Scalable Meter Data Management in Electric Power Grids

Vinod Namboodiri
Wichita State University
(vinod.namboodiri@wichita.edu)

PSERC Webinar
October 6, 2015
Acknowledgements

• Students: Babak Karimi, Vishnu Cherusola
• Collaborators: Murtuza Jadliwala, Visvakumar Aravinthan from Wichita State, and Anurag Srivastava from Washington State
• Industry advisors of PSERC project S-54.
• PSERC for support for S-54: Towards a Privacy-Aware Information-Sharing Framework for Advanced Metering Infrastructures
Presentation Outline

• Data volume as an issue

• Overview of solution approaches

• Data concentration for AMI

• Application-quality aware data aggregation for AMI

• Future work
Data and Smart Grids

• Collect, store and manage huge quantities of data -- from millions of smart meters
  • Advanced Metering Infrastructure (AMI)
  • A popular tool to modernize the electric grid (efficiency, renewable integration)
  • Facilitates two-way communication between smart meter and control center
  • Captures and transmits energy-use on a hourly or sub-hourly basis
  • As opposed to the current meters that provide daily energy usage and monthly bill
  • Expected to facilitate consumer participation in the Smart Grid
Data and Smart Grids

• Collect, store and manage huge quantities of data -- high-frequency samples from phasor measurement units (PMUs)

• Wide-area, Monitoring, Protection, and Control (WAMPAC)
• Synchrophasors or Phasor Measurement Units (PMUs) take 60 Hz snapshots of the state of the grid and send over the communications network
Data Stream – An Abstract View

Control

Data Processing/Storage → Communications Network → Data Generators
Why is Data Volume an Issue?

• Scaled up to many thousands of meters or synchrophasors sending periodic data, the volume of data can be a burden
  • Limited-bandwidth last-mile networks
  • Perhaps even finer granularity data collection in future

• Network capacity is a precious resource for electric utilities because they are
  • leasing networks from third-party providers or
  • building infrastructure themselves and leasing bandwidth out

• W. Luan, D. Sharp, and S. Lancashire, “Smart grid communication network capacity planning for power utilities”
• “Arcadian’s Smart Grid: Licensed Spectrum Network to Own or Rent,” Greentechgrid.
• M. Kennedy, “Leveraging investment in fiber optic communications,” IEEE Smart Grid.
Calls for a Data Volume Concentration Process

Electric Utility Control Center

Substation running distribution automation application

Substation running demand response application

Data Concentrator GW

quality = low

quality = medium

quality = high

1 Mbps → 1 Mbps → 3 Mbps
Data Stream – An Abstract View

Data Processing/Storage → Communications Network → Data Generators

Need In-Network Data Processing/Analytics
Some Solution Approaches

- **Message Concatenation**: group multiple messages together with a common protocol header as opposed to each message having its own header.
- **Aggregation**: apply aggregation functions as appropriate based on requirements to combine data from several measurement.
- **Compress**: exploit statistical redundancy to represent data more concisely without losing information.
Role of Data Concentrator Units

• Approaches listed have the following two features in common.
  • They must leverage the information needs of grid operators for minimizing information volume while meeting any QoS constraints at the same time.
  • They require mechanisms at intermediate relay points to execute algorithms to meet their information volume concentration objectives.
Power Distribution System
Position in Existing Literature

• Data aggregation in the field of WSNs
  • Access to electric power at all times with backup batteries
  • Shifts the focus of the problem from battery life of nodes to reduction of network capacity utilization

• Data concentration for WAMPAC
  • Similar work missing related to power distribution system and metering
Smart Metering Message Concatenation (SMMC)

Useful data (blue)
Packet header (orange)

Packet header overhead can be reduced significantly!

