

[Work in Progress] 研究報告

HTTP 圧縮ストリームの逐次的な GZIP 復号における メモリ容量の削減

八巻 隼人^{†1, a)} 西 宏章^{†2}

Memory Size Reduction for Sequential GZIP Decoding of HTTP Compressed Streams

近年のインターネット関連技術の急速な進歩により、侵入検知システムといった高度な機能を提供するサービスの需要が高まっている。そこで、ペイロード込みでパケットを解析する Deep Packet Inspection (DPI) が盛んに研究されている[1]。特に、ネットワーク機器が DPI 機能を有することで、セキュリティのみならず、負荷分散、QoS 制御といった様々なサービスを高度に提供できる可能性がある[2]。

DPI の主な解析プロトコルである Hyper Text Transfer Protocol (HTTP) は、HTTP/1.1 から転送データを GZIP 圧縮することが一般となりつつある。このように圧縮化されたデータをそのまま解析することはできず、一旦復号する必要があった。しかしながら、従来の GZIP 復号処理手法は圧縮データ全体の到着を待ってから復号を開始するため、メモリ使用量の点からネットワーク経路上での実行には不向きであった。これに対し、パケットの逐次的な復号を可能とするコンテキストスイッチ機構が提案されている[3]。

コンテキストスイッチ機構は、後続パケットの復号に要するコンテキスト情報を HTTP 圧縮ストリーム毎に管理し、パケット毎にこれを読み込むことで逐次的な GZIP 復号を実現する。コンテキスト情報として、数百 Byte の符号表、辞書と呼ばれる圧縮対象の直前 32KB 分のデータ、パケット境界等の 66Byte のデータが含まれる。この場合、1 ストリーム毎に 33KB 程度のコンテキスト情報を確保するため、メモリ使用量が容易に増大するという問題があった。そこで、本報告ではコンテキスト情報のデータサイズの大部分を占める辞書データ量の削減手法について検討する。

GZIP では、データ中の同一文字列の繰り返しを検知し、圧縮するために、圧縮対象の直前 32KB 分のデータを辞書として用いる。しかしながら、ネットワークにおける HTTP 圧縮ストリームの大部分は 32KB に満たない。図 1 は、10Gbps 回線を持つ国内の大規模ネットワークを解析し、

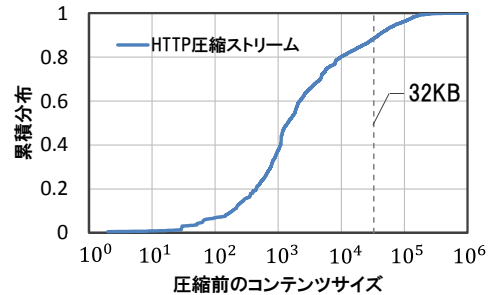


図 1 HTTP 圧縮ストリームにおける圧縮前コンテンツサイズの分布

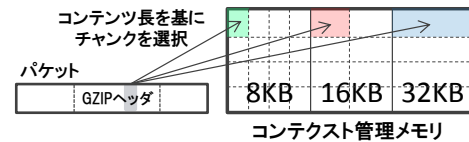


図 2 チャンク分割によるコンテキスト管理の概要

HTTP 圧縮ストリームを対象として圧縮前のコンテンツサイズの分布を計測した結果を示している。図 1 によると、90% のストリームが 32KB 以内であり、大部分は 1KB 程度である。GZIP 圧縮データの圧縮前のコンテンツサイズは、GZIP ヘッダの解析によって予め知ることが可能である。そこで、スラブアロケータを応用したコンテキスト管理手法を提案する。本手法では、図 2 に示すように、メモリを複数のチャンクに分け、ストリームのコンテンツサイズに応じたチャンクを選ぶことで、大部分の HTTP 圧縮ストリームの辞書サイズを削減する。例えば、8KB、16KB、32KB の 3 つのチャンクによって理想的にコンテキストを管理できた場合、従来手法に対してメモリ使用量は 34.9% まで削減される。本手法ではチャンクサイズ、エントリ数をいかに決定するかが重要な課題となる。

参考文献

- 1) Lin, Y. et al.: Guest Editorial Deep Packet Inspection: Algorithms, Hardware, and Applications, IEEE J-SAC, Vol.32, No.10, pp.1781-1783 (2014)
- 2) Bremler-Barr, A. et al.: Deep Packet Inspection as a Service, Proc. ACM CoNEXT'14, New York, USA, pp.271-282 (2014)
- 3) 八巻隼人, 中村優一, 高際兼一, 松井加奈絵, 西宏章: インターネットルータにおける HTTP 圧縮ストリームの高速展開処理機構の提案, 電子情報通信学会論文誌 B, Vol.J98-B, No.10, pp.1104-1114 (2015)

†1 慶應義塾大学大学院理工学研究科, 神奈川
Graduate School of Science and Technology, Keio University, Yagami
Campus, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522,
Japan
†2 慶應義塾大学理工学部システムデザイン工学科, 神奈川
Department of System Design Engineering, Keio University, Yagami
Campus, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522,
Japan
a) yamaki@west.sd.keio.ac.jp