様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年５月１日現在

機関番号：１１３０１
研究種目：挑戦的萌芽研究
研究期間：2010 ～ 2011
課題番号：22650010
研究課題名（和文）
あらゆる不正転送を検知可能な超軽量型ストリーミング流通監視システム
研究課題名（英文）
Streaming Content Leakage Detection System for Encrypted Streams
研究代表者
加藤　寧（KATO NEI）
東北大学・大学院情報科学研究科・教授
研究者番号：00236168

研究成果の概要（和文）：
本研究グループでは、動画特有のトラヒックパターンの利用によって動画が視聴されているか否かを検知する技術を研究してきた。しかし、近年ストリーミング動画にも暗号化が施される場合がほとんどとなり、既存の技術では暗号化へのロバスト性が低く、新たな手法の考案が大きな課題であった。そこで本研究では、暗号化ストリーミングに対応できる新たな視聴検知技術を確立し、その性能と有効性を評価した。

研究成果の概要（英文）：
Our research group had previously proposed streaming content leakage detection technology based on the fact that different contents exhibit different traffic patterns when they are being delivered over the networks. By comparing traffic patterns obtained near the content server with those measured at egress nodes, these conventional systems are able to detect the stream leakage to the external network. Although our fundamental techniques can be enhanced by adopting an advanced traffic pattern generation algorithm, their robustness to encrypted streams requires further improvements. In order to address this problem, we envision an enhanced mechanism for dealing with encrypted traffic. We also empirically evaluate the performance of the proposed scheme and validate its effectiveness.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２０１０年度	1,100,000	0	1,100,000
２０１１年度	1,100,000	330,000	1,430,000
総　計	2,200,000	330,000	2,530,000

研究分野：総合領域
科研費の分科・細目：情報学、計算機システム・ネットワーク
キーワード：ストリーミング、動画視聴、コンテンツ、トラヒックパターン、暗号化

１．研究開始当初の背景

近年のネットワークの発展により、大容量なリアルタイムコンテンツが流通するようになってきている。利用者が視聴するにあたっては、予め利用登録を行い、利用時には個人認証を行った上、コンテンツを暗号化して配信するのが一般的である。しかし、利用者の意図的あるいは不注意による ID やパスワ

ードの漏洩により、リアルタイムストリーミングの健全な流通に対し大きな障害となっている。この問題に対処するための方法として、ネットワーク上でストリーミングを監視し、パケット内の情報を利用して、視聴が不正かどうかを監視する仕組みが用意されているが、近年ストリーミングトラヒックの殆どが暗号化されており、パケットの中身を参照することは不可能となってきている。また、リアルタイムコンテンツがネットワーク上のフローとし高速に流れているため、ネットワークのエッジルータなどで中身を解析するには負荷が高いという問題も存在する。

２．研究の目的

本研究では、あらゆるストリーミング動画を視聴検知可能な次世代コンテンツ流通監視技術について提案する。あらゆるストリーミングに対応するためには、近年普及している暗号化トラヒックへの対応が必須である。ネットワーク上で流れるリアルタイムな動画、つまりストリーミングはこれまで一旦暗号化されてしまうと外からそれは何の動画であるかは知ることは出来なかった。暗号化はコンテンツの著作権保護に必要な技術であるが、しかし裏を返せば、コンテンツの流通の状況を隠蔽する事にもなり、悪用された時問題となる。そこで本研究では、これまでになかった暗号ストリーミングの可監視性に挑戦する。具体的には、暗号化されたストリーミングにおいても、流通しているトラヒックが何の動画であるかを正確に推測する技術を開発し、ストリーミングコンテンツの悪用防止技術の向上に貢献することを目標とする。

３．研究の方法

秘匿すべきコンテンツの外部流出を防ぐためには、パケットフィルタリングが利用されている。しかし、悪意のあるユーザによって宛先 IP アドレスやポート番号を変更され、コンテンツが外部ネットワークへ再配信される場合、トラヒックの遮断は困難である。そこで、以上のようなコンテンツの外部流出対策として、トラヒック量に関する情報のみを用いてストリーミング動画の流出を検出する手法が提案されている。従来手法では、動画配信サーバに近い場所にあるルータとエッジルータにおいて観測されるトラヒック量の変化を、トラヒックパターンとして生成する。そして、サーバ側及びエッジルータ側のトラヒックパターンを比較し、その結果に基づいて動画の流出判定を行う。

