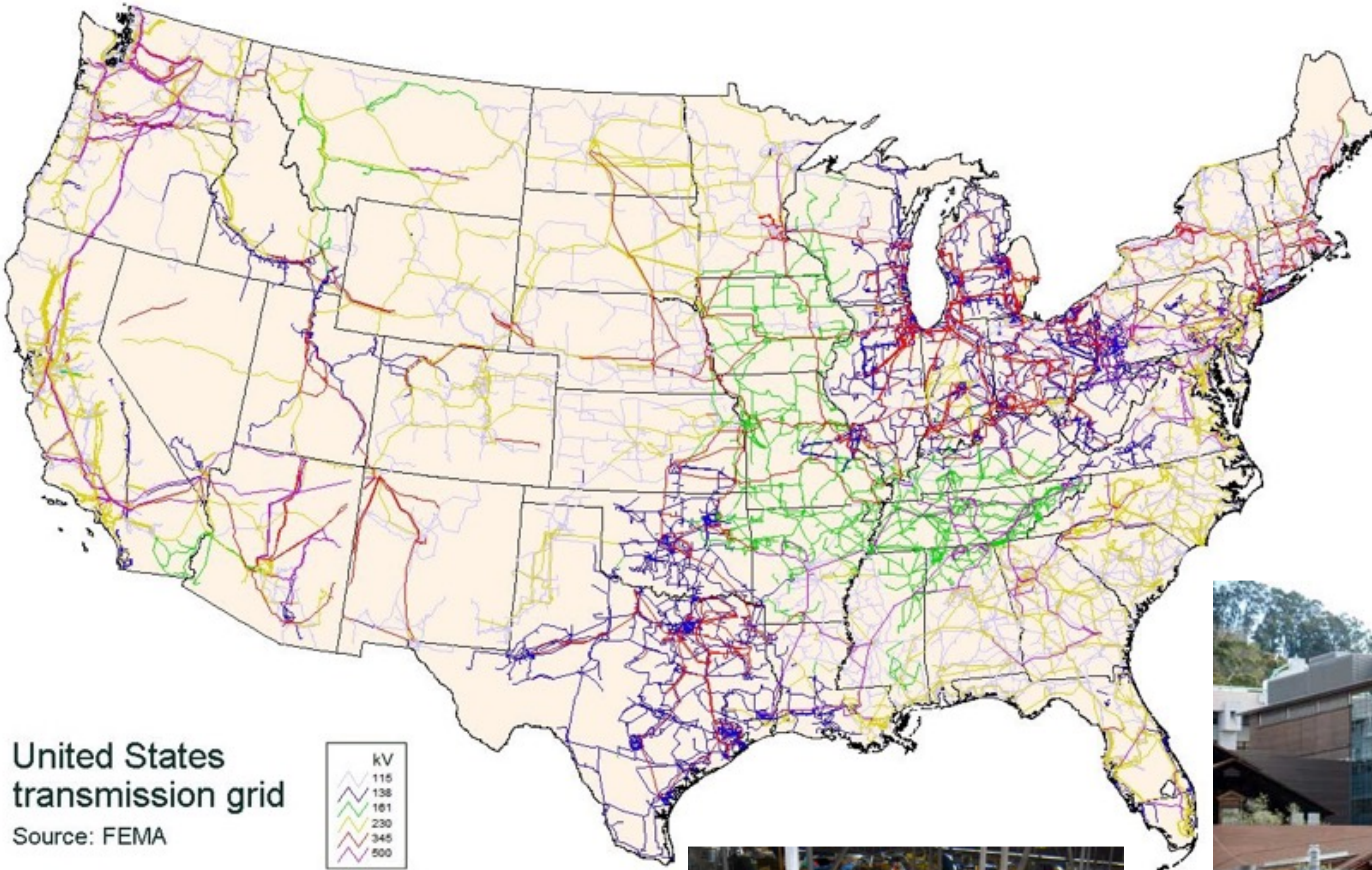
Efficient Analytics with BTrDB:

From Grids to Buildings

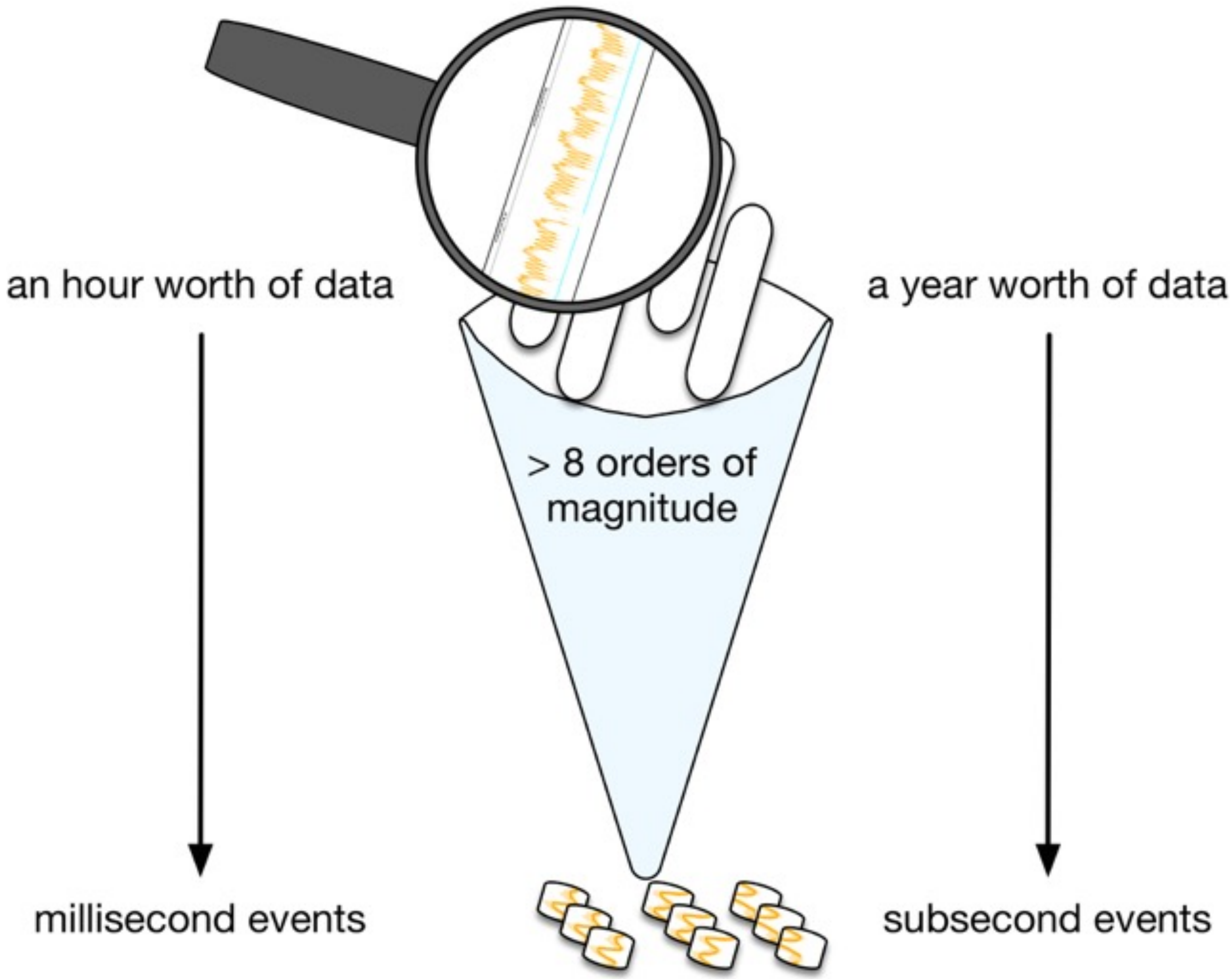
Omid Ardakanian
ardakanian@berkeley.edu

SDB Winter Retreat


United States
transmission grid
Source: FEMA
kV
115
138
161
230
345
500


209
rapidKL


305



an hour worth of data
a year worth of data
> 8 orders of magnitude
millisecond events
subsecond events

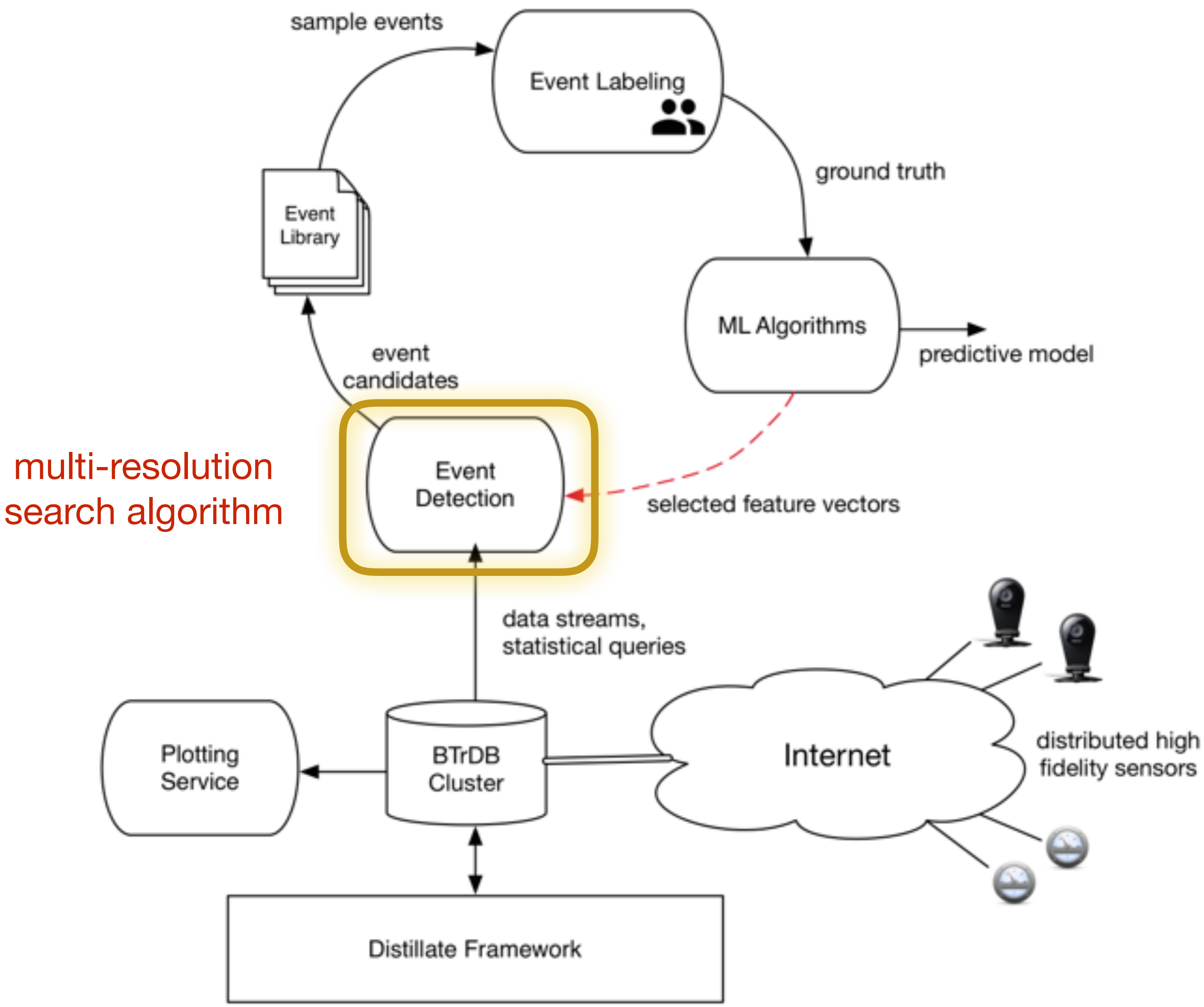
Introduction

- High-precision high-sample-rate data from distributed high fidelity high bandwidth sensors
  - ✳ timescale and size
- Finding anomalies in these systems is the holy grail
  - ✳ failing to identify and react to critical events in a timely manner may cost millions of dollars
- Energy data analytics (both real-time and historical) is critical yet computationally expensive
  - ✳ the ability to detect, analyze, and control with a limited time budget

Objectives

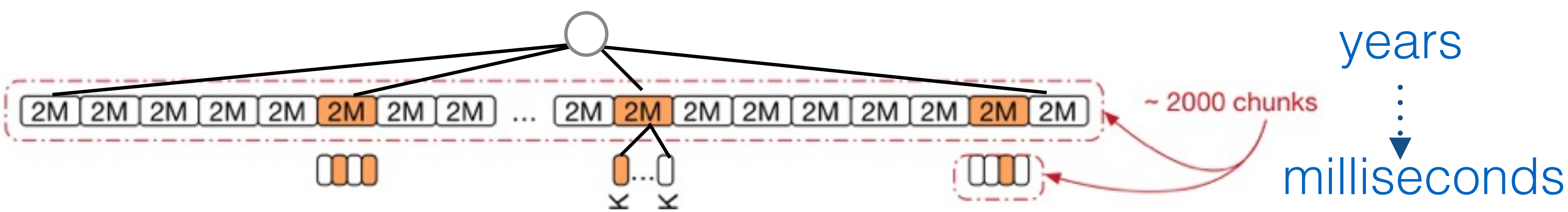
- Detect: identify rare events
    - ✳ using an efficient search algorithm that is logarithmic in the size of the data set and linear in the number of events that are found
- Analyze: run compute intensive tasks on smaller chunks of data
- Control: take corrective/preventive actions (in real-time applications)

System Architecture



Statistical Summaries

- statistical summaries (max, min, average, and count) are stored at different temporal resolutions



