Compromised through Compression

Privacy Implications of Smart Meter Traffic Analysis

Pol Van Aubel

pol.vanaubel@cs.ru.nl

https://www.polvanaubel.com/

September 6–9, 2021

iCIS—Digital Security

Radboud University

Researchers

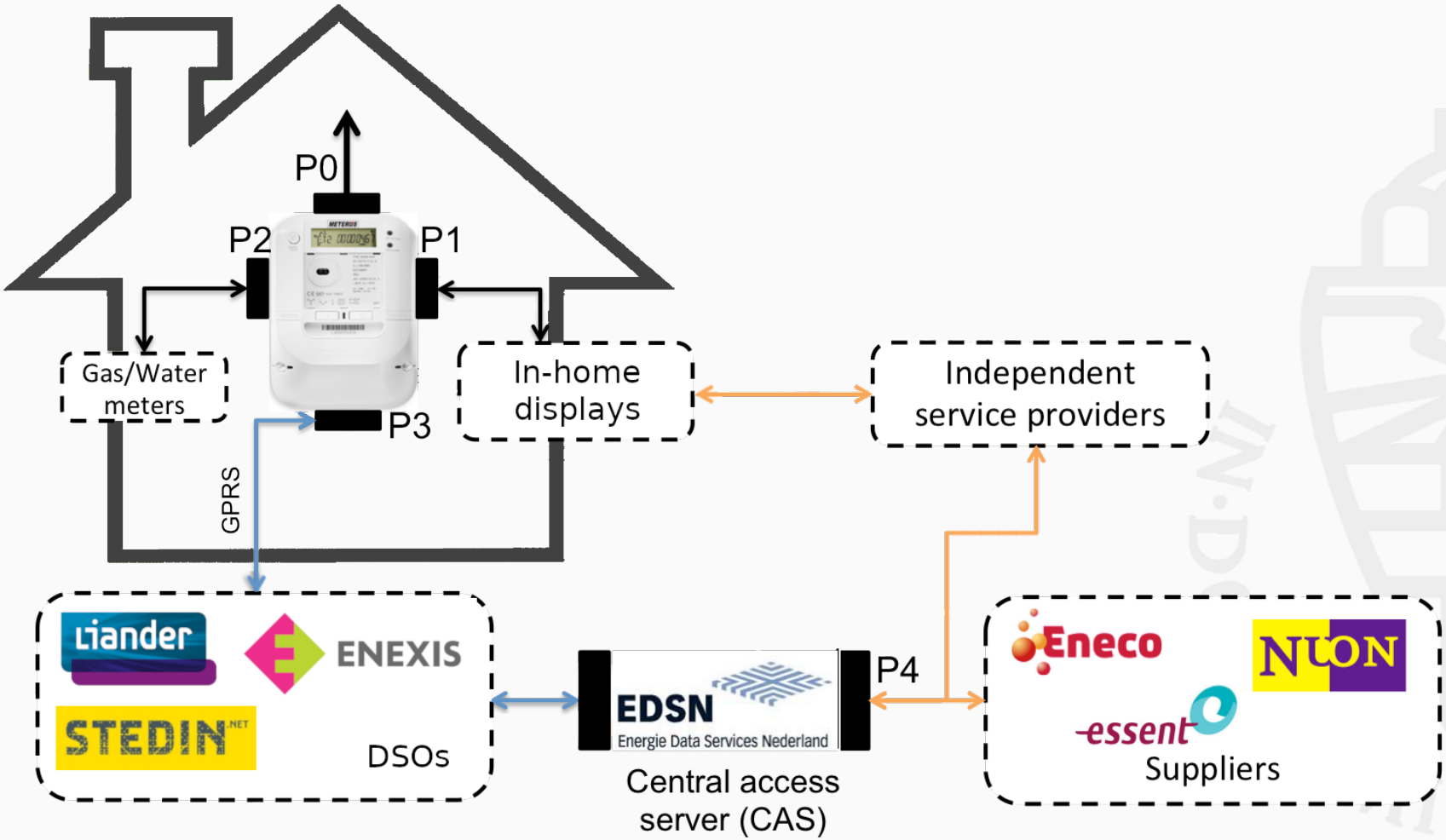
Pol Van Aubel

pol.vanaubel@cs.ru.nl
https://www.polvanaubel.com/

Erik Poll

erikpoll@cs.ru.nl
https://www.cs.ru.nl/~erikpoll/

Actors in the Dutch energy grid



Article 8

Right to respect for private and family life

(1) Everyone has the right to respect for his private and family life, his home and his correspondence.

(2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 8

Right to respect for private and family life

(1) Everyone has the right to respect for his private and family life, his home and his correspondence.

(2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

This is personal information that should be protected

Protocol used to transmit measurements daily is **DLMS/COSEM**

DLMS/COSEM allows for **encryption**

So an **attacker** observing this traffic **cannot see measurements**

This is personal information that should be protected

Protocol used to transmit measurements daily is **DLMS/COSEM**

DLMS/COSEM allows for **encryption**

So an **attacker** observing this traffic **cannot see measurements**

Encrypted messages **leak** the **message length** of their plaintext

Message length is a known side-channel

- An observable side-effect that leaks information about the encrypted message

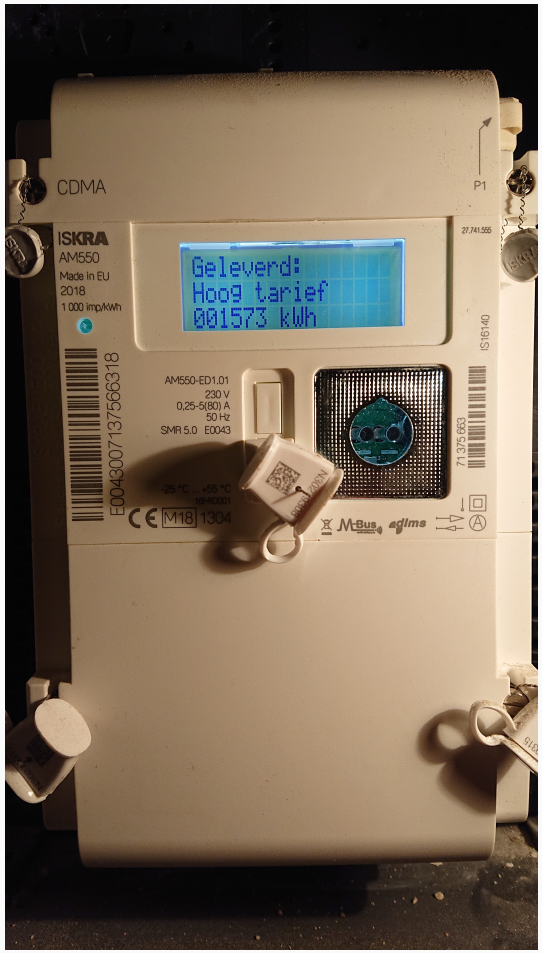
Message length is a known side-channel

- An observable side-effect that leaks information about the encrypted message
- Length identified as a problem by Kelsey already in 2002
- Hypothesized as a problem for SPDY by Langley in 2011

Message length is a known side-channel

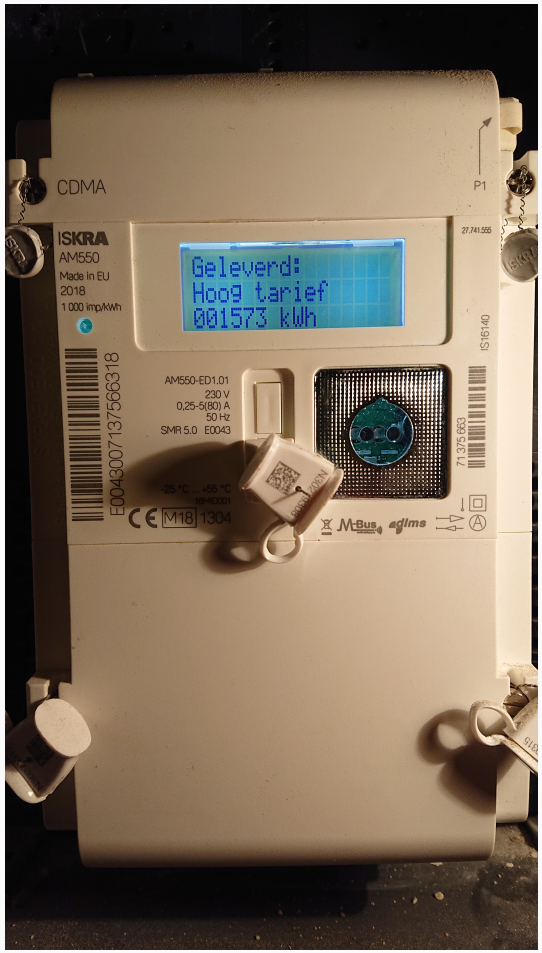
- An observable side-effect that leaks information about the encrypted message
- Length identified as a problem by Kelsey already in 2002
- Hypothesized as a problem for SPDY by Langley in 2011
- Used in CRIME against HTTPS, SPDY, general TLS in 2012 (Rizzo, Duong)
- Followed by BREACH against HTTP in 2013 (Prado, Harris, Gluck)

Bytes cost money, fewer bytes cost less money



Bytes cost money, fewer bytes cost less money

1573000 Wh $\rightarrow$ 32-bit



Bytes cost money, fewer bytes cost less money

1573000 Wh $\rightarrow$ 32-bit

Bytes cost money, fewer bytes cost less money

1573000 Wh $\rightarrow$ 32-bit

1576000 Wh $\rightarrow$ 32-bit

Bytes cost money, fewer bytes cost less money

1573000 Wh $\rightarrow$ 32-bit

$$\begin{array}{r} 1576000 \\ 1573000- \\ \hline 3000 \end{array}$$

1576000 Wh $\rightarrow$ 32-bit

Bytes cost money, fewer bytes cost less money

1573000 Wh $\rightarrow$ 32-bit

1576000

1573000−

3000

Delta Types

1576000 Wh $\rightarrow$ 32-bit

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Delta: Replace all but the first measurement with a Delta Type

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Delta: Replace all but the first measurement with a Delta Type

Min-length: Use the smallest type possible for each measurement

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Delta: Replace all but the first measurement with a Delta Type

Min-length: Use the smallest type possible for each measurement

32-bit: Always use a 32-bit Delta for each measurement

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Delta: Replace all but the first measurement with a Delta Type

Min-length: Use the smallest type possible for each measurement

32-bit: Always use a 32-bit Delta for each measurement

16-bit: Always use a 16-bit Delta for each measurement

Possible ways to encode messages

- Distilled by us from the DLMS/COSEM standard & tests

Normal: No smart encoding at all

NULL: Replace *candidates* with `NULL`

- All timestamps except the first
- All measurements that are 0

  (Few measurements in our dataset are 0)

Delta: Replace all but the first measurement with a Delta Type

Min-length: Use the smallest type possible for each measurement

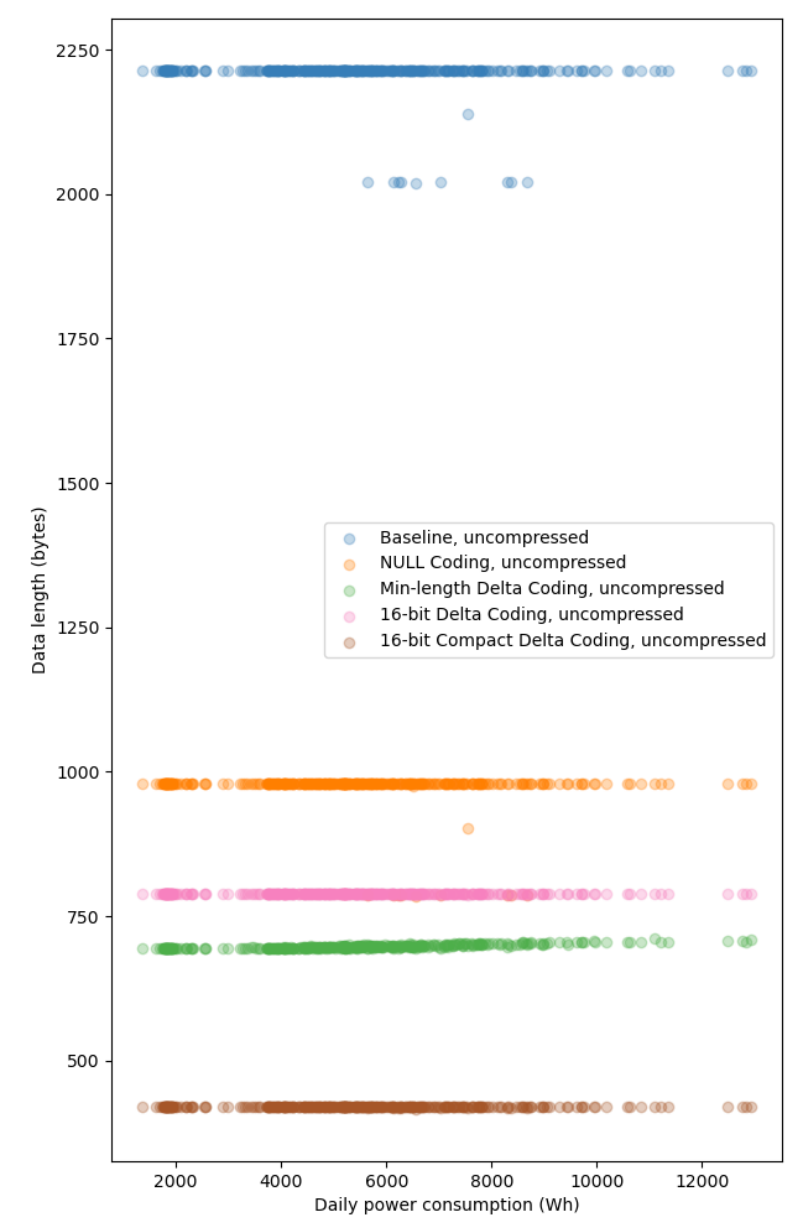
32-bit: Always use a 32-bit Delta for each measurement

16-bit: Always use a 16-bit Delta for each measurement

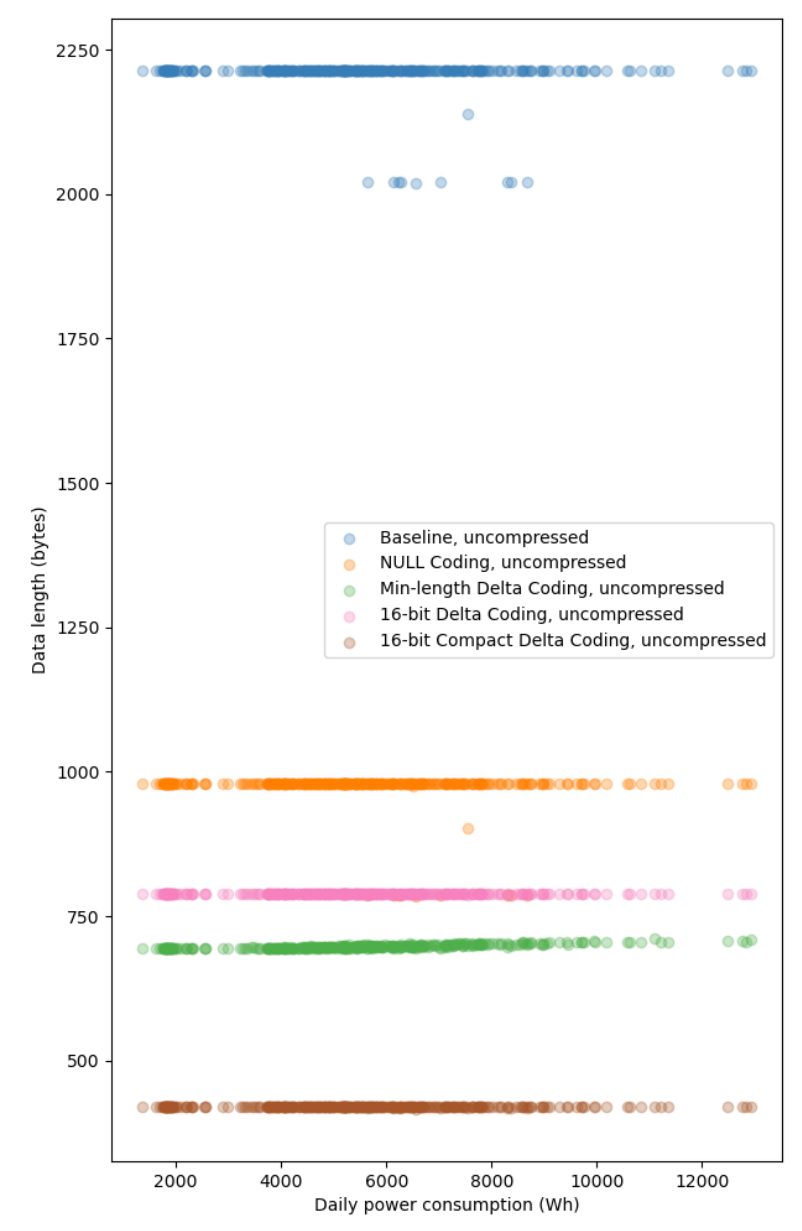
16-bit Compact: The same, except using a compacter array structure

Proposed by us, not distilled from the standard

Do these encodings result in a problematic correlation?

