L-016

トラフィック特性を利用したネットワークフォレンジックシステム向け高速圧縮手法の提案

Proposal of Fast Compression Method Using Traffic Characteristics for Network Forensic System

井上喬視† 中島潤‡
Takashi Inoue Jun Nakajima

1. はじめに

ネットワークフォレンジックシステムは，組織活動における証拠保全・収集のためにトラフィックデータを収集・記録し，その通信内容を再現するシステムである．しかしながら組織内 LAN に流れるトラフィックは膨大であるため，全てのトラフィックデータを保存しておくためには膨大な記憶領域が必要とされる．ネットワークフォレンジックは，突発的に発生するセキュリティインシデントに備えるためのシステムであるため，出来る限り長期間に渡ってトラフィックデータを保存しておくことが望まれるが，膨大な容量のストレージや磁気テープなどのバックアップメディアは総じて高価であり，企業にとっては導入または運用コストが問題となっている．

この問題に対する解決策の１つとして圧縮技術の適用が挙げられるが，ギガビットイーサネットに代表される高速ネットワーク環境においては，gzip や bzip2 に代表される汎用的なファイル圧縮手法ではトラフィックワイヤレートに対して圧縮速度が低速であり，トラフィックに対してのリアルタイム圧縮処理は困難である．

そこで本研究では，ネットワークフォレンジックシステムが主に運用される LAN 環境下において，トラフィックデータのリアルタイム圧縮処理を実現すべくトラフィックに含まれる各ネットワークフロー内の類似性を利用した圧縮手法の提案を行う．また同時に，ネットワークフォレンジックで必要とされる証拠保全能力を担保した圧縮ファイルフォーマットの提案を行う．

2. トラフィックデータの記録に対する要求

ネットワークフォレンジックに最も求められる要求は収集したトラフィックデータの原本性である．収集したトラフィックデータには法的証拠能力が求められるため，原本性の確保はネットワークフォレンジックシステムにおいて重要な要素の１つである．原本性確保の要件には機密性，完全性，見読性の３つが挙げられる．さらにネットワークフォレンジックシステムではこれら３つの原本性確保要件に加え，原本性の確保に対する即時性が求められる．これは、ネットワークフォレンジックシステムで取得したトラフィックデータには個人情報が多く含まれており，またシステム管理者など権限のある者はトラフィックデータの改ざんも可能であるため，トラフィックワイヤレートに対してリアルタイムでの原本性確保が行われなければ法的証拠能力が損なわれる恐れがある為である．本研究ではこれらを考慮し，トラフィックデータの見読性を維持しつつ機密性と完全性を付与し，かつ有限である記憶領域を有効に利用する為の圧縮処理をトラフィックワイヤレートに対してリアルタイムで行うことが要求される．

また本稿で提案する圧縮手法の対象は，標準的なトラフィックデータ保存フォーマットである PCAP フォーマット[2]とした．PCAP フォーマットは，データリンクフレーム到着時間とデータ長そしてフレームデータが連続して配置されるデータフォーマットであり，本研究では組織内 LAN 環境下におけるネットワークフォレンジックを対象としているため，データリンクタイプがイーサネットの PCAP ファイルに限定している．

3. トラフィック特性を利用した圧縮手法の提案

本研究は，PCAP ファイル内のデータリンクフレーム間での類似性，そして同一ネットワークフローに属するパケット間の類似性に着目した．同一ネットワークフローに属するパケットは同一プロトコルで構成されることになる為ペイロード内容に類似点が多く，圧縮効果が高い．本提案ではこの性質を圧縮に用いる為，ネットワークフロー単位で圧縮を行うことを提案する．

また[1]の研究により，総処理時間のうち大部分を圧縮処理時間が占め，被圧縮サイズを小さくし圧縮処理時間を減らすことが必要であることが判明している．そのため本提案では，被圧縮サイズを減らすために，パケットに含まれるヘッダ部分の圧縮を行うなど，被圧縮サイズを減少させて圧縮処理速度の向上を図ることを試みた．

また圧縮を行うためのアルゴリズムには圧縮速度が高速であることが求められる．サンプルとして収集したトラフィックデータを複数のアルゴリズムを用いて圧縮し比較したところ，圧縮速度に関しては QuickLZ アルゴリズム[3]が他と比べて高速であるという結果が得られた．よって本提案においては QuickLZ アルゴリズムを用いて圧縮を行うこととした．

3.1 圧縮単位

ネットワークフロー単位で圧縮を行うことを前に述べたが，圧縮対象としている PCAP フォーマット内には同一ネットワークフローに属するパケットが断片的に格納されている為，ネットワークフローを分類する必要がある．ネットワークフローは送信元 IP アドレス，宛先 IP アドレス，送信元ポート番号，宛先ポート番号で一意に識別される為，このタプルから計算されたハッシュ値を用いてネットワークフローを分類する．このように分類されたパケット間で圧縮器を共有して圧縮することでネットワークフロー単位の圧縮を実現する．結果として一度に圧縮を行うサイズが小さくなり並列処理性が生まれる為，処理速度の向上が見込まれる．また，ネットワークフロー単位で圧縮を行うことによりネットワークフローごとの抽出が可能になるため，

† 北海道情報大学 大学院
Graduate School of Hokkaido Information University
‡ 北海道情報大学 情報メディア学科
Faculty of Information Media, Hokkaido Information University

本フォーマットを用いた場合に通信内容の再現を行いやすいという利点もある．

また SSH プロトコルや SSL プロトコルなど暗号化された通信は，対応する鍵が無ければ復号不可能であり通信の再現をすることが困難である．そのため指定されたポート番号のペイロード部分を削除可能とした．暗号化されたペイロードを削除することにより被圧縮サイズが削減され，圧縮処理速度の高速化につながる．