Example Query

Find 5-second intervals that contain at least one value greater than a threshold

Example Query

Find 5-second intervals that contain a value greater than a threshold

- Query **max** at the given temporal resolution
- Dive down if $max_{resolution}$ > threshold
- Repeat for the next temporal resolution until the desired resolution is reached

Multi-Resolution Search

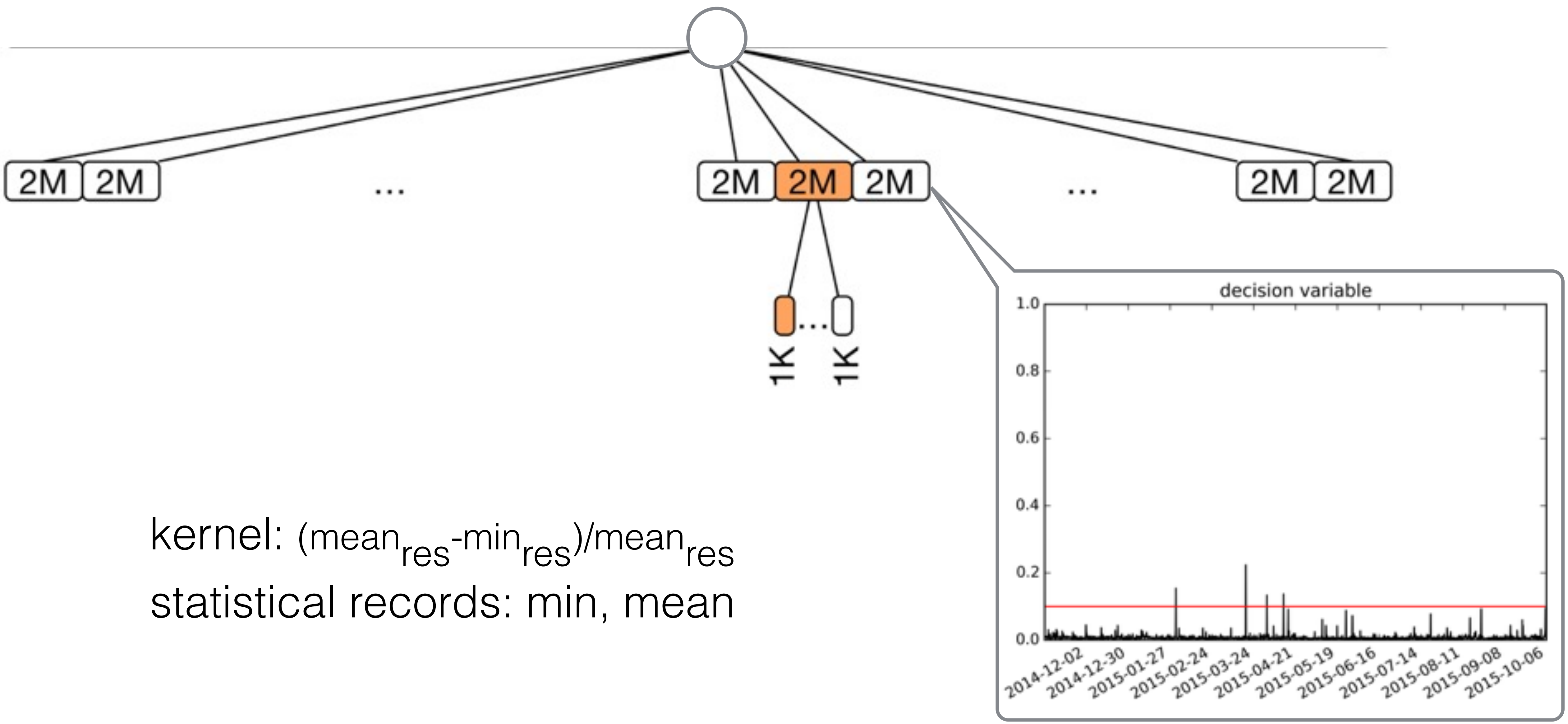
- Query statistical summaries of data at a given temporal resolution
- Compare a function of these statistical summaries against a threshold
- Dive down if the condition is satisfied
- Query raw data when the desired resolution is reached and run your algorithm on a small chunk of data

Grid Energy Analytics

Critical Events

- voltage sags
  - voltage magnitude (raw stream)
- tap changing events
  - angle difference (distillate)
- reverse flows
  - real power or displacement power factor (distillate)
- switching events
- …