本システムにおいて、ストリーミング動画は可変ビットレート形式によって配信されていると仮定される。このため、トラヒックパターンは各動画において固有のものとなり、サーバ側及びエッジルータ側におけるパターンの比較によってストリーミング動画の流出を検出することができる。図 3.1 に動画検出の概要を示す。このシステムを用いた従来手法として T-TRAT (Time slot-based TRAitor Tracing) 、 P-TRAT (Packet size-based TRAitor Tracing)、及び DP-TRAT (DP matchaed TRAitor Tracing)が提案されており、これらの従来手法は、トラヒックパターンの生成及びパターン同士の類似度の計算方法が異なっている。図 3.1 において、秘匿すべきコンテンツが正規ユーザに配信されており、同時に正規ユーザが P2P ストリーミングソフトウェアなどを用いて、外部ネットワークの非正規ユーザへコンテンツを再配信している。この時、サーバ側のルータ及びエッジルータにおいてトラヒックを観測し、その情報を管理サーバへ送信する。管理サーバでは得られたトラヒック情報を用いてトラヒックパターンを生成し、パターン同士のマッチングを行うことによって動画の流出判定を行う。

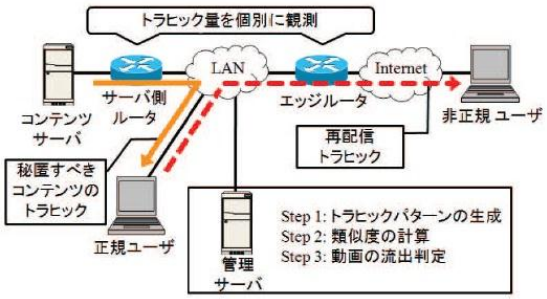


図 3.1: ストリーミング動画検出システム

トラヒック量を特定の条件に基づいて区切ることにより、トラヒックパターンを生成する。ここで、区切られた各トラヒック量をチャンクと表現する。

各手法におけるトラヒック量の区切りの条件を説明する。T-TRAT では一定時間によりトラヒック量を区切る。しかしこの方法では、パケットが区切りとなる時間以上遅延した場合、本来入るべきチャンクにパケットが入らず、パターンに歪みが発生して検出性能が低下する。遅延変動により性能が低下する問題を解決した手法が P-TRAT 及び DP-TRAT であり、トラヒック量の区切りを特定パケットの到着時と変更している。ここで特定パケットの例として、200byte 以下のパケットが利用されている。この手法ではトラヒック量の区切りがパケットの到着時間に依存しないため、遅延変動による影響を無効化したパターンが生成可能である。

以上のように生成されるパターンを行列式で表すと、式 (3.1) となる。

$$X_N = (x_1, x_2, \cdots, x_N)^t \qquad (3.1)$$

ここで、x_i は i 番目のチャンクにおけるトラ

ヒック量であり、N は全チャンク数を示している。また、パターンマッチングを行う際にパターンの長さが揃っている必要はなく、2つのパターンに重なる部分が存在すれば動画の判定は可能である。ルータによるトラヒックの観測時間を減らすことは、検出性能を低下させる可能性があるが、同時に観測による計算負荷も減少させる。そこで、多くのトラヒックが通過すると考えられるエッジルータでは、サーバ側のルータと比較して観測時間を短くしている。

