



МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ
И МАТЕМАТИКИ ИМ. А.Н. ТИХОНОВА



Исследование и разработка метода предобработки и передачи данных в удаленном интернете вещей

Карнаухов А. Ю. ayukarnaukhov@edu.hse.ru

Москва 2023



Предметная область

Удаленный интернет вещей (IoT) – концепция сетей передачи данных между физическими объектами, которые оснащены средствами взаимодействия с внешней средой и друг с другом и находятся в удаленных регионах с плохо развитой инфраструктурой сети.



Проблематика

1. Малая пропускная способность
2. Высокая стоимость отправки сообщения
3. Энергоэффективность [1]

Цель:

Разработка метода преодоления ограничений при передаче данных по низкоскоростному каналу сообщения, в частности **уменьшения размера сообщений** в спутниковых сетях удаленного интернет вещей с помощью методов предобработки данных

Задачи:

- Провести анализ предметной области
- Рассмотреть существующие методы предобработки данных
- Разработать **новый метод** предобработки данных
- Провести **экспериментальные исследования**

Методы предобработки данных:

- Сериализация данных
- Сжатие данных



50 KB



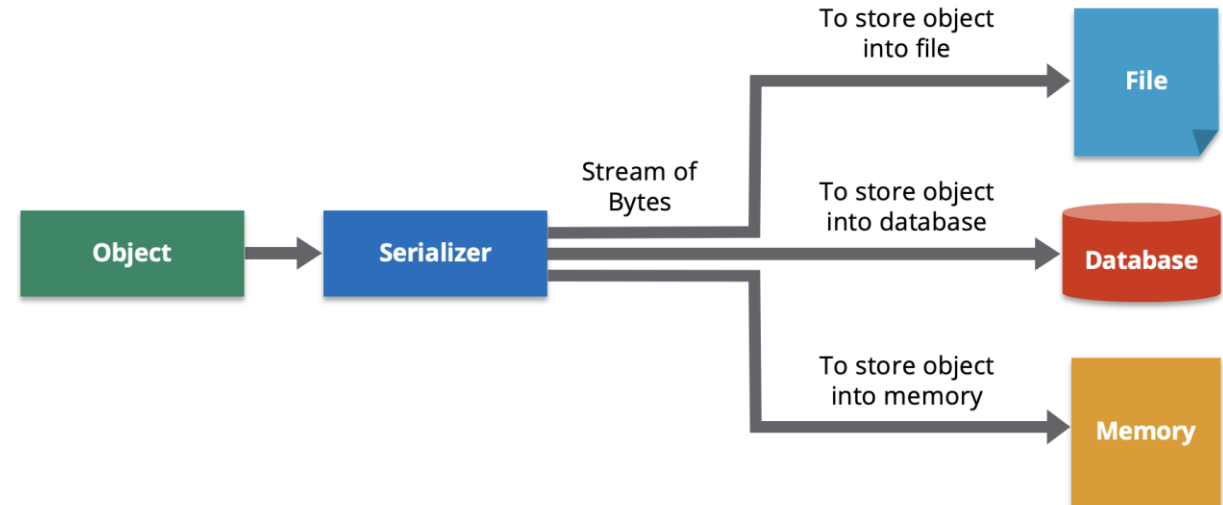
20 KB

Сериализация данных

- процесс перевода структуры данных в битовую последовательность/другой формат представления данных

Особенности:

- Извлечение данных требует полного считывания
- Возможное нарушение инкапсуляции данных

