

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4955107号

(P4955107)

(45) 発行日 平成24年6月20日 (2012. 6. 20)

(24) 登録日 平成24年3月23日 (2012. 3. 23)

(51) Int. Cl. F I
H O 4 L 12/56 (2006. 01) H O 4 L 12/56 4 O O B
H O 4 M 3/00 (2006. 01) H O 4 M 3/00 E

請求項の数 13 (全 13 頁)

(21) 出願番号	特願2010-519464 (P2010-519464)	(73) 特許権者	391030332
(86) (22) 出願日	平成20年8月6日 (2008. 8. 6)		アルカテルルーセント
(65) 公表番号	特表2010-536216 (P2010-536216A)		フランス国、75 007・パリ、 アブニ
(43) 公表日	平成22年11月25日 (2010. 11. 25)		ユ・オクターブ・グレアール、 3
(86) 国際出願番号	PCT/EP2008/060363	(74) 代理人	100062007
(87) 国際公開番号	W02009/021892		弁理士 川口 義雄
(87) 国際公開日	平成21年2月19日 (2009. 2. 19)	(74) 代理人	100140523
審査請求日	平成22年4月8日 (2010. 4. 8)		弁理士 渡邊 千尋
(31) 優先権主張番号	07290995.5	(74) 代理人	100103920
(32) 優先日	平成19年8月10日 (2007. 8. 10)		弁理士 大崎 勝真
(33) 優先権主張国	欧州特許庁 (EP)	(74) 代理人	100124855
			弁理士 坪倉 道明

最終頁に続く

(54) 【発明の名称】 I Pネットワーク内のトラフィックを分類するための方法およびユニット

(57) 【特許請求の範囲】

【請求項 1】

トラフィックを取得する段階および詳細なパケット分析の段階を含む、I P電気通信または企業ネットワーク上のトラフィックを分類するための方法であって、統計的に生成される決定木を使用してトラフィックを統計的に分類する段階と、詳細なパケット分析によって特定されていないトラフィックが、統計的な分類によって、既知のアプリケーション、具体的には暗号化されていないアプリケーションに属するものと認識された場合、トラフィックを統計的に分類する段階から、詳細なパケット分析の段階に、前記既知のアプリケーションを知らせる段階とを含み、詳細なパケット分析の段階は、取得されたトラフィックのパケットの内容に基づいて、特定のピアツーピアプロトコルを特定することを特徴とする、方法。

【請求項 2】

トラフィックを統計的に分類する段階が、詳細なパケット分析の後に実行され、このパケット分析によって特定されていないトラフィック、具体的には暗号化されたトラフィックのみに関係することを特徴とする、請求項 1 に記載の方法。

【請求項 3】

取得されたトラフィックから所定のパラメータが抽出され、前記パラメータが、決定木の少なくとも 1 つの節用の分岐基準を定義するために使用されることを特徴とする、請求項 1 または 2 に記載の方法。

【請求項 4】

10

20

所定のパラメータが、パケットサイズ、パケット間の時間間隔、パケット数、使用されるポート番号、パケット番号、所与のIPアドレスに関して異なるIPアドレスの数、を含む群から選択されることを特徴とする、請求項3に記載の方法。

【請求項5】

詳細なパケット分析の段階が、パケット間の間隔値とともに、対話型トラフィックを認識したか否かを、詳細なパケット分析の段階から、統計的に分類する段階に知らせる段階をさらに含む、請求項1から4のいずれか一項に記載の方法。

【請求項6】

決定木が二分木でないことを特徴とする、請求項1から5のいずれか一項に記載の方法。

10

【請求項7】

分岐基準としてエントロピが使用されることを特徴とする、請求項1から6のいずれか一項に記載の方法。

【請求項8】

誤り率を減らすことにつながらない条件を削除することによって、冗長な規則を削除するとともに各規則を一般化することによって、決定木を規則に変換する段階を含むことを特徴とする、請求項1から7のいずれか一項に記載の方法。

【請求項9】

決定木がC4.5またはC5.0のタイプのものであることを特徴とする、請求項1から8のいずれか一項に記載の方法。

20

【請求項10】

取得が、具体的にはルータ上で、パケットスニッファソフトウェアまたはパケット取得ソフトウェアを使用して実行されることを特徴とする、請求項1から9のいずれか一項に記載の方法。