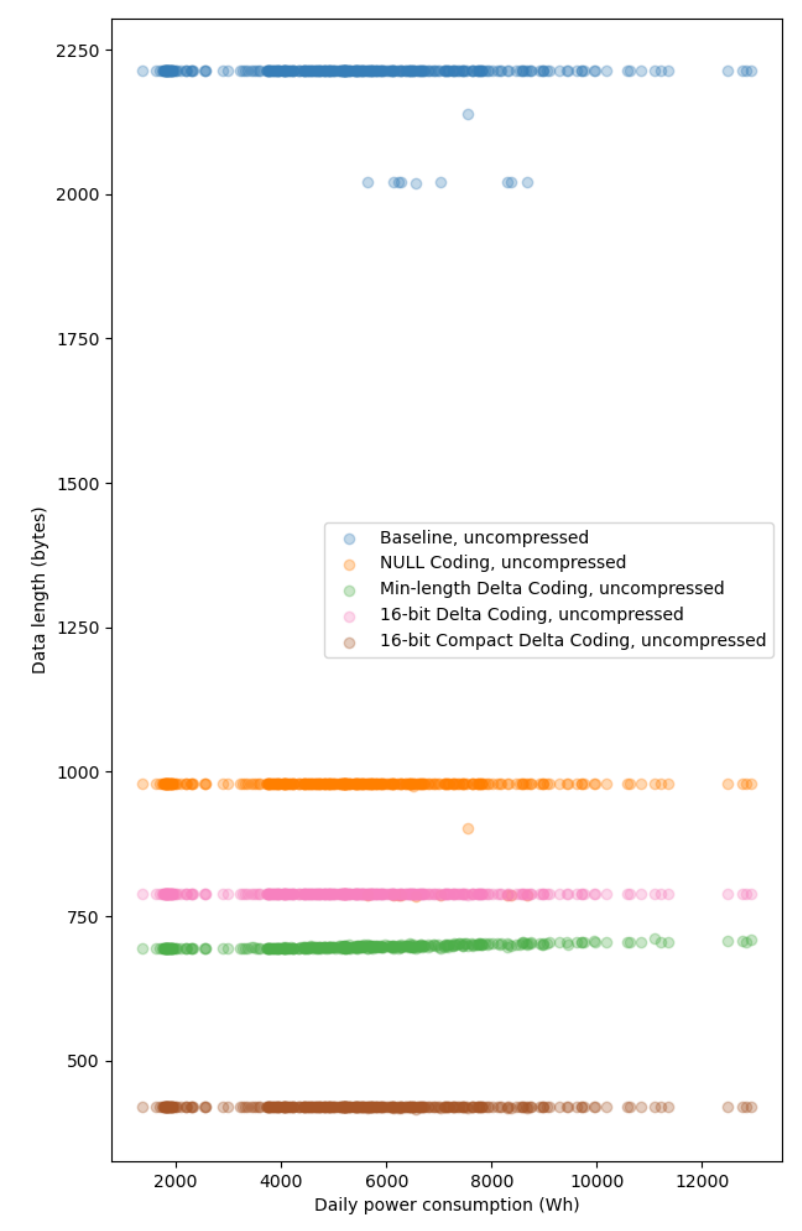


Do these encodings result in a problematic correlation?



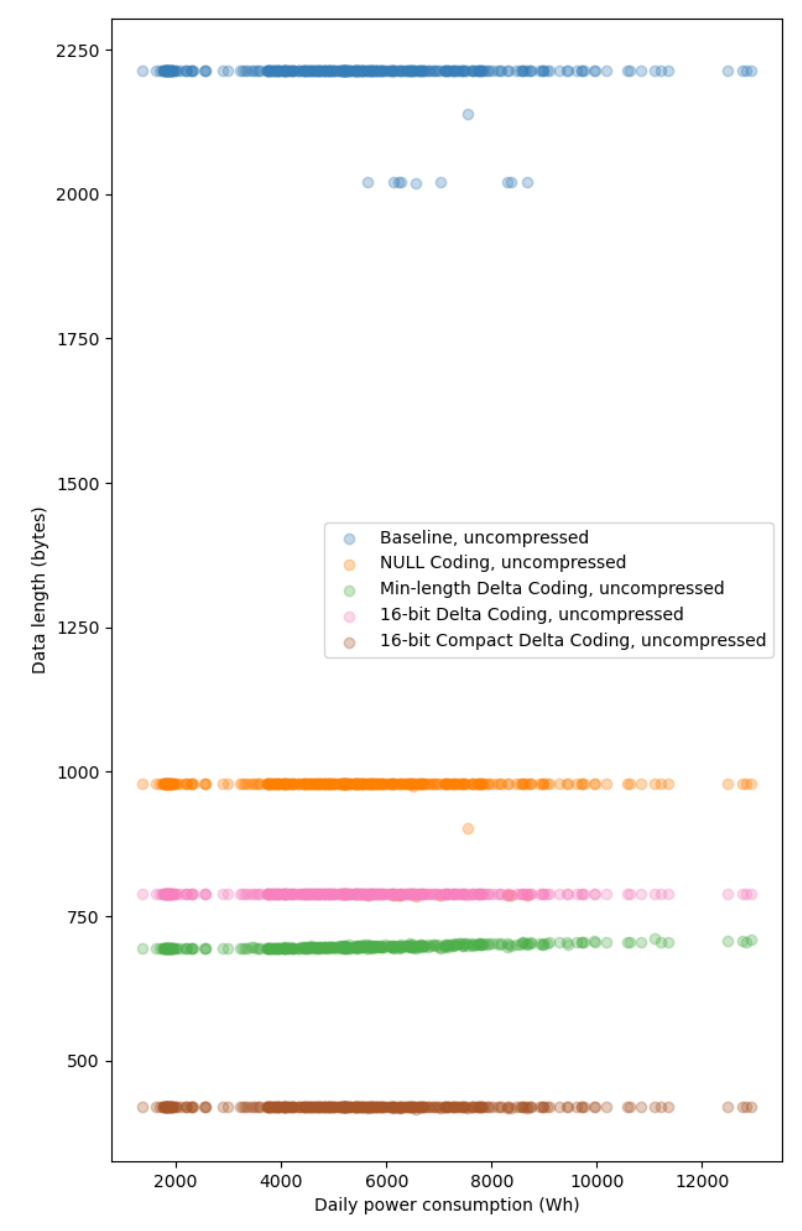
- All except Minimum-length Delta Coding do not introduce a correlation between message length and energy use

Do these encodings result in a problematic correlation?



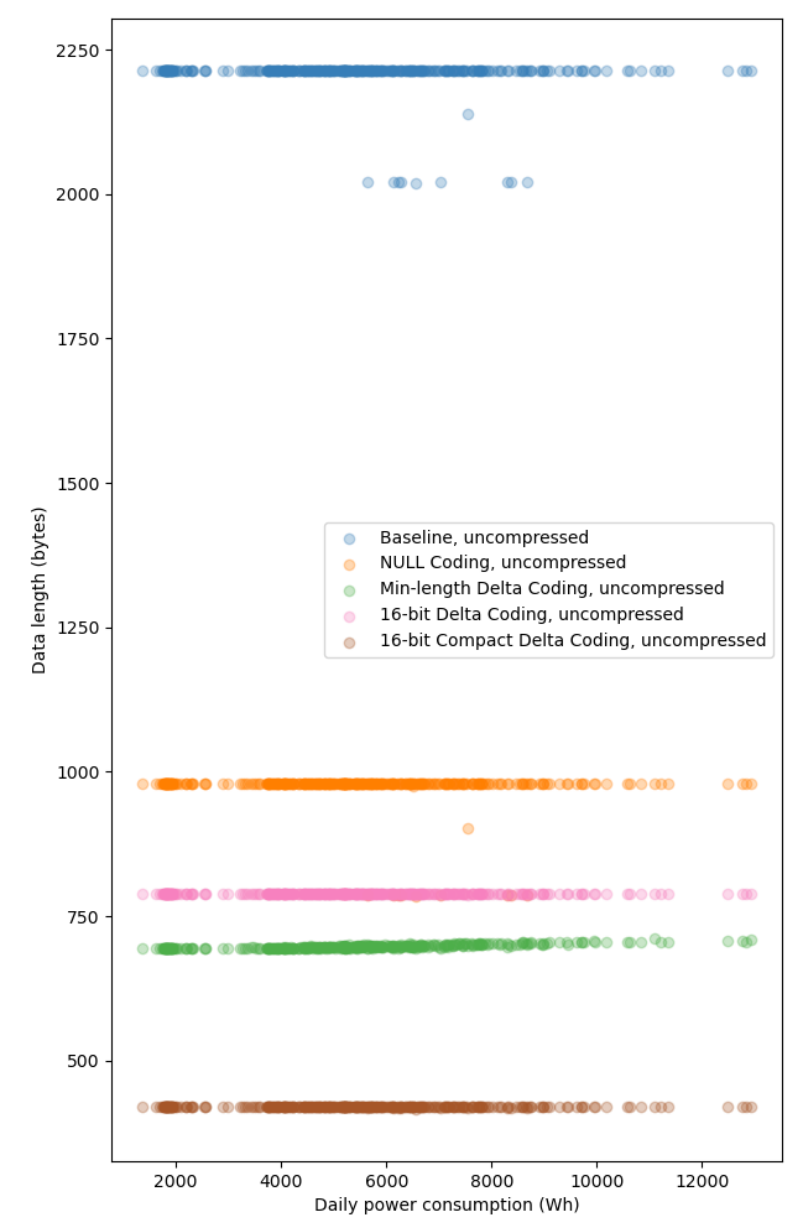
- All except Minimum-length Delta Coding do not introduce a correlation between message length and energy use
- Minimum-length Delta Coding may *look* safe, but that correlation is problematic

Do these encodings result in a problematic correlation?



- All except Minimum-length Delta Coding do not introduce a correlation between message length and energy use
- Minimum-length Delta Coding may *look* safe, but that correlation is problematic
- Fortunately, 16-bit Delta Coding is almost as good

Do these encodings result in a problematic correlation?



- All except Minimum-length Delta Coding do not introduce a correlation between message length and energy use
- Minimum-length Delta Coding may *look* safe, but that correlation is problematic
- Fortunately, 16-bit Delta Coding is almost as good
- and 16-bit **Compact** Delta Coding outperforms it significantly

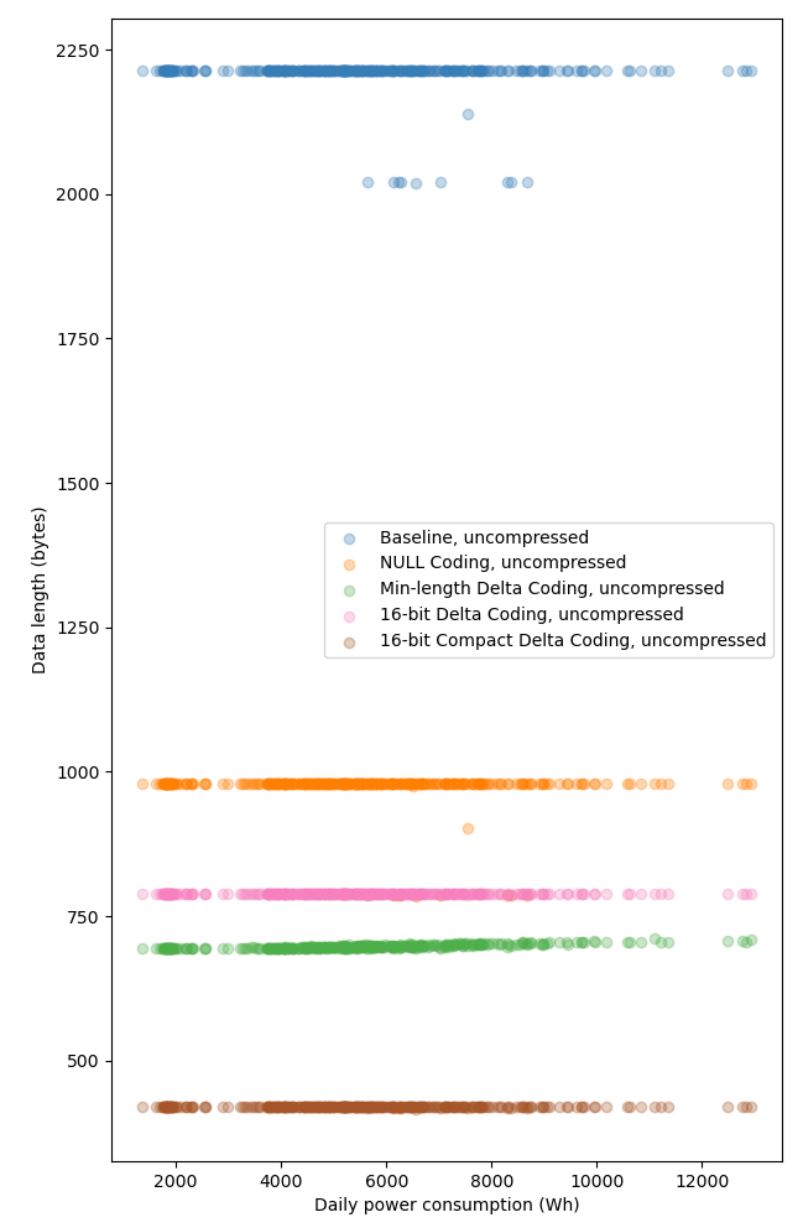
What about compression?

- The **existing** standard defines a **compression** mechanism
  - Not to be confused with encoding
  - Used on the message as a whole

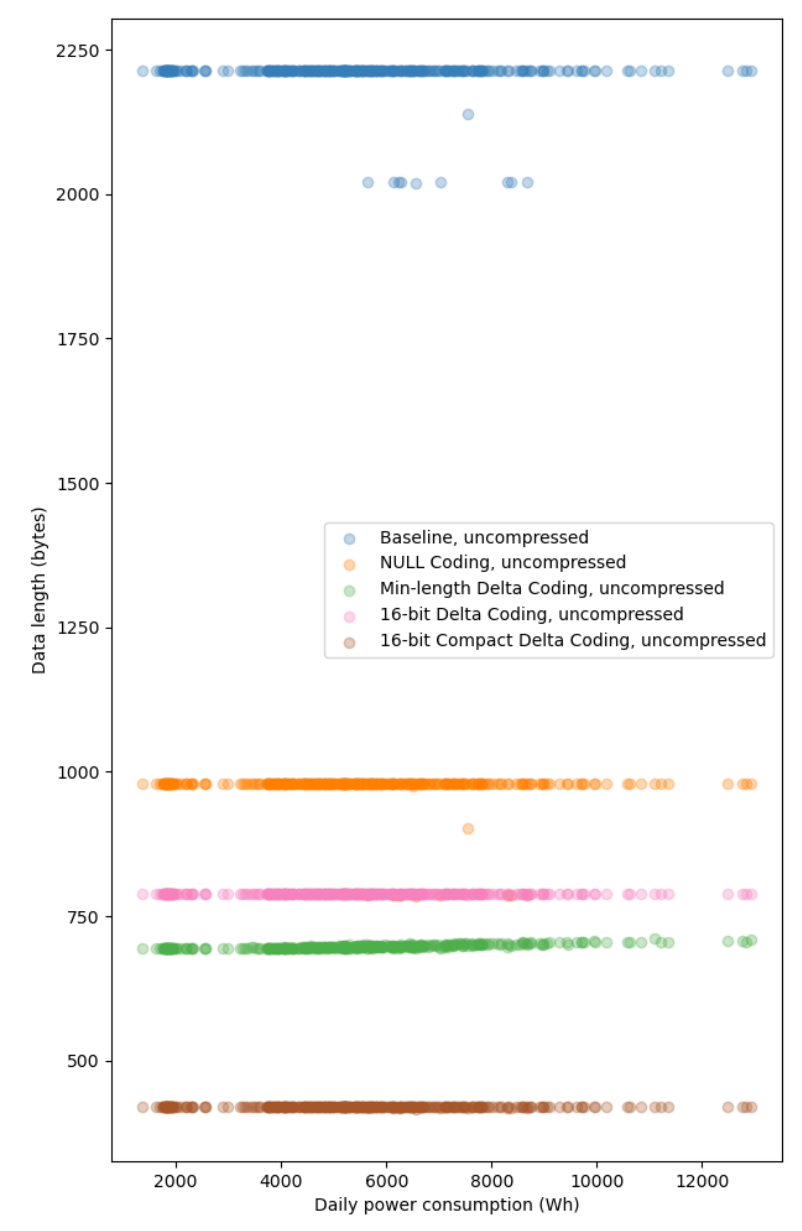
What about compression?