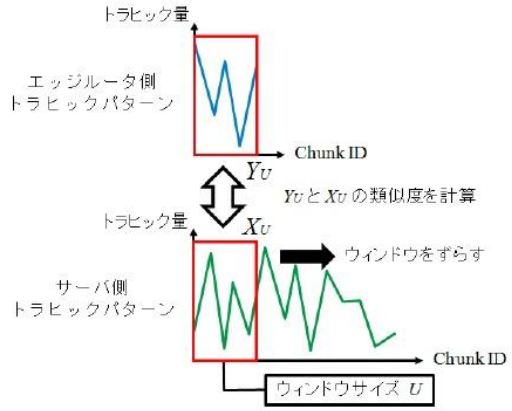


図 3.2: トラヒックパターンのマッチング

図 3.2 にトラヒックパターンのマッチング方法を示す。前述の通り、一般的にサーバ側トラヒックパターン X_s とエッジルータ側トラヒックパターン Y_u の長さは異なっており、$U \leq S$ となっている。そこで、スライディングウィンドウ方式によるサーバ側パターンの切り出し及びパターンのマッチングを行う。この方法では、初めにエッジルータ側パターンと同じ長さのウィンドウを作成する。これを用いて X_S から X_U を切り出し、エッジルータ側パターンとマッチングを行う。以上の操作を、ウィンドウをずらしながら $(S - U + 1)$ 回の全ての組み合わせにおいて繰り返す。この時、パターン同士のマッチングにおいて類似度が高い組み合わせが存在する場合、秘匿すべき動画の流出が発生していると判断する。

次に、各従来手法におけるパターン同士の類似度の計算及び流出判定の方法について説明する。T-TRAT 及び P-TRAT では、類似度の計算に相互相関係数を用いている。エッジルータのトラヒックパターン Y_U と、伐り出されたサーバ側のパターン X_U の相互相関係数 R_{XY} は以下に示す式(3.2)の通り定義される。

$$R_{XY} = \frac{X_U'^{t} \cdot Y_U'}{\sqrt{\|X_U'\|^2 \cdot \|Y_U'\|^2}}, -1 \leq R_{XY} \leq 1 \quad (3.2)$$

ここで、X_U’ 及び Y_U’ は平均 0、分散 1 に標準化されたパターンであり、式(3.3)と定義される。

$$X_U' = \begin{pmatrix} (x_1 - \bar{x})/s_x \\ (x_2 - \bar{x})/s_x \\ \vdots \\ (x_U - \bar{x})/s_x \end{pmatrix}, Y_U' = \begin{pmatrix} (y_1 - \bar{y})/s_y \\ (y_2 - \bar{y})/s_y \\ \vdots \\ (y_U - \bar{y})/s_y \end{pmatrix} \quad (3.3)$$

ここで、x 及び y は各パターン X_U、Y_U における平均の値を示しており、s_x 及び s_y は標準偏差を示している。また、R_{XY} の範囲は $-1 \leq R_{XY} \leq 1$ となっており、値が大きいほどパターンが類似しているといえる。そこで判定閾値 θ を導入し、この θ を越えるような R_{XY} が存在する場合に動画の流出が発生したと判断する。判定閾値について、T-TRAT では動画の判定ミスを減らすため、チェビシェフの不等式に基づいて動的に決定している。チェビシェフの不等式によると、ある分布において平均 μ から標準偏差 σ の k 倍以上離れた値は、全体数の $1/k^2$ を越えることはない。この考えに基づき、平均値の周りに標準偏差の 4 倍の幅を想定すると、その中には全要素の約 94%が含まれることになる。そこで、類似度の平均 μ_R にその標準偏差 σ_R の 4 倍の値を判定閾値としている。しかし、相互相関係数による類似度の定義域は $-1 \leq R_{XY} \leq 1$ であるため、判定閾値 θ は以下の式(3.5)のように定義される。

$$\theta = \min(\mu_R + 4\sigma_R, 1.0). \quad (3.5)$$

一方で P-TRAT においては、異なる動画のマッチングの際に相互相関係数の値が極めて小さい値になるため、判定ミスが発生しにくく、判定閾値として固定の値を用いている。