【請求項11】

トラフィックを取得するための手段および詳細なパケット分析のための手段を含む、電気通信ネットワーク上のトラフィックを分類するための装置であって、統計的に生成される決定木を使用してトラフィックを統計的に分類するための手段と、詳細なパケット分析によって特定されていないトラフィックが、統計的な分類によって、既知のアプリケーション、具体的には暗号化されていないアプリケーションに属するものと認識された場合、トラフィックを統計的に分類する手段から、詳細なパケット分析のための手段に、前記既知のアプリケーションを知らせる手段とを含み、詳細なパケット分析のための手段は、取得されたトラフィックのパケットの内容に基づいて、特定のピアツーピアプロトコルを特定することを特徴とする、装置。

30

【請求項12】

内部ネットワークと少なくとも1つの外部ネットワークとの間のカットオフ位置に配置されることを特徴とする、請求項11に記載の装置。

【請求項13】

詳細なパケット分析のための手段が、パケット間の間隔値とともに、対話型トラフィックを認識したか否かを、詳細なパケット分析のための手段から、統計的に分類するための手段に知らせる手段をさらに含む、請求項11または12に記載の装置。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明はIP電気通信または企業ネットワークの技術的領域に関し、より具体的にはアプリケーションを特定、分類、およびフィルタリングするためのトラフィック制御に関するものである。

【背景技術】

【0002】

本発明は特に、暗号化されたストリームを用いるアプリケーションを分類するための技

50

法に関する：それは例えばピアツーピア型アプリケーション（特にV o I P S k y p e）や、h t t p sトンネル用の企業V P N（仮想私設ネットワーク）アプリケーションである。

【0003】

技術的には可能ではあるが、著作権によって保護される著作物にオープン、自由、かつ無制限にアクセスすることは、その性質上違法であり、経済的に危険であり、よって争う必要がある。N a p s t e rはMP3形式の音楽ファイルを共有することを許可しなかったが、ピアツーピアネットワークは現在特にビデオ、ゲーム、およびソフトウェアを交換することを可能にしている。高速であることは、1時間の映画をほんの数分でダウンロードできることを意味する。

10

【0004】

違法ダウンロードの証明を確立することは法律上の理由および技術上の理由から困難であり、それについては以下のように要約することができる。

【0005】

まず、ピアツーピア（すなわちP2P）交換ソフトウェアを使用するインターネットユーザは、著作財産および芸術的財産に関係する法律に関して必ずしも違反を犯しているわけではない。特定の著作物（公的行為、報道資料）はその性質上またはその意図する目的により、パブリックドメイン内にあり得る。さらに、経過した期間が原因で多数の著作物がパブリックドメイン内にある。ダウンロードは、商業的に利用する権利を必ずしも含意するわけではない見る権利または聞く権利に結びつけることができる。例えば、D a i l y M o t i o nのサイトは、インターネットユーザが自身のビデオを投稿し、他のインターネットユーザのビデオを自由に見ることを可能にする。ピアツーピアネットワークは、グリッドコンピューティング、テレビ会議、インスタントメッセージングなど、多様かつ完全に合法的な活動を可能にする。

20

【0006】

第2に、潜在的に違法なダウンロードの数は極めて多い。国庫経済政策総局（T r e a s u r y a n d e c o n o m i c p o l i c y g e n e r a l d i r e c t o r a t e）によれば、2005年にフランスでは、ピアツーピアネットワークの不定期ユーザが800万人、常用ユーザが750,000人いた。フランスでは、CNC（国立映画センタ）が2005年に違法な音声映像のダウンロード数を1日あたり100万と推定し、この数は映画館に行く観客数の2倍にあたる。OECDによれば、2004年には2003年の30%増である約1,000万人がピアツーピアネットワークを使用した（O E C D I n f o r m a t i o n T e c h n o l o g y O u t l o o k 2 0 0 4 : P e e r t o P e e r n e t w o r k s i n O E C D c o u n t r i e s）。

30

【0007】

第3に、今日最も一般的に使用されているソフトウェアは、インターネットユーザの匿名コミュニティによって作成されるオープンソースソフトウェアである。このソフトウェアに関してはもはや編集者はおらず、この状況は無論、ピアツーピアソフトウェア編集者の刑事上の責任が審理された最近の法的評決に起因している（オランダ最高裁判所2003年12月19日、B u m a / S t e m r a v . K a Z a A、米国最高裁判所2005年6月27日、M e t r o - G o l d w y n - M a y e r v . G r o k s t e r）。K a m e l e o n、M u t e、S h a r e、A n t s、F r e e n e t、G N U n e t、I 2 Pなどのこのソフトウェアは、ユーザをフィルタおよび特定することを非常に困難にする暗号化システムが備わっている。例えばT O R（The Onion Ring）など、匿名での接続を可能にするシステムもある。

