(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6895972号
(P6895972)
(45)発行日　令和3年6月30日(2021.6.30)　(24)登録日　令和3年6月10日(2021.6.10)

(51)Int.Cl.　F I
G06F 21/16 (2013.01)　G06F 21/16
G06F 16/20 (2019.01)　G06F 16/20
G06F 16/28 (2019.01)　G06F 16/28

請求項の数 12　(全 14 頁)

(21)出願番号　特願2018-532787(P2018-532787)
(86)(22)出願日　平成28年12月19日(2016.12.19)
(65)公表番号　特表2019-508779(P2019-508779A)
(43)公表日　平成31年3月28日(2019.3.28)
(86)国際出願番号　PCT/CN2016/110714
(87)国際公開番号　WO2017/114209
(87)国際公開日　平成29年7月6日(2017.7.6)
審査請求日　令和1年12月13日(2019.12.13)
(31)優先権主張番号　201511028180.5
(32)優先日　平成27年12月31日(2015.12.31)
(33)優先権主張国・地域又は機関
中国(CN)

(73)特許権者　511050697
アリババ　グループ　ホウルディング　リミテッド
英国領ケイマン諸島　グランド　ケイマン　ジョージ　タウン　ピーオーボックス　８４７　ワン　キャピタル　プレイス　フォース　フロア
(74)代理人　100079108
弁理士　稲葉　良幸
(74)代理人　100109346
弁理士　大貫　敏史
(74)代理人　100117189
弁理士　江口　昭彦
(74)代理人　100134120
弁理士　内藤　和彦

最終頁に続く

(54)【発明の名称】ラベルデータ漏洩チャネル検出方法および装置

(57)【特許請求の範囲】
【請求項１】
ユーザラベルデータの漏洩チャネルを検出するためのラベルデータ漏洩チャネル検出方法であって、
ラベルデータ漏洩チャネル検出装置が、ユーザが閲覧したウェップページに基づいて生成される前記ユーザの好みを示すラベル（以下、「正常ラベル」という）に基づいて、前記ユーザの好みを表さないラベル（以下、「検出ラベル」という）を決定することと、
前記ラベルデータ漏洩チャネル検出装置が、前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することと、
前記ラベルデータ漏洩チャネル検出装置が、前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることと、
前記ラベルデータ漏洩チャネル検出装置が、前記検出ラベルに起因するプッシュ情報の識別に応答して、前記チャネルインデックスに従って、前記プッシュ情報が、前記検出ラベルとの関連性を有するチャネルからのものかどうかを検出することと、
前記ラベルデータ漏洩チャネル検出装置が、前記プッシュ情報が、前記検出ラベルとの関連性を有する前記チャネルと異なる他のチャネルからのものであることに応答して、前記他のチャネルを漏洩疑いチャネルとして識別することと
を含む、方法。
【請求項２】

前記ラベルデータ漏洩チャネル検出装置が、前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることが、
前記ラベルデータ漏洩チャネル検出装置が、広告元と前記正常ラベルとの間の整合度に従って計算される、プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することと、
前記ラベルデータ漏洩チャネル検出装置が、前記プッシュ情報が前記検出ラベルに基づいて生成される確率に従って、前記傍受したプッシュ情報をスクリーニングすることと
を含む、請求項１に記載のラベルデータ漏洩チャネル検出方法。
【請求項３】
前記ラベルデータ漏洩チャネル検出装置が、プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することが、
前記ラベルデータ漏洩チャネル検出装置が、前記プッシュ情報が前記正常ラベルに基づいて生成される前記確率が予め設定された閾値より低いことに応答して、傍受を実行すること
を含む、請求項２に記載のラベルデータ漏洩チャネル検出方法。
【請求項４】
前記ラベルデータ漏洩チャネル検出装置が、前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することが、
前記ラベルデータ漏洩チャネル検出装置が、前記チャネルの挙動履歴に従って前記チャネルの信頼性値を決定することと、
前記ラベルデータ漏洩チャネル検出装置が、前記チャネルの前記信頼性値に基づいてユーザグループをサンプリングすることと、
前記ラベルデータ漏洩チャネル検出装置が、サンプリングを通じて得られた前記ユーザグループの各ユーザに対し、前記ユーザの検出ラベルから１つの検出ラベルを選択することと、
前記ラベルデータ漏洩チャネル検出装置が、前記選択された検出ラベルを、前記チャネルと、前記ユーザに関連付けられた前記ユーザＩＤとに関連付けること
を含む、請求項１に記載のラベルデータ漏洩チャネル検出方法。
【請求項５】
ユーザラベルデータの漏洩チャネルを検出するためのラベルデータ漏洩チャネル検出装置であって、
命令のセットを記憶するメモリと、
プロセッサと
を含み、前記プロセッサは、前記命令のセットを実行して、前記ラベルデータ漏洩チャネル検出装置に、
ユーザが閲覧したウェッブページに基づいて生成される前記ユーザの好みを示すラベル（以下、「正常ラベル」という）に基づいて、前記ユーザの好みを表さないラベル（以下、「検出ラベル」という）を決定することと、
前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することと、
前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることと、
前記検出ラベルに起因するプッシュ情報の識別に応答して、前記チャネルインデックスに従って、前記プッシュ情報が、前記検出ラベルとの関連性を有するチャネルからのものかどうかを検出することと、
前記プッシュ情報が、前記検出ラベルとの関連性を有する前記チャネルと異なる他のチャネルからのものであることに応答して、前記他のチャネルを漏洩疑いチャネルとして