また、DP-TRAT では、類似度の計算に DP マッチングを用いている。この方法はパターンマッチングの際に、一方のパターンを伸縮させ、最も整合した時点における類似度をマッチング結果とする。これにより、パターンのずれを考慮した精度の高い類似度の計算が可能となる。DP マッチングによる類似度の計算の概要について説明する。切り出されたサーバ側パターン及びエッジ側パターン

$$\begin{cases} X_U = (x_1, x_2, \cdots, x_i, \cdots, x_U) \\ Y_U = (y_1, y_2, \cdots, y_j, \cdots, y_U), \end{cases} \quad (3.6)$$

について考えると、DP マッチングではパターン同士の距離 D(X,Y)を類似度として定義している。ここで、パターン同士の距離をマッチングコストと表現する。D(X,Y)を求める際、要素 x_i と y_i のベクトル間の局所距離 $d(i,j) = |x_i - y_i|$ をコストとして用いる。さらに、パターンのずれに対してコスト r を加算することを考えると、X_U から Y_U に至る経路は多数存在するため、これらの中で最小コストのものをパターン同士のマッチングコスト D(X,Y)

と定義する。二つのパターン、$x_1x_2\cdots x_i$ と $y_1y_2\cdots y_j$ の距離を g(i, j) とすると、D(X, Y)は動的計画法の最適性の原理を用いて以下の漸化式で与えられる。

(i) 初期条件 g(0, 0) = d(0, 0)
g(i, 0) = g(i-1, 0) + r + d(i, 0)
for i = 1, 2, …, U
g(0, j) = g(0, j-1) + r + d(0, j)
for j = 1, 2, …, U

(ii) i=1, 2, …, U に関して(iv)を実行

(iii) j=1, 2, …, U に関して(iv)を実行

(iv) $$g(i,j) = \min \begin{cases} g(i-1,j) + r + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i,j-1) + r + d(i+j) \end{cases}$$

(v) D(X, Y) = g(U, U)/2U

ここで、トラヒックパターンの全体の長さは各コンテンツによって異なるため、局所距離の加算回数で正規化が行われている。以上のように求まるマッチングコストは、パターン同士が類似しているほど小さい値になる。そこで、P-TRAT と同様に固定の値の判定閾値を用い、それを下回るような値が存在する場合に動画の流通が発生したと判断する。しかしながら DP-TRAT の問題点として、再配信トラヒックの暗号化が発生する環境において、検出性能が低下するということが考えられる。トラヒックが暗号化される場合、新たにヘッダの追加や認証データの付加によってパケットサイズが増加し、MTU (Maximum Transfer Unit)を越えるとパケットのフラグメントが発生する。また、フラグメントによって発生するサイズの小さいパケットが区切りパケットになる可能性もあり、トラヒックパターンが大きくゆがむと考えられる。以上のように、再配信トラヒックが暗号化される場合、正確な流出判定は困難になると考えられる。そこで本研究では DP-TRAT を改良することによって暗号化へ対応した。

４．研究成果

前節より、再配信トラヒックが暗号化される場合、従来手法では正確な動画検出が行えないことは明らかである。この問題を改善するためには、フラグメントによって発生する小さいパケットがトラヒック量を区切ることを防止し、パケットサイズの増加分をトラヒックパターンの生成時に取り除く必要がある。これ以降、フラグメントによって発生する小さいパケットをフラグメントパケットと表現する。本研究では、以上の方針に基づいて DP-TRAT を改良した手法した手法を提案する。ここで他の従来手法と比べ性能が良い DP-TRAT を用いる。第一に、フラグメントパケットがトラヒック量の区切りとならないように、フラグメントパケットが取りうる大きさから区切りパケットの範囲を十分に離す。例として、従来は区切りパケットとして 200 byte 以下のパケットが用いられていたが、これを 600 から 900 byte のパケットに変更するといったことが考えられる。第二に、暗号化されているエッジ側の各パケット情報から、ヘッダの追加などによる増加分を減らし、トラヒックパターンを生成する。本研究では、暗号化によるパケットサイズの増加分を δ と表現する。ここで、暗号化によるパケットサイズの増加分は多々ある暗号化技術及び暗号化の回数に依存する。これらの情報をエッジルータにて取得することは困難であるため、δ の値を動的に予測することが必要になる。