40

【0008】

第4に、インターネットユーザは、私的コピーするために合法的な情報源から著作物をダウンロードできなければならないが、通常はその情報源が合法的であることを確認する手段を有さない。したがって、例えばパリ大審裁判所（上級裁判所）（案件番号0504

50

090091)は、2005年12月8日の判決で、10億を超える音楽ファイルへのアクセスを可能にしているピアツーピアソフトウェアKazaaは、著作権によって保護された著作物のファイルとパブリックドメイン内にある著作物のファイルとの間の区別をできなくすることを認めた。

【0009】

第5に、インターネットストリームを全般的に処理することは、プライバシーを保護する法規に違反してはならない。例えば、保護された著作物を利用できるようにしているインターネットユーザのIPアドレスの集合は、原則として法的手続きとの関連でしか明示可能にすることができない。広範囲の侵害(何千もの著作物のダウンロード)との闘いは、個人データを収集することを正当化し得る。データが匿名の状態にされる限り、統計的10
目的でピアツーピアネットワークを継続的かつ自動的にスキャンすることも可能である。法的または行政的な通信傍受も当然可能である。しかし、ピアツーピアネットワークを自動的、網羅的、かつ非匿名的に処理することはプライバシーの尊重に反する。

【0010】

最後に、IPアドレスを知ることは、必ずしもインターネットユーザを特定することを意味するとは限らない。

【0011】

違法ダウンロード、具体的には著作権を無視した著作物のダウンロードには2つの異なる側面がある：

- 商業的な目的での大量ダウンロード。20
- 個人レベルまたはコミュニティレベルでの不定期のダウンロード。

【0012】

商業的な目的での大量ダウンロードは、特定の権利侵害では間違いなく一般的な制止技法によって撲滅に取り組むことができる。

【0013】

しかし、不定期のダウンロードに関しては、切り離してとると限られた問題しかもたらない多数の違反を適合処理するための技術的方策を見つける必要があり、これらの技術的方策はプライバシーを保護する施行中の法律に準拠する必要がある。

【0014】

ポート認識は推測的に想像でき、ISP(インターネットサービスプロバイダ)のネットワーク上に導入されるルータがこの特徴を提供する(Cisco、Juniper、Extreme Network、Foundryのルータ)。例えば、ポート1214はFasttrack(Kazaa)のデフォルトポートであり、ポート4661、4665、4672はeDonkeyおよびeMuleアプリケーションによりデフォルトで使用され、ポート6346はBearShare、Gnutella、LimeWire、およびMorpheusのデフォルトポートである。デフォルトポートはDirect Connect、WinMx、BitTorrent、MP2Pアプリケーションにも関連する。しかし、ポート認識はピアツーピアトラフィックを特定するには十分ではない：ピアツーピアアプリケーションでの構成可能ポートの使用、ポートの動的割り当て、ピアツーピアアプリケーションへの標準ポート(例えばポート53DNSやポート80HTTP)の使用。ほとんどのP2Pアプリケーションは、どのポートをP2Pトラフィックに割り当てることに決めるのかを自らのユーザに手動で選択する権限を与える。P2Pアプリケーションは、内部管理者が空けておかなければならないポート、例えばデフォルトではウェブサイト専用のポート80などをしばしば使用する。30

【0015】

従来技術で提案された主なインターネットトラフィックフィルタリングの技術的方策は、3つの主要なカテゴリによって分類することができる：プロトコルのフィルタリング、コンテンツのフィルタリング、およびインターネットユーザの端末(station)に対するフィルタリングである。40

【0016】

プロトコルフィルタリング解決策は、例えばそれがピアツーピアストリームであるかどうかを判定するために、インターネットユーザのクライアント端末から交換されるネットワークフレーム内のシグネチャを認識することに基づく。プロトコルは、それに従ってアプリケーションまたはサービスがネットワーク上でデータを交換する規則を定義する。これらの規則は、一連の特徴的ビットがエンベロープ（ヘッダ）を越えて各パケット内に配置されることをもたらす。この順序はパケットの性質に応じて、しかしコンテンツからは独立して可変である。このプロトコルフィルタリングは、例えば下記の事項が区別されることを可能にしなければならない：