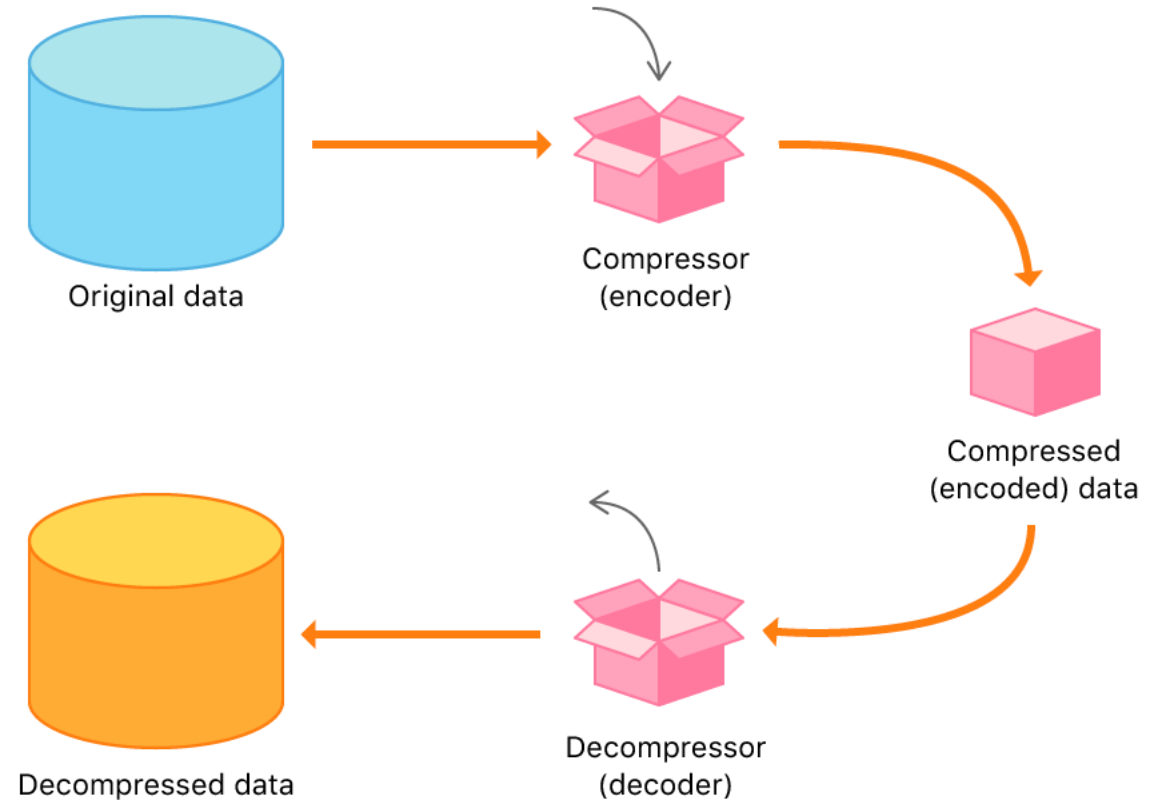


Сжатие данных

- процесс преобразования данных, цель которого является уменьшения занимаемого объема полезной информацией

Особенности:

- Основано на избыточности данных
- Вида сжатия :
 - С потерями
 - Без потерь



Метод GDEP:

- Сериализация данных в формате JSON
- Сериализация JSON методом Protobuff
- Сжатие алгоритмом Хаффмана

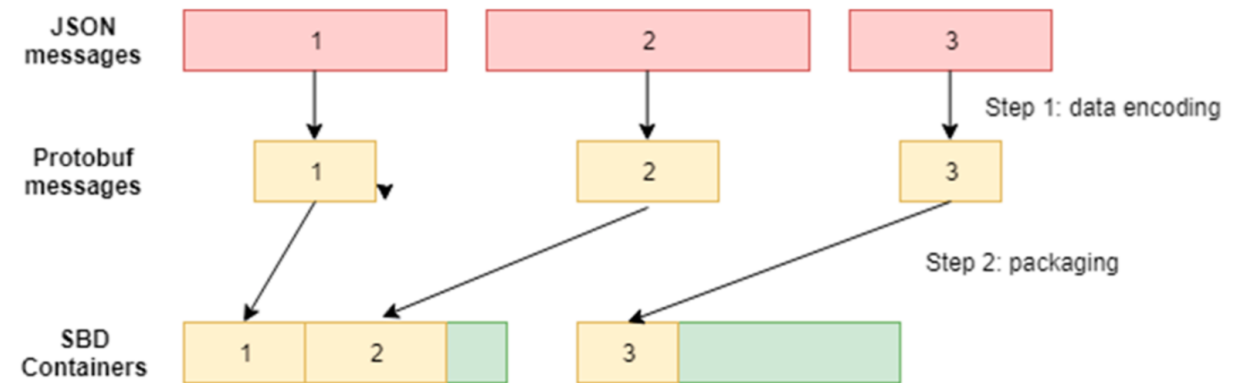


Рисунок 6. Схема алгоритма кодирования и упаковки шлюза[2]

Алгоритм Хаффмана:

Свойства:

- Обладает свойством префиксности
- Алгоритм сжатия без потерь
- Энтропийное кодирование
- Имеет большой коэффициент сжатия коротких сообщений по сравнению с другими архиваторами

| Arch. program/compr. algorithm | Compression ratio $K = \frac{N_{in}}{N_{out}}$ |
|----------------------------------|--|
| cmix | 0,86 |
| paq8hp12 | 0,62 |
| nanozip | 0,34 |
| gzip | 0,56 |
| bzip2 | 0,51 |
| 7zip | 0,24 |
| Huffman | 1,33 |
| Huffman with extended dictionary | 2,13 |