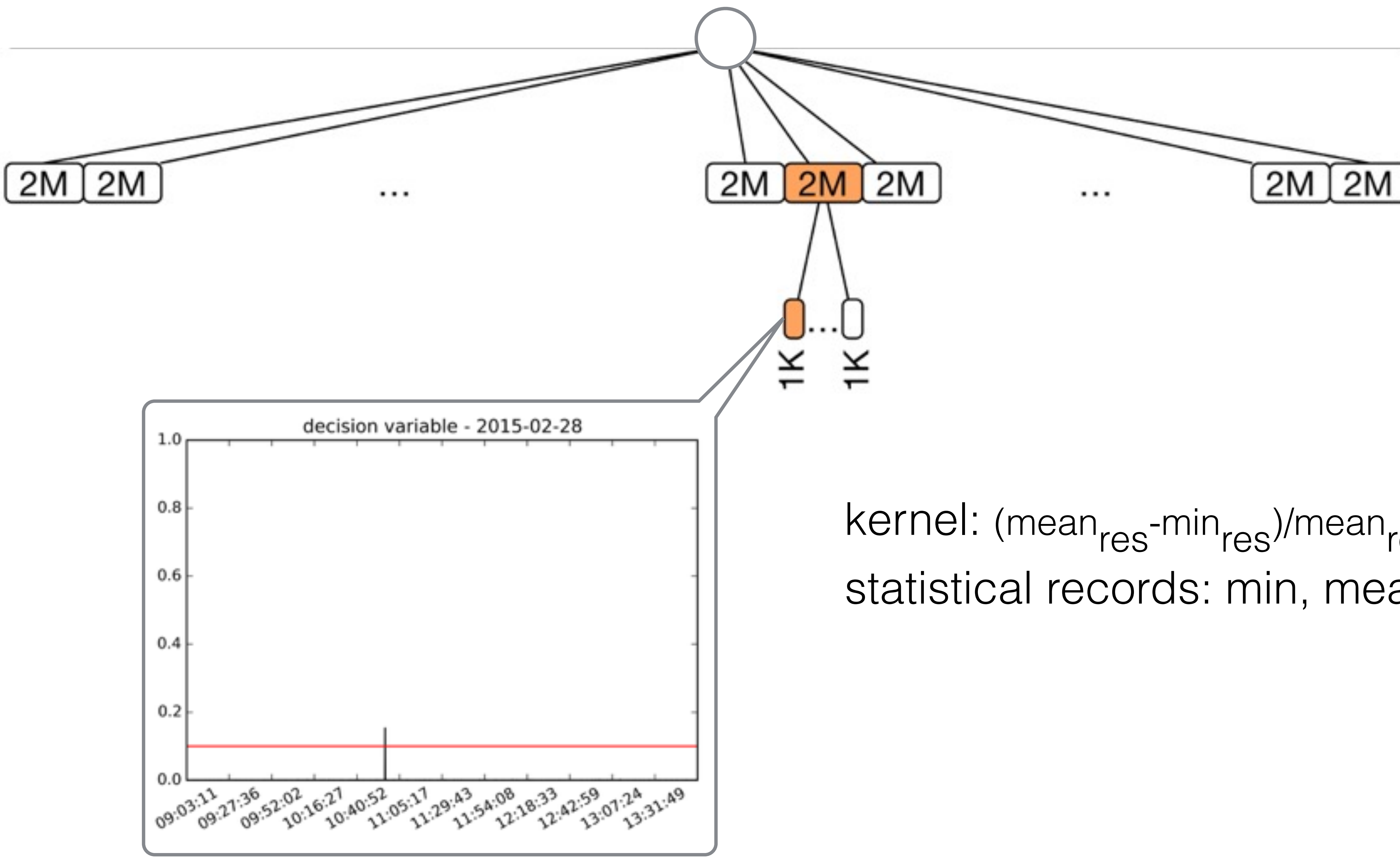
Case Study: Voltage Sag Detection



kernel: $(mean_{res} - min_{res}) / mean_{res}$

statistical records: min, mean

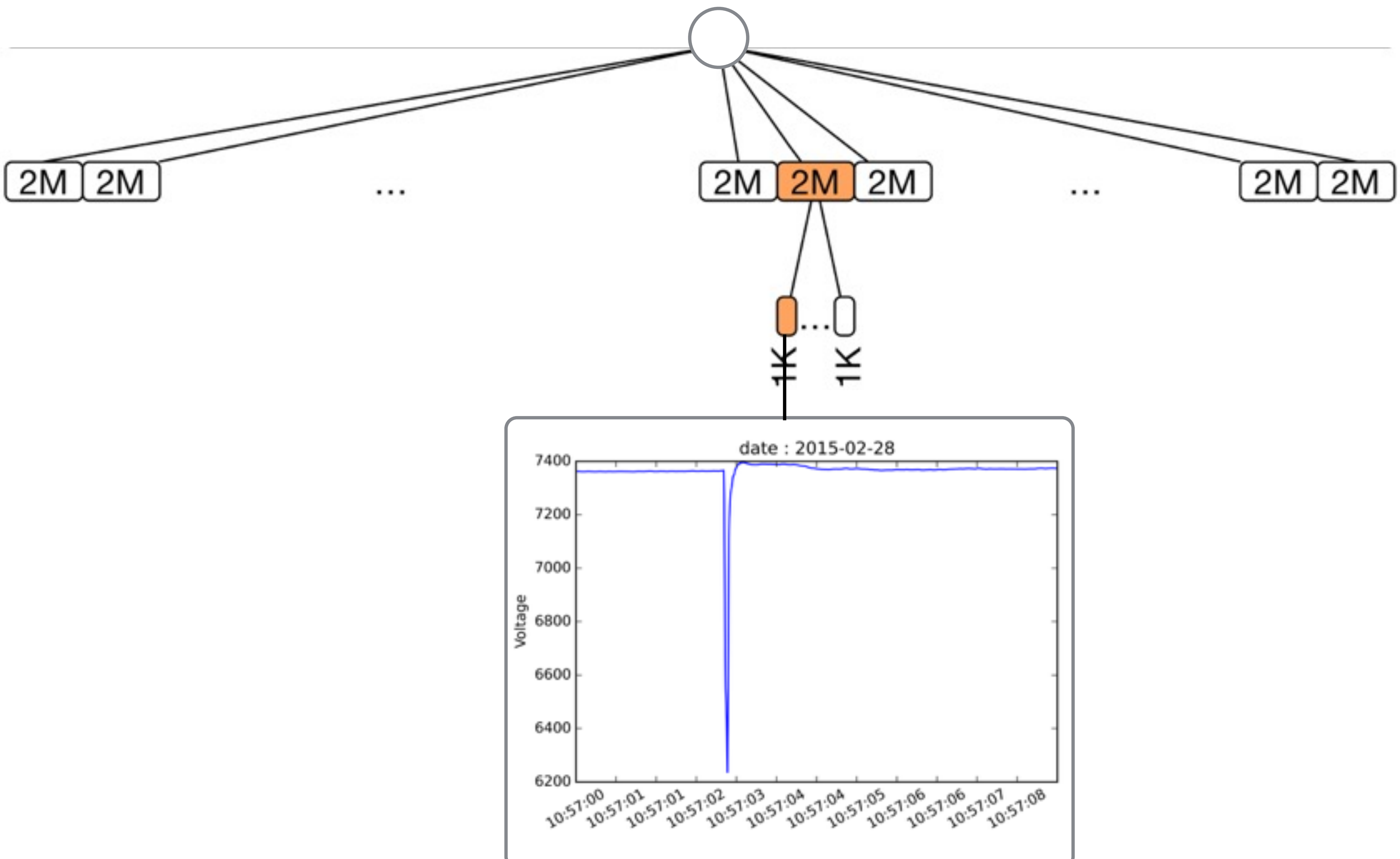
Dive Down



kernel: $(mean_{res}-min_{res})/mean_{res}$

statistical records: min, mean

Query Raw Data



Evaluation

Latency of multi-resolution search
~ 3B
~ 2M
~ 1K
11 voltage sags
Chunk size (no. points)
Latency (ms)
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
0
50
100
150
200

Example Result

	no. events (0.05)	runtime (ms)	no. events (0.1)	runtime (ms)	no. events (0.15)	runtime (ms)	no. events (0.2)	runtime (ms)	days
/clean/GP_BUS1/L1MAG	9	431.77	4	237.13	0	76.78	0	88.41	135
/clean/GP_BUS1/L2MAG	10	394.39	4	226.85	1	115.30	0	70.55	135
/clean/GP_BUS1/L3MAG	5	309.07	2	163.25	1	118.95	0	77.08	135
/clean/switch_a6/L1MAG	14	666.59	6	273.01	3	194.95	1	132.75	330
/clean/switch_a6/L2MAG	21	947.24	11	523.78	4	235.44	3	190.83	330
/clean/switch_a6/L3MAG	11	608.94	4	318.44	2	213.57	0	90.06	330
/clean/RPU/CE_CERT_Bld_1200/L1MAG	8	312.53	2	68.41	1	64.93	1	66.55	86
/clean/RPU/CE_CERT_Bld_1200/L2MAG	12	379.19	4	163.71	3	119.51	2	112.95	86
/clean/RPU/CE_CERT_Bld_1200/L3MAG	12	627.72	4	228.18	2	111.41	2	133.00	86