識別することと
を実行させるように構成される、装置。
【請求項６】
前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることが、
広告元と前記正常ラベルとの間の整合度に従って計算される、プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することと、
前記プッシュ情報が前記検出ラベルに基づいて生成される確率に従って、前記傍受したプッシュ情報をスクリーニングすることと
を含む、請求項５に記載のラベルデータ漏洩チャネル検出装置。
【請求項７】
プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することが、
前記プッシュ情報が前記正常ラベルに基づいて生成される前記確率が予め設定された閾値より低い場合、傍受を実行すること
を含む、請求項６に記載のラベルデータ漏洩チャネル検出装置。
【請求項８】
前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することが、
前記チャネルの挙動履歴に従って前記チャネルの信頼性値を決定することと、
前記チャネルの前記信頼性値に基づいてユーザグループをサンプリングすることと、
サンプリングを通じて得られた前記ユーザグループの各ユーザに対し、前記ユーザの検出ラベルから１つの検出ラベルを選択することと、
前記選択された検出ラベルを、前記チャネルと、前記ユーザに関連付けられた前記ユーザＩＤとに関連付けることと
を含む、請求項５に記載のラベルデータ漏洩チャネル検出装置。
【請求項９】
命令のセットを記憶する非一時的コンピュータ可読媒体であって、前記命令のセットは、ラベルデータ漏洩チャネル検出装置にラベルデータ漏洩チャネル検出方法を実行させるように、前記装置の少なくとも１つのプロセッサによって実行可能であり、前記方法は、
ユーザが閲覧したウェッブページに基づいて生成される前記ユーザの好みを示すラベル（以下、「正常ラベル」という）に基づいて、前記ユーザの好みを表さないラベル（以下、「検出ラベル」という）を決定することと、
前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することと、
前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることと、
前記検出ラベルに起因するプッシュ情報の識別に応答して、前記チャネルインデックスに従って、前記プッシュ情報が、前記検出ラベルとの関連性を有するチャネルからのものかどうかを検出することと、
前記プッシュ情報が、前記検出ラベルとの関連性を有する前記チャネルと異なる他のチャネルからのものであることに応答して、前記他のチャネルを漏洩疑いチャネルとして識別することと
を含む、非一時的コンピュータ可読媒体。
【請求項１０】
前記検出ラベルに起因するプッシュ情報を識別するために、前記ユーザによって受信されたプッシュ情報をモニタすることが、

広告元と前記正常ラベルとの間の整合度に従って計算される、プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することと、
前記プッシュ情報が前記検出ラベルに基づいて生成される確率に従って、前記傍受したプッシュ情報をスクリーニングすることと
を含む、請求項９に記載の非一時的コンピュータ可読媒体。
【請求項１１】
プッシュ情報が前記正常ラベルに基づいて生成される確率に従って、前記ユーザによって受信された前記プッシュ情報を傍受することが、
前記プッシュ情報が前記正常ラベルに基づいて生成される前記確率が予め設定された閾値より低い場合、傍受を実行すること
を含む、請求項１０に記載の非一時的コンピュータ可読媒体。
【請求項１２】
前記検出ラベルを、ユーザデータに基づいてプッシュ情報を送信するチャネルのチャネルＩＤと、前記ユーザに関連付けられたユーザＩＤとに関連付けて、チャネルインデックスとして記録することが、
前記チャネルの挙動履歴に従って前記チャネルの信頼性値を決定することと、
前記チャネルの前記信頼性値に基づいてユーザグループをサンプリングすることと、
サンプリングを通じて得られた前記ユーザグループの各ユーザに対し、前記ユーザの検出ラベルから１つの検出ラベルを選択することと、
前記選択された検出ラベルを、前記チャネルと、前記ユーザに関連付けられた前記ユーザＩＤとに関連付けることと
を含む、請求項９に記載の非一時的コンピュータ可読媒体。