Таблица.1 Сравнение алгоритмов сжатия при сжатии сообщений размером 30-40 байт [3]

| Изначальный размер файла (байт) - S_o | Сжатый размер файла (байт) - S_c | Алгоритм сжатия/ архиватор | Коэффициент сжатия: $k = \frac{S_o}{S_c}$ |
|--|------------------------------------|-------------------------------|--|
| 23379947 | 13493020 | Алгоритм Хаффмана | 1,7327438187 |
| | 6082151 | 7zip | 4,0782287330 |
| | 5732868 | gzip | 3,8440260691 |
| | 4959630 | bzip2 | 4,7140506449 |
| 19607352 | 11884448 | Алгоритм Хаффмана | 1,6498327899 |
| | 4988678 | 7zip | 4,1044801396 |
| | 4777061 | gzip | 3,9303703306 |
| | 4030069 | bzip2 | 4,8652645898 |
| 57610583 | 33670117 | Алгоритм Хаффмана | 1,7110300805 |
| | 14814007 | 7zip | 4,1106368019 |
| | 14015002 | gzip | 3,8889264059 |
| | 12119167 | bzip2 | 4,7536751495 |

Таблица 2. Сравнение коэффициентов сжатия
Больших сообщений

| Изначальный размер файла (байт) $- S_o$ | Сжатый размер файла (байт) - S_c | Алгоритм сжатия/ архиватор | Коэффициент сжатия: $k = \frac{S_o}{S_c}$ |
|---|---------------------------------------|-------------------------------|--|
| 32 | 18 | Алгоритм Хаффмана | 1,7777777778 |
| | 188 | 7zip | 0,1702127660 |
| | 64 | gzip | 0,5000000000 |
| | 69 | bzip2 | 0,4637681159 |
| 316 | 181 | Алгоритм Хаффмана | 1,7458563536 |
| | 337 | 7zip | 0,9376854599 |
| | 210 | gzip | 1,5047619048 |
| | 234 | bzip2 | 1,3504273504 |

Таблица 3. Сравнение коэффициентов сжатия коротких сообщений

| Изначальный размер файла (байт) - S_o | Сжатый размер файла (байт) - S_c | Формат данных | Коэффициент сжатия: $k = \frac{S_o}{S_c}$ |
|--|---------------------------------------|---------------|---|
| 109 | 67 | csv | 1,6268656716 |
| 1445 | 953 | html | 1,5162644281 |
| 165 | 96 | json | 1,7187500000 |
| 150 | 77 | txt | 1,9480519481 |
| 423 | 254 | xml | 1,6653543307 |
| 6948127 | 4311236 | csv | 1,6116322558 |
| 23542313 | 14857620 | html | 1,5845278719 |
| 15927843 | 9207210 | json | 1,7299315428 |
| 7104971 | 4258862 | txt | 1,6682792258 |
| 11104146 | 6418485 | xml | 1,7300260108 |

Таблица 4. Сравнение коэффициентов сжатия
Сообщений разного формата алгоритмом Хаффмана

Полученные результаты:

- Разработана программа реализующая метод сжатия Хаффмана
- Проведены экспериментальные исследования сравнения алгоритмов сжатия
- Проведены экспериментальные исследования сравнения зависимости коэффициента сжатия от формата данных

Дальнейшее развитие работы:

- Провести сравнительный анализ данных различного формата сериализованные методом protobuf
- Сравнить коэффициенты сжатия данных различного формата сериализованные методом protobuf, сжатых алгоритмом Хаффмана
- Провести сравнительный анализ данных различного формата сериализованные другими методами
- Сравнить коэффициенты сжатия данных различного формата сериализованные другими методами, сжатых алгоритмом Хаффмана
- Разработка алгоритма Хаффмана с расширенным словарем

Список литературы:

1. Ivan Lysogor, Voskov L., Rolich A., Efremov S. G. Study of Data Transfer in a Heterogeneous LoRa-Satellite Network for the Internet of Remote 4. Things // Sensors. 2019. Vol. 19. No. 15. P. 1-17. doi
2. Voskov L., Rolich A., Bakanov Gleb, Podkopaeva Polina. Gateway Data Encoding, Packaging and Compression method for heterogeneous IoT-satellite network, in: 2021 XVII International Symposium "Problems of Redundancy in Information and Control Systems" (REDUNDANCY). IEEE, 2021. doi P. 34-38. doi
3. Lysogor I., Voskov L., Rolich A., Efremov S. G. Energy efficient method of data transmission in a heterogeneous network of the Internet of things for remote areas, in: 2019 International Siberian Conference on Control and Communications (SIBCON). Proceedings. Tomsk: Tomsk State University of Control Systems and Radioelectronics (TUSUR), 2019. P. 1-6. doi



МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ
И МАТЕМАТИКИ ИМ. А.Н. ТИХОНОВА



Исследование и разработка метода предобработки и передачи данных в удаленном интернете вещей

Карнаухов А. Ю. ayukarnaukhov@edu.hse.ru

Москва 2023