(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6094056号
(P6094056)

(45)発行日　平成29年3月15日(2017.3.15)　(24)登録日　平成29年2月24日(2017.2.24)

(51)Int.Cl.		F I		
H04L 12/58	(2006.01)	H04L	12/58	100Z
G06F 13/00	(2006.01)	G06F	13/00	610Q
H04M 3/53	(2006.01)	H04M	3/53	

請求項の数 11　(全 28 頁)

(21)出願番号　特願2012-108491 (P2012-108491)
(22)出願日　平成24年5月10日 (2012.5.10)
(65)公開番号　特開2013-236308 (P2013-236308A)
(43)公開日　平成25年11月21日 (2013.11.21)
審査請求日　平成27年3月19日 (2015.3.19)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100094525
弁理士　土井　健二
(74)代理人　100094514
弁理士　林　恒徳
(72)発明者　片山　佳則
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　高　杰
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

最終頁に続く

(54)【発明の名称】メールチェック方法、メールチェック装置、及び、メールチェックプログラム

(57)【特許請求の範囲】

【請求項１】

受信対象メールから警戒メール候補を検出するメールチェック方法であって、

複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶する記憶工程と、

記憶された前記特徴情報を、直近の一定期間内の受信済みメール、または、直近の一定量の受信済みメールのいずれかまたは両方が含む特徴情報に更新する更新工程と、

前記受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との類似度が第１の基準値未満の場合、前記受信対象メールを前記警戒メール候補として検出する検出工程と、を有するメールチェック方法。

【請求項２】

受信対象メールから警戒メール候補を検出するメールチェック方法であって、

複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶する記憶工程と、

前記受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との一致度を、前記特徴情報に対応する重み係数にしたがって重み付けして類似度を求め、前記類似度が第１の基準値未満の場合、前記受信対象メールを前記警戒メール候補として検出する検出工程と、

前記複数の受信済みメールの前記特徴情報のばらつき度合いが第１の値の重み係数は、前記ばらつき度合いが前記第１の値より大きい第２の値の重み係数より大きい、メールチェック方法。

【請求項３】

受信対象メールから警戒メール候補を検出するメールチェック方法であって、

複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶する記憶工程と、

前記受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との一致度を、前記特徴情報に対応する重み係数にしたがって重み付けして類似度を求め、前記類似度が第１の基準値未満の場合、前記受信対象メールを前記警戒メール候補として検出する検出工程と、

前記特徴情報は、送信元メールサーバのＩＰアドレス、前記送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別を含む第１特徴情報群のうちいずれか１つまたは複数と、送信曜日、送信時刻、タイトル情報におけるパターン情報、宛先情報を含む第２特徴情報群のうちいずれか１つまたは複数とを有し、

前記第１特徴情報群の特徴情報の重み係数は、前記第２特徴情報群の特徴情報の重み係数より大きい、メールチェック方法。

【請求項４】

受信対象メールから警戒メール候補を検出するメールチェック方法であって、

複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶する記憶工程と、

前記受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との一致度を、前記特徴情報に対応する重み係数にしたがって重み付けして類似度を求め、前記類似度が第１の基準値未満の場合、前記受信対象メールを前記警戒メール候補として検出する検出工程と、

前記特徴情報は、送信元メールサーバのＩＰアドレス、前記送信元メールサーバのドメイン情報、タイムゾーン情報を含む送信元情報群のうち複数を組み合わせた組み合わせ特徴情報を含み、

前記組み合わせ特徴情報の重み係数は、前記送信元情報群の各特徴情報の重み係数より大きい、メールチェック方法。

【請求項５】

請求項１、２、４のいずれかにおいて、

前記特徴情報は、送信元メールサーバのＩＰアドレス、前記送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別のうちいずれか１つまたは複数を含むメールチェック方法。

【請求項６】

請求項５において、

前記特徴情報は、さらに、送信曜日、送信時刻、タイトル情報におけるパターン情報、宛先情報のうちいずれか１つまたは複数を含むメールチェック方法。

【請求項７】

請求項２において、

前記重み係数は前記送信元アドレス毎に設けられ、

前記特徴情報のばらつき度合いは、前記送信元アドレスが同一の前記複数の受信済みメールにおける前記特徴情報のばらつき度合いであるメールチェック方法。

【請求項８】

請求項２乃至４のいずれかにおいて、

前記類似度は、前記一致度を前記重み付けした値が第２の基準値を超える受信済みメール数を示すメールチェック方法。

【請求項９】
請求項４または５において、
前記送信元メールサーバのＩＰアドレスは、ＩＰアドレスのうち、基準上位レベルのＩＰアドレスであるメールチェック方法。
【請求項１０】
受信対象メールから警戒メール候補を検出するメールチェック装置であって、
複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶する記憶手段と、
記憶された前記特徴情報を、直近の一定期間内の受信済みメール、または、直近の一定量の受信済みメールのいずれかまたは両方が含む特徴情報に更新する更新手段と、
前記受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との類似度が第１の基準値未満の場合、前記受信対象メールを前記警戒メール候補として検出する検出手段と、を有するメールチェック装置。
【請求項１１】
複数の受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し記憶し、
記憶された前記特徴情報を、直近の一定期間内の受信済みメール、または、直近の一定量の受信済みメールのいずれかまたは両方が含む特徴情報に更新し、
受信対象メールのメールヘッダーが含む複数の特徴情報と、前記記憶した特徴情報を含む複数の受信済みメールのうち、当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との類似度が第１の基準値未満の場合、前記受信対象メールを警戒メール候補として検出する、処理をコンピュータに実行させるメールチェックプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、メールチェック方法、メールチェック装置、及び、メールチェックプログラムに関する。
【背景技術】
【０００２】
近年、特定の組織や個人のコンピュータに格納された情報の窃盗を目的とした標的型攻撃メールが急増している。標的型攻撃メールは、不特定多数に対する攻撃ではなく、ある特定の対象を定めて攻撃が行われる特徴を有する。例えば、標的型攻撃メールは、実在の組織や個人の発信元を詐称し、正当な業務や依頼であるかのように見せかける件名や本文のメールとして送りつけられる。そして、受信者に添付ファイルを開くことによるウイルス感染や、特定のサイトへの誘導によるウイルス送信を誘発させる。
【０００３】
このように、標的型攻撃メールは実在の送信元を装い、正規のメールであるかのように見せかけられて送信される。しかしながら、受信者それぞれが、受信対象のメールについて、メールヘッダー、添付ファイル、本文、送信者アドレス等の整合性を逐一チェックし、標的型攻撃メールを判別する処理には限界がある。このため、メーラーによるメール受信前に、メールチェッカーによって標的型攻撃メールが自動検出されることが望ましい。
【０００４】
近年のメールチェッカーでは、高頻度で出現する特徴的な単語等を、人手や自動学習によって指定することで、スパムメールの検出を実現している（例えば、特許文献１）。また、例えば、受信メールサーバが、送信元メールアドレスのドメインの検証し、当該ドメインが正規のサーバであるか否かを判定することによって、非正規のメールを検出する方法がある。これにより、標的型攻撃メールやスパムメール等の非正規のメールが検出される。
【先行技術文献】

【特許文献】
【０００５】
【特許文献１】特願２０１０-５０１８６４号公報
【発明の概要】
【発明が解決しようとする課題】
【０００６】
しかしながら、標的型攻撃メールはスパムメールと異なり数が少ないことから、メールチェッカーが、標的型攻撃メールから得られる情報に基づいて特徴を学習することは困難である。また、受信メールサーバがドメインの検証を行うためには、メールサーバそれぞれについて、ドメイン検証機能を搭載する必要があり、容易ではない。このように、標的型攻撃メールを効率的に検出することが困難であった。
【０００７】
本発明は、標的型攻撃メール候補を効率的に検出するメールチェック方法、メールチェック装置、及び、メールチェックプログラムを提供することにある。
【課題を解決するための手段】
【０００８】
第１の側面は、新たに受信する対象メールから警戒メール候補を検出するメールチェック方法であって、正規の複数の受信済みメールについて、前記受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報データーベース（以下、ＤＢ）を生成する特徴情報ＤＢ生成工程と、受信対象メールの前記メールヘッダーが含む第１の複数の特徴情報と、前記特徴情報ＤＢから参照される当該受信対象メールと送信元アドレスが同一の前記複数の受信済みメールの第２の複数の特徴情報との類似度が基準類似度未満の場合、前記受信対象メールを警戒メール候補として検出する。
【発明の効果】
【０００９】
第１の側面によれば、標的型攻撃メール候補を効率的に検出する。
【図面の簡単な説明】
【００１０】
【図１】本実施の形態例におけるメールチェック装置の構成の一例を示す図である。
【図２】メールヘッダーの一例を示す図である。
【図３】受信済みメールから抽出された送信日時（特徴情報）の具体例を示す図である。
【図４】タイトル情報に含まれるパターン情報、表記揺れ辞書ｄ３を示す例図である。
【図５】受信済みメールに基づいて取得される宛先情報の具体例を示す図である。
【図６】送信元サーバのＩＰアドレス、経由タイムゾーンの具体例を示す図である。
【図７】組み合わせ特徴情報の具体例を示す図である。
【図８】特徴情報ＤＢの具体例を示す図である。
【図９】類似判定処理の流れを説明するフローチャート図である。
【図１０】総一致量値ｓｉｍの算出処理を説明するフローチャート図である。
【図１１】送信元アドレス毎の特徴情報の重み係数の一例を示す図である。
【発明を実施するための形態】
【００１１】
以下、図面にしたがって本発明の実施の形態について説明する。ただし、本発明の技術的範囲はこれらの実施の形態に限定されず、特許請求の範囲に記載された事項とその均等物まで及ぶものである。
【００１２】
［メールチェック装置２００の構成］
図１は、本実施の形態例におけるメールチェック装置２００の構成の一例を示す図である。本実施の形態例におけるメールチェック装置２００は、例えば、メールチェック部１０、メーラー２１、アラートユーザインターフェース２２を有する。メールチェック装置２００は、インターネットを介して、他のメールサーバ１００と接続される。なお、本実

施の形態例におけるメールチェック装置２００は、サーバ側、クライアント側のいずれにあってもよい。

【００１３】

本実施の形態例におけるメールチェック部１０は、受信対象のメールを検出すると、受信対象のメールと正規の受信済みのメールとの特徴情報を比較して類似度を求める。そして、メールチェック部１０は、類似度が低い受信対象メールについて、非正規のメール、即ち、標的型攻撃メールである可能性があるとして、アラートユーザインターフェース２２等を介して、表示画面等に警告情報を通知する。受信者（以下、ユーザ）は、受信対象のメールについて警告情報が通知されると、受信対象メールを確認する。そして、ユーザは、受信対象メールについて、標的型攻撃メールであると判定した場合は受信を許可しない。

【００１４】

図１のメールチェック部１０は、例えば、メールサーバ１１、メール受信装置１２、メール特徴抽出部１３、標的型攻撃メール検出部１４、アラート発生部１５、特徴情報ＤＢｄ１、送信者毎の重み係数情報ｄ２、表記揺れ辞書ｄ３を有する。メール受信装置１２は、受信対象のメールを検知し、メールサーバ１１に通知する。メールサーバ１１は、メールの送受信に係る制御を行うと共に、受信対象のメールが検知されると、標的型攻撃メール検出部１４を呼び出す。

【００１５】

標的型攻撃メール検出部１４は、受信対象のメールのメールヘッダーに含まれる特徴情報を抽出すると共に、特徴情報ＤＢｄ１を参照して受信対象メールと同一の送信元アドレスの正規の受信済みメールの特徴情報を読み出す。そして、標的型攻撃メール検出部１４は、受信対象のメールと受信済みのメールとの特徴情報の類似度を求め、類似度が基準類似度未満の場合、アラート発生部１５に受信対象メールを警戒メール候補として通知する。アラート発生部１５は、通知された警戒メール候補の情報をアラートユーザインターフェース２２に表示させる。

【００１６】

メール特徴抽出部１３は、正規の受信済みメール毎に、各メールのメールヘッダーに含まれる特徴情報を抽出し特徴情報ＤＢｄ１を生成する。また、メール特徴抽出部１３は、表記揺れ語句のリストである表記揺れ辞書ｄ３を有する。メール特徴抽出部１３は、表記揺れ辞書ｄ３を参照して、メールヘッダーに含まれるタイトル情報が有する表記揺れ語句を特徴情報として検出する。詳細については後述する。

