Wan Accelerators: Optimizing Network Traffic with Compression

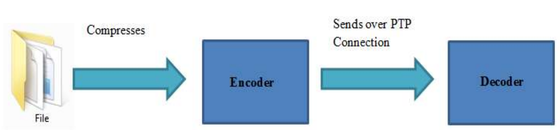
Bartosz Agas, Marvin Germar, & Christopher Tran
Rutgers School of Engineering

Introduction & Motivation

- A WAN accelerator is an appliance that can maximize the services of a point-to-point(PTP) connection providing high quality performance using minimal resources.
- There are various methods that a WAN accelerator uses to optimize resources such as caching repetitive data, compression, and latency optimization. With such means, a WAN accelerator can analyze the packets of data being transferred over the PTP connection and only send the critical or compressed differentials.
- Consider two local networks where network_1 is connected to network_2 via a PTP connection and the networks frequently send each other similar data packets such as IP headers, emails, and documents. A lot of these data packets contain repetitive data segments that do not need to utilize the resources of the PTP connection.
- Via the aforementioned methods, a WAN accelerator significantly improves the transfer process and minimizes monetary expenses. Investors can save money required for higher PTP connections.

Approach & Programs

- Use a compression algorithm to create an evolving dictionary with an encoder and decoder
- Encoder reads in a file and updates the dictionary to reflect new probabilities of more frequent string of data.
- The compressed data is sent over a PTP connection.
- Decoder will read metadata bit by bit and decompress the files using the trained dictionary from the encoder.



Huffman Coding

Huffman coding is a technique that uses lossless data compression to reduce the amount of bits required to represent a string of symbols. The algorithm accomplishes its goals by allowing symbols to vary in length. Shorter codes are assigned to the most frequently used symbols, and longer codes to the symbols which appear less frequently in the string.

Dictionary: How It's Built

- A Huffman tree is created to develop a dictionary. This dictionary adapts to the data being transferred giving preferential treatment to data that is sent more frequently.
- A WAN accelerator's dictionary learns from all data transfer and is updated unlike normal compression methods that deletes it dictionary after it outputs a file.
- By using an updateable dictionary, a company can use a WAN accelerator to optimize their network because the dictionary is trained specifically for the company's common files and Huffman coding produces a dictionary that persist over numerous data transmissions.

Given a 6 symbol alphabet with the following symbol probabilities:

A = 1, B = 2, C = 4, D = 8, E = 16, F = 32

Step 1. Combine A and B into AB with a probability of 3.
Step 2. Combine AB and C into ABC with a probability of 7.
Step 3. Combine ABC and D into ABCD with a probability of 15.
Step 4. Combine ABCD and E into ABCDE with a probability of 31.
Step 5. Combine ABCDE and F into ABCDEF with a probability of 63.

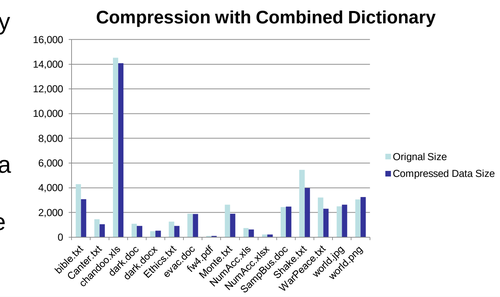
The Following tree results:

```
ABCDEF
  /   \
(0)F   ABCDE
         /   \
      (10)E   ABCD
                /   \
            (110)D   ABC
                     /   \
          (1110)C   AB
                    /   \
              (11110)A   B(11111)
```

To summarize this example, 'F' has the highest probability from the letters in this alphabet so it is represented with the shortest bit string possible and in this case just '0'.

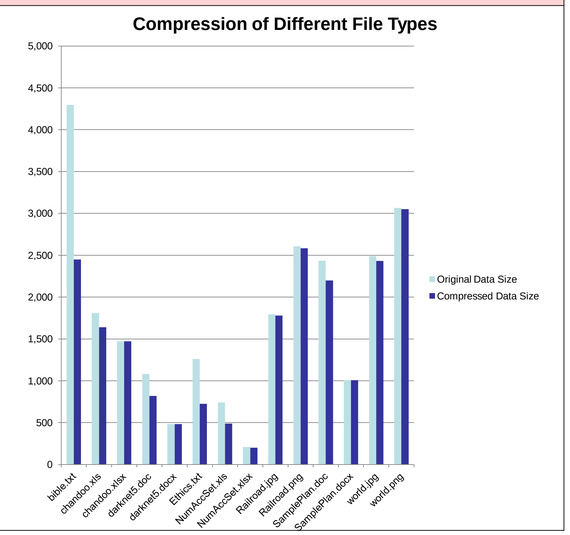
Results for Trained Files

- Results show that a company using a WAN accelerator will have reduced size file transmissions, ultimately saving money.
- The results shown here use a dictionary trained on a batch of files, which can reduce the network's overall traffic activity



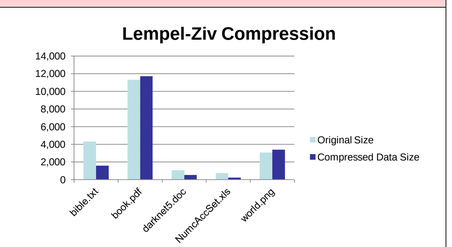
Results for Specific File Types

- Huffman coding is tested on different file types to determine its effects.
- Certain file types compressed better than others so the data type being compressed matters.
- Text files and document files such as .doc or .xls compressed better.
- Other file types such as .docx. .xlsx, and .png are already compressed and recompressing them does not make them smaller.



Another Compression Method

- Lempel-Ziv-Welch is another lossless data compression algorithm.
- It is more efficient than Huffman coding in optimizing network traffic.
- Similar to Huffman coding, the data type being compressed matters



Citations

- Dipperstein, Michael. "Huffman Code Discussion and Implementation." Michael.dipperstein.com. Michael Dipperstein. Web. 2 Apr. 2012. <http://michael.dipperstein.com/>.
- Nelson, Mark and Jean-Loup Gailly. *The Data Compression Book, Second Edition*. Cambridge, MA: IDG Books Worldwide, Inc., 1995.
- Schumacher III, James. "Lempel Ziv Compression." PlanetSourceCode. Exhedra Solutions, Inc., 18 Apr. 2010. Web. 30 Mar. 2012. <http://www.planet-source-code.com/>.