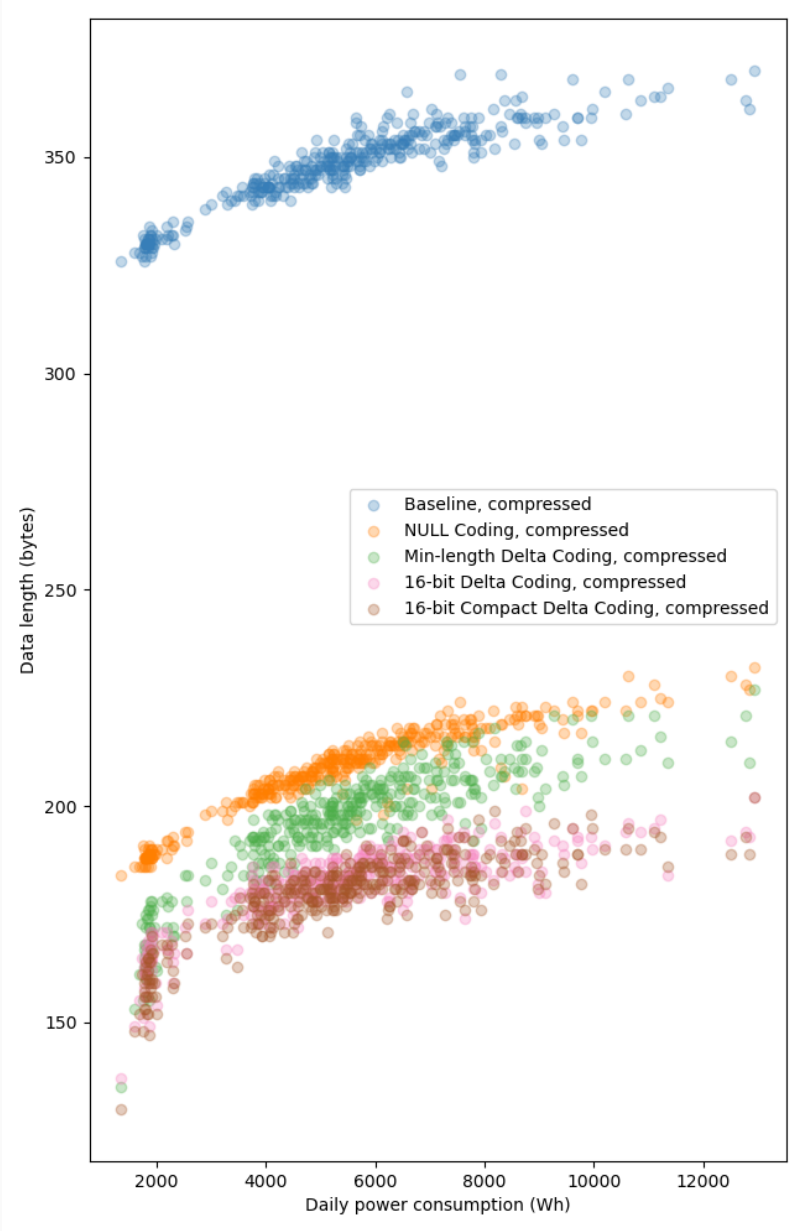
- The **existing** standard defines a **compression** mechanism
  - Not to be confused with encoding
  - Used on the message as a whole
- Compression is *intended* to *significantly decrease size*
  - Actual size after compression depends a lot on the contents of the message

Does compression result in a problematic correlation?

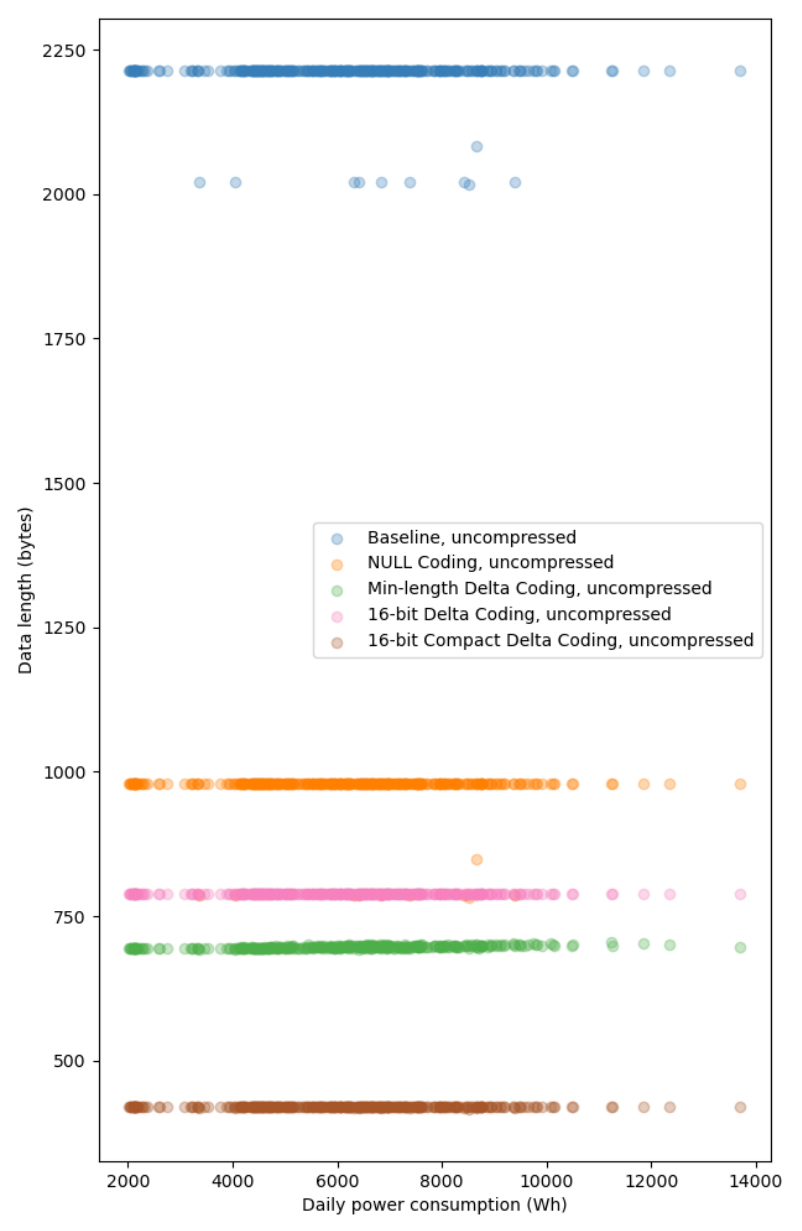


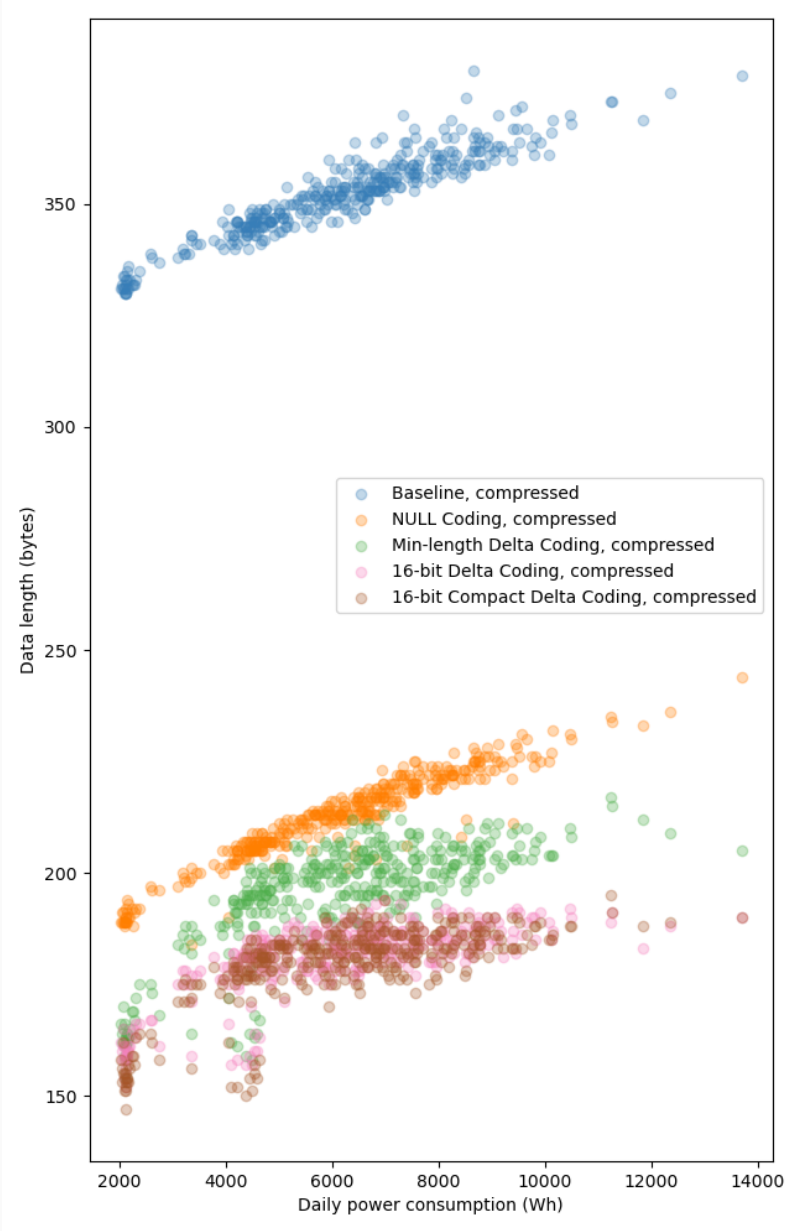
Does compression result in a problematic correlation?



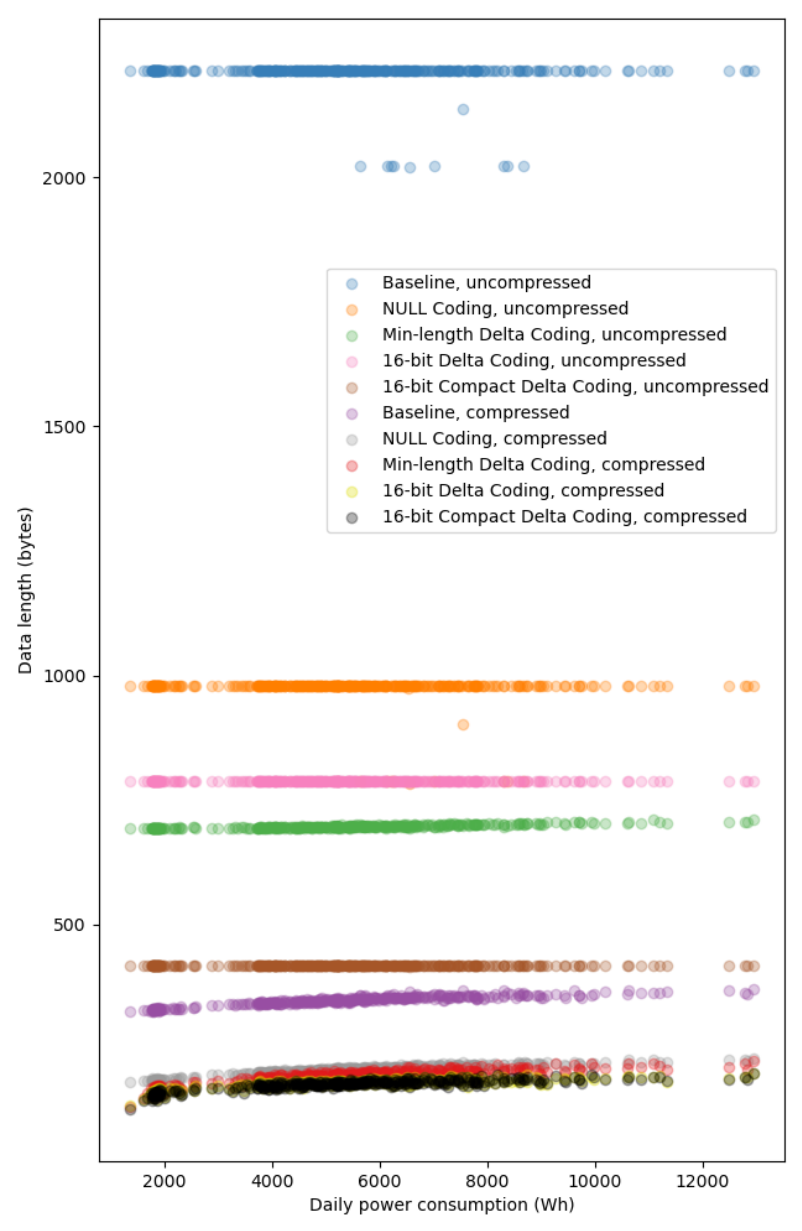


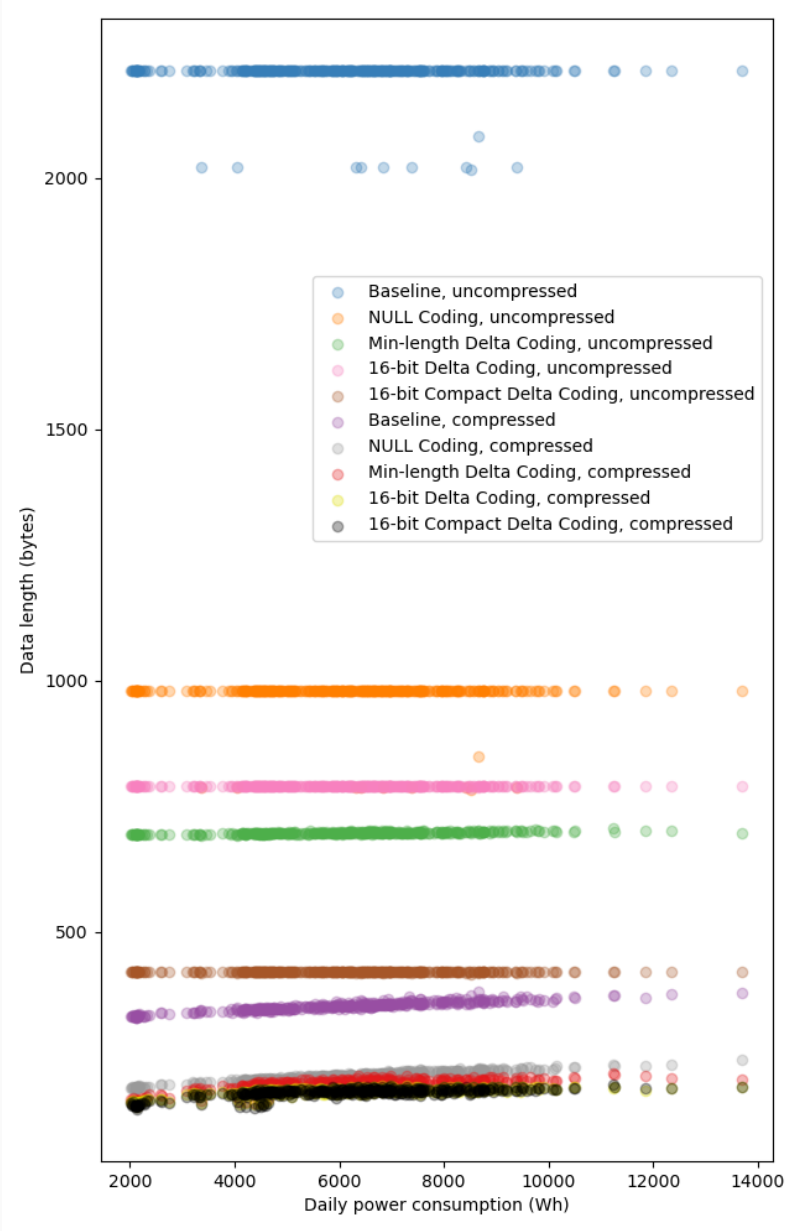
This correlation is visible for most households in our dataset





