



所属 システム情報系
 氏名 山際 伸一 准教授
 キーワード 高性能計算
 HP システム

ビッグデータ時代に対応した新しいロスレスデータ圧縮技術

概要

本技術は、九州工業大学 坂本比呂志教授と共同で開発した、ビッグデータ時代を見据えた新しいロスレスデータ圧縮技術LCA-DLT (Lowest Common Ancestor-Dynamic Lookup Table) です。

本技術は、専用ハードウェア (LSIチップ) によるデータ圧縮技術として、データの出現傾向を自動認識する、新しいヒストグラム管理技術を実現しています。さらに、圧縮されたデータに、圧縮の規則を割り当てた変換表を復元する情報を埋め込むことによって、次々と圧縮されたデータが復号側に送られていき、それを受け取った時から、順次復号化が可能な技術を確認しています。これにより、従来は圧縮データと別々に復号側に送られていたデータ変換規則を送る必要がなく、流れるデータを連続して圧縮・復号できます。

この技術はハードウェアとの親和性が高く、最大50%の圧縮が可能なモジュールを多段接続することができ、4段接続で、最大10%のサイズにまでデータ圧縮が可能です。このように、ハードウェア量によって圧縮率を自由に調整できるため資源コストが選べるという特徴を有し、さらに、ZIP形式などソフトウェアによるデータ圧縮よりも少ない電力で高速処理ができる、といったメリットがあります。

従来技術・競合技術との比較、新技術の特徴

- 従来はソフトウェアによりデータ圧縮複合していましたが、本技術ではハードウェア化により高速化が実現しています。
- 従来は、圧縮されたデータと変換表の全体を受け取ってからでないと複合を開始できませんでした。本技術ではデータの圧縮複合を並行して処理できるため、高性能なストリームデータの圧縮復元が可能になります。

企業への提案

ハードウェア化により高性能圧縮を実現した画期的技術です。本技術を活用するビジネスを提案します。

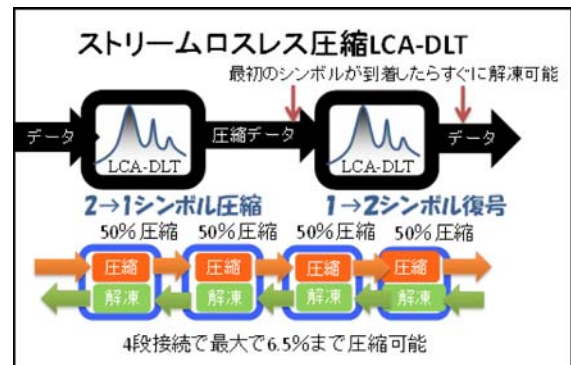


図 新圧縮技術の特徴。データが圧縮器に入力されると次々に圧縮されたデータが出力され、復号側に伝搬し、復号化側では圧縮データを1つでも受け取ると、圧縮側で作られた変換テーブルが復元され、復号されていきます。さらに、1段で50%圧縮 (2→1シンボル圧縮) 可能なモジュールを多段接続することで、ハードウェア量と圧縮率を選択できます。

想定される用途

- ネットワークあるいはデータ伝送ルートの伝送元と伝送先の装置に本技術を組み込み、ネットワーク全体のトラフィック負荷軽減する。
- 単一装置に本技術を組み込み、入力部でデータの圧縮、出力部で複合を行い内部のデータ転送不可系気により処理性能を向上させる。
- 本技術をストレージ装置の入出力部に組み込み、圧縮後のデータを格納することにより物理的な容量の2倍のストレージ装置の実現させる。

特許 (出願済) 主な論文 (以下)

- Shinichi Yamagiwa, Koichi Marumo and Hiroshi Sakamoto, Stream-based Lossless Data Compression Hardware using Adaptive Frequency Table Management, In Proceedings of VLDB 2015/BPOE-6, Springer, Sep 2015.(Best paper award受賞)
- Shinichi Yamagiwa and Hiroshi Sakamoto, A Reconfigurable Stream Compression Hardware based on Static Symbol-Lookup Table, Proceedings of IEEE BigData/BPOE 2013, IEEE, October 2013
- ETアワード2015「特別賞」受賞技術