【発明の詳細な説明】
【技術分野】
【０００１】
技術分野
本発明は、データセキュリティ技術の分野に関し、具体的には、ラベルデータ漏洩チャネル検出方法および装置に関する。
【背景技術】
【０００２】
背景技術
インターネットコンテンツ組織形態として、ラベルは、オブジェクトエンティティのプロパティと強く関わりがあるキーワードである。ラベルは、コンテンツを容易に説明および分類する上で役立ち、回収および共有を促進する。ラベルによって表される大量のユーザの好みデータは、インターネットの開発において蓄積してきており、データは、インターネット広告、推奨および他の製品の基盤を構成する。他方では、その価値により、データは、ユーザの他の個人を識別できる情報（ＰＩＩ）と共にデータ漏洩の標的となり、違法な取得および転売が行われる。既存のデータセキュリティ技術は、暗号化、システム強化、アクセス制御および監査モニタリングを使用して、データ所有者の制御可能な環境からのデータ漏洩を防ぐ。しかし、データ連携事業のシナリオでは、データは、通常、データ所有者の制御可能な環境を離れ、制御不可能なパートナー環境に入る。そのシナリオでは、従来のデータベース透かし技術および従来のデータ軌跡追跡技術は、膨大な量の動的なユーザラベルデータからの課題を解決することはできない。
【０００３】
従来のデータベース透かし技術および従来のデータ軌跡追跡技術は、ユーザラベルとして数値フィールドを欠くようなデータの透かしを効果的に生成することはできない。第２に、ラベルデータは、一般に、分散方式で使用されており、それにより、透かしの検出を難しくしている。それに加えて、ラベルデータの膨大な量の動的な特徴により、透かしの

更新および検出が厳しく要求される。ラベルデータの値は、一般に、非常にありふれたものであり、それらをインターネット上で追跡することは非常に難しい。

【発明の概要】

【発明が解決しようとする課題】

【０００４】

発明の概要

本発明の目的は、既存の技術的解決法においてラベルデータの追跡および検出が難しいという技術的問題を解決するために、想定されるデータ漏洩チャネルを効果的に検出することができるラベルデータ漏洩チャネル検出方法および装置を提供することである。

【課題を解決するための手段】

【０００５】

前述の目的を達成するため、本発明の技術的解決法は、以下の通りである。

【０００６】

ユーザラベルデータの漏洩チャネルを検出するためのラベルデータ漏洩チャネル検出方法であって、

ユーザラベルデータセットを生成するために、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加することと、

ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当て、ユーザＩＤ、検出ラベルおよびチャネルＩＤと関連付けられたチャネルインデックスを確立することと、

プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受することと、

プッシュ情報がユーザの検出ラベルから生成される確率に従って、傍受したプッシュ情報をスクリーニングし、プッシュ情報がユーザの検出ラベルから生成される確率が所定の閾値より高い場合は、ユーザの検出ラベルを漏洩疑いラベルセットに追加することと、

漏洩疑いチャネルＩＤの対応するリストを得るために、漏洩疑いラベルセットに従ってチャネルインデックスを検索することと、

プッシュ情報が、見つかったチャネルからのものかどうかを検出し、そうである場合は、対応するチャネルを削除し、残りのチャネルを漏洩疑いチャネルとして出力することとを含む、検出方法。

【０００７】

さらに、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加するステップは、

設定された第１の閾値より低い、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率

を含む。

【０００８】

さらに、ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当て、ユーザＩＤ、検出ラベルおよびチャネルＩＤと関連付けられたチャネルインデックスを確立するステップは、

所定のチャネルの挙動履歴に従って当該所定のチャネルの信頼性を計算することと、

チャネルのチャネルＩＤを変数として取り入れることによって設定された設定ハッシュ関数からハッシュ関数を選択することと、

チャネルの信頼性に基づいてユーザグループをサンプリングすることと、

サンプリングを通じて得られたユーザグループの各ユーザに対し、ユーザＩＤを変数として用いて、選択したハッシュ関数に従ってユーザの検出ラベルからチャネルに対応する検出ラベルを選択することと、