【００１７】

また、送信者重み係数情報ｄ２は、送信元アドレス毎の複数の特徴情報それぞれについて重み係数を有する。標的型攻撃メール検出部１４は、それぞれの特徴情報に対応する重み係数情報ｄ２を参照し類似度の算出時に反映する。

【００１８】

このように、本実施の形態例におけるメールチェック部１０は、正規の複数の受信済みメールについて、メールのメールヘッダーに含まれる複数の特徴情報を抽出し、特徴情報ＤＢｄ１を生成する。そして、メールチェック部１０は、受信する対象メールのメールヘッダーに含まれる特徴情報と、特徴情報ＤＢｄ１から参照される受信対象メールと送信元アドレスが同一の複数の受信済みメールの特徴情報との類似度を判定する。そして、メールチェック部１０は、類似度が基準類似度未満の場合に、受信対象のメールを警戒メール候補として検出する。

【００１９】

ここで、特徴情報について説明する。

【００２０】

［特徴情報］

特徴情報とは、メールのメールヘッダーに含まれる情報の一部を示す。本実施の形態例では、特徴情報とは、例えば、送信元メールサーバのＩＰアドレス、送信元メールサーバ

のドメイン情報、タイムゾーン情報、送信者が使用するメーラーのソフトウェア種別、送信曜日、送信時刻、タイトル情報におけるパターン情報、宛先情報等を示す。これらの特徴情報のうちいずれかの複数の特徴情報が、受信対象メールと受信済みメールとの間で比較される。比較処理の具体例については、フローチャート図に基づいて後述する。

【００２１】

続いて、ここでメールヘッダーの具体例について説明する。

【００２２】

［メールヘッダー］

図２は、メールヘッダーＭｈの一例を示す図である。同図のメールヘッダーＭｈは、例えば、アドレス「ｂｂｂｂｂ@ｄｏｍａｉｎ２．ｃｏｍ」から、アドレス「ａａａａａ@ｄｏｍａｉｎ１．ｃｏｍ」に送信された受信対象メールのメールヘッダーＭｈの一例である。

【００２３】

図２のように、メールヘッダーＭｈは、例えば、「Ｆｒｏｍ」、「Ｓｅｎｄｅｒ」、「Ｔｏ」、「Ｓｕｂｊｅｃｔ」、「Ｍｅｓｓａｇｅ－ＩＤ」、「Ｄａｔｅ」、「Ｒｅｃｅｉｖｅｄ」、「Ｒｅｐｌｙ－Ｔｏ」、「Ｘ－Ｍａｉｌｅｒ」等の項目を含む。ただし、メールヘッダーＭｈに含まれる情報及びその書式は、メーラーのソフトウェアの種別によって異なるため、図２の例に限定されるものではない。

【００２４】

具体的に、Ｆｒｏｍ情報は送信元アドレスを、Ｔｏ情報は宛先アドレスを示す。Ｒｅｐｌｙ－Ｔｏ情報はメールの返信先を示し、Ｓｕｂｊｅｃｔ情報はメールのタイトル情報を示す。また、Ｍｅｓｓａｇｅ－ＩＤ情報は、メールを特定するためのユニークな識別ＩＤであって、送信日付、時刻、送信元メールサーバのドメイン等の情報に基づいて生成される情報を示す。また、Ｄａｔｅ情報はメールの送信日時、Ｘ－Ｍａｉｌｅｒ情報は、メール送信者が使用するメーラーのソフトウェア名を示す。

【００２５】

そして、Ｒｅｃｅｉｖｅｄ情報は、メールの経路サーバのＩＰアドレス及びドメイン名、送信元サーバのＩＰアドレス及びドメイン名の情報を示す。Ｒｅｃｅｉｖｅｄ情報は、経路したサーバの数分、下から順に記録される。つまり、一番下のＲｅｃｅｉｖｅｄ情報が送信元で、一番上のＲｅｃｅｉｖｅｄ情報が自身のメールサーバを示す。この例において、１行目のＲｅｃｅｉｖｅｄ情報は、受信サーバの記録を、２行目のＲｅｃｅｉｖｅｄ情報は、送信サーバの記録を示す。Ｒｅｃｅｉｖｅｄ情報のフォーマットは、例えば、「Ｒｅｃｅｉｖｅｄ：ｆｒｏｍサーバ［ＩＰアドレス］ｂｙ受信サーバ名［ｗｉｔｈ転送プロトコル］ｉｄユニークＩＤ ｆｏｒ宛先メールアドレス：処理日時」である。処理日時には、経由サーバの送信曜日情報、送信時刻情報、タイムゾーン情報が含まれる。

【００２６】

図２のメールヘッダーＭｈの例によると、Ｆｒｏｍ情報Ｈ１に基づいて送信元アドレスが「ｂｂｂｂｂ@ｄｏｍａｉｎ２．ｃｏｍ」、Ｔｏ情報Ｈ３に基づいて送信先アドレスが「ａａａａａ@ｄｏｍａｉｎ１．ｃｏｍ」であることがわかる。また、Ｄａｔｅ情報Ｈ８に基づいて、送信日時が２０１０年１１月２２日（月）の１０：０２：１３（中国時間）であることが判明する。同様にして、Ｓｕｂｊｅｃｔ情報Ｈ２に基づいて、メールのタイトルが「次回－打ち合わせについて」であり、X-Ｍａｉｌｅｒ情報Ｈ８に基づいて、メール送信時に使用されたソフトウェアがＦｏｘｍａｉｌであることがわかる。また、Ｍｅｓｓａｇｅ－ＩＤ情報Ｈ７に基づいて、送信元のメールサーバのドメインがｄｏｍａｉｎ２．ｃｏｍであることがわかる。

【００２７】

また、図２のメールヘッダーＭｈにおける1番目のＲｅｃｅｉｖｅｄ情報Ｈ１１に基づいて、２０１０年１１月２２日（月）の１０：０２：１３（日本時間）に、宛先ａａａａａ@ｄｏｍａｉｎ１.ｃｏｍのメールを、ＩＰアドレスｘ０．２６.ｘｘｘ.ｘｘｘの受信サーバｒｅｃｅｉｖｅｈｏｓｔ.ｃｏ.ｊｐから受信したことがわかる。そして、２番目の

Ｒｅｃｅｉｖｅｄ情報Ｈ１２に基づいて、２０１０年１１月２２日（月）の１０：０２：１３（中国時間Ｈ４）に、サーバｒｅｃｅｉｖｅｈｏｓｔ．ｃｏ．ｊｐによって、ＩＰアドレス２１０．ｘｘｘ．ｘｘｘ．２２６（Ｈ５）の送信元サーバｓｅｎｄｈｏｓｔ．ｃｏ．ｊｐからメールを受信したことがわかる。

【００２８】

このように、図２で示したようなメールのメールヘッダーＭｈから、複数の種別の特徴情報が検出される。続いて、それぞれの特徴情報について、具体例に基づいて説明する。

【００２９】

［特徴情報：送信日時］

図３は、受信済みメールから抽出された送信日時（特徴情報）の具体例を示す図である。同図のリストＬ１は、３つの送信元アドレスｓｕｚｕｋｉ、ｔａｎａｋａ、ｉｔｐｍｌ２の受信済みメールのメールヘッダーに含まれる送信日時の具体例を示す。この例において、送信元アドレスは、例えば、ｓｕｚｕｋｉのように、メールアドレスにおけるローカル部（@以前の情報）として表される。図２で前述したとおり、送信日時は、Ｄａｔｅ情報に基づいて取得される。

【００３０】

例えば、図３の日時リストＬ１における情報Ｌ１－１は、送信元アドレスｉｔｐｍｌ２の受信済みメールが、２０１２年３月１日の１２時４１分に送信されたことを示す。同様にして、日時リストＬ１における情報Ｌ１－２は、送信元アドレスｓｕｚｕｋｉの受信済みメールが、２０１２年３月１日の１０時１４分に送信されたことを示す。このように、それぞれの受信済みメールに基づいて、送信日時、時刻に含まれる送信時刻、送信曜日等が特徴情報として抽出される。

【００３１】

また、図３のグラフ時刻ｔｇ１～ｔｇ３は、同図のリストＬ１に基づいて、送信元アドレス別に、送信時刻毎の受信済みメール数がグラフ化された図である。時刻グラフｔｇ１～ｔｇ３の横軸は送信時刻を、縦軸は受信メール数を示す。そして、同図の曜日グラフｗｇ１～ｗｇ３は、リストＬ１に基づいて、送信元アドレス別に、送信曜日毎の受信済みメール数がグラフ化された図である。曜日グラフｗｇ１～ｗｇ３の横軸は送信曜日を、縦軸は受信メール数を示す。

【００３２】

まず、送信元アドレスｔａｎａｋａの送信日時について説明する。送信元アドレスｔａｎａｋａの時刻グラフｔｇ１によると、送信元アドレスｔａｎａｋａからの受信済みメールは、全て、８時から１０時近くに送信されている。一方、送信元アドレスｔａｎａｋａの曜日グラフｗｇ１によると、送信元アドレスｔａｎａｋａからの受信済みメールは、月曜日から金曜日まで偏りなく送信されている。つまり、送信元アドレスｔａｎａｋａからの受信済みメールは、月曜日から金曜日のいずれかの曜日に、８時から１０時近くに送信される傾向を有する。

【００３３】

続いて、送信元アドレスｓｕｚｕｋｉの送信日時について説明する。送信元アドレスｓｕｚｕｋｉの時刻グラフｔｇ２によると、送信元アドレスｓｕｚｕｋｉからの受信済みメールは、時刻に偏りなく送信されている。また、送信元アドレスｓｕｚｕｋｉの曜日グラフｗｇ２によると、送信元アドレスｓｕｚｕｋｉからの受信済みメールは、月曜日から土曜日の間、偏りなく送信されている。このため、送信元アドレスｓｕｚｕｋｉの受信済みメールについて、送信時刻、送信曜日のばらつき度合いは大きい。

【００３４】

続いて、送信元アドレスｉｔｐｍｌ２の送信日時について説明する。送信元アドレスｉｔｐｍｌ２の時刻グラフｔｇ３によると、送信元アドレスｉｔｐｍｌ２からの受信済みメールは、全て１２時に送信されている。また、送信元アドレスｉｔｐｍｌ２の曜日グラフｗｇ３によると、送信元アドレスｉｔｐｍｌ２からの受信済みメールは、必ず水曜日または金曜日かに送信されている。このため、送信元アドレスｉｔｐｍｌ２の受信済みメール

について、送信時刻、送信曜日のばらつき度合いは小さい。送信元アドレスｉｔｐｍｌ２の受信済みメールは、例えば、メールマガジン等の定期送信メールである。

【００３５】

このように、それぞれの受信済みメールから、メールヘッダーＭｈに基づいて、送信日時に含まれる送信時刻や送信曜日が特徴情報として取得される。送信時刻、送信曜日の特徴的傾向や情報のばらつき度合いは、送信者、即ち、送信元アドレスによって異なる。また、送信日時は、第３者によってなりすましし難い情報である。このため、タイトル情報に含まれる送信日時情報が特徴情報として比較の対象とされることによって、より細やかな類似判定を可能にする。

【００３６】

なお、図３の例では、送信時刻を時間単位の情報として抽出する例について述べたが、送信時刻は時間帯単位の情報として抽出されてもよい。この場合、送信時刻は、例えば、就業時間内、就業時間外の時間帯、深夜時間帯のように、時間帯の情報として取得される。

【００３７】

［特徴情報：タイトル情報におけるパターン情報］

図４は、受信済みメールに基づいて取得されるタイトル情報に含まれるパターン情報の具体例と表記揺れ辞書ｄ３の例を示す図である。同図のリストＬ２－１～Ｌ２－３は、３つの送信元アドレスｓｕｚｕｋｉ、ｔａｎａｋａ、ｉｔｐｍｌ２の受信済みメールのメールヘッダーが含むタイトル情報の具体例を示す。図２で前述したとおり、タイトル情報はＳｕｂｊｅｃｔ情報に基づいて取得される。パターン情報は、例えば、表記揺れ語句や、接頭、接尾語句を示す。

【００３８】

図４における表記揺れ辞書ｄ３は、表記揺れの発生し易い語句が予め登録される辞書である。表記揺れとは、同音、同意味の１つの語句について異なる文字表記が存在することを指す。表記揺れ語句とは、例えば、括弧の種別や記号種別、記号の全角、半角の揺れ等である。具体的に、表記揺れ語句は、「サーバー」と「サーバ」、「打ち合わせ」と「打合せ」、「－（全角）」と「-（半角）」、「『』」と「［］」等である。