16-bit Compact Delta Coding gets very close to compression

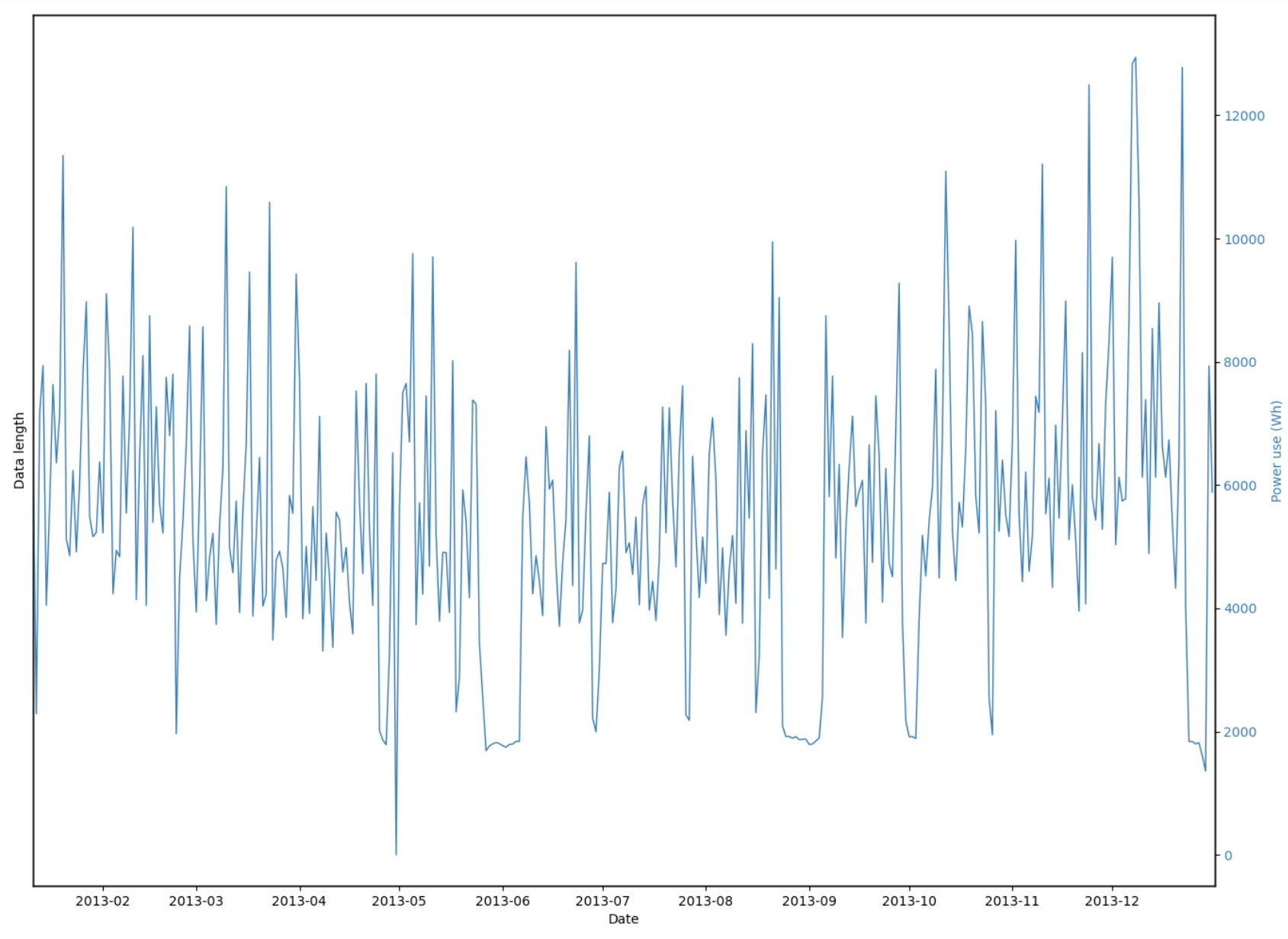




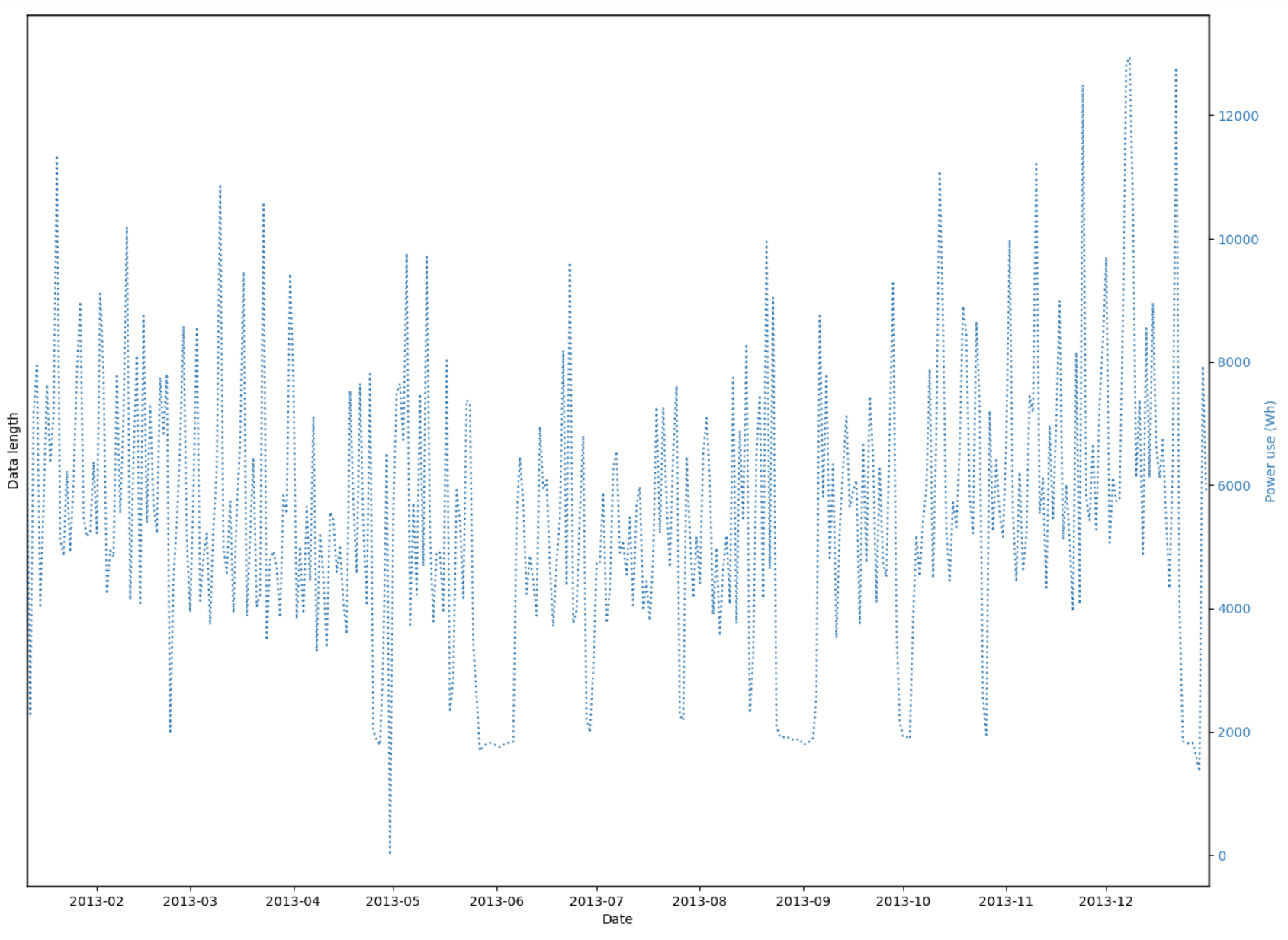
What can we get from this side-channel?

Subject 17

Yearly power use of “17”

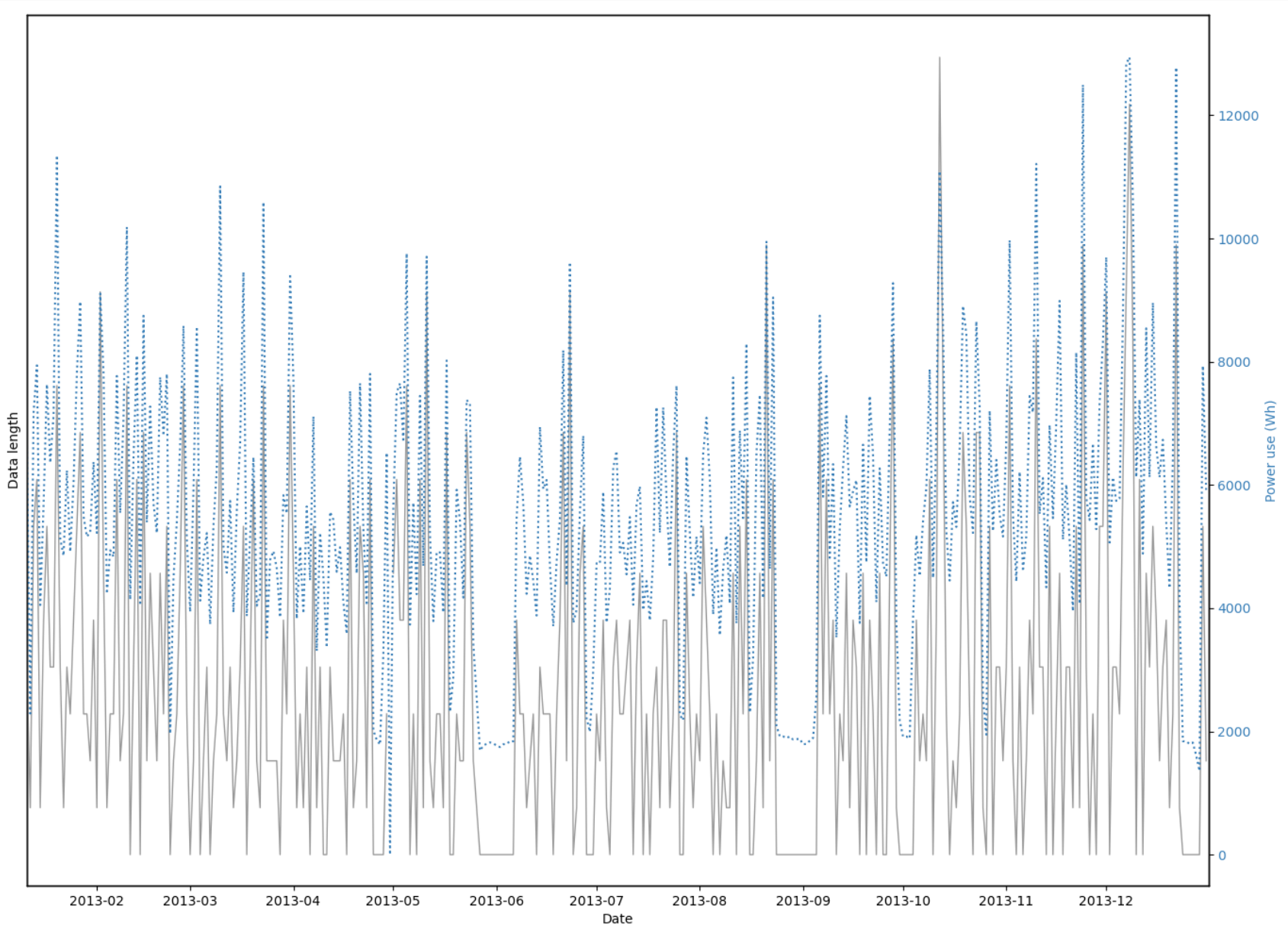


Yearly power use of “17”

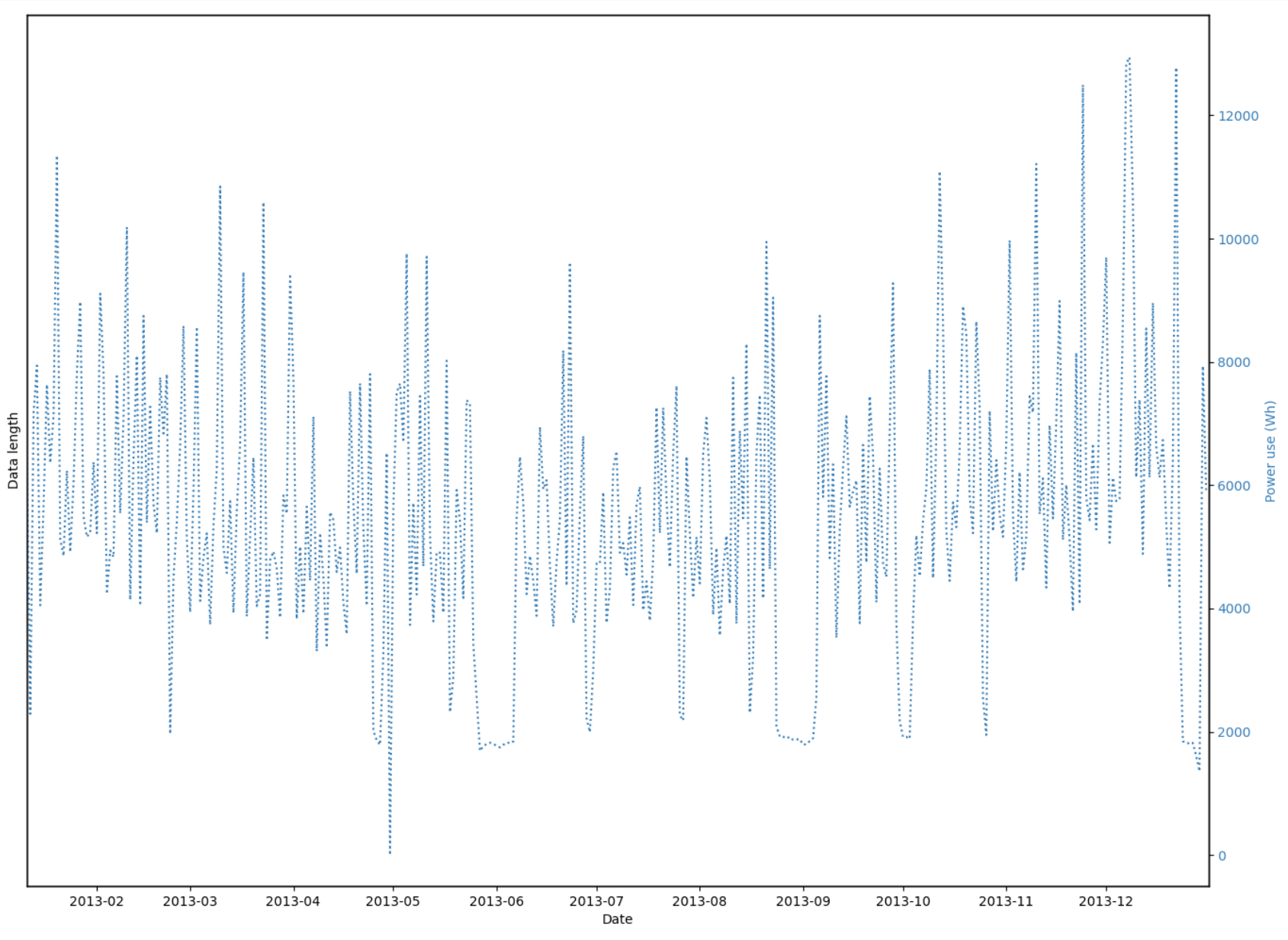


Yearly power use of “17”

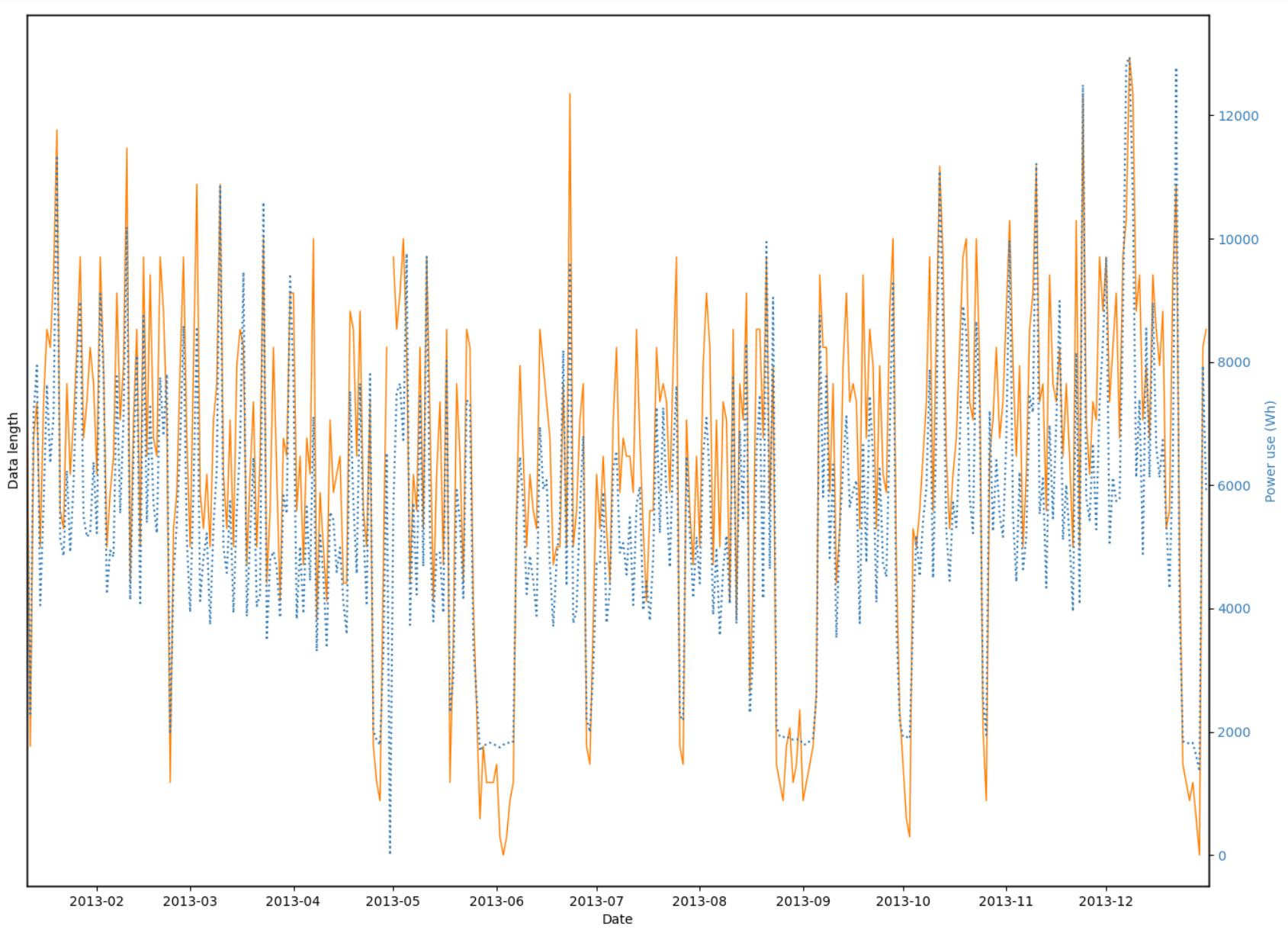
Yearly power use of “17”