［ユーザＩＤ，検出ラベル］からチャネルＩＤへのチャネルインデックスを確立することと

を含む。

【０００９】
さらに、プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受するステップは、
プッシュ情報が正常なラベルから生成される確率が、設定された第２の閾値より低い場合は傍受を実行し、そうでない場合はユーザにプッシュ情報を表示すること
を含む。
【００１０】
さらに、検出方法は、ユーザの正常なラベルの変化に従ってユーザの検出ラベルを更新するステップをさらに含み、当該ステップは、具体的には、
新しい正常なラベルが既存の検出ラベルと同時に発生する確率に従って、ユーザの当該新しい正常なラベルとの同時発生の高い確率を有する検出ラベルを削除することと、
第１の閾値より低い、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率、ユーザの新しい検出ラベルを追加することと
を含む。
【００１１】
さらに、検出方法は、
削除した検出ラベルの関連アイテムをチャネルインデックスから取り除くこと
をさらに含む。
【００１２】
本発明は、ユーザラベルデータの漏洩チャネルを検出するためのラベルデータ漏洩チャネル検出装置であって、
ユーザラベルデータセットを生成するために、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加するように構成された検出ラベル追加モジュールと、
ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当て、ユーザＩＤ、検出ラベルおよびチャネルＩＤと関連付けられたチャネルインデックスを確立するように構成されたチャネル関連付けモジュールと、
プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受するように構成された傍受モジュールと、
プッシュ情報がユーザの検出ラベルから生成される確率に従って、傍受したプッシュ情報をスクリーニングし、プッシュ情報がユーザの検出ラベルから生成される確率が所定の閾値より高い場合は、ユーザの検出ラベルを漏洩疑いラベルセットに追加するように構成された傍受情報分析モジュールと、
漏洩疑いチャネルＩＤの対応するリストを得るために、漏洩疑いラベルセットに従ってチャネルインデックスを検索するように構成されたチャネル検索モジュールと、
プッシュ情報が、見つかったチャネルからのものかどうかを検出し、そうである場合は、対応するチャネルを削除し、残りのチャネルを漏洩疑いチャネルとして出力するように構成された出力モジュールと
を含む、検出装置をさらに提案する。
【００１３】
さらに、検出ラベル追加モジュールが、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加する場合、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率は、設定された第１の閾値より低い。
【００１４】
さらに、ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当てる場合、チャネル関連付けモジュールは、以下の動作、すなわち、
所定のチャネルの挙動履歴に従って所定のチャネルの信頼性を計算することと、
チャネルのチャネルＩＤを変数として取り入れることによって設定された設定ハッシュ関数からハッシュ関数を選択することと、
チャネルの信頼性に基づいてユーザグループをサンプリングすることと、
サンプリングを通じて得られたユーザグループの各ユーザに対し、ユーザＩＤを変数と

して用いて、選択したハッシュ関数に従ってユーザの検出ラベルからチャネルに対応する検出ラベルを選択することと、
［ユーザＩＤ，検出ラベル］からチャネルＩＤへのチャネルインデックスを確立することと
を実行する。
【００１５】
さらに、プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受する場合、傍受モジュールは、以下の動作、すなわち、
プッシュ情報が正常なラベルから生成される確率が、設定された第２の閾値より低い場合は傍受を実行し、そうでない場合はユーザにプッシュ情報を表示すること
を実行する。
【００１６】
さらに、検出ラベル追加モジュールは、ユーザの正常なラベルの変化に従ってユーザの検出ラベルを更新するようにさらに構成され、具体的には、以下のステップ、すなわち、
新しい正常なラベルが既存の検出ラベルと同時に発生する確率に従って、ユーザの当該新しい正常なラベルとの同時発生の高い確率を有する検出ラベルを削除することと、
第１の閾値より低い、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率、ユーザの新しい検出ラベルを追加することと
を実行する。
【００１７】
さらに、チャネル関連付けモジュールは、削除した検出ラベルの関連アイテムをチャネルインデックスから取り除くようにさらに構成される。
【００１８】
本発明は、ラベルデータ漏洩チャネル検出方法および装置を提案し、同方法および装置は、同じユーザのラベルの異なる発生確率に従って異なるデータ使用チャネルのための異なる検出ラベルを生成し、次いで、検出ラベルの使用を間接的に検出し、最終的に、膨大な量のデータのインデックス作成および検索技術に基づいて、想定されるデータ漏洩チャネルを効果的に検出する。検出方法は、高い検出効率を有し、膨大な量の動的なユーザラベルデータを処理することができる。
【図面の簡単な説明】
【００１９】
【図１】本発明による、ラベルデータ漏洩チャネル検出方法のフローチャートである。
【図２】本発明による、ラベルデータ漏洩チャネル検出装置の概略構造図である。
【発明を実施するための形態】
【００２０】
詳細な説明
本発明の技術的解決法は、添付の図面および実施形態を参照して、以下でさらに詳細に説明する。以下の実施形態は、本発明を制限しない。
【００２１】
ユーザがインターネットをブラウズする際、ブラウズされたウェブページは、ユーザのためにユーザの好みを示すラベルを生成することができる。ラベルによって表される大量のユーザの好みデータは、インターネットの開発において蓄積してきている。本発明は、ユーザが所有する正常なラベルに基づいて、各ユーザに対するある特定の数の検出ラベルを追加する。検出ラベルによるプッシュ情報が見つかった際には、ユーザラベルデータの漏洩チャネルは、プッシュ情報に従って検索することができる。この実施形態におけるプッシュ情報は、広告、プッシュウェブページおよび同様のものを含み得る。以下では、例として広告を採用することによって説明が行われる。
【００２２】
この実施形態は、ラベルデータ漏洩チャネル検出方法を提供する。図１に示されるよう

に、方法は、以下のステップを含む。

【００２３】

ステップＳ１．ユーザラベルデータセットを生成するために、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルが追加される。

【００２４】

この実施形態では、ユーザのインターネットサーフィンから生成され、ユーザの好みを識別するラベルは、正常なラベルと呼ばれる。ユーザのためにこのステップを通じて生成され、後続の検出のために使用されるラベルは、検出ラベルと呼ばれる。わかるように、検出ラベルは、ユーザの好みを表さず、後続の検出のためだけに使用される。ユーザラベルデータセットは、正常なラベルおよび検出ラベルを含む。