【００３９】

同一送信元アドレスの正規の受信済みメールにおけるタイトル情報は、送信者の語句の選択の傾向的特性を有する。このため、同一送信元アドレスの正規の受信済みメールにおけるタイトル情報には、同一の表記揺れ語句や接頭、接尾語句が存在することが多い。また、同一送信元アドレスのメールマガジン等のタイトル情報にも、共通の接頭、接尾語句が存在する。

【００４０】

まず、送信元アドレスｔａｎａｋａからの受信済みメールのタイトル情報リストＬ２－１について説明する。タイトル情報のリストＬ２－１におけるタイトル情報Ｔ１は、表記揺れ辞書ｄ３に含まれる表記揺れ語句「－（半角ハイフン）」を有する。そこで、タイトル情報Ｔ１から半角文字のハイフンが特徴情報として抽出される。また、タイトル情報Ｔ１には、他の受信済みメールと共通の接頭語句「☆業務」が含まれる。そこで、タイトル情報Ｔ１に基づいて、接頭語句「☆業務」についても特徴情報として抽出される。他の同一送信元アドレスの受信済みメールについても、同様にして、それぞれのタイトル情報に基づいて特徴情報が抽出される。

【００４１】

続いて、送信元アドレスｓｕｚｕｋｉからの受信済みメールのタイトル情報リストＬ２－２に基づいて説明する。タイトル情報のリストＬ２－２におけるタイトル情報Ｔ２には、表記揺れ語句「打合せ」、「:（半角コロン）」、「－（半角ハイフン）」が含まれる。そこで、タイトル情報Ｔ２から、「打合せ」、「:（半角コロン）」「－（半角ハイフン）」が特徴情報として抽出される。他の受信済みメールについても、同様にして、タイトル情報に基づいて特徴情報が抽出される。

【００４２】
そして、送信元アドレスｉｔｐｍｌ２からの受信済みメールのタイトル情報リストＬ２－３に基づいて説明する。タイトル情報のリストＬ２－３におけるタイトル情報Ｔ３には、表記揺れ語句（『』）、及び、接尾語句「Ｘ通信Ｎｅｗｓ－号」が含まれる。そこで、タイトル情報Ｔ３から、「『』」、「Ｘ通信Ｎｅｗｓ－号」が特徴情報として抽出される。他の受信済みメールについても、同様にして、それぞれのタイトル情報に含まれるパターン情報が特徴情報として抽出される。
【００４３】
このように、それぞれの受信済みメールに基づいて、タイトル情報に含まれる表記揺れや接頭、接尾語句等のパターン情報が特徴情報として取得される。また、タイトル情報におけるパターン情報は、送信者によって傾向が異なるものの、第３者によってなりすまし難い情報である。このため、タイトル情報におけるパターン情報が特徴情報として比較の対象とされることによって、より綿密で柔軟な類似判定が可能となる。
【００４４】
［特徴情報：宛先情報］
図５は、受信済みメールに基づいて取得される宛先情報の具体例を示す図である。同図の宛先情報リストＬ３－１～Ｌ３－３は、３つの送信元アドレスｓｕｚｕｋｉ、ｔａｎａｋａ、ｉｔｐｍｌ２の受信済みメールのメールヘッダーが含む宛先情報の具体例を示す。宛先情報はメールヘッダーにおけるＴｏ情報、Ｃｃ情報に基づいて取得される。同図の例では、宛先を示すアドレスは、例えば、佐藤、山田Ｂ、Ｍｇｍ－ｍｌのように、識別し易い名前によって表されている。
【００４５】
具体的に、送信元アドレスｔａｎａｋａからの受信済みメールの宛先情報リストＬ３－１における一番目の宛先情報によると、Ｔｏに佐藤、Ｃｃに山田ＢとＭｇｍ－ｍｌが指定される。同様にして、送信元アドレスｓｕｚｕｋｉからの受信済みメールの宛先情報リストＬ３－２における一番目の宛先情報によると、Ｔｏに佐藤、Ｃｃに山本Ｂと鈴木（送信者自身）が指定される。
【００４６】
図５の宛先情報リストＬ４－１～Ｌ４－３は、宛先情報リストＬ３－１～Ｌ３－３が有する宛先情報のパターンを示す。例えば、送信元アドレスｔａｎａｋａの宛先情報リストＬ３－１は、Ｔｏに佐藤、Ｃｃに山田ＢとＭｇｍ－ｍｌが指定される宛先情報を多数有する。また、送信元アドレスｓｕｚｕｋｉの宛先情報リストＬ３－２は、Ｔｏに佐藤、Ｃｃに山本Ｂと鈴木が指定される宛先情報を多数有する。同様にして、送信元アドレスｉｔｐｍｌ２の宛先情報リストＬ３－３は、Ｔｏに佐藤、田中、鈴木、Ｃｃに加藤が指定される宛先情報を多数有する。
【００４７】
このように、それぞれの受信済みメールに基づいて、宛先情報に含まれるＴｏ情報、Ｃｃ情報が特徴情報として取得される。また、宛先情報には、送信元アドレス毎に、高頻度に指定される宛先情報のパターンが存在することがある。
【００４８】
なお、宛先情報リストＬ３－２において、Ｃｃに指定された鈴木は、送信者自身のメールアドレスを示す。これは、送信者自身にメールを送信する場合を示す。一般的に、標的型攻撃メールでは、Ｃｃに送信者のメールアドレスが指定され難い。送信者が詐称される標的型攻撃メールは、Ｃｃに送信元メールアドレスや他のアドレスが指定されることにより、早期に検知される可能性を高めてしまうためである。このため、本実施の形態例では、Ｔｏ情報に加えてＣｃ情報についても、特徴情報として比較判定の対象とされることによって、標的型攻撃メールの検知の精度が向上する。
【００４９】
［特徴情報：送信元ＩＰアドレス］
図６は、受信済みメールに基づいて取得される送信元サーバのＩＰアドレスの具体例Ｌ

５、及び、経由タイムゾーンの具体例Ｌ６を示す図である。同図のリストＬ５は、３つの送信元アドレスｓｕｚｕｋｉ、ｔａｎａｋａ、ｉｔｐｍｌ２の受信済みメールのメールヘッダーが含む送信元サーバのＩＰアドレスの具体例を示す。図２で前述したとおり、２行目のＲｅｃｅｉｖｅｄ情報は送信サーバの記録を示し、送信サーバのＩＰアドレス情報が含まれる。

【００５０】

例えば、図６のＩＰアドレスリストＬ５の情報Ｌ５－１は、送信元アドレスｉｔｐｍｌ２の受信済みメールが、ＩＰアドレス「ｘ０．２６．０．＊＊＊（上位３レベル）」のサーバから送信されたことを示す。この例において、ＩＰアドレスにおける＊は、比較対象外の値を示す。同様にして、ＩＰアドレスリストＬ５の情報Ｌ５－２は、送信元アドレスｓｕｚｕｋｉの受信済みメールが、ＩＰアドレス「ｘ４．６５．５．＊＊＊」のサーバから送信されたことを示す。

【００５１】

図６の例では、送信サーバのＩＰアドレスは、上位の３レベルの情報が特徴情報として取得される。このように、本実施の形態例において、送信サーバのＩＰアドレスにおける上位の所定レベルが特徴情報として取得される。例えば、「ｘ４．６５．＊＊＊．＊＊＊」のように、上位の２レベルの情報が特徴情報として取得されてもよい。送信サーバのＩＰアドレスの下位レベルは、ＩＰアドレスが動的に割り当てられる場合や、送信サーバが分散運用される場合、変動し易いためである。

【００５２】

また、会社と自宅のように複数の送信サーバによって、同一の送信元アドレスからメールが送信される場合、正規の受信済みメールであっても送信元サーバのＩＰアドレスは異なる。ただし、このような場合、送信元サーバのＩＰアドレスは、会社と自宅から送信する場合における２パターンのＩＰアドレスに限定されることになる。

【００５３】

このように、それぞれの受信済みメールに基づいて、Ｒｅｃｅｉｖｅ情報に含まれる送信元サーバのＩＰアドレス情報（この例では、上位３レベル）が特徴情報として抽出される。また、ＩＰアドレス情報は、送信元アドレス毎に、特定のパターンに限定され易い情報であるため、特徴情報として比較の対象とされることによって、標的型攻撃メールの検知の精度が向上する。

【００５４】

［特徴情報：経由タイムゾーン］

また、図６のリストＬ６は、３つの送信元アドレスｓｕｚｕｋｉ、ｔａｎａｋａ、ｉｔｐｍｌ２の受信済みメールのメールヘッダーが含む経由タイムゾーンの具体例を示す。経由タイムゾーンは、発信元の送信サーバから受信サーバまでに経由される１つまたは複数のサーバのＲｅｃｅｉｖｅｄ情報に含まれるタイムゾーン情報に基づいて取得される。

【００５５】

例えば、図６の経由タイムゾーンリストＬ６の情報Ｌ６－１は、送信元アドレスｓｕｚｕｋｉの受信済みメールが、＋０８００（中国）のタイムゾーンのサーバを経由して送信されたことを示す。同様にして、経由タイムゾーンリストＬ６の情報Ｌ６－２は、送信元アドレスｓｕｚｕｋｉの受信済みメールが、タイムゾーン＋０８００（中国）のサーバ、及び、タイムゾーン＋０９００（日本）のサーバを経由して送信されたことを示す。

【００５６】

このように、それぞれの受信済みメールに基づいて、経由サーバ情報（Ｒｅｃｅｉｖｅ情報）に含まれる各タイムゾーン情報が特徴情報として抽出される。標的型攻撃メールは、海外のサーバを経由して送信される場合がある。このため、受信対象メールの経由サーバのタイムゾーンが特徴情報として比較の対象とされることによって、標的型攻撃メールの検知の精度が向上する。

【００５７】

［特徴情報：使用メーラー］

図２で前述したとおり、使用メーラーの情報はＸ－Ｍａｉｌｅｒ情報に基づいて取得される。メール送信時に使用するメーラーのソフトウェアは、送信者によって、ほとんど変化しないことが一般的である。このため、同一の送信元アドレスの受信済みメールに基づいて取得される使用メーラーの情報は、同一である可能性が高い。このため、使用メーラー情報が特徴情報として比較の対象とされることによって、標的型攻撃メールの検知の精度が向上する。

【００５８】

［特徴情報：組み合わせ情報］

図７は、受信済みメールに基づいて取得される組み合わせ特徴情報の具体例を示す図である。同図のメールヘッダーＭｈは、図２のメールヘッダーＭｈと同一である。同図における送信元サーバのドメイン情報Ｈ７と、送信元サーバのＩＰアドレスＨ５と、送信元サーバのタイムゾーンＨ４は、メールの送信元環境に係る情報（送信元情報）を示す。同図では、これらの送信元情報のうち、複数の送信元情報を組み合わせた特徴情報（組み合わせ特徴情報）について説明する。

【００５９】

標的型攻撃メールは正規の送信者を詐称して送信される。このため、標的型攻撃メールと正規の受信済みメールとでは、組み合わせ特徴情報のうち一部の送信元情報が一致したとしても、複数の送信元情報の組み合わせは一致し難い。一方、送信元アドレスが同一の正規の受信済みメール間では、複数の送信元情報の組み合わせは一致する。また、たとえ、自宅と会社のように、複数の拠点のサーバによって同一の送信元アドレスから送信された正規の受信済みメールであっても、送信元情報の組み合わせは特定パターンに限定される。このため、特徴情報として、組み合わせ特徴情報が比較の対象とされ類似度に反映されることによって、標的型攻撃メールの検知の精度が向上する。

【００６０】

図７の表Ｌ７は、組み合わせ特徴情報として、送信元サーバのドメイン、ＩＰアドレス、タイムゾーンの組み合わせの一例を示す。例えば、表Ｌ７におけるＬ７－１には、タイムゾーンが日本であって、ドメインがａａａ．ｘｘ．ｃｏｍ、ＩＰアドレスが「ｘ２１．２３．０１.＊＊＊（上位３レベル）」の組み合わせが例示されている。このように、送信元アドレスに対して、送信元サーバのドメイン、ＩＰアドレス、タイムゾーンは特定パターンに限定される。このため、例えば、受信対象メールのメールヘッダーＭｈにおいて、送信元サーバのドメイン、ＩＰアドレス、タイムゾーンの一部が詐称されている場合であっても、組み合わせ特徴情報として一致せず類似度に反映される。