Yearly power use of “17”

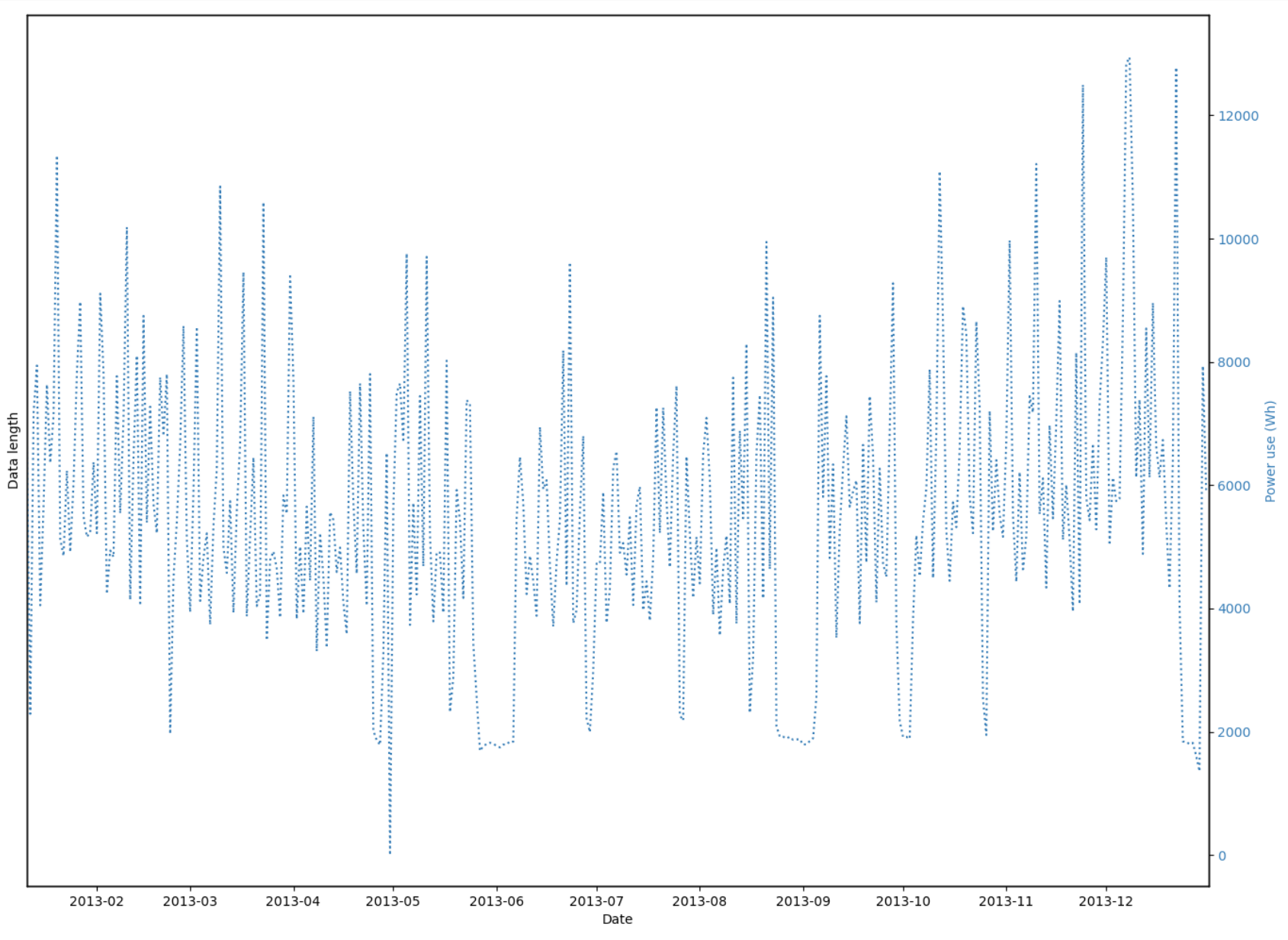


Yearly power use of “17”

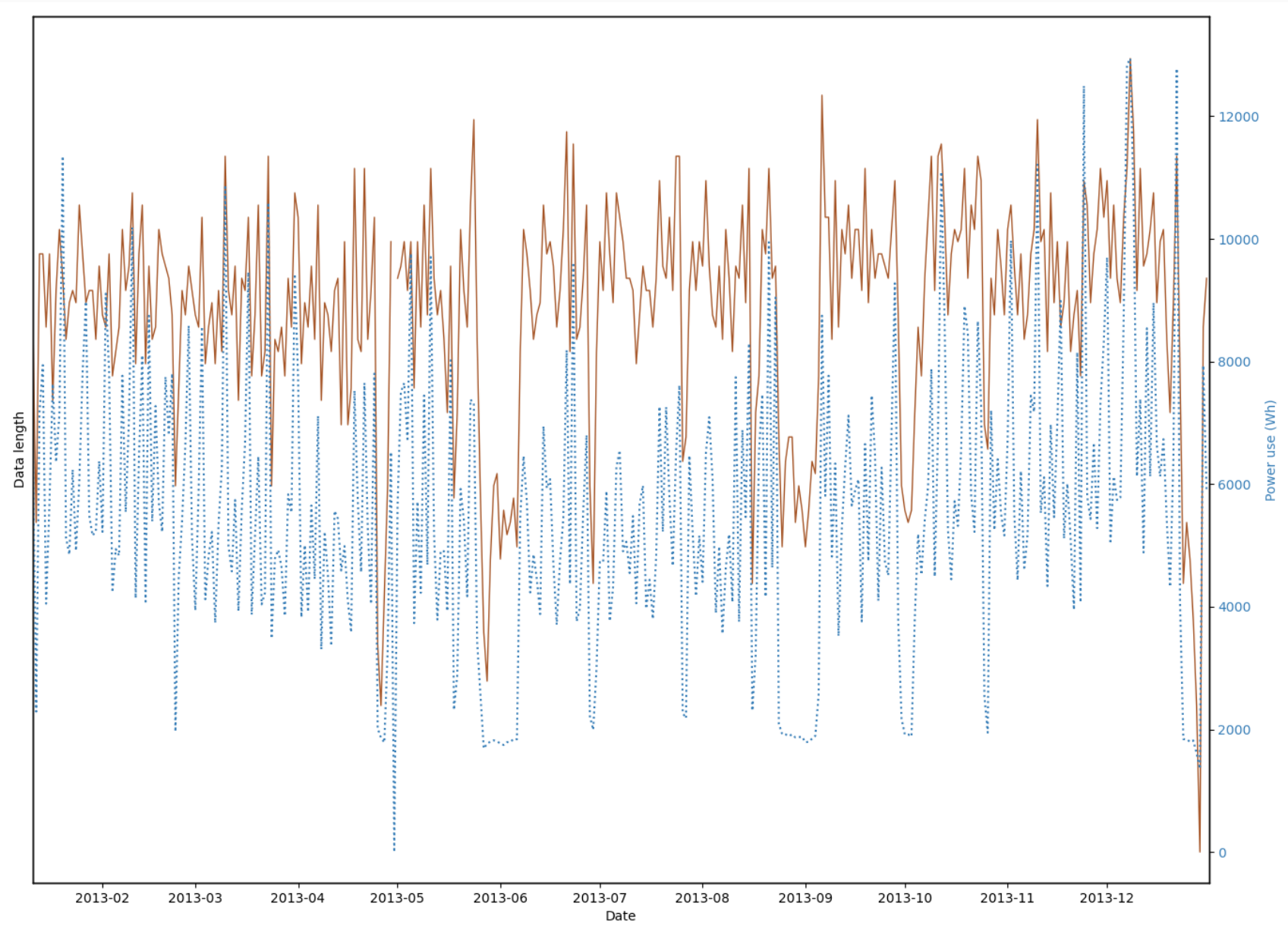


Yearly power use of “17”

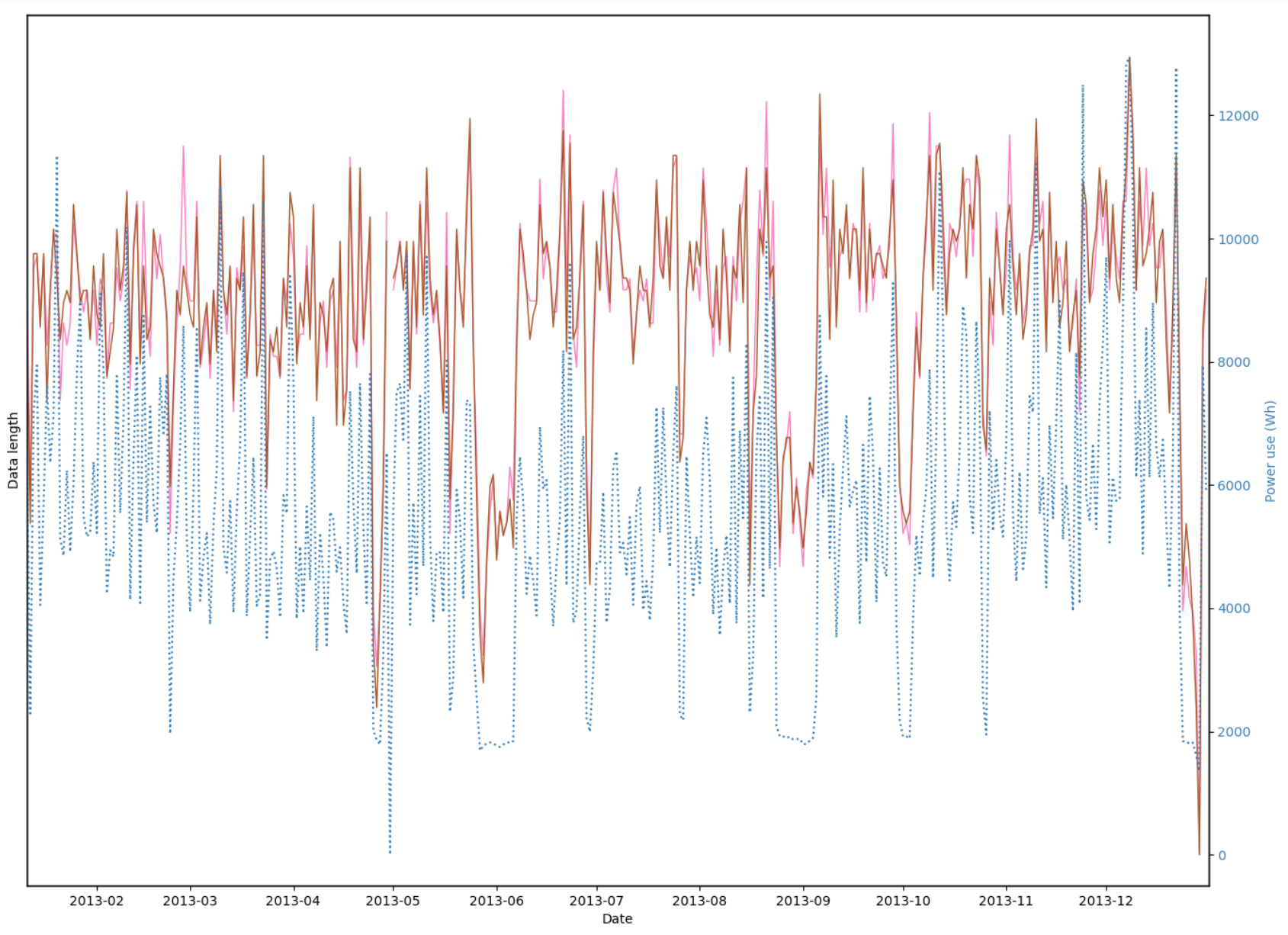
Yearly power use of “17”



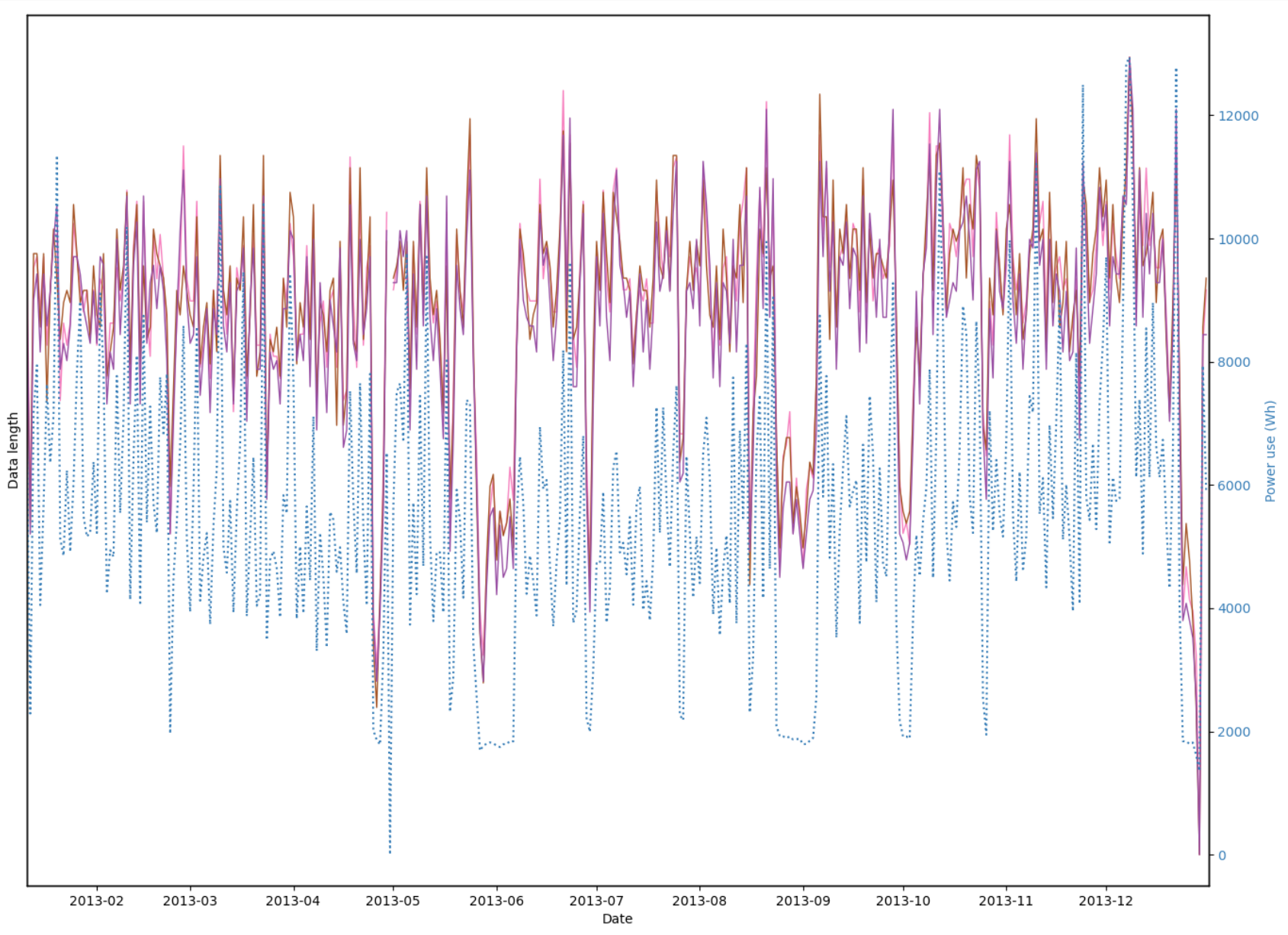
Yearly power use of “17”