－ 標準的プロトコル：s m t p（簡易メール転送プロトコル）、h t t p（ハイパーテキスト転送プロトコル）。

10

－ 従来型ピアツーピアプロトコル：（企業M e t a M a c h i n eによって世に出され、2006年9月以降即座に閉鎖された）e D o n k e y、B i t T o r r e n t、F a s t t r a c k（K a z a a、K a z a a L i t e、I M e s h）。

－ 暗号化されたP2Pプロトコル：F r e e n e t、S o f t E t h e r、E a r t h S t a t i o n 5、F i l e t o p i a。

【0017】

現在、プロトコルフィルタリングは、詳細なパケット分析技法、具体的にはD P Iすなわちディープパケットインスペクションによって実施されている。このパケット分析は、C i s c oによるP D M LルータまたはJ u n i p e rによるN e t s c r e e n - I D Pにおいてネイティブ状態で提案されている。特定の企業は、I S Pが、そのネットワークに関するカットオフ位置に挿入される追加のボックスを使用することを提案する（A l l o t b o x N e t e n f o r c e r K A C 1 0 2 0、P a c k e t e e r P S 8 5 0 0 I S P）。C i s c oもボックスを提供する（C i s c o P _ C u b e）。

20

【0018】

クライアントソースコードの可用性は、特にオープンソースモードまたは同等のモードでの開発に関して、これらのプロトコルが実装される方法を分析するために使用され、妥当な場合、プロトコルの上流部分（接続、ネゴシエーション、暗号化モードへの転換）についての認識を可能であれば導入するために使用される（例えばe M u l e スクランプルバージョンに関しては可能ではない）。そのような解決策は例えばA l l o tによって導入されており、A l l o tはとりわけS o f t E t h e r、E a r t h S t a t i o n 5、およびF i l e t o p i aのプロトコルをフィルタするように主張する。

30

【0019】

プロトコルによるフィルタリングにはいくつかの不都合がある。

【0020】

まず、このフィルタリングの標的とされるプロトコルは、合法データおよび違法データのどちらも運ぶことができるため、必ずしも違法活動の兆候ではない。

【0021】

さらに、暗号化を実施することは、ネットワークフレームの検出を動作不能にし、またははるかに複雑にする可能性がある。この暗号化は、例えば接続フレームまたはファイルの接尾辞をアップグレードすることにより、ピアツーピアプロトコルを修正することによって導入することができ、これはインターネットユーザのワークステーションおよびサーバ上にインストールされたクライアントアプリケーション（K a z a a、e D o n k e y、t r a c k e r s B i t T o r r e n t）を修正することを意味する。暗号化は、例えばS S L / H T T P SまたはS S H（セキュアシェル）型のトンネリングプロトコルを使用することによっても導入できる。特定のピアツーピアプロトコル、具体的にはF r e e N e t（日本のプログラムW i n n y）、S S L（S o f t E t h e r、E a r t h S t a t i o n 5、F i l e t o p i a）、S S H（S o f t E t h e r）は既に暗号化されている。

40

【0022】

さらに、I P V 6に向けたインターネットプロトコルの進化は、利用可能なアドレス指

50

定範囲の拡張とは別に、具体的には I P S e c プロトコルおよび暗号化機能が一般化されながら、T C P / I P のセキュリティおよび認証機能についての進化をもたらす。

【0023】

コンテンツフィルタリングは、ストリームをコンテンツレベルの要素に基づいて特定し、必要な場合はフィルタするために使用される：

- － 生の音楽ファイル、例えばW A V（ウェーブフォーム）、M P 3、M P C。
- － D R Mソリューションにリンクされた形式の音楽ファイル（例えばA A C、W M A、A t r a c +）。
- － C Dの画像または生の音楽ファイルのセットを含むアーカイブ（例えばZ I P、R A R、A C E）。

10

【0024】

企業A u d i b l e M a g i cは、コンテンツフィルタリングツール（C o p y S e n s e b o x）を提供する。企業A d v e s t i g oも、文献F R 2 8 8 7 3 8 5に記載のコンテンツフィルタリング技法を提供する。

【0025】

インターネットユーザの端末上でのフィルタリングは、インターネットユーザの端末上の1組の機能へのアクセスが特定され、必要な場合は禁止されることを可能にする。これらの機能は以下のレベルにあり得る：

- － ネットワーク、例えば特定のポートの閉鎖や、D N S名のリストまたは指標付きI Pアドレスとの交換の禁止。
- － コンテンツ、例えばアプリケーション（ダウンロード後のP 2 Pクライアント）によりM P 3形式のファイルが作成される場合における検出およびアラート／禁止。
- － アプリケーション、例えばクライアント端末（例えばe M u l eクライアント）上で特定のアプリケーションが起動されることの検出およびアラートまたは禁止。

