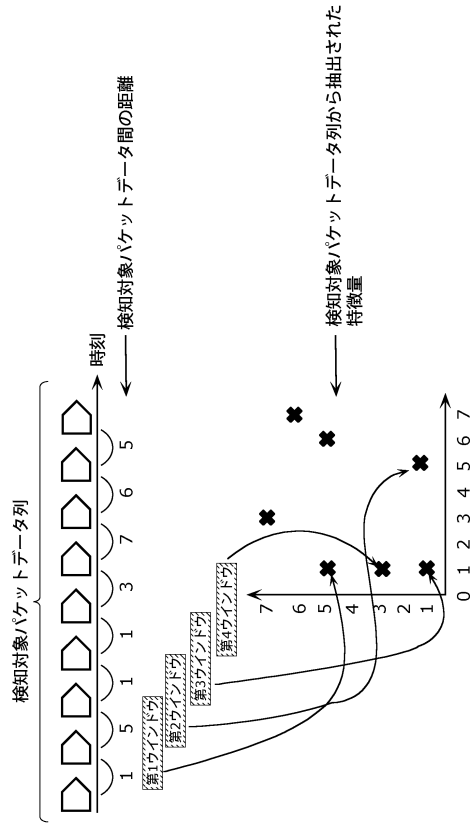
40

50

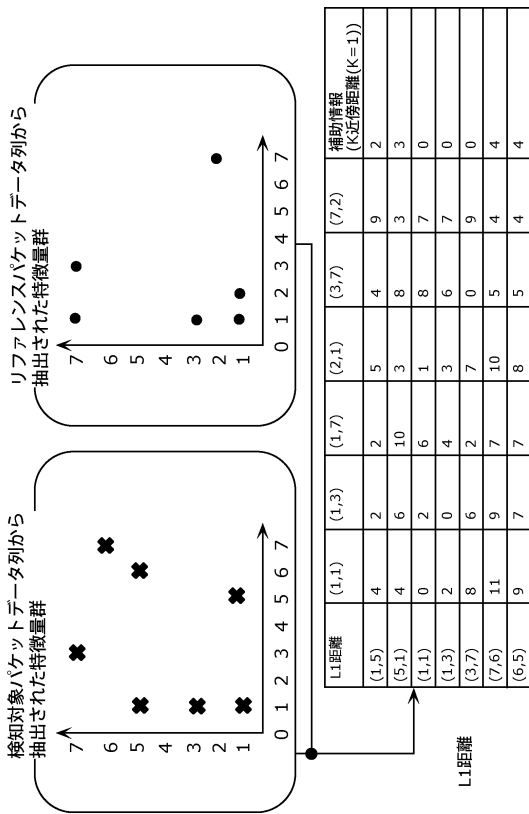
【図 1 1】



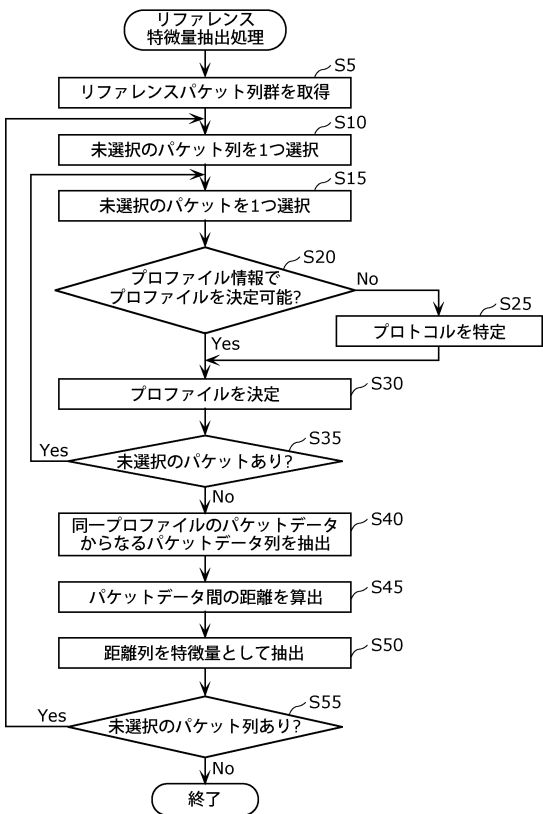
【図 1 2】



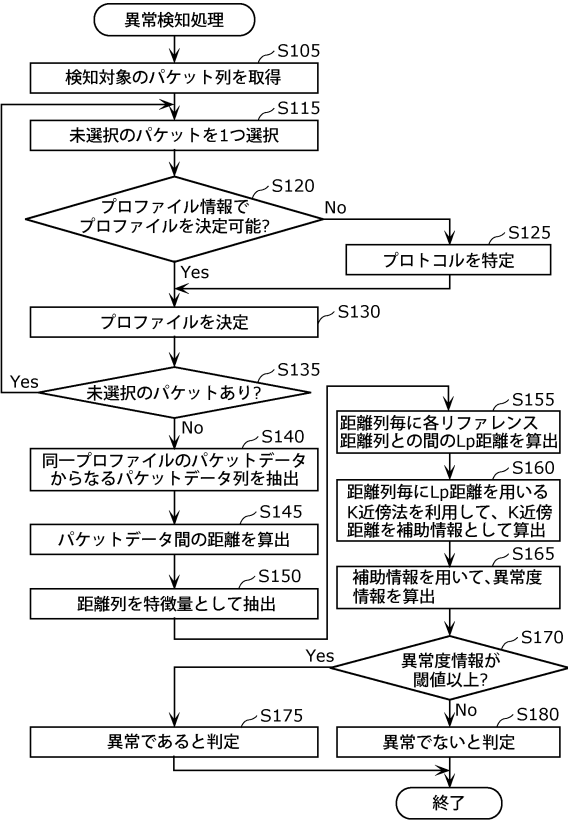
【図 1 3】



【図 1 4】



【図 1 5】



【図 1 6】

Protocol	ICMP	ICMP	TCP	UDP	TCP	UDP
Destination	192.168.1.20	192.168.1.20	192.168.1.20	192.168.1.20	192.168.1.20	192.168.1.20
Source	192.168.1.10	192.168.1.10	192.168.1.10	192.168.1.10	192.168.1.10	192.168.1.10
Description	ICMP PING	ICMP PING undefined code	SCAN nmap XMAS	SHELLCODE x86 inc ebx NOOP	SCAN nmap XMAS	SHELLCODE x86 inc ebx NOOP
Severity	low	low	medium	high	medium	high
Date	2019/1/1 10:01:22	2019/1/1 10:01:22	2019/1/1 10:01:22	2019/1/1 10:01:22	2019/1/1 10:01:22	2019/1/1 10:01:22

10

20

30

40

50

フロントページの続き

- (72)発明者 大庭 達海
日本国大阪府門真市大字門真１００６番地 パナソニック株式会社内
- 審査官 長谷川 未貴
- (56)参考文献 特開２０１９－１１０５１３（ＪＰ，Ａ）
国際公開第２０１８／１６３３４２（ＷＯ，Ａ１）
特開２０１７－７６８４１（ＪＰ，Ａ）
中国特許出願公開第１０９５４７４０９（ＣＮ，Ａ）
特開２０１７－１０８２２０（ＪＰ，Ａ）
- (58)調査した分野 (Int.Cl.，ＤＢ名)
Ｈ０４Ｌ ４１／００
Ｈ０４Ｌ ４３／００