New Opportunities For PODC?: Massive, Volatile, but Highly Predictable Resources

Andrew A. Chien
University of Chicago and Argonne National Laboratory

ACM PODC
July 26, 2016


Andrew A. Chien
University of Chicago and Argonne National Laboratory
w/ Fan Yang, Rich Wolski (UCSB), Victor Zavala (UW), Kibaek Kim (Argonne)

ACM PODC
July 26, 2016
Climate change due to Anthropomorphic Carbon Emissions

- => drives a radical change to Renewable-based power generation

Dramatic ICT Power Consumption Growth

- Data centers a growing fraction of ICT
- Computing becomes environmentally limited

Scopes of Energy Efficiency

- Rack, Data center
- Power Distribution, Power Grid
- Generators

Talk Focus

Outline

- Can Power be Zero-Carbon?
- Understanding Stranded Power
- Dispatchable Computing Loads
- Surprise! This is good for the Power Grid
- Properties of Massive, Intermittent Resources
- Some Challenging New Problems
- Q&A
Stranded Power

- Grid economic management mechanisms can cause renewable power to be “stranded”
- Ramps, Transmission Congestion or Shortage, Correlated Production
- MISO “down dispatched” 2.2TWH of wind in 2014
  - Equivalent to 251 MW average; ~2%, MISO: Success!
  - Real opportunity is 7.7TWh!

Curtailment power is a global issue! 2014, United States

- 2014, China, 15TWH
- 2015, China, 34TWH

2013, Denmark, Spain, Sweden, Canada, Ireland, Italy, Japan, Portugal,
Understanding Stranded Power

Mid-continent Independent System Operator (MISO)

- 15 states, serves 42M people
- 130 GW peak load, >500TWh/year
- $37B Power (2014)
  - 413 Market participants
  - >2,000 pricing nodes
- >65,000 miles transmission
  - $2.2B transmission charges
- ~10% Wind generated
MISO Market Properties

- Day-ahead, Real-time
- Dispatches power in 5-minute intervals
- Prices vary from -$999 to +$999 / MWh
  - Vary over time/weather
  - Vary over locations (topology and distance)
  - Vary with quality of prediction (and hedging)

- Swings are wide and fast...

MISO Time-lapse Power Prices Movie

- MISO Contour Map - Movie
MISO Extremes – High Prices

July 18, 18:00

July 19, 12:50

MISO Extremes – Low Prices

July 19, 22:55

July 20, 06:30
Example Site – Timeline (zoom)

$$/\text{MWh} against Hours

- Price vs. Time, Unit 8238, January 2015

Example Site – Stranded Power

data points: 2363
Unit 6406

- EcoMax (potential) vs Delivered [2015: Jan-April]
Zero-Carbon Cloud Approach

- Volatile Zero-carbon Computing facilities running on Stranded Power
- Distributed throughout the power grid
  - Green, Zero-carbon
  - Good for Grid stability
- Scale computing and Renewable Power


Making Zero-Carbon Cloud Work

- Can Stranded Power be Usable?
- Can Intermittent Resources Productive?
- Is ZCCloud good for the Power Grid?
## Analysis of a Power Grid Market

**MISO Market Records**

<table>
<thead>
<tr>
<th>Period</th>
<th>Jan 1, 2013-Apr 14, 2015</th>
<th>28.5 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation Sites</td>
<td>1,259 Total</td>
<td>200 Wind</td>
</tr>
<tr>
<td>5-min Intervals</td>
<td>76.9M Total</td>
<td>36.6M Wind</td>
</tr>
<tr>
<td>Power</td>
<td>1,188 TWh Total</td>
<td>89 TWh Wind</td>
</tr>
<tr>
<td>$$ Revenue</td>
<td>$95 Billion</td>
<td>$4 Billion</td>
</tr>
</tbody>
</table>

## Stranded Power II

- Intervals of Uneconomic Power (average price)
- **NetPrice** = average price of power over a longer period
  - A good deal for generator for the entire period
  - Idea: smooth out the fluctuations

\[
NetPrice = \frac{\sum_{\text{period}} LMP \cdot Power}{\sum_{\text{period}} Power}
\]

- **NetPrice(C)**: NetPrice < C (C = $0,$1,...,$5)
Wind Sites, SP Duty Factor: NetPrice0

NetPrice Stranded Power vs Price Threshold ($/MWh)
NetPrice Stranded Power: Intervals (by time)

Stranded power is usable, duty factor and long intervals on a single site.

Idea: Dispatchable Computing Loads to Exploit Stranded Power
ZCCloud Approach

- Volatile Zero-carbon Computing facilities running on Stranded Power
- Distributed throughout the power grid
- Complement traditional computing resources ...
- Soak up the growing demand
- Stabilize the grid?

How much of future computing can be supported by Zero-carbon Cloud?