20

【0026】

クライアント端末上でフィルタリングするための多様なツールが利用可能である：ファイアウォール、C i s c o C S AやS k y R e c o n型のセキュリティソリューション、C y b e r P a t r o l型のペアレナタルコントロールソリューション。

【0027】

要求に応じてのフィルタリングにはいくつかの不都合がある。

30

【0028】

P 2 Pソフトウェアのユーザは、必ずしも自身をフィルタしたいと思うわけではなく、親は自分の子供に対してフィルタ済みのサブスクリプションを強いるのが困難である。要求に応じてI S Pレベルでフィルタリングすることは、すべてのフィルタ済みのサブスクリプションを処理できるプラットフォームへのタップ（例えばルーティングやトンネル）を作り出すことを意味する。インターネットユーザの端末上でのフィルタリングは、トラフィックの観測および分析、ストリームの体系的なフィルタリング、またはレーダのポジショニングを許さない。

【先行技術文献】

【特許文献】

40

【0029】

【特許文献1】 仏国特許第2 8 8 7 3 8 5号明細書

【非特許文献】

【0030】

【非特許文献1】 O E C D I n f o r m a t i o n T e c h n o l o g y O u t l o o k 2 0 0 4 : P e e r t o P e e r n e t w o r k s i n O E C D c o u n t r i e s

【非特許文献2】 S u b h a b r a t a S e nら、「A c c u r a t e , s c a l a b l e i n - n e t w o r k i d e n t i f i c a t i o n o f t h e P 2 P t r a f f i c u s i n g a p p l i c a t i o n s i g n a t u r e s」（P

50

proceedings of the 13th international conference on world wide web, New York 2004)
【非特許文献3】Karagiannisら、「Transport layer identification of P2P traffic」(Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, Taormina, Italy, 2004)

【発明の概要】

【発明が解決しようとする課題】

【0031】

本発明は、言及した問題の少なくとも一部に対する解決策を提供することを目的とする。本発明は具体的には、IP電気通信ネットワークおよび企業ネットワークにおけるサービス品質QoSの問題に対する解決策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】

【0032】

このために、本発明は第1の態様によれば、IPネットワーク／電気通信または企業ネットワーク上のトラフィックを分類するための方法に関し、前述の方法はトラフィックを取得する段階および詳細なパケット分析（具体的にはDPI）の段階を含み、前述の方法は統計的に生成される決定木を使用してトラフィックを統計的に分類する段階を含む。

【0033】

有利には、トラフィックシグネチャを統計的に最適化することに基づき得るこのトラフィックを統計的に分類する段階は、詳細なパケット分析（具体的にはDPI）の後に実行され、このパケット分析によって特定されていないトラフィック、具体的には例えば暗号化されたピアツーピアプロトコルを実装する暗号化されたトラフィックのみに関係する。

【0034】

有利には、この方法は、詳細なパケット分析によって特定されないトラフィックが、統計的に最適化されたシグネチャに基づき得る統計的分析により、知られているアプリケーション、具体的には暗号化されていないアプリケーションに属すると認識される場合、トラフィックシグネチャを最適化するために、詳細なパケット分析の段階と、決定木によりトラフィックを統計的に分析する段階との間に、情報を交換する段階を含む。

【0035】

有利には、一方ではこの決定木は二分木ではなく（このことは区別を最適化し）、他方では分岐基準としてエントロピが使用される。一実装形態では、この決定木はC4.5またはC5.0のタイプのものである。有利には、この木は決定木を規則に変換する段階を含む。

【0036】

有利には、この方法は、トラフィックを分類することを可能にする前述の規則を統計的に最適化する段階を含む。

【0037】

取得は、具体的には例えばルータ上で、パケットスニッフアソフトウェアを使用して、またはデータベース内にコピーすることによって実行される。取得された要素から所定のパラメータが抽出され、次いでこれらのパラメータは、決定木の少なくとも1つのノード用の分岐基準を定義するために使用される。この所定のパラメータは、次のものを含む群から選択される：パケットサイズ、パケット間の時間間隔、パケット数、使用されるポート番号、パケット番号、所与のIPアドレスに関して異なるIPアドレスの数。

【0038】

本発明は第2の態様によれば、IPネットワーク／電気通信または企業ネットワーク上のトラフィックを分類するための装置に関し、前述の装置はトラフィックを取得する手段および詳細なパケット分析のための手段を含み、前述の装置は統計的に生成される決定木を使用してトラフィックを統計的に分類するための手段を含む。トラフィックは、統計的