【００２５】

後続の分析を容易にするため、各ユーザは、異なるチャネルに対応できるほど十分な検出ラベルを有する必要がある。この目的のため、ユーザが十分な検出ラベルを有さない場合は、ユーザの検出ラベルが設定量に達するように、ユーザに対する検出ラベルが生成される。

【００２６】

例えば、ユーザＵ１は、２つの正常なラベルを有し、それらは、テレビを見ることおよびジャンクフードのそれぞれである。この実施形態は、２つの検出ラベルを必要とする。従って、例えば、野菜およびハイキング用の靴など、ユーザＵ１に対して２つの検出ラベルが生成される。

【００２７】

ユーザの検出ラベルを生成するための具体的なプロセスは、以下の通り、すなわち、

ユーザラベルデータセットに指定数の検出ラベルが存在するかどうかを判定し、検出ラベルが指定数に達した場合は終了し、そうでない場合は次のステップに進むことと、

ユーザの既存のラベルとの同時発生の確率が、設定された第１の閾値より低いラベルを生成し、ユーザの検出ラベルとしてユーザラベルデータセットにラベルを追加することとである。

【００２８】

新しい検出ラベルの生成のあいだ、ユーザの既存の正常なラベルおよび既存の検出ラベルとの同時発生の比較的低い確率を有するラベルを共通のラベルから見つけることが必要である。すなわち、新しく生成された検出ラベルは、ユーザラベルセットのいかなる既存のラベルとも似ていない。新しく生成された検出ラベルおよび既存のラベルは、互いに異なり、同時発生の低い確率を有する。

【００２９】

ステップＳ２．ユーザラベルデータセットに従って検出ラベルが所定のチャネルに割り当てられ、ユーザＩＤ、検出ラベルおよびチャネルＩＤと関連付けられたチャネルインデックスが確立される。

【００３０】

所定のチャネルの信頼性は、その挙動履歴に従って計算することができる。この実施形態では、チャネルは、ユーザデータを使用するチャネルを指す。例えば、ネットワークプラットフォームは、広告者に対してそのユーザデータを提供することができる。広告者は、ネットワークプラットフォームの顧客であり、ユーザデータを使用するチャネルでもある。チャネルの信頼性は、チャネルによってユーザデータに基づいて広告を送信することの信頼性を指す。チャネルが、ユーザデータに基づいて広告をプッシュしないが、ユーザが関心を持たない広告をユーザにプッシュする場合は、チャネルは、信頼できるものではない。さらに、チャネルの一意ＩＤは、設定ハッシュ関数のセットからハッシュ関数Ｈ１を選択するために、可変キーとし使用することができる。次に、チャネルの信頼性に基づいて、ユーザグループがサンプリングされる。高い信頼性を有するチャネルの場合は、サンプリングされたユーザグループはより小さいものであり得る。次いで、サンプリングされたユーザグループの各ユーザに対し、ユーザＩＤをキーとして用いて、Ｈ１関数に従っ

てユーザの検出ラベルセットからチャネルに対応する検出ラベルが選択される。

【００３１】

例えば、所定のチャネル１のサンプリングされたユーザは、ユーザＵ１を含む。乱数値は、Ｈ１関数およびユーザＵ１のユーザＩＤに従って計算される。検出ラベルは、乱数値に従ってユーザＵ１のすべての検出ラベルから選択され、チャネル１に割り当てられる。例えば、チャネル１に対し、Ｈ１関数に従って計算された乱数値が１である場合は、ユーザ１の検出ラベルのソーティングに従って、第１の検出ラベルが選択され、チャネル１に割り当てられる。ユーザＵ１の「野菜」という検出ラベルがチャネル１に割り当てられると仮定する。

【００３２】

同様に、ユーザＵ１の「ハイキング用の靴」という検出ラベルがチャネル２に割り当てられる。

【００３３】

従って、［ユーザＩＤ，検出ラベル］からチャネルＩＤへのチャネルインデックスを確立することができる。すなわち、チャネルインデックスにおいて記録が確立される。例えば、表１に示されるようなチャネルインデックスが確立される。

【００３４】

【表１】

シリアル番号	[ユーザ ID, 検出ラベル]	チャネル ID
1	[U1, 野菜]	チャネル 1
2	[U1, ハイキング用の靴]	チャネル 2

表 1

【００３５】

検出ラベルは、ユーザラベルデータセットに追加され、チャネルに対応する検出ラベルのみが、対応するチャネルに割り当てられる。例えば、［Ｕ１，ハイキング用の靴］は、チャネル２に割り当てられる。チャネル２がユーザラベルデータセットに従って広告をプッシュする場合は、広告が正常なラベルに従って送信されるかまたは［Ｕ１，ハイキング用の靴］という検出ラベルに従って送信されるかにかかわらず、広告は、安全なものと見なされる。違法なユーザが漏洩ユーザラベルデータを得た後、ハイキング用の靴などの広告も、ユーザに送信され、違法なチャネルがチャネルインデックスのチャネル２ではないことがチャネルインデックスに従って分かった場合、ユーザラベルデータが漏洩したと見なされる。