サーバ側のトラヒック情報は暗号化前のものであり、エッジ側のトラヒック情報は暗号化後となっている。そこでフラグメントが発生していない場合、エッジ側の最小パケットは、サーバ側の最小パケットが暗号化されたものと考えられる。そのため、サーバ側及びエッジ側の最小パケットサイズの差分を計算することで δ は求められる。しかしフラグメントが発生する場合、エッジ側の最小パケットは、フラグメントパケットが暗号化されたものとなる可能性がある。そのため、正確な δ を求めることは困難である。

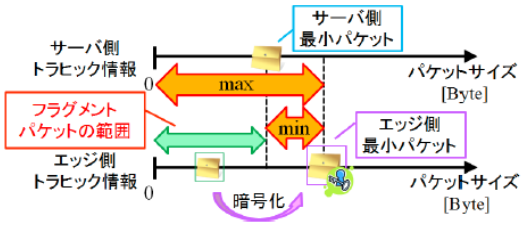


図 4.1: フラグメントパケットの範囲

しかしこの場合、図 4.1 で示す通り、δ の算出のために想定するフラグメントパケットのサイズの範囲は、0 からサーバ側の最小パケットまでとなると考えられる。これは、フラグメントパケットがサーバ側の最小パケットより大きくなる場合、エッジ側の最小パケットはそれらのフラグメントパケットの影響を受けずに、サーバ側の最小パケットが暗号化されたものとなるためである。これはサーバ側の最小パケットを越えるサイズのフラグメントパケットが暗号化される場合、サーバ側の最小パケットが暗号化されたパケットのサイズより大きくなり、エッジ側における最小パケットとならないと考えられるからである。以上より図 4.1 のように、フラグメントパケットの範囲とエッジ側の最小パケットから δ の範囲を求めることができ、最小値 δ(min)がサーバ側及びエッジ側の最小パケットサイズの差分、最大値 δ(max)がエッジ側の最小パケットサイズとなる。

以上のようにδの範囲が求まるので、この中から正確なδを選ぶアルゴリズムが必要になる。そこで、どのようなアルゴリズムを用いればよいか判断するため、δを範囲内において変化させ、暗号化前後のトラヒックパターンにおけるマッチングコストを調査した。実験で用いたトポロジは図 4.2 であり、実験環境を表 4.1 に示す。

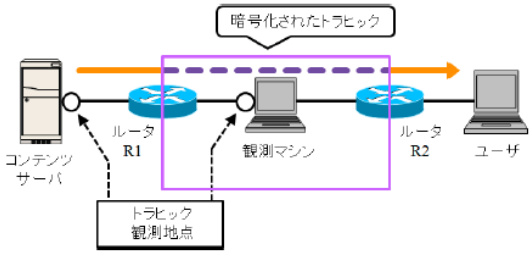


図 4.2: 評価実験のトポロジ

表 4.1: 実験環境

OS	Debian Linux 4.0
ストリーミングサーバ	Helix Streaming Server
メディアプレーヤ	RealPlayer v10.0.9
動画ビットレート	1 Mbps
トラヒック観測	Libpcap v0.7.2 Library
ユーザ側パケット数	2500
T-TRAT: 区切り時間	0.1 s
P-TRAT, DP-TRAT: 区切りパケットサイズ	200 byte 以下のパケット
T-TRAT: 判定閾値	チェビシェフの不等式
P-TRAT: 判定閾値	0.7
DP-TRAT: 判定閾値	700

サーバ側及び観測マシンにおいて暗号化前後のトラヒックを観測し、それらのマッチングコストを算出する。このとき、δの値をδ(min)からδ(max)まで変化させ、観測マシン側のパケット情報から減らしてトラヒックパターンは生成される。パターンマッチングに用いる手法は DP-TRAT を改変したものであり、区切りとなるパケットを 600 から 900 byte へ変更している。またフラグメントが発生する環境を想定しているため、動画の最大パケットサイズは 1500 byte としており、ドキュメンタリ、バラエティ及びアニメについて実験を行った。