Meter Data Concentrator/Aggregator

SMMC Problem and Desired Result
## Smart Meter Data Messages Types

<table>
<thead>
<tr>
<th>Message/ Traffic Description</th>
<th>Size (Bytes)</th>
<th>Delay Obj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meter clock synchronization</td>
<td>64</td>
<td>2 secs</td>
</tr>
<tr>
<td>Interval data read</td>
<td>480</td>
<td>Best effort</td>
</tr>
<tr>
<td>Firmware patch/upgrade confirmation/ acknowledge</td>
<td>20</td>
<td>Best effort</td>
</tr>
<tr>
<td>Meter ping (on demand read)</td>
<td>64</td>
<td>2 secs</td>
</tr>
<tr>
<td>Meter remote diagnostic</td>
<td>500</td>
<td>2 secs</td>
</tr>
<tr>
<td>Tamper notification</td>
<td>64</td>
<td>5 secs</td>
</tr>
<tr>
<td>Meter remote disconnect/ reconnect response</td>
<td>500</td>
<td>2 secs</td>
</tr>
</tbody>
</table>

Smart Meter Data Message

- Meter id, Equipment status, Type of message, metered data (ANSI C12 standard)
- The size of the packet header depends on the communication protocol used.
  - It can add from 10 to 50 bytes per packet of data sent.

### Transport Protocols

<table>
<thead>
<tr>
<th>Protocols:</th>
<th>Estimated Protocol frame size (bytes)</th>
<th>Used For</th>
<th>Physical Carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Remote AMR</td>
<td>AMI</td>
</tr>
<tr>
<td>IEC 61334 (S-FSK, FSK, OFDM)</td>
<td>45 bytes(^7)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>IEC 62056-31 Euridis(^9)</td>
<td>45 bytes(^10)</td>
<td>Y</td>
<td>-</td>
</tr>
<tr>
<td>EN 13757 M-Bus(^11)</td>
<td>27 bytes</td>
<td>-</td>
<td>Y(^12)</td>
</tr>
<tr>
<td>TCP/IP protocol(^13)</td>
<td>50 bytes</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>SITRED(^14)</td>
<td>45 bytes(^15)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PRIME</td>
<td>8 bytes(^16)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>EverBlu</td>
<td>unknown(^17)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>OPERA/UPA(^18)</td>
<td>24 bytes(^19)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ZigBee(^20)</td>
<td>15 bytes</td>
<td>Y</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^7\) 45 bytes is the standard size for IEC 61334.
\(^9\) IEC 62056-31 standard.
\(^10\) 45 bytes is the standard size for IEC 62056-31.
\(^11\) EN 13757 M-Bus standard.
\(^12\) 27 bytes is the standard size for EN 13757 M-Bus.
\(^13\) TCP/IP protocol standard.
\(^14\) SITRED standard.
\(^15\) 45 bytes is the standard size for SITRED.
\(^16\) 8 bytes is the standard size for PRIME.
\(^17\) Unknown size for EverBlu.
\(^18\) OPERA/UPA standard.
\(^19\) 24 bytes is the standard size for OPERA/UPA.
\(^20\) ZigBee standard.
Heuristic Approach

• Earliest Deadline First (EDF)

• DCU queues all arriving packets
• When time = time of deadline for message with earliest deadline
  • Insert message into newly created packet
  • Fill it with other messages to maximize packet size using one of the schemes presented next

• Deadline at DCU can be created by subtracting off some estimate of network and processing latencies from end-point deadline.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF-DKB</td>
<td>Inserts deadline messages as much as possible inside the packet and the remaining space will be filled through knapsack selection over best-effort messages that have been queued.</td>
</tr>
<tr>
<td>EDF-SDKB</td>
<td>Only a single deadline message sits inside the packet with any available space filled with non-deadline messages in the non-deadline queue through knapsack selection.</td>
</tr>
<tr>
<td>EDF-FCFS</td>
<td>Messages will be placed in the packet according to their arrival sequence from a common queue of deadline and non-deadline messages on a first-come first-served basis.</td>
</tr>
<tr>
<td>EDF-KN</td>
<td>Messages are chosen from a common pool of deadline and best-effort messages selected through the knapsack algorithm.</td>
</tr>
<tr>
<td>EDF-KDKB</td>
<td>A sequence of knapsack selections first on all queued deadline messages and then over the queued best-effort messages if needed to fill the packet.</td>
</tr>
<tr>
<td>EDF-KBKD</td>
<td>Reverse order of knapsack process in EDF-KDKB working first on the queued best-effort messages and then on the deadline messages if needed.</td>
</tr>
</tbody>
</table>
Evaluation Methodology