【００３６】

ステップＳ３．プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報が傍受される。

【００３７】

正常な状況の下では、インターネットサーフィンのためのユーザ端末は一般にユーザ側にあるため、ユーザによって受信された広告は、ユーザ端末に反映される。広告の検出は、ユーザ端末のクライアントにおいて最初に実行することができる。例えば、多くのパーソナルコンピュータおよびスマートフォンには、現在、セキュリティアシスタントがインストールされており、既存のセキュリティアシスタントを直接使用して、ユーザ端末上の広告を傍受することができる。また、ユーザ端末上の広告を検出するために、特定のクライアントを開発することもできる。

【００３８】

広告傍受する際、広告が正常なラベルから生成される確率が、設定された第２の閾値より低い場合に広告は傍受され、そうでない場合は、広告はユーザに表示される。

【００３９】

ユーザ端末上の既存のセキュリティアシスタントが使用される場合は、ステップＳ２に

おいて、セキュリティアシスタントがインストールされていないユーザは、ユーザラベルデータセットから最初にフィルタ除去されるべきであることを、理解することは容易である。すなわち、セキュリティアシスタントがインストールされているユーザのみがサンプリングされ、セキュリティアシスタントがインストールされていないユーザは考慮されない。従って、追加のクライアントを開発することは不要であり、ユーザのセキュリティアシスタントを直接使用してユーザ端末側の広告がフィルタリングされる。

【００４０】

具体的には、広告はフィルタリングされる。すなわち、ユーザによって受信された広告は、広告がユーザの正常なラベルから生成される確率に従って傍受される。広告が正常なラベルから生成される確率が、設定された閾値より低い場合は、次の処理ステップが実行され、そうでない場合は、広告はユーザに表示される。

【００４１】

ユーザの正常なラベルは、広告が正常なラベルから生成される確率に従ってセキュリティアシスタントが傍受を実行するように、ユーザのユーザ側のセキュリティアシスタントと同期させる必要があることに留意すべきである。広告が正常なラベルから生成される確率は、一般に、広告源とユーザの正常なラベルとの間の整合度に従ってセキュリティアシスタントによって計算されるが、それについては、本明細書ではさらなる説明は行わない。正常なラベルから生成される確率が、設定された閾値より低い広告は、次の処理ステップのために、傍受され、専用のバックエンドサーバに送信される。

【００４２】

ステップＳ４．プッシュ情報がユーザの検出ラベルから生成される確率に従って、傍受したプッシュ情報がスクリーニングされ、プッシュ情報がユーザの検出ラベルから生成される確率が所定の閾値より高い場合は、ユーザの検出ラベルが漏洩疑いラベルセットに追加される。

【００４３】

バックエンドサーバに送信された広告は、広告がユーザの検出ラベルから生成される確率に従ってさらにスクリーニングされる。広告がユーザのある特定の検出ラベルから生成される確率が所定の閾値より高い場合は、ユーザの検出ラベルは、漏洩疑いラベルセットに追加される。

【００４４】

例えば、ユーザＵ１に送信されたトレッキングポールの広告は、広告が「テレビを見ること」および「ジャンクフード」という正常なラベルから生成される比較的低い確率に従って、バックエンドサーバに送信される。しかし、ユーザＵ１の「ハイキング用の靴」という検出ラベルの場合、広告が「ハイキング用の靴」から生成される確率は比較的高いため、［Ｕ１，ハイキング用の靴］が漏洩疑いラベルセットに追加される。

【００４５】

ステップＳ５．漏洩疑いチャネルＩＤの対応するリストを得るために、漏洩疑いラベルセットに従ってチャネルインデックスが検索される。

【００４６】

次に、疑いラベルが漏洩疑いラベルセットから抽出され、想定されるチャネルＩＤのソートされたリストを得るために、チャネルインデックスにおいて検索が行われる。

【００４７】

前述の例のように、［Ｕ１，ハイキング用の靴］という漏洩疑いラベルは、漏洩疑いラベルセットから抽出され、チャネルインデックスのチャネル２の検出ラベルは「ハイキング用の靴」を含むため、チャネル２は、漏洩疑いチャネルＩＤのリストに追加される。

【００４８】

ステップＳ６．プッシュ情報が、見つかったチャネルからのものかどうかが検出され、そうである場合は、対応するチャネルが削除され、残りのチャネルが漏洩疑いチャネルとして出力される。