結果を図 4.3 に示す。同図からδを変化させた場合のマッチングコストは、全ての動画において区間内に唯一の極小点を持つ凸性のグラフとなっていることが分かった。そこで本研究では、少ない試行回数でマッチングコストが最小値となるδの探索が行えるアルゴリズムとして、黄金分割法に着目した。黄金分割法は区間[A,B]内において、ある凸性のグラフ f(x)の最小値を求めるアルゴリズムであり、以下の操作を行う。

(i) 以下の式を用いて、点 C 及び点 D を計算する。

$$C = A + R(B - A) \quad (4.1)$$

$$D = B - R(B - A) \quad (4.2)$$

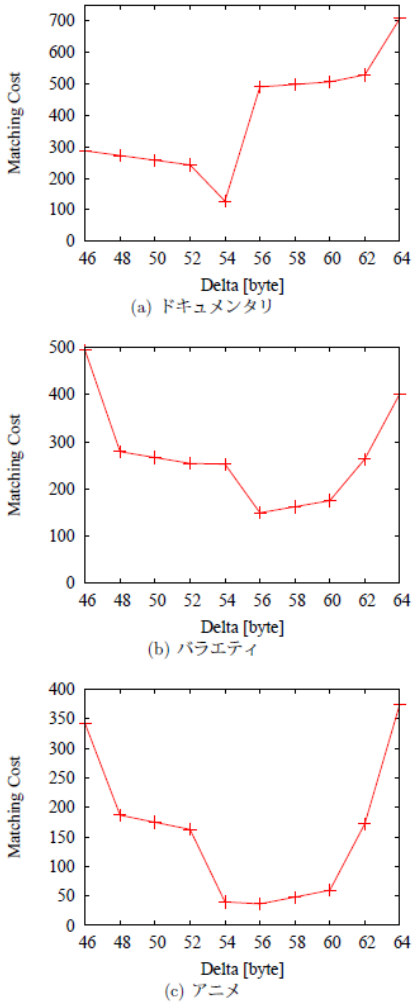
$$R = 0.382 \quad (4.3)$$

(ii) f(C)及び f(D)を計算し、比較する。

(iii) f(C) > f(D)の場合

点 D を点 B に、点 C を点 D に変更し、式

図 4.3: δに対するマッチングコストの変化



(4.1)を用いて点 C を計算する。

(iv) f(C) ≤ f(D)の場合

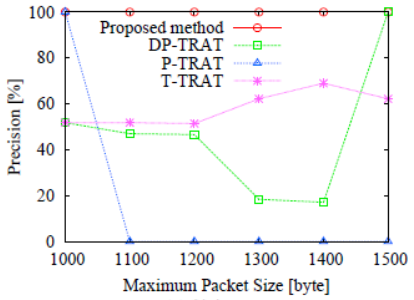
点 C を点 A に、点 D を点 C に変更し、式(4.2)を用いて点 D を計算する。

以降、(ii)から(iv)を繰り返す。

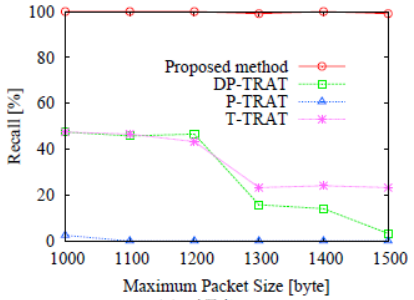
以上のように、区間を徐々に挟めていくことで最小値を求めるアルゴリズムである。区間が十分に狭くなった時、C または D の値が最小値を示すと考えられる。

黄金分割法を用いた提案手法の流れについて説明する。初めに、フラグメントが発生していない場合を考慮し、δ(min)をδと決定してエッジ側のパケット情報から減らし、パターンの生成及びマッチングを行う。マッチングコストが閾値以上だった場合、フラグメントが発生している可能性があるため、δ(min)を A、δ(max)を B として黄金分割法を用いる。このとき、黄金分割法の過程として順に求まる C または D の値をδと決定し、パターンの生成及びマッチングを行う。以上の