10% drop

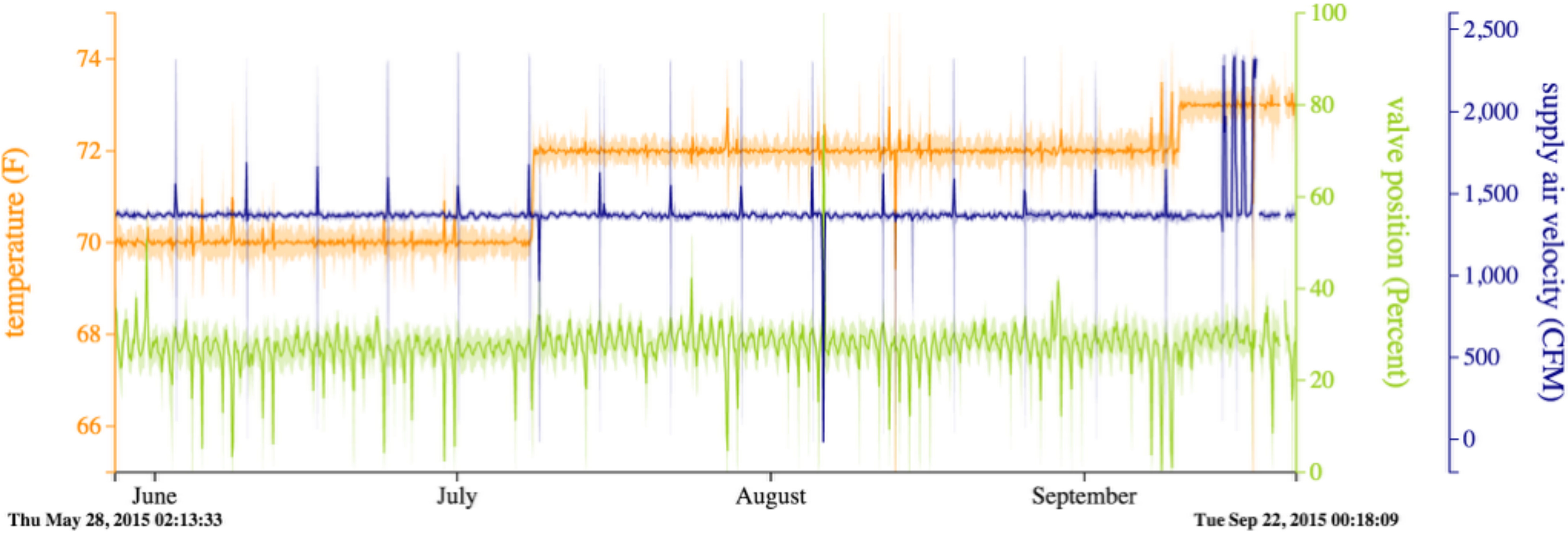
logarithmic in the size of the data set and linear in the number of events that are found