【００４９】

最後に、ユーザ端末の広告源がチャネル２であるかどうかを検出する必要がある。そうである場合は、承認を示し、チャネルリストからチャネル２が削除される。

【００５０】

最終的なチャネルリストは、すべての想定されるラベルデータ漏洩チャネルを含む。これらのチャネルに対し、証拠を収集するために、連携データへのモニタされたおとり（ハニーポット）データの追加、オフライン調査との組合せおよび他の手段など、さらなる調査手段を講じることができる。

【００５１】

さらに、ユーザの正常なラベルは更新される場合が多いため、ユーザの検出ラベルは、ユーザの正常なラベルが更新された後に更新する必要がある。この実施形態では、ユーザの検出ラベルを更新するためのプロセスは、以下の通り、すなわち、

新しい正常なラベルが既存の検出ラベルと同時に発生する確率に従って、ユーザの新しい正常なラベルとの同時発生の高い確率を有する検出ラベルを削除することと、

第１の閾値より低い、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率、ユーザの新しい検出ラベルを追加すること。

【００５２】

それに応じて、チャネルインデックスをさらに更新する必要がある。すなわち、

削除した検出ラベルの関連アイテムをチャネルインデックスから取り除く。

【００５３】

従って、次の広告傍受において漏洩疑いチャネルを検出するために新しいチャネルインデックスが使用されるように、チャネルインデックスが更新される。

【００５４】

図２は、ユーザラベルデータの漏洩チャネルを検出するためのラベルデータ漏洩チャネル検出装置であって、

ユーザラベルデータセットを生成するために、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加するように構成された検出ラベル追加モジュールと、

ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当て、ユーザＩＤ、検出ラベルおよびチャネルＩＤと関連付けられたチャネルインデックスを確立するように構成されたチャネル関連付けモジュールと、

プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受するように構成された傍受モジュールと、

プッシュ情報がユーザの検出ラベルから生成される確率に従って、傍受したプッシュ情報をスクリーニングし、プッシュ情報がユーザの検出ラベルから生成される確率が所定の閾値より高い場合は、ユーザの検出ラベルを漏洩疑いラベルセットに追加するように構成された傍受情報分析モジュールと、

漏洩疑いチャネルＩＤの対応するリストを得るために、漏洩疑いラベルセットに従ってチャネルインデックスを検索するように構成されたチャネル検索モジュールと、

プッシュ情報が、見つかったチャネルからのものかどうかを検出し、そうである場合は、対応するチャネルを削除し、残りのチャネルを漏洩疑いチャネルとして出力するように構成された出力モジュールと

を含む、検出装置を示す。

【００５５】

この実施形態の装置は、アプリケーションシステムのバックエンドサーバに適用できることを理解することは容易である。傍受モジュールは、ユーザ端末において統合することができ、ユーザ端末側で傍受を実行することができる。傍受モジュールは、第三者クライアント（例えば、セキュリティアシスタントまたは専用クライアント）を使用することによって傍受を実行することができる。

【００５６】

この実施形態では、検出ラベル追加モジュールが、ユーザが所有する正常なラベルに基づいてユーザの検出ラベルを追加する場合、追加された検出ラベルがユーザの既存のラベ

ルと同時に発生する確率は、設定された第１の閾値より低い。すなわち、新しく生成された検出ラベルは、ユーザラベルセットのいかなる既存のラベルとも似ていない。新しく生成された検出ラベルおよび既存のラベルは、互いに異なり、同時発生の低い確率を有し、従って、互いに影響しない。

【００５７】

この実施形態では、ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当てる場合、チャネル関連付けモジュールは、以下の動作、すなわち、

所定のチャネルの挙動履歴に従って所定のチャネルの信頼性を計算することと、

チャネルのチャネルＩＤを変数として取り入れることによって設定された設定ハッシュ関数からハッシュ関数を選択することと、

チャネルの信頼性に基づいてユーザグループをサンプリングすることと、

サンプリングを通じて得られたユーザグループの各ユーザに対し、ユーザＩＤを変数として用いて、選択したハッシュ関数に従ってユーザの検出ラベルからチャネルに対応する検出ラベルを選択することと、

［ユーザＩＤ，検出ラベル］からチャネルＩＤへのチャネルインデックスを確立することと

を実行する。

【００５８】

この実施形態では、プッシュ情報がユーザの正常なラベルから生成される確率に従って、ユーザによって受信されたプッシュ情報を傍受する場合、傍受モジュールは、以下の動作、すなわち、

プッシュ情報が正常なラベルから生成される確率が、設定された第２の閾値より低い場合は傍受を実行し、そうでない場合はユーザにプッシュ情報を表示すること

を実行する。

【００５９】

この実施形態では、検出ラベル追加モジュールは、ユーザの正常なラベルの変化に従ってユーザの検出ラベルを更新するようにさらに構成され、具体的には、以下のステップ、すなわち、