【００６１】

図３～図７で説明してきたとおり、受信済みメールのメールヘッダーＭｈに基づいて複数の特徴情報が抽出される。複数の正規の受信済みメールに基づいて、受信済みメールそれぞれについて特徴情報が抽出され特徴情報ＤＢｄ１が生成される。続いて、受信済みメールのメールヘッダーＭｈに基づいて生成された特徴情報ＤＢｄ１の一例について説明する。

【００６２】

［特徴情報ＤＢｄ１の具体例］

図８は、特徴情報ＤＢｄ１の具体例を示す図である。同図の特徴情報ＤＢｄ１は、例えば、送信元メールアドレス毎に、送信元サーバのＩＰアドレス、タイムゾーン、送信時間帯、使用メーラー、経由タイムゾーン、送信元サーバのドメイン情報、Ｔｏ情報、Ｃｃ情報、タイトル情報におけるパターン情報を特徴情報として有する。ただし、特徴情報ＤＢｄ１が有する特徴情報の例は、この例に限定されるものではない。

【００６３】

図８の例において、メールアドレスは、例えば、ＩＤ４、ＩＤ２０のように表わされる。また、同図の特徴情報ＤＢｄ１は、時間帯、及び使用メーラーの情報を、数値等の簡易な情報に変換して保持する。例えば、時間帯に係る値１は８時～１０時を、値２は１０時～１２時を示す。また、例えば、使用メーラーに係る値１はＦｏｘｍａｉｌ、値２はＴｈ

ｕｎｄｅｒｂｉｒｄを示す。
【００６４】
また、図８の特徴情報ＤＢｄ１は、サブジェクト情報についても、数値等の簡易な情報に変換して保持する。例えば、サブジェクト情報（１,１）における１番目の引数の値１は「－（ハイフン）」、２番目の引数の値１は全角文字であることを示す。同様にして、例えば、サブジェクト情報（１,２）における１番目の引数の値１は「（ハイフン）」、２番目の引数の値２は半角文字であることを示す。また、例えば、サブジェクト情報（７,７）は、接頭語句「Ｘ通信」を示す。
【００６５】
図８の特徴情報ＤＢｄ１において、ＩＤ＝１の特徴情報は、１つの受信済みメールのメールヘッダーに基づいて抽出された特徴情報の例を示す。具体的に、ＩＤ＝１の受信済みメールから抽出される特徴情報について、送信元アドレスはＩＤ４、送信元サーバのＩＰアドレスにおける上位３レベルは「ｘ８．１０３．１２４．＊＊＊」、時間帯は２（１０時～１２時）、使用メーラーは１（Ｆｏｘｍａｉｌ）であることを示す。さらに、経由タイムゾーンは＋０９００（日本）のみであり、送信元サーバのドメインは「ａａａａ.ｂｂｂｂ.ｃｏｍ」、Ｔｏに受信者に加えてアドレスＩＤ３１が指定され、Ｃｃにアドレス ＩＤ４（送信者自身）が指定されることを示す。また、タイトル情報に「－（全角のハイフン）」に加え、接頭語句「Ｘ通信」を含むことを示す。
【００６６】
このように、各受信済みメールに基づいて、それぞれ特徴情報が抽出される。図８の特徴情報ＤＢｄ１における他の受信済みメールに基づいて抽出される特徴情報についても同様である。
【００６７】
なお、図８の特徴情報ＤＢｄ１は、例えば、直近の一定期間内に受信したメール、または、直近の一定量の受信済みメールのいずれかまたは両方の受信済みメールに基づいて生成される。ただし、メールの送信元アドレスによって送信頻度が異なることから、直近の一定期間内に受信したメールに基づくと抽出対象の受信済みメールの数が十分ではないことがある。そのため、直近の一定量の受信済みメールが抽出の対象とされてもよい。
【００６８】
また、受信済みメールに基づいて抽出される特徴情報は、メール送信者の環境の変化によって変移することがある。例えば、メール送信者の転勤や部署異動が発生した場合、送信元サーバのＩＰアドレスや、ドメイン情報、宛先情報等の特徴情報が変化する可能性がある。そこで、特徴情報の変化が適宜反映されるように、特徴情報ＤＢｄ１は、定期的に更新されることが望ましい。これにより、適宜更新された特徴情報ＤＢｄ１に基づいて、受信対象メールと受信済みメールとの類似度がタイムリーに判定される。
【００６９】
続いて、本実施の形態例におけるメールチェック処理の流れについて、フローチャート図に基づいて説明する。
【００７０】
［フローチャート：類似判定処理の流れ］
図９は、本実施の形態例のメールチェック装置２００における類似判定処理の流れを説明するフローチャート図である。同図の処理において、メール特徴抽出部１３によって、例えば、直近の一か月の正規の受信済みメールに基づいて、特徴情報ＤＢｄ１が生成されているものとする。
【００７１】
メールチェック部１０のメール受信部が受信対象のメールを検出すると（Ｓ１１）、メールチェック部１０の標的型攻撃メール検出部１４は、受信対象メールと送信元アドレスが同一、即ち、同一送信者の受信履歴があるか否かを判定する（Ｓ１２）。受信履歴は、特徴情報ＤＢｄ１が特徴情報の抽出の対象とした受信メールのリストを示す。同一送信者の受信履歴がある場合（Ｓ１２のＹＥＳ）、標的型攻撃メール検出部１４は、受信対象の

メールと、同一送信者のＮ個（ｉ＝１～Ｎ）の受信済みのメール（ｈｉ～ｈＮ）とをそれぞれ比較する。

【００７２】

図９のフローチャート図では、受信対象メールがＮ個の受信済みメールそれぞれと比較され、受信対象メールとの総一致量値ｓｉｍが基準一致量値ｔｈを超える受信済みメール（類似受信済みメール）の数ＳＮが類似度として判定される。そして、類似受信済みメールの数ＳＮが基準類似度ｍｍを越える場合に、受信対象メールが正規メールと判定される。基準一致量値ｔｈ、及び、基準数ｍｍは、予め定められる（Ｓ１３）。また、初め、類似受信対象メールの数ｍｍは０個に初期化されている（Ｓ１４）。

【００７３】

標的型攻撃メール検出部１４は、受信対象メールと送信元アドレスが同一の複数の受信済みメールとをそれぞれ比較して、総一致量値ｓｉｍを算出する（Ｓ１５）。続いて、標的型攻撃メール検出部１４は、算出した総一致量値ｓｉｍが基準一致量値ｔｈを超えるか否かを判定する（Ｓ１６）。基準一致量値ｔｈを超える場合（Ｓ１６のＹＥＳ）、標的型攻撃メール検出部１４は、基準一致量値ｔｈを超える類似受信済みメールカウントＳＮをカウントアップする（Ｓ１７）。次に、標的型攻撃メール検出部１４は、類似受信済みメール数ＳＮが基準類似度ｍｍを超えるか否かを判定する（Ｓ１８）。

【００７４】

類似受信済みメール数ＳＮが基準類似度を超えない場合（Ｓ１８のＮＯ）、または、総一致量値ｓｉｍが基準一致量値ｔｈを超えない場合（Ｓ１６のＮＯ）、標的型攻撃メール検出部１４は、変数ｉをインクリメントし、送信元アドレスが同一の次の受信済みメールを判定の対象とする（Ｓ１９）。判定済みの受信済みメール数（ｉ）がＮ個に達していない場合（Ｓ２０のＮＯ）、標的型攻撃メール検出部１４は、同様にして、受信済みメールと受信対象メールとの類似度を算出し（Ｓ１５）、類似度の判定を行う。

【００７５】

一方、類似受信済みメール数ＳＮが基準類似度ｍｍを超える場合（Ｓ１８のＹＥＳ）、正規の受信済みメールとの特徴情報の類似度が高いことから、正規メールであると判定される。そこで、標的型攻撃メール検出部１４は、当該受信対象メールの情報を受信履歴情報に追加し、特徴情報ＤＢｄ１の抽出対象とする（Ｓ２４）。そして、メールサーバは、受信対象メールを受信する（Ｓ２５）。

【００７６】

工程Ｓ１２の判定に戻り、同一送信者の受信履歴がない場合（Ｓ１２のＮＯ）、特徴情報ＤＢｄ１に同一送信者の受信済みメールの特徴情報がないことを示す。即ち、受信対象メールが新規送信者からメールであることを示す。そこで、アラート発生部１５は、新規送信者からのメールであり、ユーザによる確認が必要なメールであるとして警告情報を表示する（Ｓ２１）。このように、本実施の形態例において新規送信者からのメールの場合、特徴情報ＤＢに受信済みメールの特徴情報がないことから警告情報が表示されるが、新規送信者からのメールに対する警告情報の表示の抑制方法については別途検討される。

【００７７】

また、工程Ｓ２０において、判定済みの受信済みメール数がＮ個に達した場合（Ｓ２０のＹＥＳ）、類似受信済みメール数ＳＮが基準類似度ｍｍを超えない間に（Ｓ１８のＮＯ）、即ち、正規メールであると判定されない間に、送信元アドレスが同一の全ての受信済みメールについて判定されたことを示す。そこで、アラート発生部１５は、受信対象メールと類似する受信済みメールが存在していない旨の警告情報を表示する（Ｓ２２）。

【００７８】

受信対象メールが新規送信者からのメールである旨の警告情報（Ｓ２１）、または、受信対象メールが類似度を満たしていない旨の警告情報（Ｓ２２）が表示画面等に表示されると、ユーザは、受信対象メールの受信を許可するか否かを判定する（Ｓ２３）。受信対象メールの受信が許可された場合（Ｓ２３のＹＥＳ）、標的型攻撃メール検出部１４は、当該受信対象メールの情報を受信履歴情報に追加し（Ｓ２４）、メールサーバは受信対象

メールを受信する（Ｓ２５）。一方、受信対象メールの受信が許可されない場合（Ｓ２３のＮＯ）、標的型攻撃メール検出部１４は、受信対象メールを破棄する（Ｓ２６）。ただし、標的型攻撃メール検出部１４は、受信対象メールに対して安全化処理を行った上で、安全化処理後の受信対象メールを受信してもよい。安全化処理とは、例えば、添付ファイルの削除や本文への警告文の追加、本文におけるリンク削除等である。

【００７９】

続いて、図９のフローチャート図における工程Ｓ１５における総一致量値ｓｉｍの算出処理について、次のフローチャート図に基づいて説明する。

【００８０】

［フローチャート：総一致量値ｓｉｍの算出処理の流れ］

図１０は、受信対象メールと受信済みメールとの特徴情報の総一致量値ｓｉｍの算出処理を説明するフローチャート図である。初めに、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれるＤａｔｅ情報に基づいて、送信日時情報を取得する（Ｓ３２）。続いて、標的型攻撃メール検出部１４は、取得した送信日時情報における送信曜日が、比較対象の受信済みメールから抽出された送信曜日と一致するか否かを判定する（Ｓ３３）。

【００８１】

標的型攻撃メール検出部１４は、一致する場合（Ｓ３３のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ３４）、一致しない場合（Ｓ３３のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。続いて、標的型攻撃メール検出部１４は、取得した送信日時情報における送信時間帯が、比較対象の受信済みメールにおける送信時間帯と一致するか否かを判定する（Ｓ３５）。標的型攻撃メール検出部１４は、一致する場合（Ｓ３５のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ３６）、一致しない場合（Ｓ３５のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８２】

次に、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれるＳｕｂｊｅｃｔ情報に基づいて、タイトル情報におけるパターン情報を取得する（Ｓ３８）。続いて、標的型攻撃メール検出部１４は、取得したパターン情報が、比較対象の受信済みメールから抽出されたパターン情報と一致するか否かを判定する（Ｓ３８）。標的型攻撃メール検出部１４は、一致する場合（Ｓ３８のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ３９）、一致しない場合（Ｓ３８のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８３】

次に、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれるＴｏ、Ｃｃ情報に基づいて、宛先情報を取得する（Ｓ４０）。続いて、標的型攻撃メール検出部１４は、取得した宛先情報が、比較対象の受信済みメールから抽出された宛先情報と一致するか否かを判定する（Ｓ４１）。標的型攻撃メール検出部１４は、一致する場合（Ｓ４１のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ４２）、一致しない場合（Ｓ４１のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８４】