10

20

30

40

50

に生成される決定木を変換することによりもたらされる規則によって分類される。この決定木は、分類すべきトラフィックと比較することができ、規則の形で定義されるトラフィックシグネチャを、自動的に定義するために使用される統計的ツールである。

【0039】

有利には、取得、パケット分析、および統計的分類のための手段は1つのボックスに統合される。一実装形態では、この装置は、内部ネットワークと少なくとも1つの外部ネットワークとの間のカットオフ位置に配置することができる。

【0040】

本発明の他の目的および利点は、以下の説明を添付の図面を参照して読むことにより明らかになるであろう。

【図面の簡単な説明】

【0041】

【図1】インターネットトラフィックを分類するための規則を統計的に生成するための方法の一実施形態を示す図である。

【図2】既に生成された規則を使用して通信を監視し、傍受するための方法の一実施形態を示す図である。

【発明を実施するための形態】

【0042】

図1をまず説明する。

【0043】

データフロー1、例えばインターネットトラフィックが取得される。この取得F0は、例えばtcpdumpやWiresharkなど、最も一般的なプロトコルを認識するパケットスニフファソフトウェアを使用して、またはWinpcapなどのパケット取得ソフトウェアによって実行される。妥当な場合、すべてのトラフィックが、例えばルータにおいてコピーされる。

【0044】

説明を簡潔にするために、この説明の残りの部分では、取得されるインターネットトラフィックは2種類のトラフィックを含むものとみなし、それはすなわち次のトラフィックである：

- 特徴付けられるアプリケーションのトラフィックではないトラフィック。
- 暗号化された（したがって、DPIによって調べることができない）トラフィック、またはシグネチャが知られておらず、このトラフィックを後に検出できるようにするために決定木が最初に統計的に生成されるトラフィック。これはスクランブルeMuleピアツーピアアプリケーションのトラフィックに当てはまり得る。

【0045】

取得F0の後に、取得された要素が、上述の2種類のトラフィックのうちの1つにそれぞれが対応する2つのデータベース2、3に記憶される。

【0046】

データベース2、3に記憶される取得された要素は、要約される同法を使用してF1で変換される。この取得された要素から特定の情報が抽出される。この情報は例えば以下のものである：

- OSI（開放型システム間相互接続）モデルのレイヤ3およびレイヤ4に関しては：IP、TCP、UDP、ICMP情報。
- パケットサイズ、パケット数、ポート番号、パケット間の時間間隔、パケット番号、あるIPアドレスが関係する固有のIPアドレスの量。

【0047】

抽出される情報の選択は、P2Pプロトコルについての得られた知識に起因する。例えば、eDonkey、Fasttrack、WinMx、Gnutella、MP2P、およびDirect Connectは、原則としてTCP（伝送制御プロトコル）およびUDP（ユーザデータグラムプロトコル）の両方をOSIモデルのレイヤ4用のラン

10

20

30

40

50

スポーツプロトコルとして使用する。IPアドレス／ポート番号の対の経時的な進化は、特定のP2Pプロトコルにとっては普通である、ポート番号のランダム変化を示し得る。

【0048】

詳細なパケット分析技法は、それ自体で知られている。例えば次の文献を参照されたい：Subhabrata Senら、「Accurate, scalable in-network identification of the P2P traffic using application signatures」(Proceedings of the 13th international conference on world wide web, New York 2004)、Karagiannisら、「Transport layer identification of P2P traffic」(Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, Taormina, Italy, 2004)。パケット分析は、一定のピアツーピアプロトコルを特定することを可能にする。Gnutella、eDonkey、BitTorrent、およびKazaaプロトコルに関するシグネチャの例が、上記に言及したSubhabrata Senらによって発表された文献内に示されている。ソフトウェアBlubsterまたはPiolerを使用するMP2Pネットワークでは、ランダムポート上のパケット分析は応答「SIZE<<file size in bytes」を明らかにする。

【0049】

実際、詳細なパケット分析またはポート検出が、使用されるプログラムを特定できるようにしない場合、TCP/UDP/ICMPメッセージレベルにおいてアクセス可能な特定の情報は、それでもなお関連したままでいる：この情報は通常は暗号化によって変更されない。ストリーム対称暗号化アルゴリズム（ストリーム暗号）はテキストのサイズを変更しない。TCP/UDP/ICMPメッセージを分ける時間間隔は、圧縮後は比例的に一定である。