Building Energy Analytics

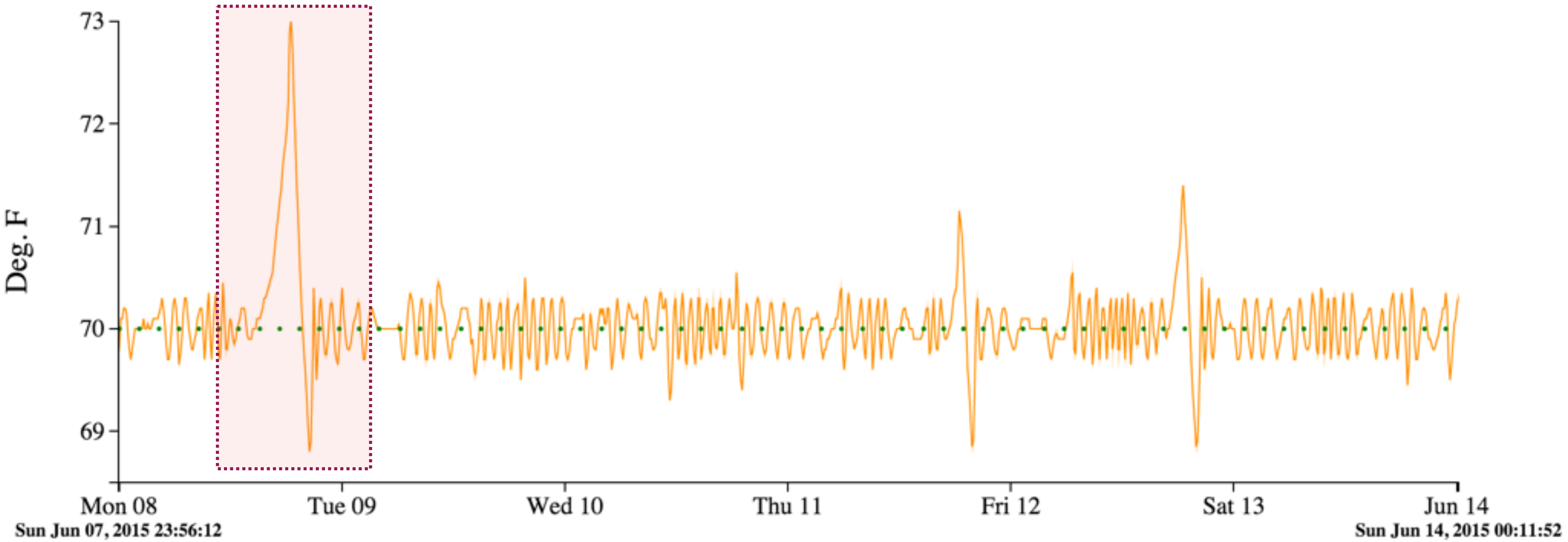
Identified Events

- HVAC anomalies
- Water leakage events
- Correlated use of large reactive loads

Anomalies in Buildings

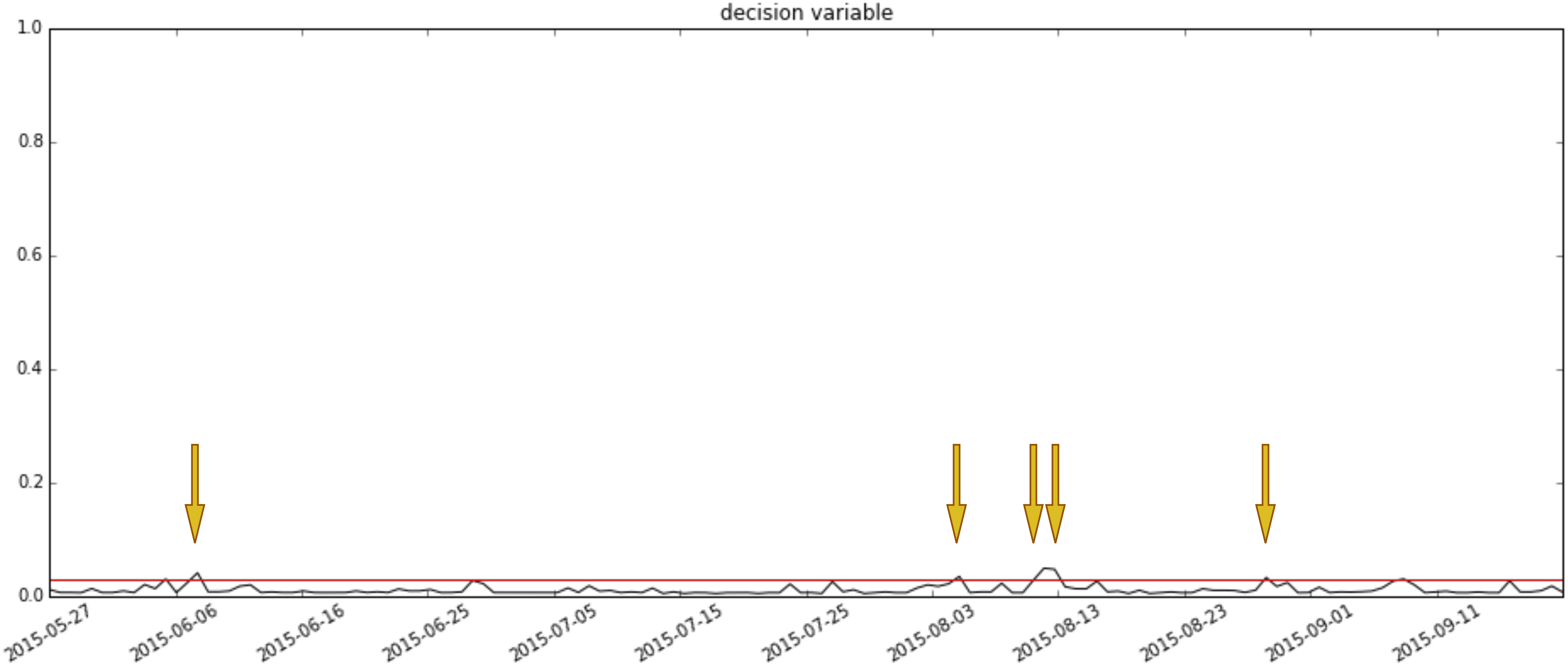


Anomalies in Buildings



find intervals in which zone temperature increases by more than X%

Case Study: Temperature Rise Detector

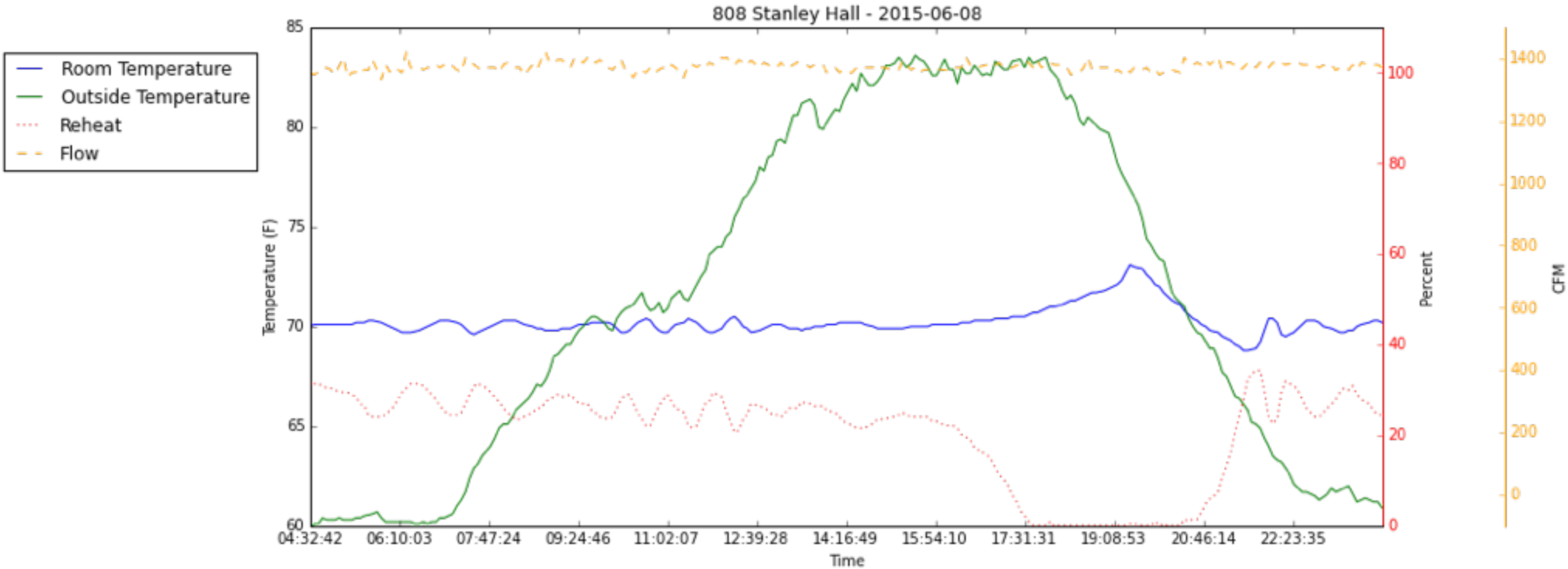


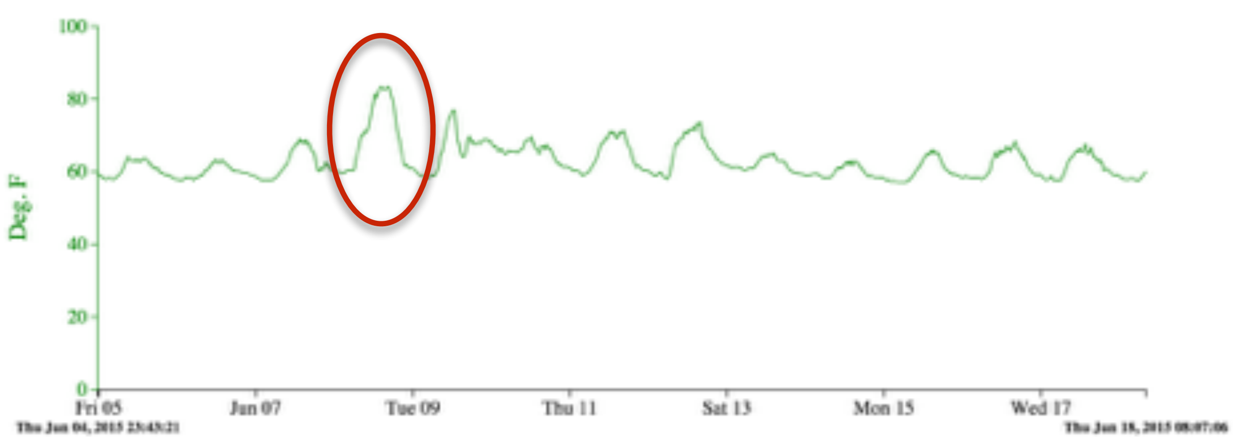
find intervals in which zone temperature increases by more than 3%