次に、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれるＸ－ｍａｉｌｅｒ情報に基づいて、使用メーラー情報を取得する（Ｓ４３）。続いて、標的型攻撃メール検出部１４は、取得した使用メーラー情報が、比較対象の受信済みメールから抽出された使用メーラー情報と一致するか否かを判定する（Ｓ４４）。標的型攻撃メール検出部１４は、一致する場合（Ｓ４４のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ４５）、一致しない場合（Ｓ４４のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８５】

次に、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる送信元サーバのＲｅｃｅｉｖｅｄ情報を取得し（Ｓ４６）、当該Ｒｅｃｅｉｖｅｄ情報に基づいて送信元サーバのＩＰアドレス情報を抽出する（Ｓ４７）。続いて、標的型攻撃メール検出部１４は、取得したＩＰアドレス情報が、比較対象の受信済みメールから抽出され

たＩＰアドレス情報と一致するか否かを判定する（Ｓ４８）。標的型攻撃メール検出部１４は、一致する場合（Ｓ４８のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ４９）、一致しない場合（Ｓ４８のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８６】

次に、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる経由送信サーバのＲｅｃｅｉｖｅｄ情報に基づく経由タイムゾーン情報が、比較対象の受信済みメールから抽出される経由タイムゾーン情報と一致するか否かを判定する（Ｓ５０）。標的型攻撃メール検出部１４は、一致する場合（Ｓ５０のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ５１）、一致しない場合（Ｓ５０のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８７】

そして、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる送信元情報を示す送信元サーバのドメイン、ＩＰアドレス、タイムゾーンの組み合わせ情報（組み合わせ特徴情報）が、比較対象の受信済みメールから抽出された組み合わせ特徴情報と一致するか否かを判定する（Ｓ５２）。標的型攻撃メール検出部１４は、一致する場合（Ｓ５２のＹＥＳ）、変数Ｓｃｏｒｅに値５を加算し（Ｓ５３）、一致しない場合（Ｓ５２のＮＯ）、変数Ｓｃｏｒｅに加算を行わない。

【００８８】

このように、累計値である変数Ｓｃｏｒｅが、受信対象メールと１つの受信済みメールとの総一致量値ｓｉｍとして算出される。そして、総一致量値ｓｉｍは、基準一致量値ｔｈと比較され（図９のＳ１６）、類似受信済みメールか否かが判定される。

【００８９】

このように、本実施の形態例におけるメールチェック部１０は、正規の複数の受信済みメールについて、受信済みメールのメールヘッダーが含む複数の特徴情報を抽出して、特徴情報ＤＢｄ１を生成する。そして、メールチェック部１０は、受信する対象メールのメールヘッダーが含む特徴情報と、特徴情報ＤＢｄ１から参照される受信対象メールと送信元アドレスが同一の複数の受信済みメールの複数の特徴情報との類似度を判定する。そして、メールチェック部１０は、類似度が基準類似度未満、即ち、類似度が低い受信対象メールを非正規メールである可能性があると判定し、警告メールとしてユーザに通知する。また、各受信対象メールについて、複数の特徴情報について比較されるため、警告メール検出の精度が向上する。なお、図１０のフローチャート図で比較される特徴情報の項目例は、一例である。メールヘッダーに含まれる複数の特徴情報のうち、いずれの特徴情報が比較され、類似度が算出されてもよい。

【００９０】

なお、図９、図１０のフローチャート図では、受信対象メールと送信元アドレスが同一であって、受信対象メールとの総一致量値ｓｉｍが基準一致量値ｔｈを超える受信済みメール（類似受信済みメール）の数ＳＮが類似度として算出され判定される。これにより、類似した受信済みメールが基準数分（基準類似度）存在する場合に、正規メールであると判定される。ただし、この例に限定されるものではない。例えば、メールチェック部１０は、送信元アドレスが同一であって、正規の所定数の受信対象メールとの総一致量値ｓｉｍの総累計値を類似度として算出し、基準類似度と比較してもよい。

【００９１】

［具体例］

図９のフローチャート図における具体例について、図８の特徴情報ＤＢｄ１に基づいて説明する。この例において、基準一致量値ｔｈは値２０、基準類似度は１０であるものとする。

【００９２】

また、受信対象メールの送信元アドレスがＩＤ４であり、送信元サーバのＩＰアドレスにおける上位３レベルが「ｘ８．１０３．１２４．＊＊＊」、時間帯が３、使用メーラーの識別値がＦｏｘｍａｉｌ、経由タイムゾーンが＋０９００（日本）のみであり、送信元

サーバのドメインが「ａａａａ.ｂｂｂｂ.ｃｏｍ」、Ｔｏに受信者に加えてＩＤ３１、ＣｃにＩＤ４が指定される場合を示す。また、タイトル情報に、半角のハイフンが含まれる。また、受信対象メールの送信曜日は、例えば、火曜日であるものとする。

【００９３】

例えば、図８の特徴情報ＤＢｄ１における、受信対象メールと送信元アドレスＩＤ４が同一の受信済みメールが比較判定の対象とされる場合を例示する。図８の特徴情報ＤＢｄ１において、送信元アドレスがＩＤ４である受信済みメールの特徴情報は、ＩＤ＝１、３の特徴情報を示す。なお、図示していないが、図８の特徴情報ＤＢｄ１には、送信元アドレスがＩＤ４である９０個分の受信済みメールの特徴情報が抽出されているものとする。

【００９４】

まず、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる送信日時情報を取得し（Ｓ３２）、取得した送信日時情報における送信曜日が、ＩＤ＝１の受信済みメールから抽出された送信曜日と一致するか否かを判定する（Ｓ３３）。図８の特徴情報ＤＢｄ１の例において、送信曜日は抽出されていないが、例えば、ＩＤ＝１の受信済みメールの送信曜日は、金曜日であるものとする。この場合、送信曜日が一致しないため（Ｓ３３のＮＯ）、変数Ｓｃｏｒｅは０のままである。

【００９５】

続いて、標的型攻撃メール検出部１４は、送信時間帯について比較する（Ｓ３５）。この例において、受信対象メールの送信時間帯は３、ＩＤ＝１の受信済みメールの送信時間帯は２であるため、送信時間帯は一致しない（Ｓ３５のＮＯ）。そのため、変数Ｓｃｏｒｅは値０のままである。次に、標的型攻撃メール検出部１４は、タイトル情報におけるパターン情報を取得し（Ｓ３７）、比較する（Ｓ３８）。この例において、受信対象メールとＩＤ＝１の受信済みメールのタイトル情報は一致しないため（Ｓ３８のＮＯ）、変数Ｓｃｏｒｅは値０のままである。

【００９６】

続いて、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる宛先情報を取得し（Ｓ４０）、比較する（Ｓ４１）。この例において、受信対象メールとＩＤ＝１の受信済みメールの宛先情報は一致するため（Ｓ３８のＹＥＳ）、Ｓｃｏｒｅに値５が加算される。続いて、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる使用メーラー情報を取得し（Ｓ４３）、比較する（Ｓ４４）。この例において、受信対象メールとＩＤ＝１の受信済みメールの使用メーラー情報は一致するため（Ｓ４４のＹＥＳ）、Ｓｃｏｒｅの値は１０に増加する。

【００９７】

続いて、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる送信サーバのＩＰ情報を抽出し（Ｓ４６、Ｓ４７）、比較する（Ｓ４８）。この例において、受信対象メールの送信元サーバと、ＩＤ＝１の受信済みメールの送信元サーバのＩＰアドレス（上位３レベル）は一致するため（Ｓ４８のＹＥＳ）、変数Ｓｃｏｒｅの値は１５に増加する。続いて、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる経由タイムゾーン情報を比較するが（Ｓ５０）、一致することにより（Ｓ５０のＹＥＳ）、変数Ｓｃｏｒｅの値は２０に増加する。

【００９８】

そして、標的型攻撃メール検出部１４は、受信対象メールのメールヘッダーに含まれる送信元サーバのドメイン、送信元サーバのＩＰアドレス、送信元サーバのタイムゾーンの組み合わせを比較する（Ｓ５２）。この例において、受信対象メールの送信元サーバのドメインは、ａａａａ.ｂｂｂｂ.ｃｏｍ、ＩＰアドレスは「ｘ８．１０３．１２４．＊＊＊（上位３レベル）」、タイムゾーンは＋０９００（日本）である。このため、受信対象メールの組み合わせ特徴情報と、図８のＩＤ＝１の組み合わせ特徴情報とは一致する。このため、変数Ｓｃｏｒｅの値は２５に増加する。

【００９９】

そして、算出された変数Ｓｃｏｒｅの値、即ち、総一致量値ｓｉｍが、基準一致致量値

ｔｈ（この例では、２０）と比較される。この例において、総一致量値ｓｉｍ（変数Ｓｃｏｒｅ＝２５）が基準一致量値ｔｈを超えるため（Ｓ１６のＹＥＳ）、類似受信済みメールカウントＳＮがインクリメントされる。続いて、受信対象メールが特徴情報ＤＢｄ１におけるＩＤ＝３の受信済みメールと比較され、同様にして総一致量値ｓｉｍが算出され、基準一致量値ｔｈと比較される。この判定が、特徴情報ＤＢｄ１に格納された、送信元アドレスがＩＤ４である他の受信済みメールについても行われる。そして、類似受信済みメールカウントＳＮが基準類似度ｍｍ（この例では、１０）を超えると（Ｓ１８のＹＥＳ）、正規メールであると判定され受信が許可される（Ｓ２４、Ｓ２５）。

【０１００】

また、例えば、具体例において、受信対象メールのメールヘッダーに含まれる使用メーラー情報が一致しない場合（Ｓ４４のＮＯ）、変数Ｓｃｏｒｅの値は２０となる。このとき、変数Ｓｃｏｒｅの値（総一致量値ｓｉｍ）は、基準一致致量値ｔｈ（この例では、２０）を超えないため（Ｓ１６のＮＯ）、類似受信済みメールカウントＳＮはインクリメントされない。このように、複数の特徴情報の一致結果に基づいて、受信対象メールの類似度が判定される。基準一致致量値ｔｈは、比較対象の特徴情報の数やばらつき度合い等に基づいて適切に設定される。

【０１０１】

なお、図１０のフローチャート図では、いずれの特徴情報の内容が一致した場合であっても、変数Ｓｃｏｒｅに常に値５が加算される場合における総一致量値ｓｉｍの算出方法について述べた。ただし、特徴情報が一致したときに加算される加算値は、特徴情報によって変更されてもよい。続いて、特徴情報に対してそれぞれ設定され、特徴情報が一致したときにおける加算値を変化させる重み係数について説明する。

【０１０２】

［重み係数］

同一送信元アドレスの複数の受信済みメールの特徴情報のばらつきの度合いは、特徴情報の種別によって異なる。つまり、特徴情報の種別によって、内容が固定化またはパターン化されやすい特徴情報についてはばらつきの度合いが小さく、内容がパターン化されにくい特徴情報についてはばらつきの度合いが大きい。

【０１０３】

例えば、使用メーラーや、送信元サーバのドメイン、タイムゾーン、経由タイムゾーン等の特徴情報等は送信環境に依存する情報であることから、同一送信元アドレスの受信済みメール間での特徴情報のばらつき度合いは小さい。つまり、これらの特徴情報については、内容が１つまたは特定パターンに限定されるため、同一の送信元アドレスの正規の受信済みメール間で一致し易い。また、送信元サーバのＩＰアドレスについては、送信者によってばらつき度合いが異なる。例えば、自宅や会社、出張先等の複数の拠点から送信する送信者の送信元サーバのＩＰアドレスのばらつき度合いは大きく、会社等の特定の拠点から送信する送信者の送信元サーバのＩＰアドレスのばらつき度合いは小さい。

【０１０４】

一方、タイトル情報におけるパターン情報や定型語句、送信時間帯、送信曜日、宛先情報等は、送信者によって人為的に指定される情報であることから、受信済みメール間での特徴情報のばらつき度合いが大きい。つまり、これらの特徴情報については、同一の送信元アドレスの正規の受信済みメール間で一致し難い。

【０１０５】

そこで、特徴情報の種別によって重み係数が付与される。重み係数は、例えば、図１０のフローチャート図におけるＳｃｏｒｅに乗算される。つまり、重み係数が３である場合、当該特徴情報に対応する一致量値は１５（＝５＊３）となる。ただし、この例に限定されるものではなく、例えば、一致量値８（＝５＋３）のように、重み係数がＳｃｏｒｅに加算されてもよい。類似度に対して重み係数が反映されればいずれの方法でもよい。