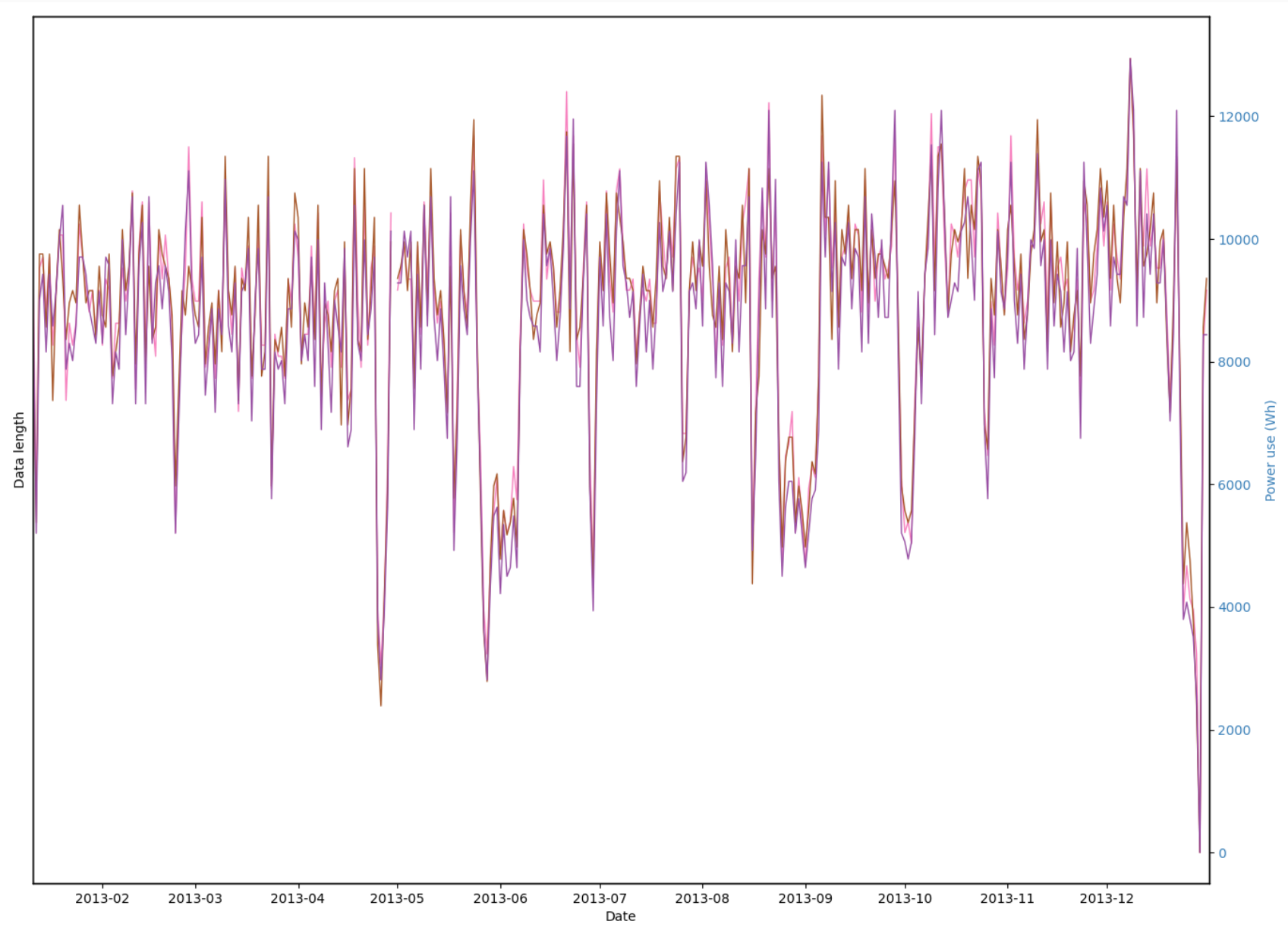
Yearly power use of “17”



Yearly power use of “17”



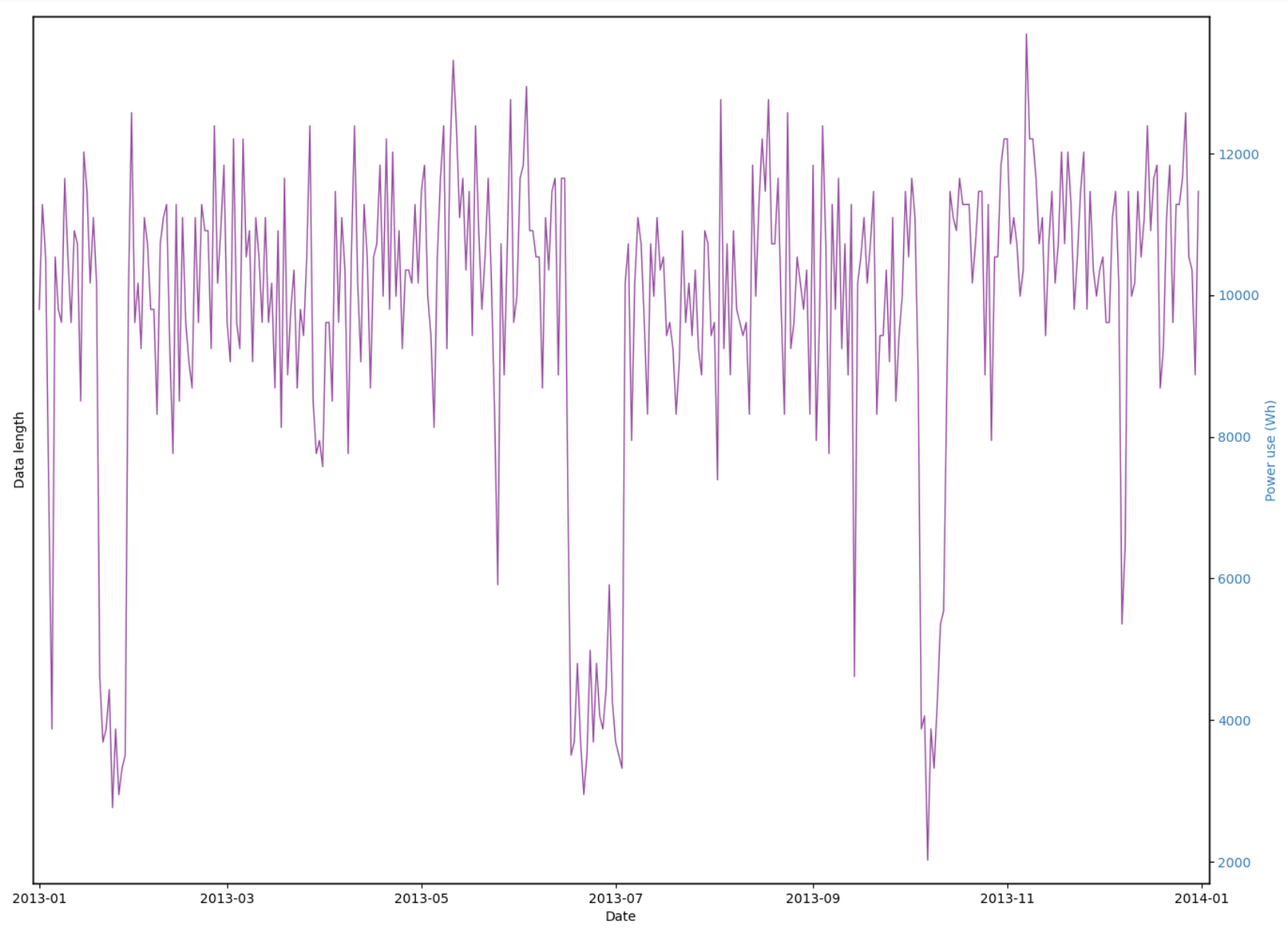
Yearly power use of “17”



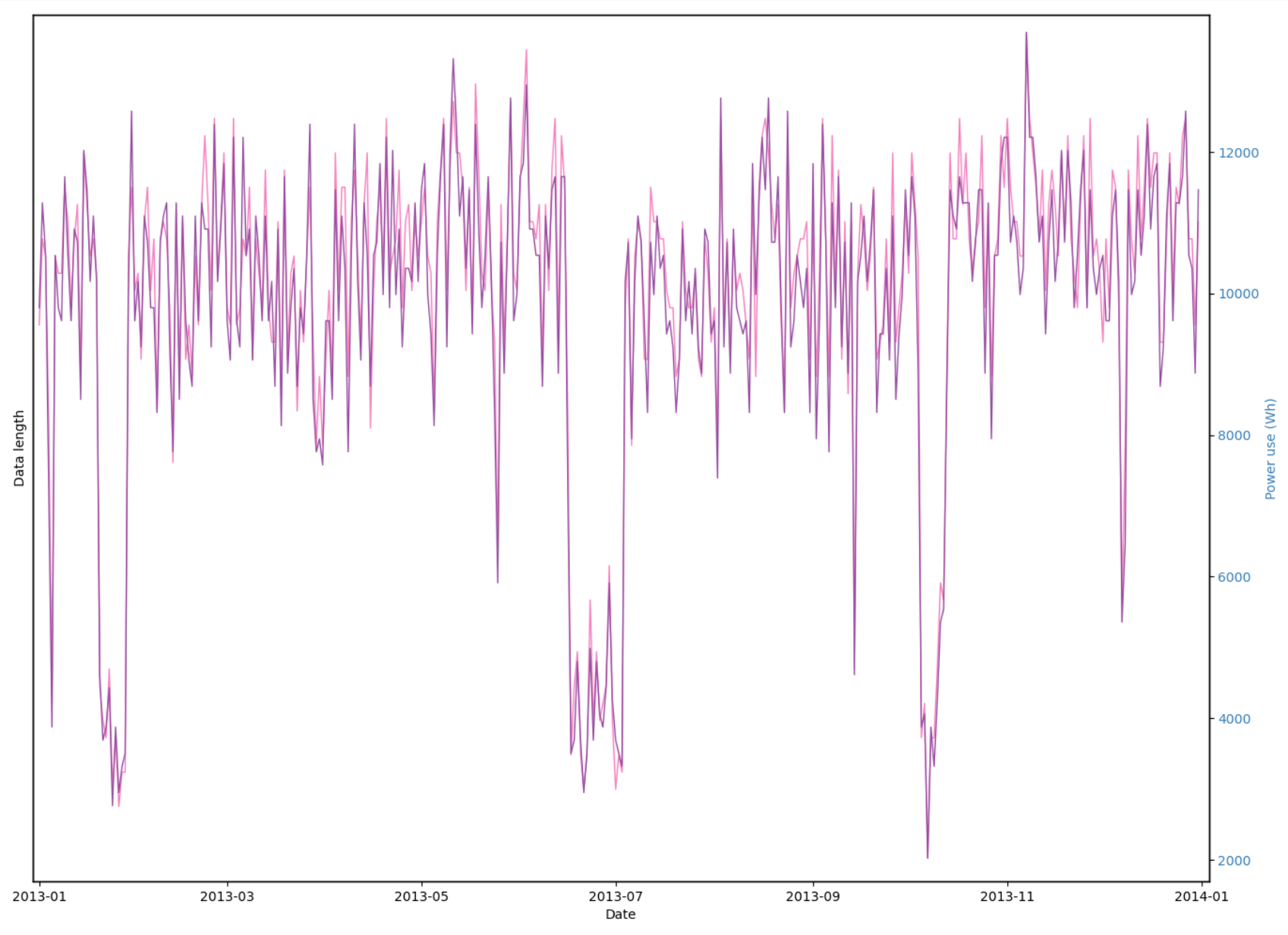
What can we get from this side-channel?

Subject 46

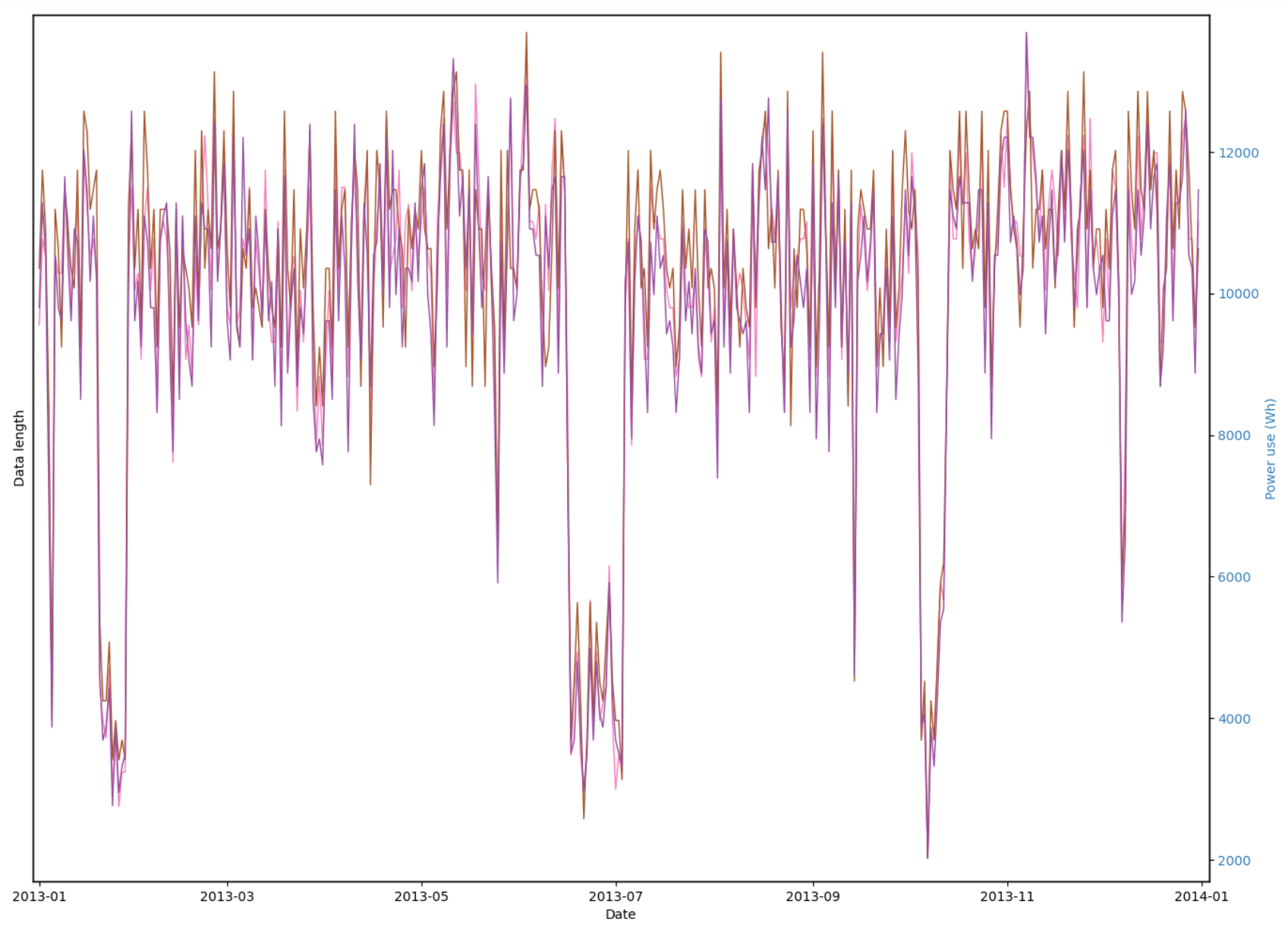
Compressed data of “46”



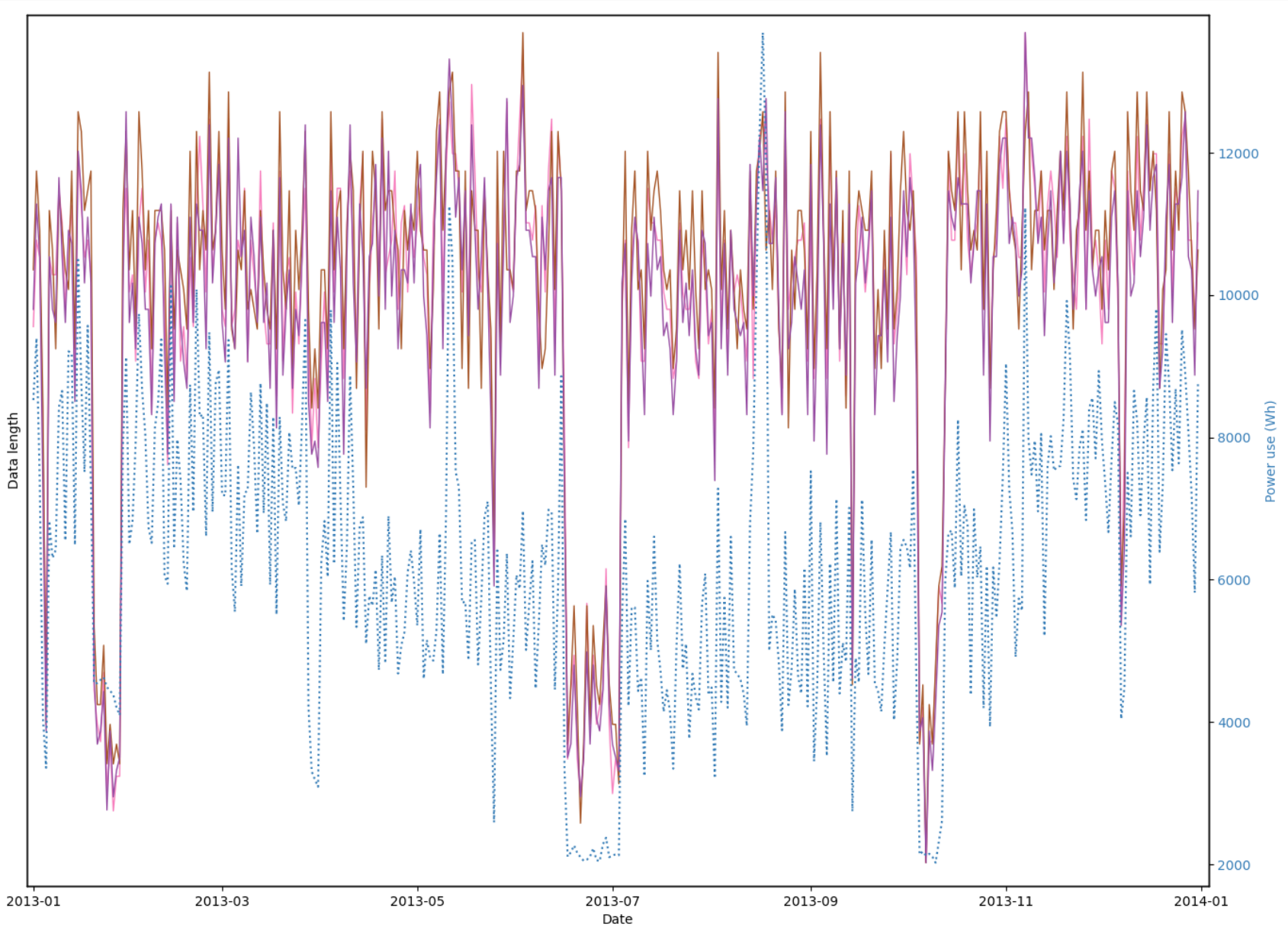
Compressed data of “46”