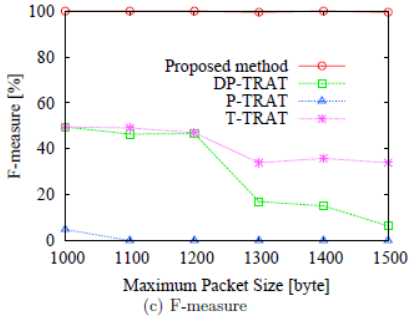
操作により、C または D がマッチングコストを最小とする δ へと近づき、暗号化によるパケットサイズの増加による影響を改善したパターンが生成され、正確なパターンマッチ



(a) 精度



(b) 再現率



(c) F-measure

図 4.4: 暗号化に関する実験結果

ングが可能になると考えられる。また、黄金分割法による δ 探索の終了条件として、マッチングコストが閾値以下となり動画の流出を検出できた場合、及び範囲[A,B]が十分に狭くなった場合としている。

提案手法の性能を評価するために評価実験を行った。本実験では、VPN ルータを用いて IPsec (IPsecurity)によるトラヒックの暗号化を行う。実験トポロジは図 4.2 に環境は表 4.1 と同様である。

図 4.2 において、ルータ R_1 がトラヒックを暗号化し、パケットに新たなヘッダや認証データが加えられる。この時、増加したパケットサイズが MTU である 1500 byte を越える場合、パケットのフラグメントが発生する。本実験では、フラグメントの発生頻度を変更するため、ストリーミングコンテンツの最大パケットサイズを 1000 から 1500 byte まで 100 byte ずつ変化させて行う。ここでパケットのフラグメントは、最大パケットサイズが 1300 byte を越える場合に発生し、最大パケットサイズが大きくなるほど多くのパケットがフラグメントされる。トラヒック量の観測はサーバ側及び観測マシンにて行い、トラヒックの暗号化前後のトラヒックパターンはサーバ側においては観測したトラヒック情報を全て用いて、ユーザ側のトラヒックパターンは観測したトラヒック情報について 2500 個のパケットを区切りとして 20 箇所にてトラヒックを切り出し生成する。評価指標として精度、再現率及び F-measure を用いる。それぞれの指標は式(4.4)、(4.5)及び(4.6)で計算される。

$$\text{精度}: P_r = \frac{C}{A} \tag{4.4}$$

$$\text{再現率}: R_e = \frac{C}{W} \tag{4.5}$$

$$\text{F-measure}: F = \frac{2 \times P_r \times R_e}{P_r + R_e} \tag{4.6}$$

ここで C は正しく流出判定ができたトラヒック数である。A は流出判定をしたトラヒック数を示し、正しい判定と誤った判定の両方を含んでいる。また、W は対象コンテンツを流出しているトラヒック数である。ここで、一般的に精度と再現率はトレードオフの関係となっていることから、両方を考慮した指標として F-measure を用いる。F-measure の範囲は 0 から 1 であり、値が大きいほど手法の性能が良いと判断することができる。

実験結果を図 4.4 に示す。図 4.4 より、提案手法はトラヒックの暗号化によるパケットサイズの増加及びフラグメントが発生しても、高い検出性能を維持できていることが確認できる。これにより提案手法の有効性を確認した。

５．主な発表論文等

〔学会発表〕（計１件）

1. 浅野敦史、西山大樹、加藤寧、The Effect of Packet Reordering and Encrypted Traffic on Streaming Content Leakage Detection 、International Conference on Computer Communication Networks (ICCCN) 2010、2010 年 8 月 4 日、チューリッヒ（スイス）

〔産業財産権〕出願状況（計１件）

名称：コンテンツ不正使用判定装置及び方法及びプログラム
発明者：高橋洋介、石橋圭介、塩本公平、加藤寧、西山大樹
権利者：東北大学、日本電信電話株式会社
種類：特許
番号：特願 2012-025573
出願年月日：2012 年 2 月 8 日
国内外の別：国内

６．研究組織

(1)研究代表者

加藤　寧（KATO NEI）
東北大学・大学院情報科学研究科・教授
研究者番号：00236168