【０１０６】

そこで、例えば、特徴情報のうち、ばらつき度合いが小さく内容が所定パターンに特定

され易い特徴情報については重み係数が大きく、ばらつき度合いが大きく内容が特定され難い特徴情報については重み係数が小さく設定される。この結果、受信対象メールと各受信済みメールとの総一致量値において、重み係数が大きい特徴情報に対応する一致量値の比重は大きくなり、重み係数が小さい特徴情報に対応する一致量値の比重は小さくなる。これにより、受信済みメールにおける内容のばらつき度合いが小さく、信頼性の高い特徴情報の比較結果がより顕著に類似度に反映され、標的型攻撃メール候補がより的確に検出される。

【０１０７】

また、前述した送信元サーバのＩＰアドレスの例のように、送信者によって、受信済みメールにおける特徴情報のばらつき度合いの傾向は異なる。このため、重み係数は、同一送信元アドレスの受信済みメールの特徴情報のばらつき度合いに基づいて、送信元アドレス単位に調整されてもよい。つまり、送信者毎のばらつき度合いの相違は、各送信者それぞれについて設定される各特徴情報の重み係数によって吸収され、対応される。なお、重み係数は、直近の所定量の受信済みメールに基づいて、適宜更新される。これにより、送信者毎の特徴情報の重み係数は、各送信者の受信済みメールの特徴情報のばらつき度合いに基づいて最適な値に設定される。

【０１０８】

これにより、同一送信元アドレスの受信済みメールにおける内容のばらつき度合いの小さい特徴情報の比較結果がより顕著に類似度に反映されることにより、標的型攻撃メール候補がより高精度に検出される。例えば、メールマガジン等のように特定の送信日時に送信されるメールの送信元アドレスについては、送信日時（送信曜日、送信時刻）に係る特徴情報の重み係数が大きな値に設定される。

【０１０９】

また、例えば、特徴情報のうち、送信元メールサーバのＩＰアドレス、送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別を示す送信環境に依存する第１特徴情報群の重み係数は大きく、送信曜日、送信時刻、タイトル情報、宛先情報を示す送信者に因る人為的な第２特徴情報群の重み係数は小さく設定される。正規の受信済みメールにおいて、第１特徴情報群の特徴情報の内容は特定のパターンに限定され易くばらつき度合いが小さく、第２特徴情報群の特徴情報の内容は、傾向が限定され難くばらつき度合いが大きいという特性を有するためである。

【０１１０】

そこで、受信対象メールと各受信済みメールとの総一致量値について、第１特徴情報群の特徴情報の重み係数を大きくして対応する一致量値の比重を大きくすると共に、第２特徴情報群の特徴情報の重み係数を小さくして対応する一致量値の比重を小さくする。これにより、受信済みメールにおける送信環境に依存する特徴情報の比較結果がより顕著に類似度に反映されるため、標的型攻撃メール候補がより高精度に検出可能となる。

【０１１１】

また、送信元メールサーバのＩＰアドレス、ドメイン情報、タイムゾーン情報を含む送信元情報群のうち複数の送信元情報を組み合わせた特徴情報（組み合わせ特徴情報）の重み係数は、送信元情報それぞれ示す各特徴情報の重み係数より大きい値に設定されてもよい。例えば、組み合わせ特徴情報の重み係数は、送信元メールサーバのＩＰアドレス、ドメイン情報等を示す個々の特徴情報の重み係数より大きな値に設定される。

【０１１２】

標的型攻撃メールは正規の送信者を詐称して送信されるため、標的型攻撃メールと正規の受信済みメールとでは、組み合わせ特徴情報のうち一部の送信元情報が一致したとしても、複数の送信元情報の組み合わせは一致し難い。そのため、信頼性の高い特徴情報を示す組み合わせ特徴情報の重み係数がより大きい値に設定され、その比較結果が類似度により顕著に反映されることにより、標的型攻撃メール候補がより高精度に検出される。

【０１１３】

図１１は、送信元アドレス毎の特徴情報の重み係数情報ｄ２の一例を示す図である。こ

の例では、送信元アドレス毎に特徴情報それぞれについて重み係数が付与されている。ただし、すべての送信元アドレスに対して、共通の値として、特徴情報それぞれについて重み係数が付与されてもよい。

【０１１４】

図１１の重み係数の例において、具体的に、送信元アドレスがＩＤ４の送信元サーバのＩＰアドレス、タイムゾーン、ドメイン、使用メーラー、経由タイムゾーンには重み係数１０が付与されている。つまり、送信環境に依存する特徴情報には、大きい値の重み係数（この例では、１０）が付与されている。一方、送信時間帯には重み係数７が、宛先のＴｏ、Ｃｃには重み係数５が、タイトル情報のパターン情報には重み係数３が付与されている。つまり、人為的に指定されることにより内容が変動しやすい特徴情報にはより小さい値の重み係数（この例では、３、５、７）が、内容のばらつき度合いが小さく信頼性の高い特徴情報にはより大きな値の重み係数（この例では、１０）が付与されている。

【０１１５】

また、送信元アドレスがＩＤ２０の各特徴情報の重み係数は、パターン情報以外大きな値が付与されている。これは、例えば、メールマガジン等の送信元アドレスの重み係数を示す。例えば、メールマガジン等では、送信時間帯や宛先情報等の特徴情報の内容が固定的であるため、大きい重み係数が付与される。なお、図１１には図示していないが、組み合わせ特徴情報についても同様に重み係数が付与される。

【０１１６】

これにより、受信対象メールが正規のメールについて、タイトル情報におけるパターン情報や、宛先情報が一致しない場合であっても、重み係数の大きい送信元サーバのＩＰアドレス等の特徴情報が一致することにより、総一致量値が基準一致量値ｔｈを超える。これにより、受信対象メールが正規メールとして判定されることになる。

【０１１７】

また、重み係数は、特徴情報ＤＢｄ１と同様にして、直近の正規の受信済みメールに基づいて適宜更新されることが望ましい。これにより、特徴情報のばらつき度合いが変化した場合であっても、各特徴情報に対応する一致量値の総一致量値における比重が調整されることにより、受信対象メールの類似度がタイムリーに判定される。

【０１１８】

以上のようにして、本実施の形態例におけるメールチェック装置２００は、正規の複数の受信済みメールについて、受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報ＤＢｄ１を生成する。そして、メールチェック装置２００は、受信対象メールのメールヘッダーが含む第１の複数の特徴情報と、特徴情報ＤＢｄ１から参照される当該受信対象メールと送信元アドレスが同一の複数の受信済みメールの第２の複数の特徴情報との類似度が基準類似度未満の場合、受信対象メールを警戒メール候補として検出する。

【０１１９】

標的型攻撃メールは正規な送信者を詐称して送信されるため人による判定は困難であることが多いが、メールチェック装置２００は、受信対象のメールと、送信者が同一の正規の受信済みメールとの類似度を判定し、類似度の低い受信対象メールを警告通知することにより、標的型攻撃メールの候補を自動検出することができる。これにより、メールチェック装置２００は、メーラーによる受信前に、標的型攻撃メールの候補を検出することができ、被害の回避を可能にする。

【０１２０】

また、本実施の形態例におけるメールチェック装置２００は、ＰＣや携帯端末等のクライアント環境での動作を可能にするため、サーバ環境で各ユーザの受信履歴情報や特徴情報を管理する必要がない。また、特徴情報は、受信済みメールのメールヘッダーに含まれる情報のうち一部の情報であるため、特定の個人が特定され難い。これにより、個人情報保持の観点においても有効である。

【０１２１】

また、本実施の形態例におけるメールチェック装置２００において、複数の特徴情報は、送信元メールサーバのＩＰアドレス、前記送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別を有する第１特徴情報群のうちいずれか１つまたは複数を有する。これにより、メールチェック装置２００は、メールの送信環境に依存することによりばらつき度合いの小さい特徴情報に基づいて、受信対象メールと受信済みメールの類似度をより高精度に判定することができる。

【０１２２】

また、本実施の形態例におけるメールチェック装置２００において、複数の特徴情報は、さらに、送信曜日、送信時刻、タイトル情報、宛先情報を有する第２特徴情報群のうちいずれか１つまたは複数を有する。これにより、メールチェック装置２００は、メールの送信環境に依存する特徴情報に加えて、さらに、送信者による人為的な傾向的特徴を有する特徴情報に基づいて、受信対象メールと受信済みメールの類似度をより綿密に判定することができる。これにより、メールチェック装置２００は、なりすまし難い特徴情報を比較対象とすることにより、メールヘッダーにおける送信環境に係る特徴情報が詐称されている場合であっても、標的型攻撃メール候補を検出することができる。

【０１２３】

また、本実施の形態例におけるメールチェック装置２００において、特徴情報ＤＢｄ１は、特徴情報毎に重み係数を有し、重み係数は、複数の受信済みメールについて、特徴情報のばらつき度合いが第１の度合い（ばらつき度合いが小さい）の場合に第１の重み係数（例えば１０）に設定され、特徴情報のばらつき度合いが第１の度合いより大きい第２の度合いの（ばらつき度合いが大きい）場合に第１の重み係数より小さい第２の重み係数（例えば５）に設定される。そして、メールチェック装置２００は、検出工程において、受信対象メールの複数の特徴情報と、送信元アドレスが同一の複数の受信済みメールにおける複数の特徴情報とがそれぞれ一致するか否かを判定し、一致した特徴情報に対応する重み係数が反映された各一致量値の累計加算値に応じて類似度を求める。

【０１２４】

このように、ばらつき度合いが小さく内容が所定パターンに特定され易い特徴情報については重み係数が大きく、ばらつき度合いが大きく内容が特定され難い特徴情報については重み係数が小さく設定される。この結果、受信対象メールと各受信済みメールとの総一致量値において、重み係数が大きい特徴情報に対応する一致量値の比重は大きくなり、重み係数が小さい特徴情報に対応する一致量値の比重は小さくなる。これにより、受信済みメールにおける内容のばらつき度合いの小さい特徴情報の比較結果がより顕著に類似度に反映され、標的型攻撃メール候補がより高精度に検出可能となる。

【０１２５】

または、本実施の形態例におけるメールチェック装置２００において、送信元メールサーバのＩＰアドレス、送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別を示す第１特徴情報群の特徴情報の重み係数は第１の重み係数（例えば、１０）に設定される。そして、送信曜日、送信時刻、タイトル情報、宛先情報を示す第２特徴情報群の特徴情報の重み係数は第１の重み係数より小さい第２の重み係数（例えば、５）に設定される。そして、メールチェック装置２００は、検出工程において、受信対象メールの複数の特徴情報と、送信元アドレスが同一の複数の受信済みメールにおける複数の特徴情報とがそれぞれ一致するか否かを判定し、一致した特徴情報に対応する重み係数が反映された各一致量値の累計加算値に応じて類似度を求める。

【０１２６】

このように、送信環境に依存することにより内容が特定パターンに限定され易い特徴情報については重み係数が大きく、送信者に依存することにより内容が特定され難い特徴情報については重み係数が小さく設定される。この結果、受信対象メールと各受信済みメールとの総一致量値において、内容が所定パターンに特定され易い特徴情報に対応する一致量値の比重は大きくなり、内容が特定され難い特徴情報に対応する一致量値の比重は小さくなる。これにより、受信対象メールが正規メールである場合、各受信済みメールとの特

徴情報の総一致量値が大きくなり、より高精度に標的型攻撃メールが検出される。

【０１２７】

さらに、本実施の形態例におけるメールチェック装置２００において、特徴情報は、送信元メールサーバのＩＰアドレス、ドメイン情報、タイムゾーン情報を含む送信元情報群のうち複数の送信元情報を組み合わせた組み合わせ特徴情報を含み、組み合わせ特徴情報の重み係数は送信元情報それぞれ示す各特徴情報の重み係数より大きい第３の重み係数（例えば、１０より大きい値）に設定される。そして、メールチェック装置２００は、検出工程において、組み合わせ特徴情報である複数の送信元情報の組み合わせが一致するか否かを判定する。

【０１２８】

標的型攻撃メールは正規の送信者を詐称して送信されるため、標的型攻撃メールと正規の受信済みメールとでは、送信環境に依存する特徴情報の組み合わせは一致し難い。そこで、メールチェック装置２００は、送信元情報を示す情報の組み合わせである組み合わせ特徴情報を比較し類似度に反映することによって、より高精度に標的型攻撃メールを検出することを可能にする。このため、メールチェック装置２００は、信頼性の高い特徴情報である組み合わせ特徴情報の重み係数を大きくし、その比較結果をより顕著に類似度に反映することにより、標的型攻撃メール候補の検出をより高精度にする。