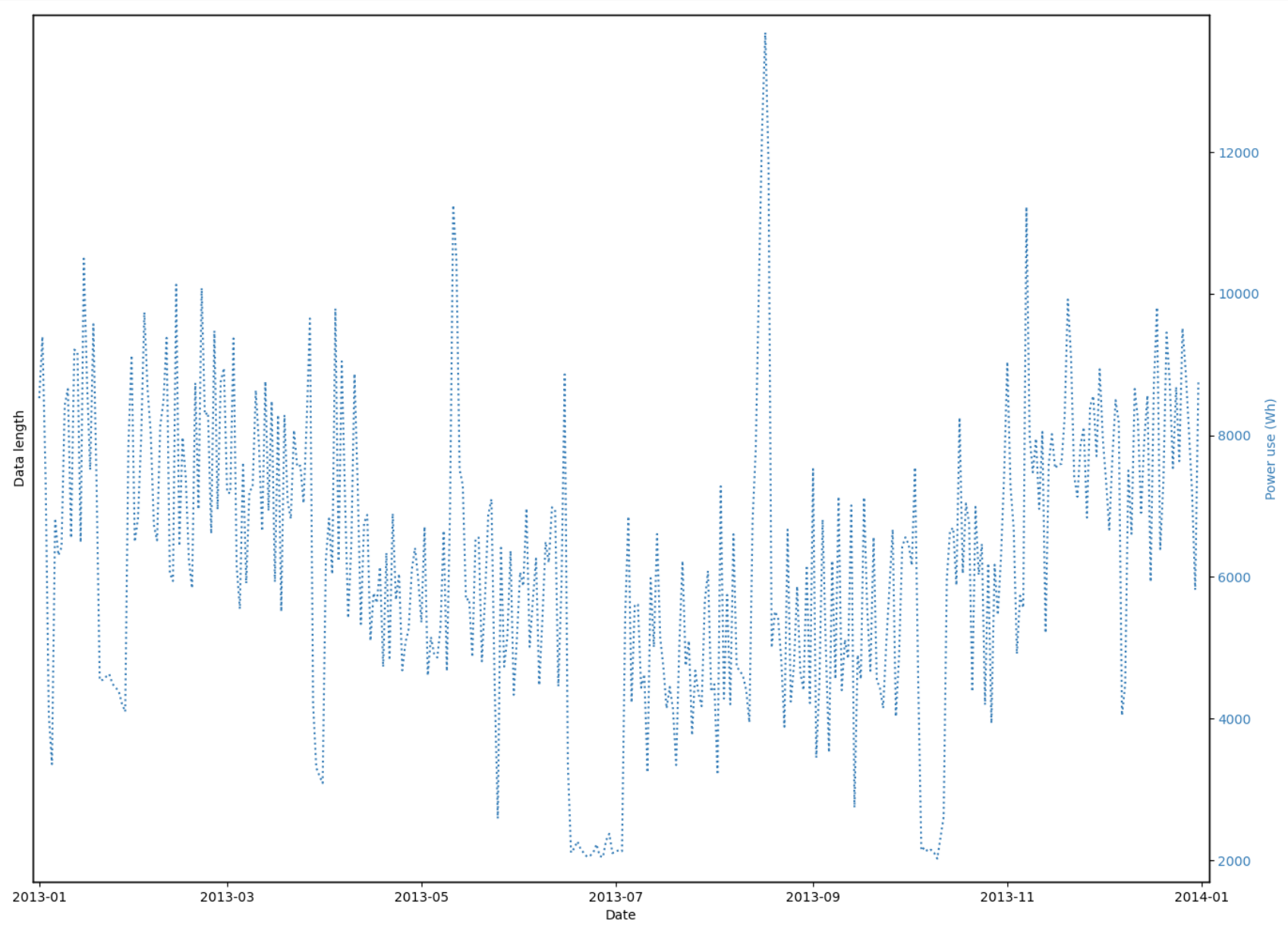
Compressed data of “46”



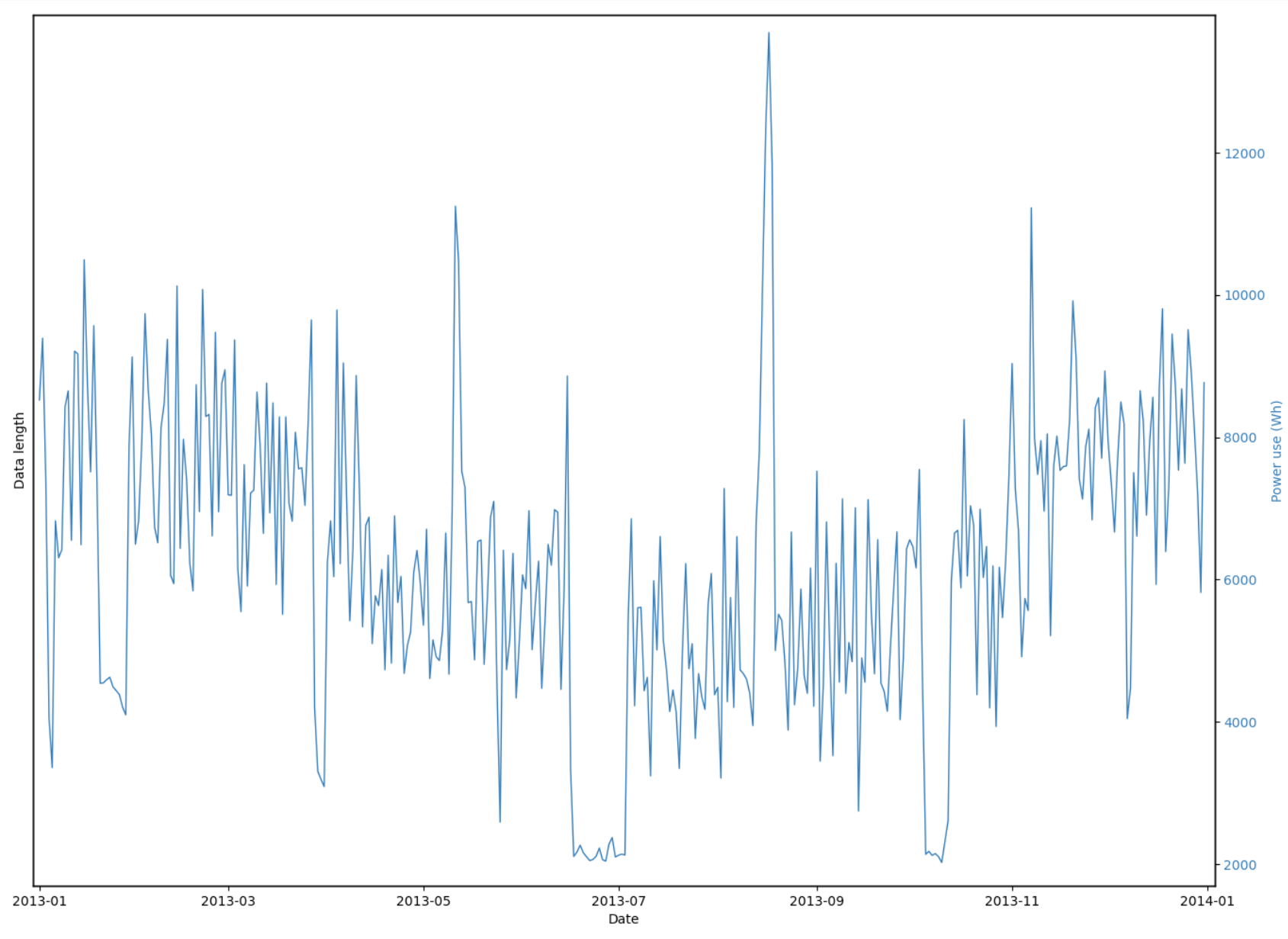
Compressed data of “46”



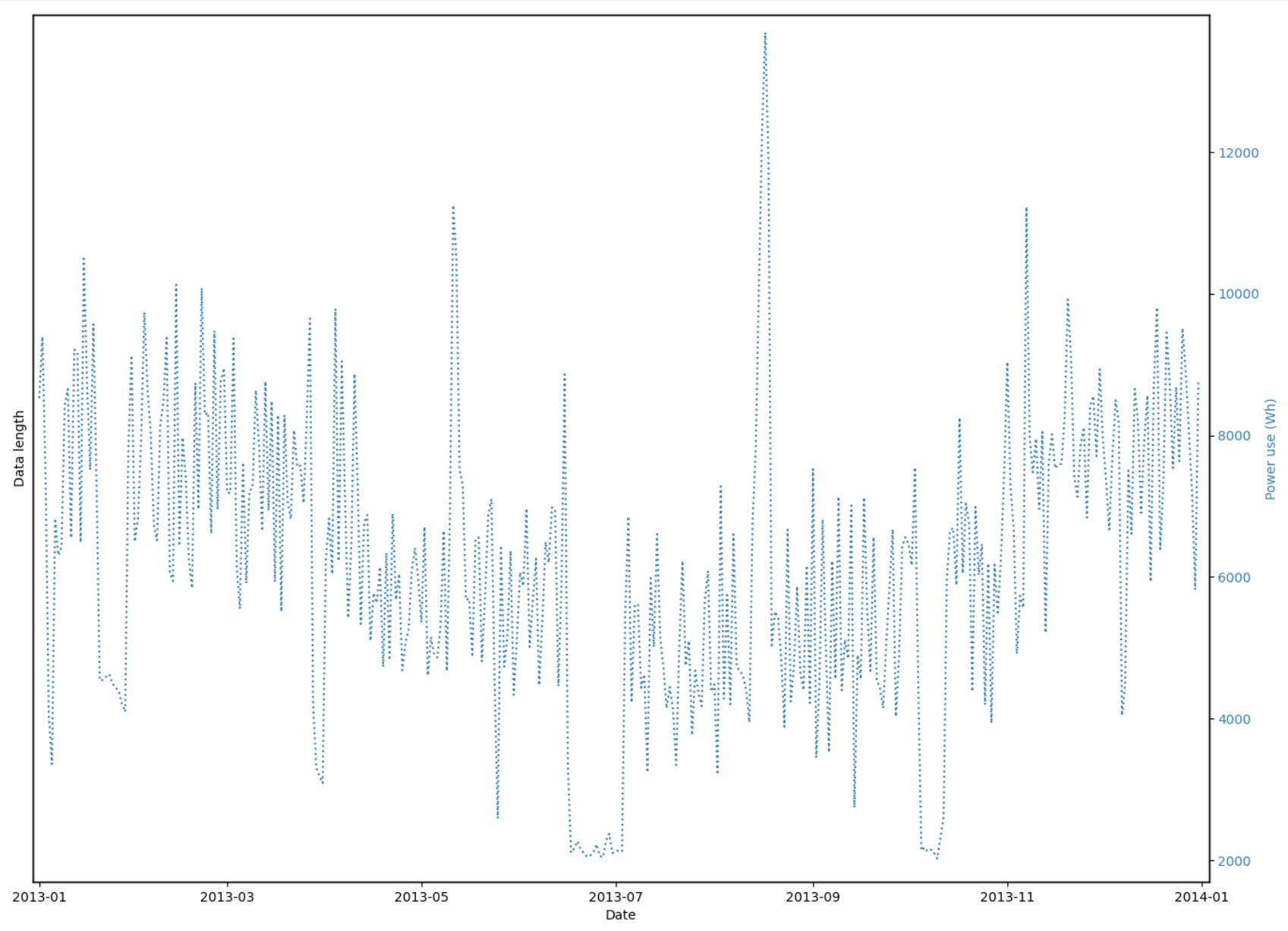
Compressed data of “46”



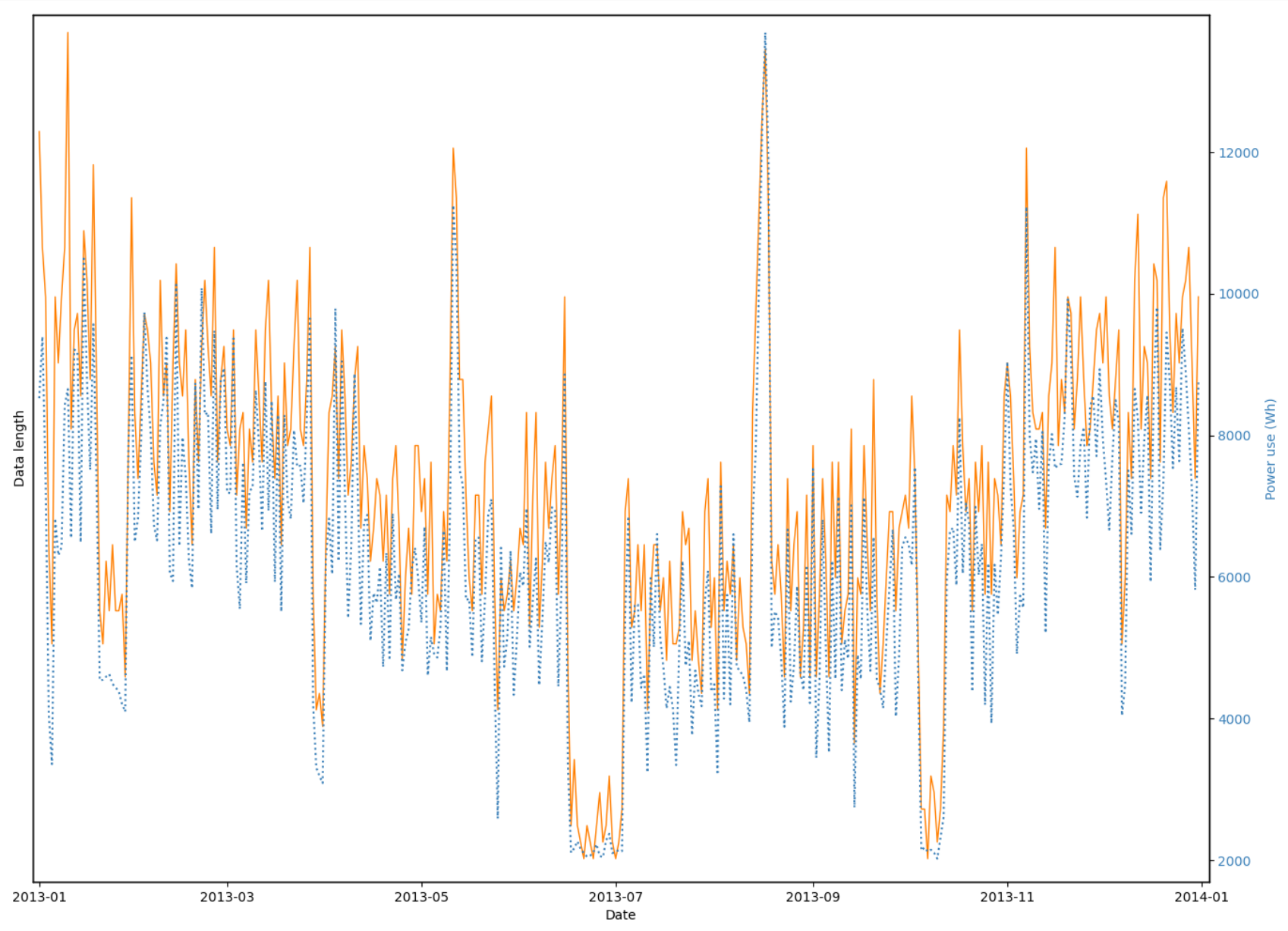
Compressed data of “46”



