

# Discussion on Huffman Encoding and Threshold

17 March, 2025

Airi Kokuryo

# Current work (summary)

- We use a static Huffman table based on the 4 seasons
- Sends data only when either temperature or humidity exceeds the threshold
- Since the Huffman table is based on data divided into several range, combine run-length compression when there is serial data range for more compression.
- Send data to the LoRaWAN using the combination of Huffman Encoding and run-length compression

# Problems

- Since we use static Huffman table created using the past data within the area, it won't be applicable outside of its observation areas.
- Due to the global warming, data such as temperature and humidity can become unstable -> can't rely on static past sources.
- Since we use thresholds, it is likely that data being sent would be near the threshold. For example, temperature and humidity would likely score high during extreme rainfall, and this may not be the most common within the Huffman table introduced in our current work.

# Proposed approach

- I. “Normal”
  - I. Collect data every 1 hour and put it in a buffer
  - II. Calculate **flood probability (FP)**
  - III. If FP is over 70%, start “Emergency”
  - IV. If FP is under 70%, go back to I.I
- II. “Emergency”
  - I. Send the first 3 data in the buffer using **Huffman table  $\alpha$**   
(sent only once at the start)
  - I. Send rest of the data in the buffer using **difference method and Huffman table  $\beta$**
  - II. Collect data every 5 minutes and put it in a buffer
  - III. If the buffer is full, send data using II.II
  - IV. Check FP and if it's under 70% twice in a row, go back to “Normal”
  - V. Check FP and if it's over 70%, go back to II.III

# Flood Probability (FP)

This is how we decide the threshold for sending data

- Clausius Clapeyron Equation

$$e_s(T) = 6.112 * \exp\left(\frac{17.67T}{T + 243.5}\right)$$

- Precipitable Water (PW)

$$PW = 0.1 * RH * e_s(T)$$

(RH is the Relative Humidity; measured using DHT11 sensor)

- Flood Probability (FP)

$$FP(\%) = k * PW * RH$$

$$k = 0.1 \sim 0.5$$

Calculation Citations:

<https://pubs.acs.org/doi/pdf/10.1021/ed028p428>

[https://www.jstage.jst.go.jp/article/jmsj/98/6/98\\_2020-058/\\_html/-char/en](https://www.jstage.jst.go.jp/article/jmsj/98/6/98_2020-058/_html/-char/en)

<https://www.weather.gov/lmk/humidity>

# Flood Probability (FP)

LoRaWANモジュール初期化...  
ネットワーク未参加! JOINを待機...  
Dragino LA66 ネットワーク未参加! JOINを待機...  
LoRaWANネットワークに接続成功!

送信データ: 32.00°C, 80.00%  
Flood probability (%): 91.38  
LoRaWAN送信コマンド発行

→ sent

Dragino LA66 ネットワーク未参加! JOINを待機...  
LoRaWANネットワークに接続成功!

送信データ: 26.30°C, 37.00%  
Flood probability (%): 14.05

→ Not sent

送信データ: 26.30°C, 36.00%  
Flood probability (%): 13.30  
送信データ: 26.30°C, 36.00%

# Huffman table $\alpha$

- Used to compress raw data
- Either use the static table proposed in the current research, or create a dynamic table given a weight to the threshold
- Since it is only 3 data that is being sent using this method, we may use another approach with less calculation required

# Huffman table $\beta$

- Taking the equation from the paper, “Data Compression in Wireless Sensor Nodes with LoRa”

$$\hat{v}_i = \frac{5}{4}v_{i-1} + \frac{1}{2}v_{i-2} - \frac{3}{4}v_{i-3}$$

- Use the equation above to calculate estimation of the next data. Compare it to the actual data and only send its difference (difference method)
- Huffman table  $\beta$  will use the likelihood for this difference

Actual Data	$v_0$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$
Estimate	-	-	-	$\hat{v}_3 = \frac{5}{4}v_0 + \frac{1}{2}v_1 - \frac{3}{4}v_2$	$\hat{v}_4 = \dots$	$\hat{v}_5 = \dots$
Send	$v_0$	$v_1$	$v_2$	$\hat{v}_3 - v_3$	$\hat{v}_4 - v_4$	$\hat{v}_5 - v_5$

# Huffman Table $\beta$

TABLE IV. THE HUFFMAN DICTIONARY.