【0050】

変換モジュール4によってF1で変換された情報は、前述の2種類のトラフィックのうちの1つにそれぞれが対応する2つのデータベース5および6に記憶される。

【0051】

データベース5および6に記憶される情報を、それらのデータベースが参照するトラフィックの種類を知らながら決定木が入力として受け取り、こうして決定木生成モジュール7によりF2で木が生成される。この木はデータベース8aに記憶される。

【0052】

決定木により分類を実施することにはいくつかの利点がある。決定木は明確な分類規則を提供し、異種のデータ、欠落した項目、および非線形効果をサポートする。

【0053】

それ自体で知られている限り、この決定木は様々なノード分岐基準を使用して構築することができる： χ^2 基準、Giniインデックス、Twoingインデックス、エントロピ。

【0054】

有利には、使用される分岐基準はエントロピであり、木の各ノードについて $\sum f_i \log(f_i)$ によって定義され、ただし f_i $i=1$ から p は p クラスのノードにおける相対頻度である。

【0055】

一例として、この木の最初のノードであるルートは4番目のパケットのサイズであり、この変数がトラフィックを2つのクラスAおよびBに分けている。

【0056】

次いで第1のクラスAが、第1のパケットと2番目のパケットとを分ける間隔の持続時間によって分類され、こうして2つのサブクラスA1およびA2を定義する。第1のサブ

10

20

30

40

50

クラスA1では、2番目のパケットのサイズにより終端ノード（リーフ）が定義され、このリーフに割り当てられる個人は、例えばeMule暗号化プロトコルを使用したファイル検索に対応する。

【0057】

第2のクラスBは、3番目のパケットのサイズによって分類され、こうして2つのサブクラスB1およびB2を定義する。第1のサブクラスB1では、第1のパケットのサイズにより終端ノード（リーフ）が定義され、このリーフに割り当てられる個人は、例えばhttpプロトコル（ハイパーテキスト転送プロトコル）のSSL（セキュアソケットレイヤ）によって保護されたバージョン、https通信プロトコルによって確立されるウェブトラフィックに対応する。第2のサブクラスB2では、5番目のパケットのサイズによって終端ノードが定義され、このリーフに割り当てられる個人は、例えばeMule暗号化プロトコルを使用してダウンロードされるファイルに対応する。

10

【0058】

一例として、使用される木は1998年からC5.0の木であり、これはJ. Ross Quinlanによる自身の以前の木ID3（1986年）およびC4.5（1993年）を完成させたものである。この木は、例えばEnterprise Miner（SAS）やClementine（SPSS）に実装される。この木は、See5という名でWindows（登録商標）プラットフォームに対しても売り込まれており、用語C5.0はUnix（登録商標）プラットフォームで使用されている。この木は学習標本から構築され、その後、各ノードの誤り率および信頼区間に基づいて剪定される。

20

【0059】

このC5.0の木は二分木ではない：各段階で1つの母集団を3つ以上の母集団に分けることができる。

【0060】

このC5.0の木も、その前のバージョンC4.5と同様に次の利点を有する：ある手順が木を規則のセットに変換する（規則作成のモジュール8bを参照されたい）。冗長な規則は削除され、このプログラムは、誤り率を減らすことにつながらない条件を削除することにより、各規則を一般化することを試みる。規則のリストはデータベース8cに記憶され、学習過程が進むにつれ、モジュール8dがこれらの規則を単純化することを確実にする。

30

【0061】

この決定木のノードは、トラフィックを分類する目的に基づいて定義される。

【0062】

例えば、目的が特定のプロトコルの使用を検出することである場合、少数の基準が必要となる。したがって企業は、暗号化されたトラフィックにおいて、VoIP（例えばSkype）によって認可された通信とそのVoIPが認可していない転送とを、認可されていないそれらの転送が特定のプロトコル（例えばeMule暗号化プロトコル）によって実行されるものとして区別することができる。

【0063】

目的が特定の種類の行動（ダウンロード、または逆のアップロード）を検出することである場合、より難解な決定木を形成するためにさらなるパラメータが必要になる。

40

【0064】

規則の数は、有利には装備されるネットワークの資源に適合させることができる。

【0065】

ある有利な実装形態では、統計的分析モジュール11は、今のところは特定されていないストリームが、モジュール10により通常検出されるアプリケーションに属している様子であることをモジュール10に知らせる。この方法および装置の性能は次のようにして最適化される：最大数のストリームが最適な段階によって処理される。一例として、暗号化されていないストリームを検知しておらず、よってモジュール11に送信した、NIDS侵入検知システム（Bro、Snort、Shadow、NFR）をモジュール10が