kernel: $(max_{res} - mean_{res}) / mean_{res}$

statistical records: max, mean

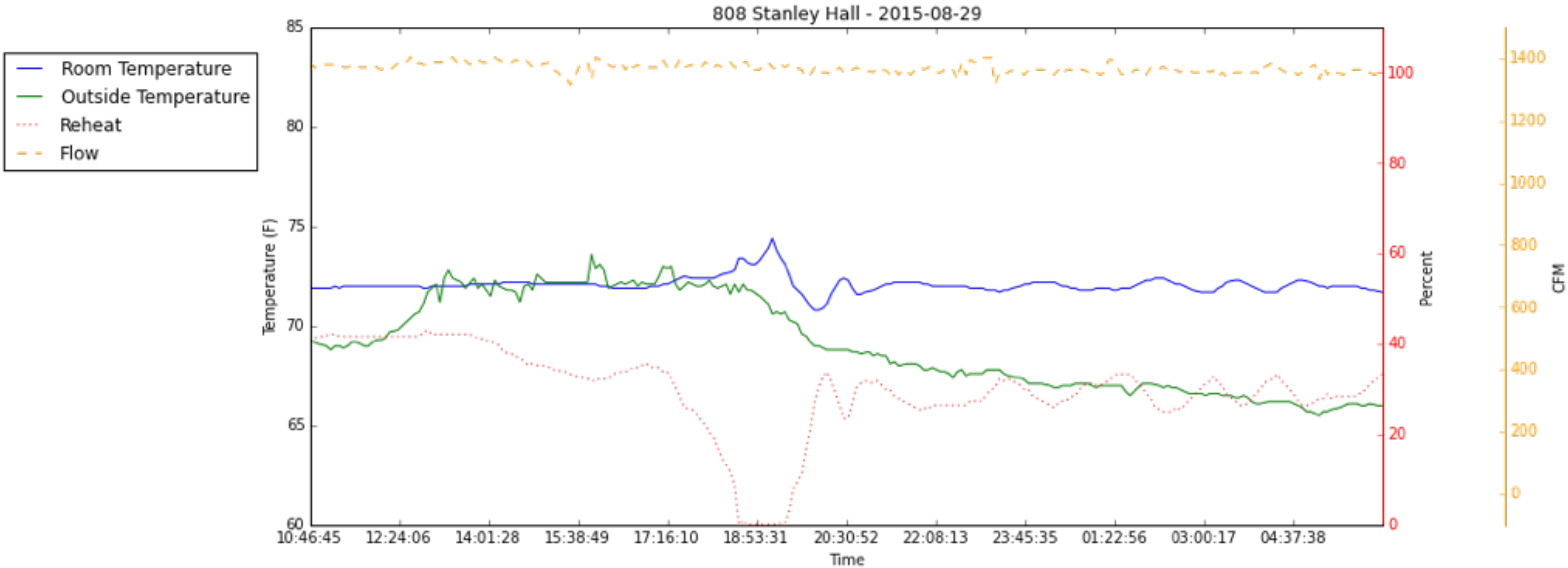
Case Study: Temperature Rise Detector





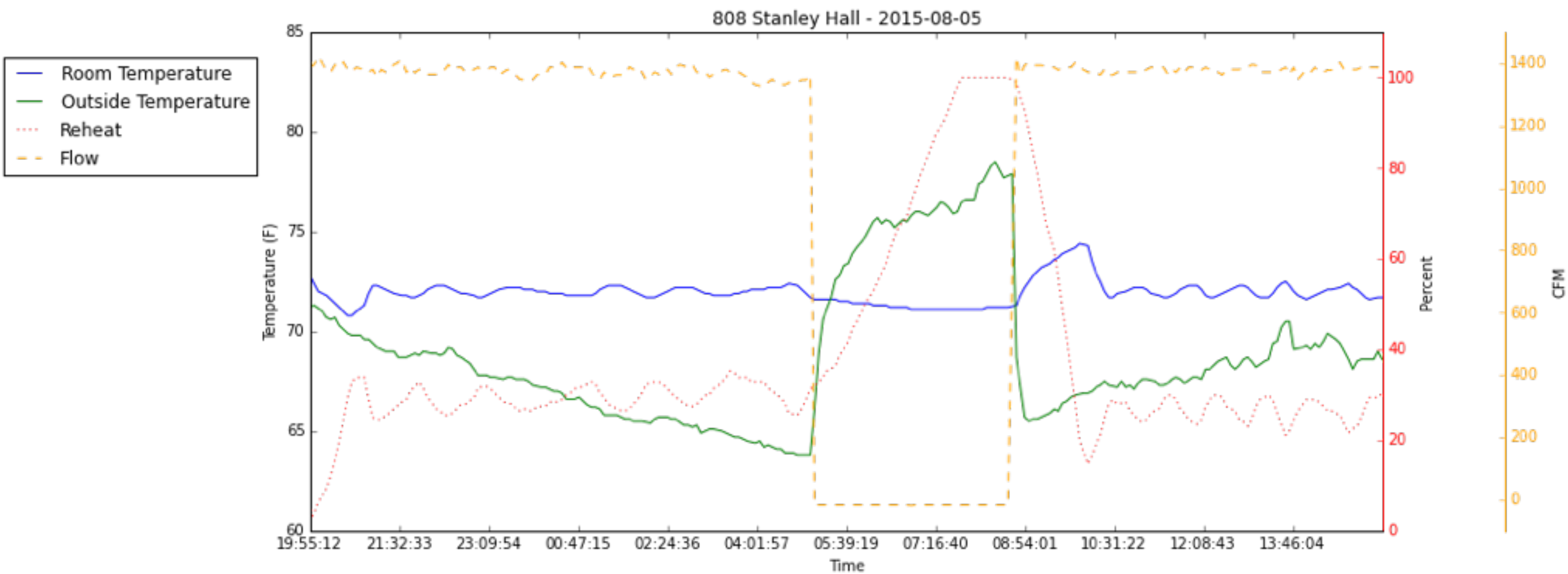
heat transfer?