$n_i$	$s_i$	$d_i$
<b>0</b>	00	0
<b>1</b>	01	-1,+1
<b>2</b>	100	-3,-2,2,3
<b>3</b>	101	-7,...,-4,+4,...,7
<b>4</b>	110	-15,...,-8,8,...,15
<b>5</b>	1110	-31,...,-16,16,...,31
<b>6</b>	11110	-63,...,-32,32,...,63
<b>7</b>	111110	-127,...,-64,64,...,127
<b>8</b>	1111110	-255,...,-128,128,...,255

## I. Normal

## II. Emergency



Buffer is the amount of data you can send at once.

Questions:

- Is this amount of data enough for predicting how the flood has happened?
- Is 70% a good indicator for the start of flood?
- Is one hour and five minutes too long/short?

# Discussion on Huffman Encoding and Threshold (日本語)

17 March, 2025

Airi Kokuryo

# 現在のデータ送信方法(概要)

- 四季に基づいた静的なハフマン表を使用
- 温度または湿度が閾値を超えた場合のみデータを送信
- ハフマン表は複数の範囲に分けられた温度・湿度データに基づいており、連続したデータがある場合は、より高い圧縮を実現するためにランレンジス圧縮を組み合わせる
- ハフマン符号化とランレンジス圧縮の組み合わせを用いてLoRaWANにデータを送信

# 問題

- 過去の観測データを使用して特定のエリア向けに作成した静的なハフマン表を利用しているため、観測エリア外では適用できない。
- 地球温暖化の影響で、温度や湿度などのデータが不安定になる可能性があり、過去の静的データに依存するのは難しい。
- 閾値を用いているため、送信されるデータは閾値付近の値が多くなる可能性がある。特に豪雨時には温度や湿度が極端に高くなることがあり、こうした異常値は各季節の通常のデータ分布とは異なるため、現在のハフマン表では適切に圧縮できない可能性がある。

# Flood Probability (FP)

閾値の設定方法

Clausius Clapeyron Equation

$$e_s(T) = 6.112 * \exp\left(\frac{17.67T}{T + 243.5}\right)$$

- Precipitable Water (PW)

$$PW = 0.1 * RH * e_s(T)$$

(RH is the Relative Humidity; measured using DHT11 sensor)

- Flood Probability (FP)

$$FP(\%) = k * PW * RH$$

$$k = 0.1 \sim 0.5$$

Calculation Citations:

<https://pubs.acs.org/doi/pdf/10.1021/ed028p428>

[https://www.jstage.jst.go.jp/article/jmsj/98/6/98\\_2020-058/\\_html/-char/en](https://www.jstage.jst.go.jp/article/jmsj/98/6/98_2020-058/_html/-char/en)

<https://www.weather.gov/lmk/humidity>

# Proposed approach

## I. 「通常」 モード

- I. 1時間ごとにデータを収集し、バッファに保存
- II. 洪水確率 (FP) を計算
- III. FPが70%を超えた場合、「緊急」 モードに移行
- IV. FPが70%未満の場合、引き続き「通常」 モードを継続

## II. 「緊急」 モード

- I. バッファ内の最初の3つのデータをハフマン表 $\alpha$ を使って送信（開始時のみ1回送信）
- II. 残りのバッファ内データは差分法とハフマン表 $\beta$ を使って送信
- III. 5分ごとにデータを収集し、バッファに保存
- IV. バッファが満杯になった場合、II.IIを使用してデータを送信
- V. FPを確認し、2回連続で70%未満の場合、「通常」 モードに戻る
- VI. FPが70%以上の場合、II.IIIに戻る

# Flood Probability (FP)

LoRaWANモジュール初期化...  
ネットワーク未参加! JOINを待機...  
Dragino LA66 ネットワーク未参加! JOINを待機...  
LoRaWANネットワークに接続成功!

送信データ: 32.00°C, 80.00%  
Flood probability (%): 91.38  
LoRaWAN送信コマンド発行

→ sent

Dragino LA66 ネットワーク未参加! JOINを待機...  
LoRaWANネットワークに接続成功!