【０１２９】

なお、特徴情報毎の重み係数は、送信元アドレス毎に設けられてもよい。このとき、送信元アドレス毎の重み係数は、送信元アドレスが同一の複数の受信済みメールにおける特徴情報のばらつき度合いに基づいて決定される。これにより、送信元アドレスによって特徴情報のばらつき度合いの傾向性が異なる場合、受信対象メールが、同一の送信元アドレスの受信済みメールと同様の特徴を有するか否かがより高精度に判定可能となる。つまり、同一送信元アドレスの受信済みメールにおける特徴情報のばらつき度合いの特性が的確に類似度に反映されることにより、標的型攻撃メール候補がより高精度に検出される。

【０１３０】

また、本実施の形態例のメールチェック装置２００は、定期的に、直近の一定期間内の受信済みメール、または、直近の一定量の受信済みメールのいずれかまたは両方における正規の複数の受信済みメールについて、受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報ＤＢｄ１を生成、更新する。これにより、メールチェック装置２００は、正規の受信済みメールに基づいて特徴情報ＤＢｄ１を適宜更新することにより、受信対象メールの類似度を最新の特徴情報に従ってタイムリーに判定することができる。このため、正規の送信者の送信環境が変化した場合でも、直近の正規の受信済みメールに基づいて特徴情報ＤＢｄ１が適宜更新されることにより、標的型攻撃メール候補が常時、適切に検出される。

【０１３１】

なお、本実施の形態例におけるメールチェック装置２００において、類似度は、一致量値の累計加算値が基準一致量値を超える受信済みメール数を示す。つまり、メールチェック装置２００は、受信対象メールと特徴情報の一致度合いが高い受信済みメールが基準の数分、検出された場合に正規メールと判定する。これにより、複数の受信済みメールとの類似度に基づいて判定されることから、メールチェック装置２００は、より高精度に標的型攻撃メールを検出することができる。

【０１３２】

なお、本実施の形態例におけるメールチェック処理は、コンピュータ読み取り可能な記録媒体にプログラムとして記憶され、当該プログラムをコンピュータが読み出して実行することによって行われてもよい。

【０１３３】

以上の実施の形態をまとめると、次の付記のとおりである。

【０１３４】

（付記１）

新たに受信する対象メールから警戒メール候補を検出するメールチェック方法であって、
正規の複数の受信済みメールについて、前記受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報データーベース（以下、ＤＢ）を生成する特徴情報ＤＢ生成工程と、
受信対象メールの前記メールヘッダーが含む第１の複数の特徴情報と、前記特徴情報ＤＢから参照される当該受信対象メールと送信元アドレスが同一の前記複数の受信済みメールの第２の複数の特徴情報との類似度が基準類似度未満の場合、前記受信対象メールを警戒メール候補として検出する検出工程と、を有するメールチェック方法。

【０１３５】
（付記２）
付記１において、
前記第１、第２の複数の特徴情報は、送信元メールサーバのＩＰアドレス、前記送信元メールサーバのドメイン情報、タイムゾーン情報、送信メーラー種別を有する第１特徴情報群のうちいずれか１つまたは複数を有するメールチェック方法。

【０１３６】
（付記３）
付記２において、
前記第１、第２の複数の特徴情報は、さらに、送信曜日、送信時刻、タイトル情報におけるパターン情報、宛先情報を有する第２特徴情報群のうちいずれか１つまたは複数を有するメールチェック方法。

【０１３７】
（付記４）
付記１乃至３のいずれかにおいて、
前記特徴情報ＤＢは、前記特徴情報毎に重み係数を有し、
前記重み係数は、前記複数の受信済みメールについて、前記特徴情報のばらつき度合いが第１の度合いの場合に第１の重み係数に設定され、前記特徴情報のばらつき度合いが前記第１の度合いより大きい第２の度合いの場合に前記第１の重み係数より小さい第２の重み係数に設定され、
前記検出工程は、前記第１の複数の特徴情報と前記第２の複数の特徴情報とがそれぞれ一致するか否かを判定し、一致した特徴情報に対応する前記重み係数が反映された各一致量値の累計加算値に応じて前記類似度を求めるメールチェック方法。

【０１３８】
（付記５）
付記３において、
前記特徴情報ＤＢは、前記特徴情報毎に重み係数を有し、
前記第１特徴情報群の特徴情報の重み係数は第１の重み係数に設定され、前記第２特徴情報群の特徴情報の重み係数は前記第１の重み係数より小さい第２の重み係数に設定され、
前記検出工程は、前記第１の複数の特徴情報と前記第２の複数の特徴情報とがそれぞれ一致するか否かを判定し、一致した特徴情報に対応する前記重み係数が反映された各一致量値の累計加算値に応じて前記類似度を求めるメールチェック方法。

【０１３９】
（付記６）
付記１乃至３のいずれかにおいて、
前記特徴情報ＤＢは、前記特徴情報毎に重み係数を有し、
前記特徴情報は、さらに、送信元メールサーバのＩＰアドレス、ドメイン情報、タイムゾーン情報を含む送信元情報群のうち複数の送信元情報を組み合わせた組み合わせ特徴情報を含み、
前記組み合わせ特徴情報の前記重み係数は、前記送信元情報それぞれ示す各特徴情報の

重み係数より大きい第３の重み係数に設定され、
前記検出工程は、前記受信対象メールについて、各前記複数の受信済みメールと前記組み合わせ特徴情報である前記複数の送信元情報の組み合わせが一致するか否かを判定し、一致した特徴情報に対応する前記重み係数が反映された各一致量値の累計加算値に応じて前記類似度を求めるメールチェック方法。
【０１４０】
（付記７）
付記４において、
前記重み係数は前記送信元アドレス毎に設けられ、
前記特徴情報のばらつき度合いは、前記送信元アドレスが同一の前記複数の受信済みメールにおける前記特徴情報のばらつき度合いであるメールチェック方法。
【０１４１】
（付記８）
付記１乃至７のいずれかにおいて、
前記複数の受信済みのメールは、直近の一定期間内の受信済みメール、または、直近の一定量の受信済みメールのいずれかまたは両方を示し、
前記特徴情報ＤＢ生成手段は、定期的に前記特徴情報ＤＢを更新するメールチェック方法。
【０１４２】
（付記９）
付記４乃至６のいずれかにおいて、
前記類似度は、前記一致量値の累計加算値が基準一致量値を超える受信済みメール数を示すメールチェック方法。
【０１４３】
（付記１０）
付記２において、
前記送信元メールサーバのＩＰアドレスは、ＩＰアドレスのうち、基準上位レベルのＩＰアドレスであるメールチェック方法。
【０１４４】
（付記１１）
新たに受信する対象メールから警戒メール候補を検出するメールチェック装置であって、
正規の複数の受信済みメールについて、前記受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報データーベース（以下、ＤＢ）を生成する特徴情報ＤＢ生成手段と、
受信対象メールの前記メールヘッダーが含む第１の複数の特徴情報と、前記特徴情報ＤＢから参照される当該受信対象メールと送信元アドレスが同一の前記複数の受信済みメールの第２の複数の特徴情報との類似度が基準類似度未満の場合、前記受信対象メールを警戒メール候補として検出する検出手段と、を有するメールチェック装置。
【０１４５】
（付記１２）
新たに受信する対象メールから警戒メール候補を検出するメールチェック処理をコンピュータに実行させるコンピュータ読み取り可能なメールチェックプログラムであって、
前記メールチェック処理は、
正規の複数の受信済みメールについて、前記受信済みメールのメールヘッダーが含む複数の特徴情報を抽出し特徴情報データーベース（以下、ＤＢ）を生成する特徴情報ＤＢ生成工程と、
受信対象メールの前記メールヘッダーが含む第１の複数の特徴情報と、前記特徴情報ＤＢから参照される当該受信対象メールと送信元アドレスが同一の前記複数の受信済みメールの第２の複数の特徴情報との類似度が基準類似度未満の場合、前記受信対象メールを警

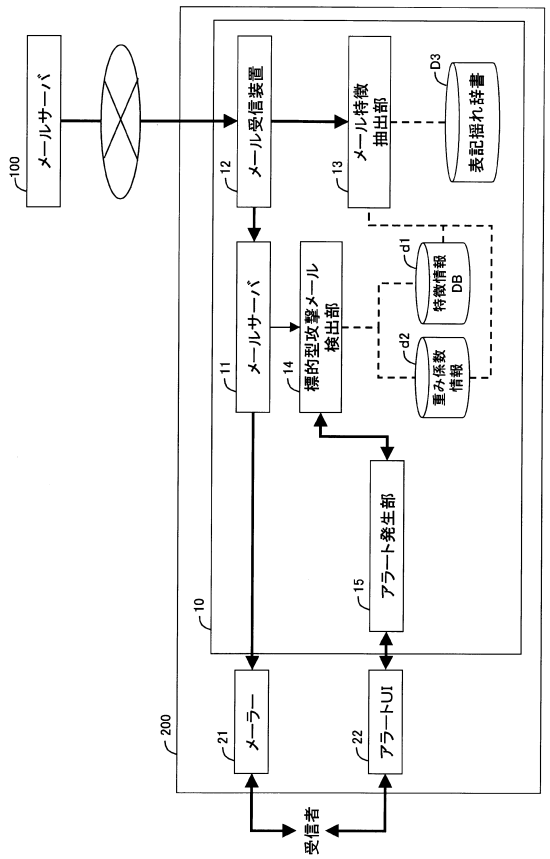
戒メール候補として検出する検出工程と、を有するメールチェックプログラム。

【符号の説明】

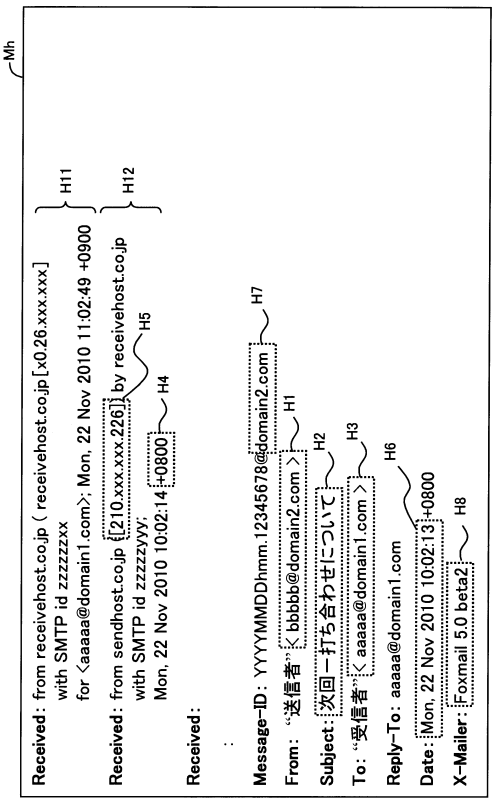
【０１４６】

２００：メールチェック装置、１０：メールチェック部、２１：メーラー、
２２：アラートユーザインターフェース、１１：メールサーバ、１２：メール受信装置、
１３：メール特徴抽出部、１４：標的型攻撃メール検出部、１５：アラート発生部１５、
ｄ１：特徴情報ＤＢ、ｄ２：重み係数情報、ｄ３：表記揺れ辞書