50

含む場合、モジュール11は、モジュール10が自らのその後の分析を微調整するための情報をモジュール10に送信することができる。

【0066】

ある有利な実装形態では、モジュール10が、統計的分析モジュール11に情報を与え、最適化された規則の選択を可能にするNIDS侵入検知器（例えばBRO）を含む。例えばモジュール10は、パケット間の間隔値を用いて、対話型のトラフィックを認識したかどうかをモジュール11に知らせる。

【0067】

本発明は、誤判定がほとんどない状態でアプリケーションを厳密に特定できるようにする。データマイニング学習フェーズ中、標的とされるアプリケーション（例えばSkypeやeMule）に対して数千の接続が行われ、決定木は、これらのアプリケーションに関連する暗号化されたトラフィックについての最も固有の情報を特定することを可能にする。

10

【0068】

トラフィックを分類するためのこの方法および装置は、有利には以下のように使用することができる。

【0069】

トラフィックブロッキング

自身の会社の内部ネットワークに接続されたボブは、ピアツーピアプロトコルeMuleを使用して友人のアリスとファイルを交換したい。eMuleを使用することは会社によって許可されていない。ボブは自分のコンピュータまたは端末上でeMuleアプリケーションを起動する。ボブのコンピュータからのトラフィック9は暗号化されている。このトラフィック9が、詳細なパケット分析モジュール10に到達する。トラフィック9は暗号化されているので、このモジュール10は、具体的にどのアプリケーションがトラフィック9に関連するのか特定できない。モジュール10はこの未知のトラフィックを、統計的分析モジュール11に宛てる。このモジュール11は、検出規則、決定木から前に確立された規則のデータベース12に接続されている。これらの検出規則は、自動規則生成モジュール13によって前に更新されている。分析モジュール11はeMuleプロトコルを認識し、送信元IPアドレス、宛先IPアドレス、送信元ポート番号、および宛先ポートを特定する。この情報はネットワーク管理システムに送信され、この管理システムはこのような場合に計画されている方策を適用する：

20

- ファイアウォール、ルータ、侵入を検知または阻止するための装置によるトラフィックの傍受、ブロック14。
- 良いサービス品質QoSを目的としたボブへのメッセージ送信15。
- または他の任意の所定の方策16。

30

【0070】

トラフィック傍受

法的判断がボブの通信を傍受することを許可した。この傍受を担う機関は、当該アプリケーションに関して決定木の規則が更新され、ボブの端末に関連するルータ上に実装されていることを確実にする。ボブの端末上で送受信されるデータストリームが、そのルータにコピーされる。この暗号化されたストリームはまず、詳細なパケット分析（具体的にはディープパケットインスペクション）にかけられる。次いで統計的分析が実行される。この分析の結果次第で、傍受を担う機関は、ボブによって送受信されるデータのコンテンツを復号するかどうか決定する。この種の使用では、トラフィックを暗号化するアプリケーションを知り、ストリームのパラメータ（ポート番号およびIPアドレス）を特定することが重要である。

40

【0071】

本発明にはいくつかの利点がある。

【0072】

本発明は、政府機関およびインターネットサービスプロバイダが効果的、迅速、かつ正

50

フロントページの続き

- (72)発明者 クルビー, ロラン
フランス国、９１６２０・ノゼー、ルート・ドウ・ビルジユスト、アルカテルルーセント・ベル
・ラブズ・フランス
- (72)発明者 マルタン, アントニー
フランス国、９１６２０・ノゼー、ルート・ドウ・ビルジユスト、アルカテルルーセント・ベル
・ラブズ・フランス
- (72)発明者 アループド, ハイサム
フランス国、９１６２０・ノゼー、ルート・ドウ・ビルジユスト、アルカテルルーセント・ベル
・ラブズ・フランス
- (72)発明者 アンスー, アルノー
フランス国、９１６２０・ノゼー、ルート・ドウ・ビルジユスト、アルカテルルーセント・ベル
・ラブズ・フランス

審査官 安藤 一道

- (56)参考文献 北村 強、静野 隆之、岡部 稔哉, フロー挙動分析技術に基づくアプリケーション識別手法,
電子情報通信学会技術研究報告 Vol. 105 No. 470, 社団法人電子情報通信学会,
2005年12月 8日, p. 13-16

(58)調査した分野(Int.Cl., DB名)

H04L 12/56

H04M 3/00