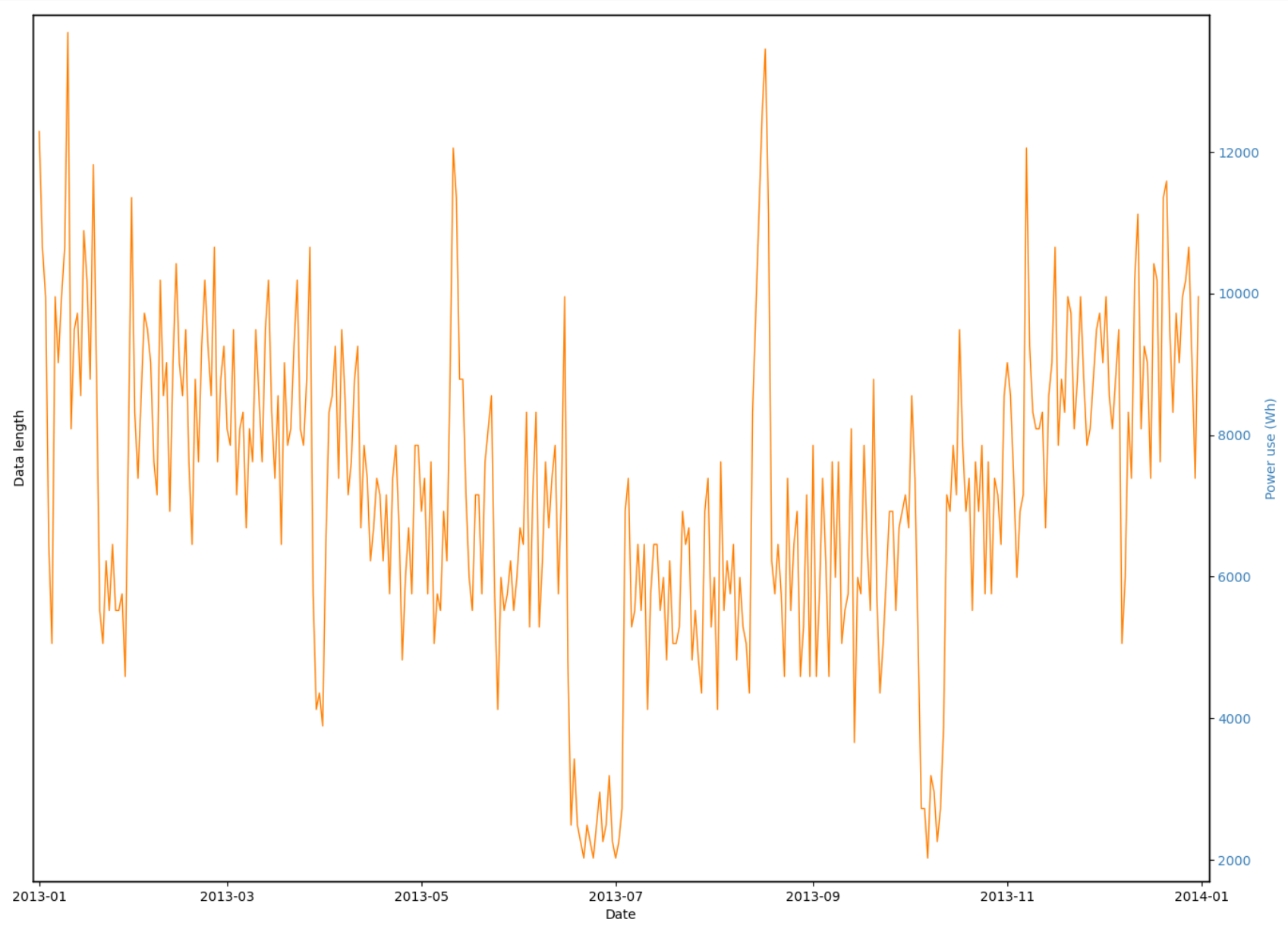
Compressed data of “46”



Compressed data of “46”



Compressed data of “46”

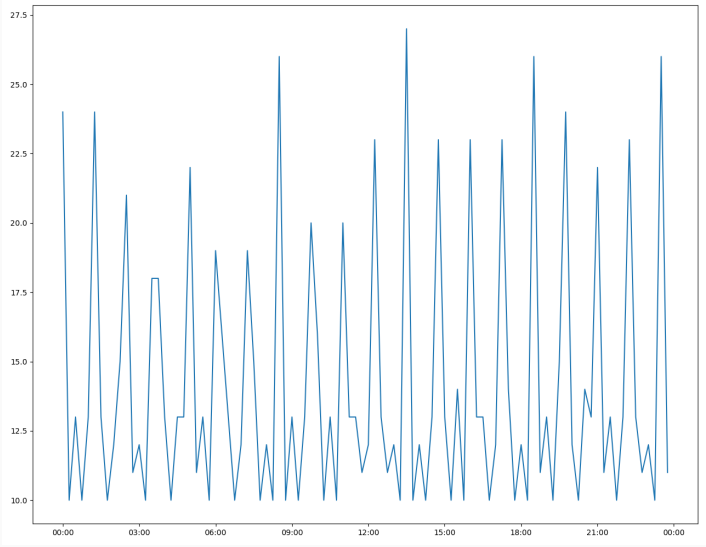


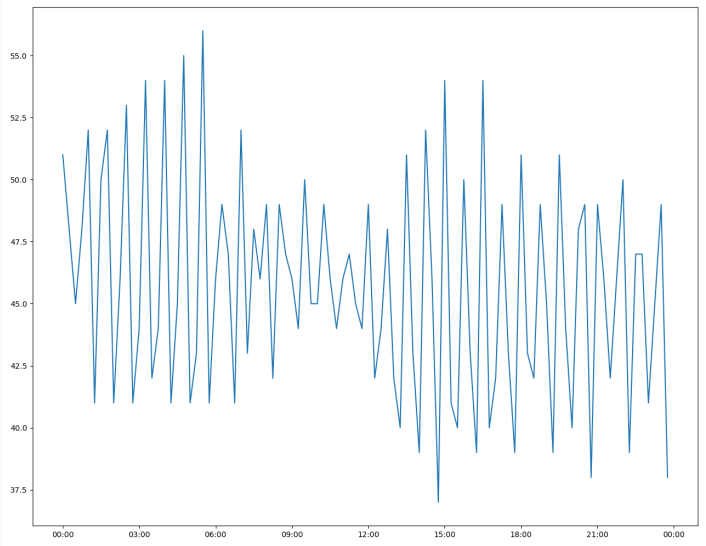
Compressed data of “46”

Compressed data of “46”

Compressibility of days

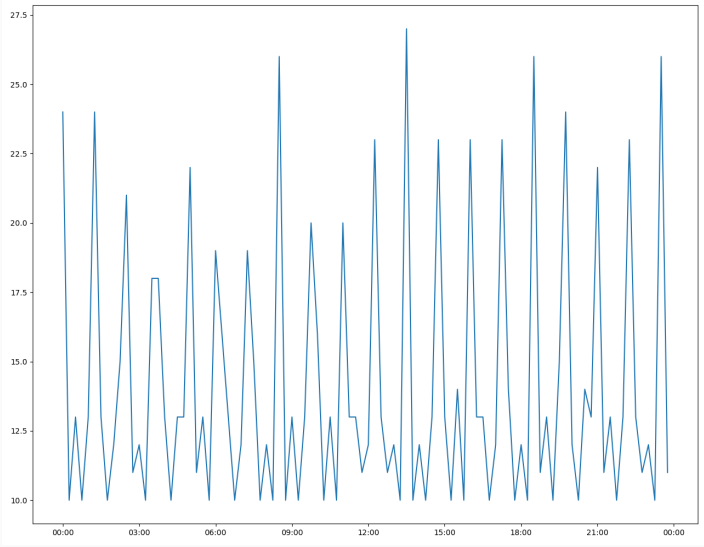
Most compressible:

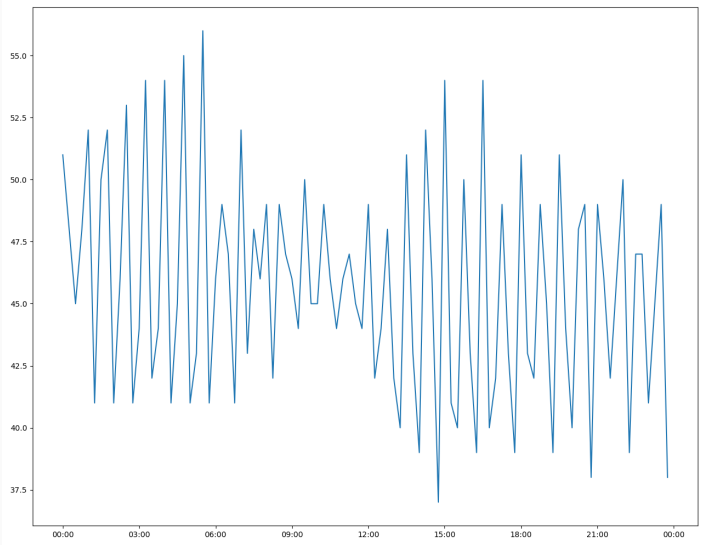




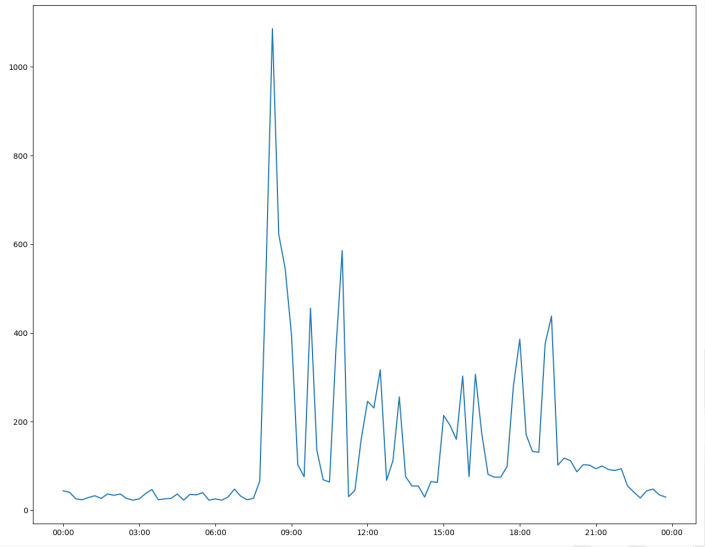
Compressibility of days

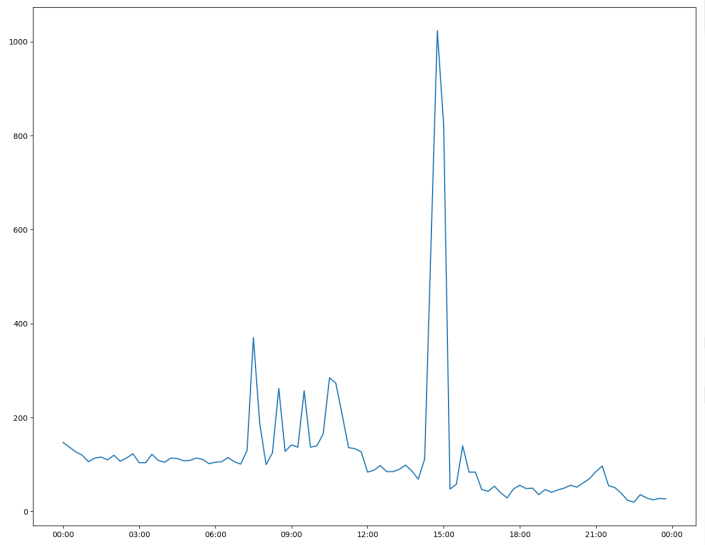
Most compressible:





Least compressible:





Conclusions

- **Compression and Minimum-length Delta Coding enable traffic analysis:**
  - absence of a day is observable
  - more frequent transmission will enable more detailed analysis

Conclusions

- **Compression and Minimum-length Delta Coding enable traffic analysis:**
  - absence of a day is observable
  - more frequent transmission will enable more detailed analysis
- **Our fixed-length 16-bit Compact Delta Coding makes traffic analysis impossible**
  - size already gets pretty close to compression
  - additional gains may be possible

Conclusions

- **Compression and Minimum-length Delta Coding enable traffic analysis:**
  - absence of a day is observable
  - more frequent transmission will enable more detailed analysis
- **Our fixed-length 16-bit Compact Delta Coding makes traffic analysis impossible**
  - size already gets pretty close to compression
  - additional gains may be possible
- **This is just a small issue in a larger ecosystem where privacy is a concern everywhere**
  - good to see that privacy-awareness is becoming mainstream

Recommendations

- **Ensure that message length is not influenced by private information**
  - E.g. when privacy-sensitive data is transmitted, do not use Minimum-length Delta encoding.

Recommendations

- **Ensure that message length is not influenced by private information**
  - E.g. when privacy-sensitive data is transmitted, do not use Minimum-length Delta encoding.
  - When privacy-sensitive data is transmitted, prohibit compression.

Recommendations

- **Ensure that message length is not influenced by private information**
  - E.g. when privacy-sensitive data is transmitted, do not use Minimum-length Delta encoding.
  - When privacy-sensitive data is transmitted, prohibit compression.
  - Consider our proposed 16-bit Compact Delta encoding: Good savings, but constant length.

Recommendations

- **Ensure that message length is not influenced by private information**
  - E.g. when privacy-sensitive data is transmitted, do not use Minimum-length Delta encoding.
  - When privacy-sensitive data is transmitted, prohibit compression.
  - Consider our proposed 16-bit Compact Delta encoding: Good savings, but constant length.
- **Review the compression mechanism you use**
  - The mechanism chosen in DLMS/COSEM (ITU-T V.44) is not a public standard and does not have publicly available implementations, which hampers direct analysis.

Recommendations

- **Ensure that message length is not influenced by private information**
  - E.g. when privacy-sensitive data is transmitted, do not use Minimum-length Delta encoding.
  - When privacy-sensitive data is transmitted, prohibit compression.
  - Consider our proposed 16-bit Compact Delta encoding: Good savings, but constant length.
- **Review the compression mechanism you use**
  - The mechanism chosen in DLMS/COSEM (ITU-T V.44) is not a public standard and does not have publicly available implementations, which hampers direct analysis.
  - We had to make assumptions on encoding options and what would actually be used in practice. Our paper makes some of these considerations concrete.

Thank you for your attention

Pol Van Aubel

pol.vanaubel@cs.ru.nl

https://www.polvanaubel.com/

iCIS—Digital Security

Radboud University

Recommendations

- **Specify the ways Delta Types can be used extensively**
  - When privacy-sensitive data is transmitted, prohibit Minimum-length Delta encoding

Recommendations

- **Specify the ways Delta Types can be used extensively**
  - When privacy-sensitive data is transmitted, prohibit Minimum-length Delta encoding
  - Consider our proposed 16-bit Compact Delta encoding
  - We had to make assumptions on encoding options and what would actually be used in practice. Our paper makes some of these options concrete.

Recommendations

- **Specify the ways Delta Types can be used extensively**
  - When privacy-sensitive data is transmitted, prohibit Minimum-length Delta encoding
  - Consider our proposed 16-bit Compact Delta encoding
  - We had to make assumptions on encoding options and what would actually be used in practice. Our paper makes some of these options concrete.
- **Review the compression mechanism used in DLMS**
  - The mechanism chosen in DLMS/COSEM (ITU-T V.44) is not a public standard and does not have publicly available implementations
  - When privacy-sensitive data is transmitted, prohibit compression

Which encodings are safe?

- Anything compressed: (UNSAFE)

Which encodings are safe?

- Anything compressed: (UNSAFE)
- Uncompressed:

Normal: No smart encoding at all (SAFE)

Which encodings are safe?

- Anything compressed: (UNSAFE)
- Uncompressed:

Normal: No smart encoding at all (SAFE)

NULL: Replace *candidates* with `NULL`

- All timestamps except the first (SAFE)
- All measurements that are 0 (PROBABLY UNSAFE)

Which encodings are safe?

- Anything compressed: (UNSAFE)
- Uncompressed:

Normal: No smart encoding at all (SAFE)

NULL: Replace *candidates* with `NULL`

- All timestamps except the first (SAFE)
- All measurements that are 0 (PROBABLY UNSAFE)

  Few measurements in our dataset are 0

  This encoding seems useless

Which encodings are safe?

- Anything compressed: (UNSAFE)
- Uncompressed:

Normal: No smart encoding at all (SAFE)

NULL: Replace *candidates* with `NULL`

- All timestamps except the first (SAFE)
- All measurements that are 0 (PROBABLY UNSAFE)

  Few measurements in our dataset are 0

  This encoding seems useless

Delta: Replace all but the first measurement with a Delta Type

Min-length: Smallest type possible each measurement (UNSAFE)

Which encodings are safe?

- Anything compressed: (UNSAFE)
- Uncompressed:

Normal: No smart encoding at all (SAFE)

NULL: Replace *candidates* with `NULL`

- All timestamps except the first (SAFE)
- All measurements that are 0 (PROBABLY UNSAFE)

  Few measurements in our dataset are 0

  This encoding seems useless

Delta: Replace all but the first measurement with a Delta Type

Min-length: Smallest type possible each measurement (UNSAFE)

32-bit: 32-bit Delta for each measurement (SAFE)

16-bit: 16-bit Delta for each measurement (SAFE)

16-bit Compact: Using a compacter array structure (SAFE)