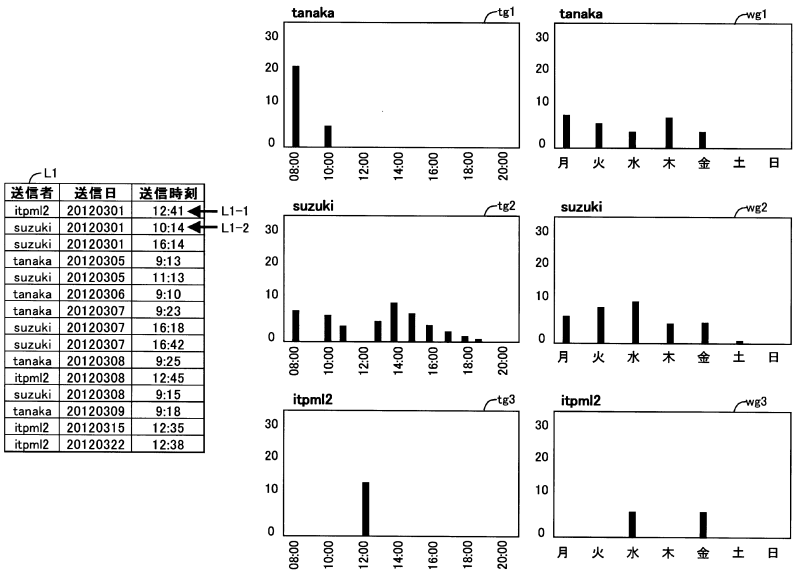
【図１】



【図２】



【図３】



L1

送信者	送信日	送信時刻	
itpml2	20120301	12:41	← L1-1
suzuki	20120301	10:14	← L1-2
suzuki	20120301	16:14	
tanaka	20120305	9:13	
suzuki	20120305	11:13	
tanaka	20120306	9:10	
tanaka	20120307	9:23	
suzuki	20120307	16:18	
suzuki	20120307	16:42	
tanaka	20120308	9:25	
itpml2	20120308	12:45	
suzuki	20120308	9:15	
tanaka	20120309	9:18	
itpml2	20120315	12:35	
itpml2	20120322	12:38	

【図４】

送信者:tanaka　L2-1

☆業務-グループ環境・・・ ← T1
☆業務-今年度の・・・
☆業務-依頼事項の・・・
☆業務-部会の時間・・・
☆業務-技術セミナー・・・
・・・

送信者:suzuki　L2-2

進捗打合せ時間(15:00-)・・・ ← T2
次回の打合せの予定(13:30-)
事前打合せとして(10:30-)・・・
これまでの打合せの時間・・・
・・・

送信者:itpml2　L2-3

「グローバルで・・・」『X通信News172号』 ← T3
「このままでは・・・」『X通信News171号』
「日本にある・・・」『X通信News170号』
「グループ各社の・・・」『X通信News169号』
・・・

d3

	種別	内容
1	長音	ー|ー|ー| …
2	送りカナ	打合せ|打ち合わせ, |打合わせ|打ち合せ| …
3	句点	。| . | . | …
4	読点	、|,|….
5	カッコ	(| (| 【|…
		……

【図５】

送信者:tanaka　L3-1

To	Cc
佐藤	山田B, Mgm-ml
佐藤	山田B, Mgm-ml
鈴木, 佐藤	山田B, Mgm-ml
佐藤, 山田B	加藤
・・・	

↓

L4-1

To	Cc
佐藤	山田B, Mgm-ml
・・・	

送信者:suzuki　L3-2

To	Cc
佐藤	山本B, 鈴木
佐藤, 清水	山本B, 鈴木
佐藤	山本B, 鈴木
佐藤, 田中	山本B, 鈴木
・・・	

↓

L4-2

To	Cc
佐藤	山本B, 鈴木
・・・	

送信者:itpml2　L3-3

To	Cc
佐藤, 田中, 鈴木	加藤
佐藤, 田中, 鈴木	加藤
佐藤, 田中, 鈴木	加藤
佐藤, 田中, 鈴木	加藤
・・・	

↓

L4-3

To	Cc
佐藤, 田中, 鈴木	加藤
・・・	

【図６】

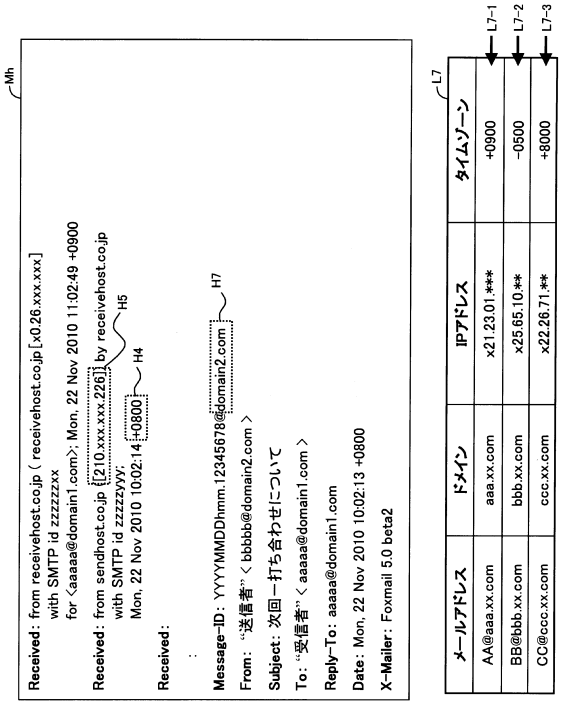
L5

送信者	IPアドレス	受信時刻	
itpml2	x0.26.0.***	20120301 12:41:00	← L5-1
suzuki	x4.65.5.***	20120301 10:14:00	← L5-2
suzuki	x4.65.5.***	20120301 16:14:00	
tanaka	x1.23.0.***	20120305 9:13:00	
suzuki	x4.65.5.***	20120305 11:13:00	
tanaka	x1.23.10.***	20120306 9:10:00	
tanaka	x1.23.10.***	20120307 9:23:00	
suzuki	x4.65.15.***	120120307 6:18:00	
suzuki	x4.65.15.***	20120307 16:42:00	
tanaka	x1.23.0.***	20120308 9:25:00	
itpml2	x0.26.0.***	20120308 12:45:00	
suzuki	x4.65.5.***	20120308 9:15:00	
tanaka	x1.23.10.***	20120309 9:18:00	
itpml2	x0.26.0.***	20120315 12:35:00	

L6

送信者	タイムゾーン	受信時刻	
itpml2	+0900	20120301 12:41:00	
suzuki	+0800	20120301 10:14:00	← L6-1
suzuki	+0900	20120301 16:14:00	
tanaka	−0500	20120305 9:13:00	
suzuki	+0800	20120305 11:13:00	
tanaka	−0500	20120306 9:10:00	
tanaka	−0500	20120307 9:23:00	
suzuki	+0800	120120307 6:18:00	
suzuki	+0900 +0800	20120307 16:42:00	← L6-2
tanaka	−0500	20120308 9:25:00	
itpml2	+0900	20120308 12:45:00	
suzuki	+0900 +0800	20120308 9:15:00	
tanaka	−0500	20120309 9:18:00	
itpml2	+0900	20120315 12:35:00	

【図７】



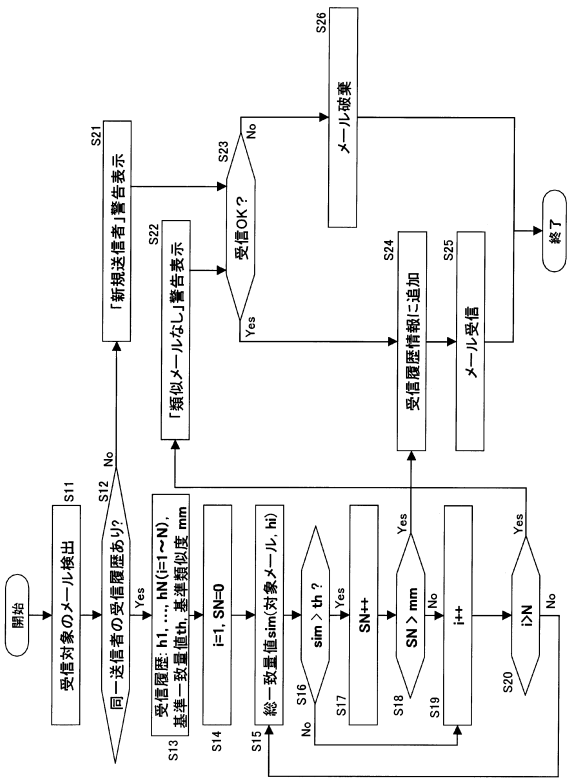
メールアドレス	ドメイン	IPアドレス	タイムゾーン
AA@aaa.xx.com	aaa.xx.com	x21.23.01.***	+0900
BB@bbb.xx.com	bbb.xx.com	x25.65.10.**	−0500
CC@ccc.xx.com	ccc.xx.com	x22.26.71.**	+8000

【図８】

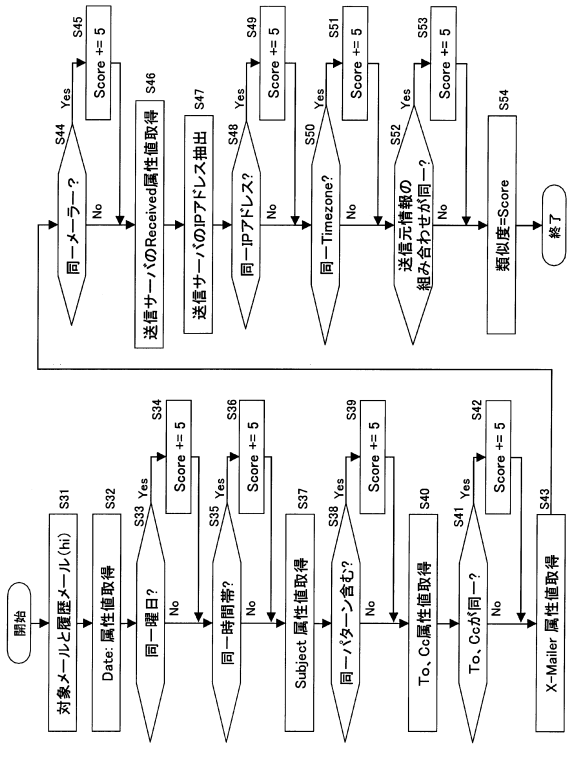
d1

ID	メールアドレス	送信元IPアドレス	送信元タイムゾーン	時間帯	使用メーラー	経由タイムゾーン	MessageID (ドメイン)	宛先 To	宛先 Cc	パターン情報
1	ID4	x8.103.124.***	+0900	2	1	+0900	*@aaaa.bbbb.com	ID31	ID4	(1,1),(7,7)
2	ID20	x2.11.130.***	+0800	5	2	+0800	*@ssss.tttt.uu.com		ID4	(1,2),(2,1),(3,2),(4,1)
3	ID4	x2.31.140.***	+0900	3	1	+0900	*@aaaa.bbbb.com	ID4		(4,2),(5,2),(6,2)
4	ID14	x0.25.149.***	+0900	1	2	+0900	*@aaaa.bbbb.com		ID4	(1,2),(2,1),(3,2),(4,1),(5,2),(6,1)
5	ID7	x0.25.149.***	+0900	2	1	+0900	*@ffff.ggg.com	ID4		(1,1),(2,2)
6	ID7	x0.25.149.***	+0900	6	1	+0900	*@ffff.ggg.com			(6,6)
:	:	:	:	:	:	:	:	:	:	:

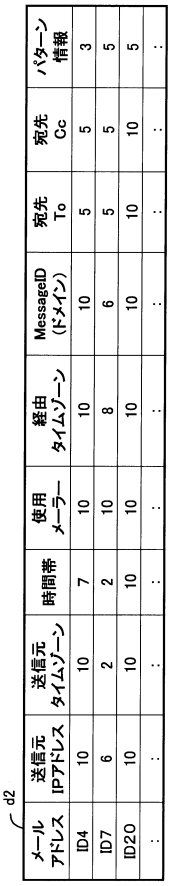
【図９】



【図１０】



【図１１】



d2

メールアドレス	送信元IPアドレス	送信元タイムゾーン	時間帯	使用メーラー	経由タイムゾーン	MessageID（ドメイン）	宛先To	宛先Cc	パターン情報
ID4	10	10	7	10	10	10	5	5	3
ID7	6	2	2	10	8	6	5	5	5
ID20	10	10	10	10	10	10	10	10	5
：	：	：	：	：	：	：	：	：	：

フロントページの続き

(72)発明者　吉岡　孝司
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　森永　正信
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　津田　宏
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

審査官　大石　博見

(56)参考文献　梅田　昴翔　Takato　Umeda, 電子メールヘッダの特徴情報を用いた標的型攻撃の検知　Targeted Attack Detection by Analyzing Characteristics of Electronic Mail Header., コンピュータセキュリティシンポジウム２０１０　論文集　［第一分冊］　Computer Security Symposium 2010 (CSS2010), 日本, 一般社団法人情報処理学会　Information Processing Society of Japan, ２０１０年１０月１２日, 第2010巻, 第１０９頁～第１１３頁

(58)調査した分野(Int.Cl., ＤＢ名)
Ｈ０４Ｌ　１２／５８
Ｇ０６Ｆ　１３／００
Ｈ０４Ｍ　３／５３