3.2 パケットヘッダ圧縮

PCAP フォーマットは，データリンクフレームごとに対してフレーム到着時間を示す GMT タイムスタンプが付与される．高速ネットワーク環境においてはフレーム間の到着時間差は僅かであるため，1 つ前のフレーム到着時間との差分を取ることでフィールド長を短縮することが出来る．

また同一ネットワークフローに属するパケットのヘッダ部分のうち，IP アドレス，TCP ポート番号などパケット間で変化しないフィールドがあり，これらパケット間で変化しないフィールドは，後続のパケットヘッダにて最初に到着したパケットヘッダを参照することにより，フィールドを省略することが出来る．またフレームチェックシーケンスや IP ヘッダチェックサム，TCP チェックサムなど再計算可能なフィールドに付いても同様に省略することが可能である．

3.3 ネットワークフローの証拠保全手法

本提案ではネットワークフォレンジックにおける原本性確保のために，ファイルフォーマットに機密性と完全性を含ませている．本提案では機密性の確保の為に AES 暗号を用いるが，暗号化処理は処理速度が低速であるため，リアルタイム圧縮を実現する場合に弊害となる．本提案では，3.1 節で述べたようにネットワークフロー単位で圧縮を行うことを利用し，ネットワークフローを構成するパケットのうち，最初に到着したパケットのみをヘッダ圧縮を含む圧縮処理の後暗号化し，後続のパケットは圧縮処理のみを行う．後続のパケットヘッダの伸張には 3.2 節で述べた様に 1 つ目のパケットデータを用いなければならない為，ネットワークフローの機密性が確保されることになる．また完全性確保の為にパケットごとに圧縮前データの MD5 値を付与し，パケットの改ざん検知を可能にしている．

4. 圧縮ファイルフォーマットの設計

本研究で提案する圧縮ファイルフォーマットの概要を図 1 に示す．Flag フィールドには圧縮パケットがどのような性質を持つものかを示す情報が格納され，図中の破線で囲まれた部分は Flag フィールドの各ビットによって出現が制御されるフィールドであることを示している．このフィールドは 1 オクテットであり，それぞれのビットに対して圧縮パケットデータが暗号化されているかどうか，ネットワークフロー分類可能であるかどうか，到着秒が変化したかどうかなどの情報が入る．Data Length フィールドには圧縮パケットデータ長が入る．これには可変長数値表現を用い，多くの場合 1 ないし 2 オクテットである．Differential PCAP Header フィールドはフレーム到着時間を示し，そのうち Δts_sec は Flag フィールドの TS ビットによって出現が制御される．Δts_usec は Δts_sec からの増分をマイクロ秒で示したフィールドであり，一つ前のタイムスタンプとの差分になる．Flag フィールドにてネットワークフロー分類可能であるとされたパケット（F ビットによって示される）については，3.2，3.3 節で述べた手法を適用し，分類不可能なパケットについてはタイムスタンプのみ差分圧縮するものとした．分類可能なパケットには各ネットワークフローに対して一意である Flow ID が付与され，伸張時のネットワークフロー識別に用いられる．また Compressed Header における破線部分のフィールドは，イーサネットフレームまたは TCP(UDP)/IP スタックにおける普遍的なフィールドをまとめたものであり，最初に到着したフレームや，内容に変更があった場合にフィールドが出現する．

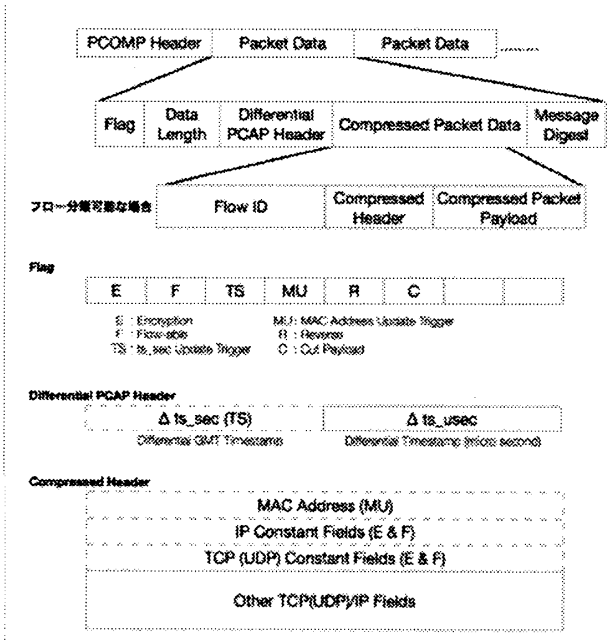

図 1 圧縮ファイルフォーマット概要図

5. おわりに

本稿では，同一ネットワークフローに属するパケット間のヘッダ共通点とペイロード内容の類似性を用いた圧縮手法と，トラフィックデータに機密性と完全性を付与する証拠保全手法、またそれらを考慮したファイルフォーマットの提案を行った．今後は本提案を実装したソフトウェアを開発し，実験を通じて圧縮処理速度・圧縮率の評価を行う予定である．

参考文献

[1] 井上番視，中島潤，“ネットワークフォレンジックシステムに適したトラフィックデータ保存形式の提案”，電子情報通信学会 2008 年総合大会，2008
[2] “TCPDUMP/LIBPCAP public repository”, http://www.tcpdump.org/
[3] Lasse Reinhold, “QuickLZ”, http://www.quicklz.com/