• Reference Algorithms
  • EDF-based MILP
  • Theoretical

• Discrete Event Simulations in MATLAB
  • 95% confidence intervals shown

• Message generation at meters assumed to follow a Poisson process with an aggregate arrival rate of \( \lambda \) messages/sec at the DCU
  • \( \lambda \) of 0.1, 0.5 and 1 studied

• Mix of message types arriving at DCU
  • Assume a Beta distribution with shape parameters \( \alpha > 0 \), and \( \beta > 0 \) that can be varied to vary the mix

• For simplicity we have assumed there is no Network and Processing delay in these evaluations.
# Message Type Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform ((\alpha = 1, \beta = 1))</td>
<td>The traffic would have almost equal percentage of all message types.</td>
</tr>
<tr>
<td>More smaller ((\alpha = 2.8, \beta = 1.9))</td>
<td>Most of the arrived messages are of the smaller size of message types.</td>
</tr>
<tr>
<td>More larger ((\alpha = 0.18, \beta = 0.25))</td>
<td>There is higher percentage of large message size and very few numbers of small size messages.</td>
</tr>
<tr>
<td>More deadline ((\alpha = 1, \beta = 1.8))</td>
<td>Most of the times there are incoming messages with deadline restriction.</td>
</tr>
<tr>
<td>More best-effort ((\alpha = 2.5, \beta = 0.5))</td>
<td>There are very few numbers of messages with a deadline and so many best-effort messages.</td>
</tr>
</tbody>
</table>
Data reduction ratio vs Different message arrival rates ($\lambda = 0.1, 0.5, \text{ and } 1$)
Application-Aware Data Aggregation

- Data from multiple samples could be represented by one sample

- Aggregation function examples \( \{2, 5, 8, 5\} \)
  - Minimum: 2
  - Maximum: 8
  - Average: 4
  - Most Repeated: 5
  - Median: 5

Data Granularity

• A measure of how well the original data samples are represented after aggregation

Assume that each smart meter $s_i$ generates $m$ data samples each of size $g_{ij}$ bytes each over a period of time $T$ and sends it to the DCU. Based on network congestion, assume that the data concentrator aggregates multiples of incoming messages from all smart meter into a total of $k$ messages $a_{i1} \cdots a_{ik}$, $1 \leq k \leq n \cdot m$ before sending them out over the backhaul network. We can then define data granularity $\gamma$ over period $T$ as

$$\gamma = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij} - \sum_{i=1}^{k} a_i}{\sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}}$$

(1)
Application: State Estimation

Average Residual $N_r$ vs Total number of Measurements

- Substation
- Switch
- Secondary transformer
- Smart meter

Total Number of Measurements

$N_r$ (Average Residual)
Application: Transformer Life Estimation
Algorithm 1 Aggregation Algorithm at DCU

1: Take algorithm execution period $T$, initial estimate of backhaul network capacity $C$, application data granularity tolerance level $\gamma_{lb}$ as inputs
2: Repeat every $T$ time units
3: Compute $\tau$ using knowledge of all samples that arrived over the last period and a current estimate of $C$
4: if $\tau > 1$ then
5: Collect a window of $\tau$ samples, apply aggregation function, then send out
6: Compute data granularity $\gamma$ using knowledge of all incoming samples and outgoing aggregate messages
7: if $\gamma < \gamma_{lb}$ then
8: Notify control center of compromise in data quality below specified threshold
9: end if
10: end if
Preliminary Results

![Graph showing Delay Comparison between scenarios: Without aggregation and With aggregation. The graph illustrates the delay in milliseconds for each packet number between 1 and 16. The blue line indicates the delay without aggregation, while the red line shows the delay with aggregation. The graph shows variations in delay across different packet numbers.]
Recap and Conclusions

- Efforts to manage data volume needed in conjunction with design of communications infrastructures and new applications
- Just message concatenation can reduce data volume of metering applications by 10-25%
- Data aggregation is another effective mechanism if the application-quality tradeoff can be managed well
Future Work

• Distributed data management and analytics, beginning with:
  • What data needs to be collected
  • What granularity data is needed where
• Data concentration and aggregation of PMU data
• Employ compression techniques in addition to aggregation and concatenation
Thank you!