Case Study: Temperature Rise Detector

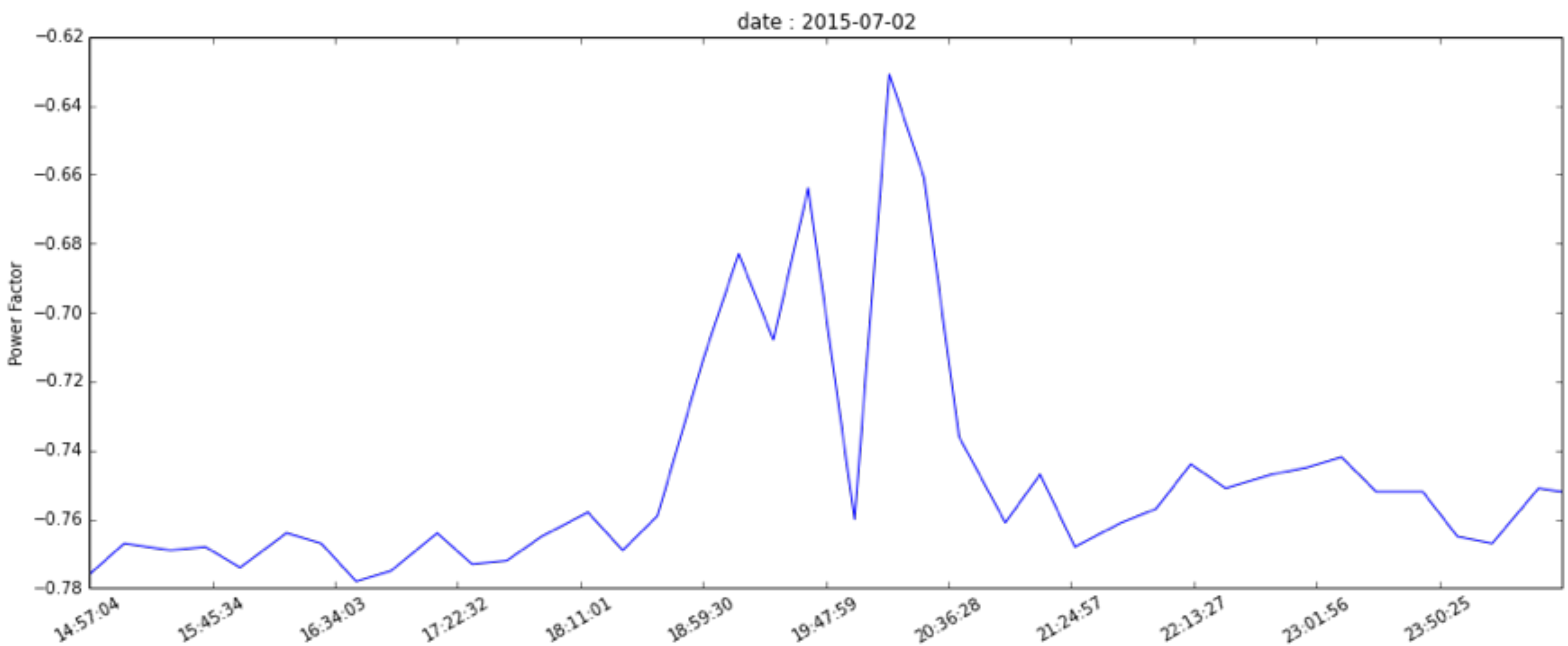


heat load due to occupancy, equipment load, direct sunlight?

Case Study: Temperature Rise Detector

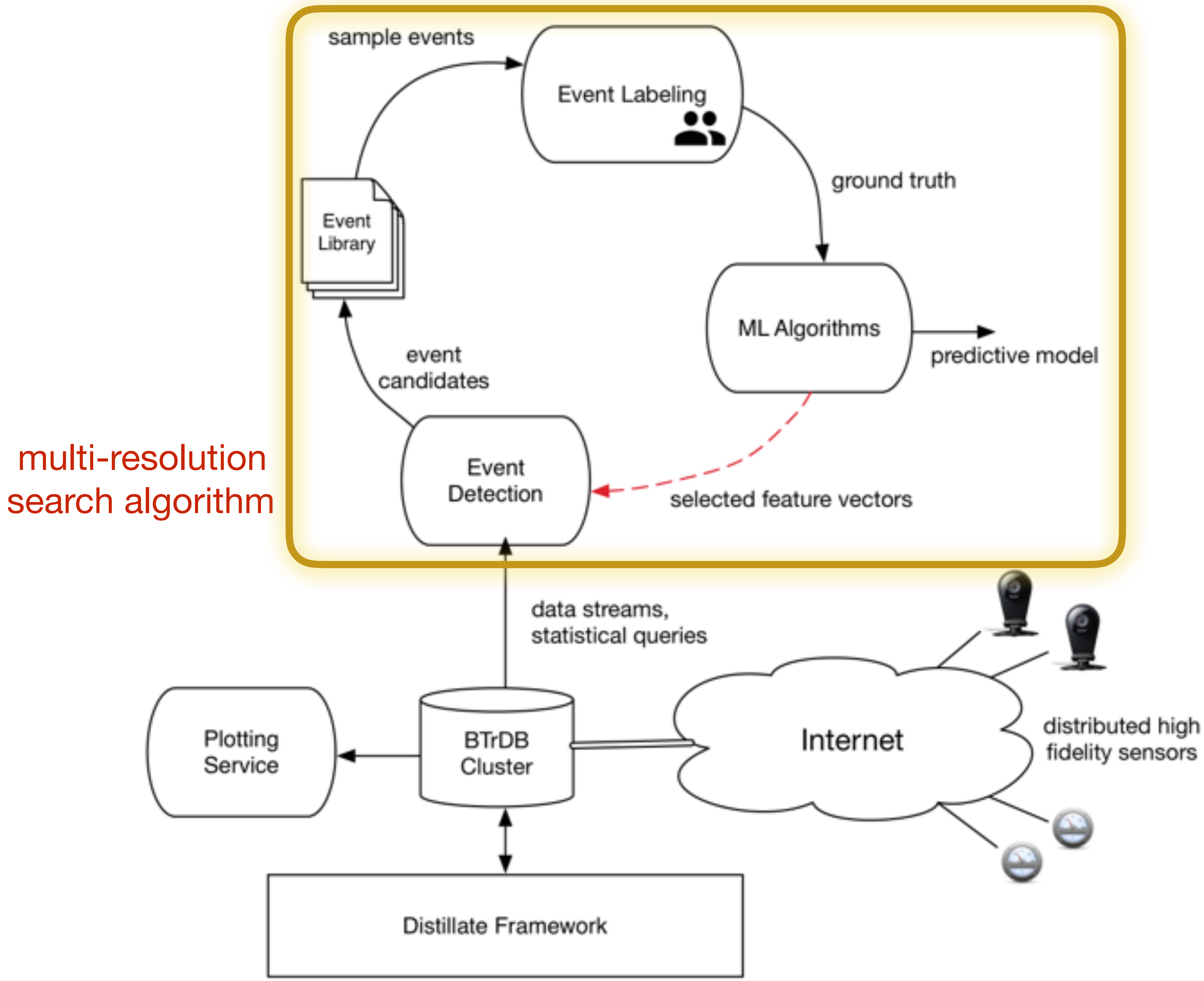


Case Study: Power Flow Reduction Detector



kernel: $\max_{res}$

statistical records: max



sample events
Event Labeling
ground truth
Event
Library
ML Algorithms
predictive model
event
candidates
Event
Detection
selected feature vectors
multi-resolution
search algorithm
data streams,
statistical queries
distributed high
fidelity sensors
Plotting
Service
BTrDB
Cluster
Internet
Distillate Framework

Takeaways

- Complexity of the search algorithm is O(nLog(L))
- Writing a kernel function can be quite challenging in some cases
- Machine learning techniques can be used to develop sophisticated detectors