送信データ: 26.30°C, 37.00%  
Flood probability (%): 14.05

→ Not sent

送信データ: 26.30°C, 36.00%  
Flood probability (%): 13.30  
送信データ: 26.30°C, 36.00%

# ハフマン符号表 $\alpha$

- 生データの圧縮に使用
- 現在の研究で提案された静的ハフマン表を使用するか、閾値に重みを与えて動的なハフマン表を作成する
- この方法で送信するのはわずか3つのデータであるため、計算負荷の少ない別の手法を使用することも検討可能

# ハフマン符号表 $\beta$

- “Data Compression in Wireless Sensor Nodes with LoRa”より式を採用

$$\hat{v}_i = \frac{5}{4}v_{i-1} + \frac{1}{2}v_{i-2} - \frac{3}{4}v_{i-3}$$

- 上記の式を用いて次のデータを推定し、実際のデータと比較してその差分のみを送信する（差分法）
- ハフマン表  $\beta$  は、この差分の発生確率を基に使用する

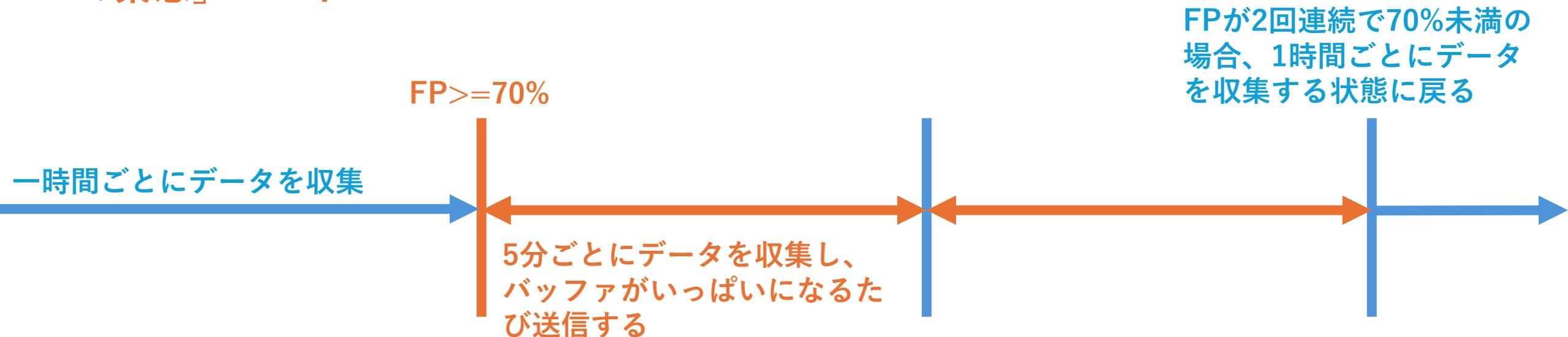
Actual Data	$v_0$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$
Estimate	-	-	-	$\widehat{v}_3 = \frac{5}{4}v_0 + \frac{1}{2}v_1 - \frac{3}{4}v_2$	$\widehat{v}_4 = \dots$	$\widehat{v}_5 = \dots$
Send	$v_0$	$v_1$	$v_2$	$\widehat{v}_3 - v_3$	$\widehat{v}_4 - v_4$	$\widehat{v}_5 - v_5$

# Huffman Table $\beta$

TABLE IV. THE HUFFMAN DICTIONARY.

$n_i$	$s_i$	$d_i$
<b>0</b>	00	0
<b>1</b>	01	-1,+1
<b>2</b>	100	-3,-2,2,3
<b>3</b>	101	-7,...,-4,+4,...,7
<b>4</b>	110	-15,...,-8,8,...,15
<b>5</b>	1110	-31,...,-16,16,...,31
<b>6</b>	11110	-63,...,-32,32,...,63
<b>7</b>	111110	-127,...,-64,64,...,127
<b>8</b>	1111110	-255,...,-128,128,...,255

I. 「通常」モード  
II. 「緊急」モード



※ここでのバッファとは、一度に送信できるデータ量のことです。

考えられる質問:

- このデータ量は洪水の発生状況を予測するのに十分か？
- 洪水の開始を判断する指標として70%は適切か？
- 1時間および5分という時間設定は長すぎるか、短すぎるか？