新しい正常なラベルが既存の検出ラベルと同時に発生する確率に従って、ユーザの新しい正常なラベルとの同時発生の高い確率を有する検出ラベルを削除することと、

第１の閾値より低い、新しく追加された検出ラベルがユーザの既存のラベルと同時に発生する確率、ユーザの新しい検出ラベルを追加することと

を実行する。

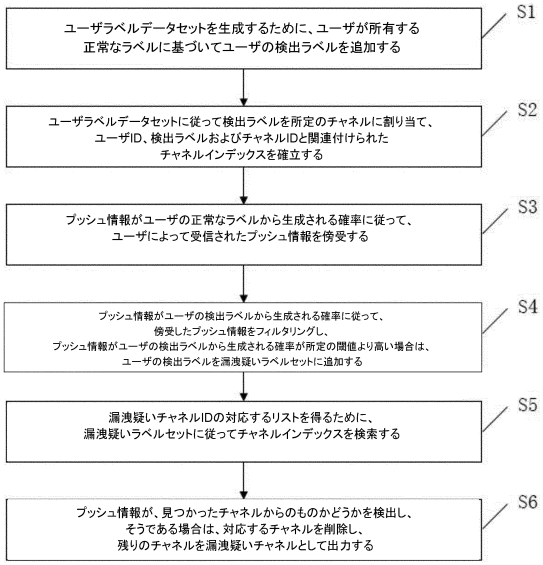
【００６０】

この実施形態では、チャネル関連付けモジュールは、削除した検出ラベルの関連アイテムをチャネルインデックスから取り除くようにさらに構成される。従って、ユーザが新しい正常なラベルを生成する場合、ユーザラベルセットは、適時に更新される。

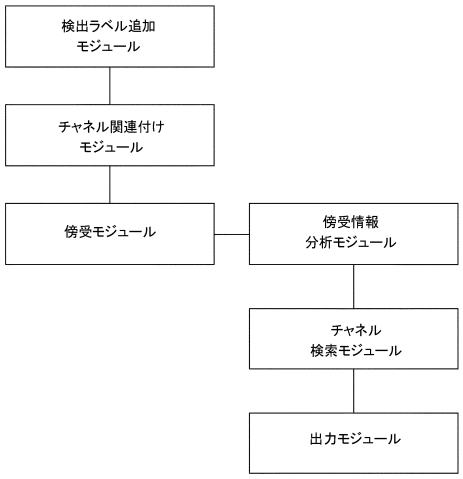
【００６１】

上記の実施形態は、本発明の技術的解決法を制限する代わりに、本発明の技術的解決法を説明するためだけに使用される。当業者であれば、本発明の精神および本質から逸脱することなく、対応する様々な変更および変形を行うことができ、対応する変更および変形はすべて、本発明の添付の請求項の保護範囲に包含されるべきである。

【図１】


ユーザラベルデータセットを生成するために、ユーザが所有する
正常なラベルに基づいてユーザの検出ラベルを追加する
S1
ユーザラベルデータセットに従って検出ラベルを所定のチャネルに割り当て、
ユーザID、検出ラベルおよびチャネルIDと関連付けられた
チャネルインデックスを確立する
S2
プッシュ情報がユーザの正常なラベルから生成される確率に従って、
ユーザによって受信されたプッシュ情報を傍受する
S3
プッシュ情報がユーザの検出ラベルから生成される確率に従って、
傍受したプッシュ情報をフィルタリングし、
プッシュ情報がユーザの検出ラベルから生成される確率が所定の閾値より高い場合は、
ユーザの検出ラベルを漏洩疑いラベルセットに追加する
S4
漏洩疑いチャネルIDの対応するリストを得るために、
漏洩疑いラベルセットに従ってチャネルインデックスを検索する
S5
プッシュ情報が、見つかったチャネルからのものかどうかを検出し、
そうである場合は、対応するチャネルを削除し、
残りのチャネルを漏洩疑いチャネルとして出力する
S6

【図２】


検出ラベル追加
モジュール
チャネル関連付け
モジュール
傍受モジュール
傍受情報
分析モジュール
チャネル
検索モジュール
出力モジュール

フロントページの続き

(72)発明者　ウェン，ジェン
中華人民共和国，ジャージャン　３１１１２１，ハンチョウ　ユ　ハン　ディストリクト，ウェスト　ウェン　イ　ロード　ナンバー９６９，ビルディング　３，５／エフ　アリババ　グループ　リーガル　デパートメント

審査官　岸野　徹

(56)参考文献　特開２０１２－１５０６５２（ＪＰ，Ａ）
特開２０１３－１６４７３９（ＪＰ，Ａ）
国際公開第２０１５／００１９６９（WO，Ａ１）
国際公開第２０１４／０２４９５９（WO，Ａ１）
特開２０１５－１７６２６３（ＪＰ，Ａ）
特開２００５－２２２１３５（ＪＰ，Ａ）
米国特許出願公開第２０１６／０３７３４４７（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／１６
Ｇ０６Ｆ　１６／２０
Ｇ０６Ｆ　１６／２８