Example: Computing Services

- Amazon Spot Instances:
  - Execute a “virtual machine” when the cost below a bid price
  - Start time + End time: dictated by the system
  - Queue of “spot instance” requests and prices

- Stateless Computing (Serverless):
  - Register a computation on an “event”
  - When fired, run to completion in 5 minutes, pay by activations and runtime (100ms)
  - Amazon Lambda, Google Cloud Functions, IBM OpenWhisk/Swift
Example: Computing Services

- Batch High-Performance Computing:
  - Execute a parallel job (array of VM's); 10K, 100K
  - Start time + End time: dictated by the system
  - Queue of "new" parallel job requests

Augmenting Mira with ZCCloud

10 Petaflops, 4MW, 786K Cores
2nd System of the Same Size
10 Petaflops, 4MW, 786K Cores
Argonne Leadership Computing Facility Workload Traces

- Mira: #5 Supercomputer in the World
- IBM Blue Gene/Q System, 768K Cores, 10 Petaflops

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of Jobs</td>
<td>78,795</td>
</tr>
<tr>
<td>Runtime (Hours)</td>
<td>Range: 0.004 to 82</td>
</tr>
<tr>
<td># Nodes</td>
<td>Range: 1 – 48K</td>
</tr>
</tbody>
</table>

Performance, Intermittent Resources

- LMP models perform worse than periodic model with equal duty
- NetPrice models can provide higher duty factors, and even better performance than periodic
Throughput, System Scale and Duty Factor

- Throughput grows with Stranded Power Models, Resources

Dispatchable load Benefits in the Renewable-based Power Grid
Growing RPS Standards and Dynamic Range

- Dynamic range for these grids approaches 50-100% of cleared load to meet aggressive RPS goals.
- CAISO adopted 50% RPS goal for 2030, excluding hydro (highest in the world), September 14, 2015
  - Dynamic Range = 100% of Peak Demand
  - San Diego: 100% Renewable by 2035 (wow!)

<table>
<thead>
<tr>
<th></th>
<th>Renewable Fraction</th>
<th>Peak Renewable @ 30% Productivity</th>
<th>Total Energy/Year</th>
<th>Peak Demand</th>
<th>Dynamic Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAISO 2010</td>
<td>10%</td>
<td>8.8GW</td>
<td>230TWH</td>
<td>40GW</td>
<td>22%</td>
</tr>
<tr>
<td>CAISO 2020</td>
<td>33%</td>
<td>29GW</td>
<td>230TWH</td>
<td>40GW</td>
<td>72%</td>
</tr>
</tbody>
</table>

Study Case: CAISO and WECC

- 1.8 million square miles
- 78 million people
- 126,285 miles of transmission (225 busses, 300 lines)
- 130 generators, 5 major import/interconnections
- ~31GW power
- 8 day types <season, wd/we>
  - Fall, winter, spring, summer
  - Weekday, weekend
- 1,000 statistical samples for each day type (generation)
Summer Day, Wind Power 15%

- Wide variation across scenarios; Coastal behavior

Rising RPS – Absorbed & Spillage

- WECC study, scaling Wind Penetration to 50%
- Real generation statistics (8,000 scenarios)
- Real topology, import, ramps, etc.
As Wind Generation scales, so does Stranded Power

- Economic situation for Wind Generators is poor
- Absorbed power with Economic value does not increase.

Data Center Strategy Impact on Total Power Cost

- Case 1: Base
- Case 2: 20x200MW DC
- Case 3: 20x200MW DC +Wind on-site
- Case 4: 20x200MW DC Dispatchable Loads (ZCCloud)
**DC Strategy Impact on Achieved RPS**

- Increased load (DC’s) aid in achieving high RPS
- Dispatchable loads best, DC+Windfarm Worst!

![Graph showing DC strategy impact on achieved RPS]

**Dispatchable Load Summary**

- WECC case study, increasing wind penetration likely to produce dramatic quantities of stranded power
- Dispatchable loads (computing) can harvest large quantities of this power
- Dispatchable loads (computing) can provide grid flexibility to manage penalty clauses, reducing costs significantly

---

*Kim, Yang, Zavala, and Chien, Data Centers as Dispatchable Loads to Harness Stranded Power, IEEE Transactions on Sustainable Energy, 2016.*
Related Work

- Characterizing Stranded Power (Curtailment, Uneconomic)
  - [NREL Bird2013, Bird2014], [Wilson 2014], ISO’s. -- Total, no fine-grained
  - Wind Vision 2015 [DOE]
- Demand-Response to Stabilize high RPS grids [many]
  - Data Center Demand-Response [Zavala, Wierman2013,2015], 15% and Economic Incentive
- Data Center SolarPV, Long-term Purchase for Wind
  - Energy Proportional [Barroso09]
  - Offset, some support of new renewable generation [Google, FB, Msft, Amazon]
  - Colocation: load counter-variation with renewables in the grid [Apple]
  - EE, Green power, Dynamic Management [Bianchini Parasol, 2011, 13, ...]
  - Net-zero Data Center [HP/Bash, Patel, NREL/Hammond]

Zero-carbon Cloud Summary

- Stranded power is a growing phenomenon in power grids around the world
  - Duty factor and quantity -> Useful
- Computational Dispatchable loads
  - Produces intermittent resources of massive scale
  - These resources are predictable
  - May be beneficial for the power grid at high RPS
The Future Cloud: Traditional+ZCCloud

- Volatile Resources running on Stranded Power complement Traditional
- Computing load shaped to be “compliant”
  - Time shifting
  - Location-shifting
- All that can be is shifted
  - Economic
  - Environmental

How are these Resources different?

- Massive scale – 10-100MW data centers
  - 10,000’s of servers, 100,000’s of cores
- Intermittent [100’s of outages, 100H’s downtime]/year
  - A single “9”
  - 1000x traditional rates
  - But – “green”, low cost power, cheaper to construct
Cloud vs. ZCCloud

- Outages: Rare (0.999x reliable), Unpredictable
- Outages: Common (0.9x reliable), Predictable?

P2P vs. ZCCloud

- Outages: Common (50% reliable), Unpredictable, Small resources
- Outages: Common (90% reliable), Predictable?
  - Massive => Big Opportunity
**Cloud Workloads – Traditional**

- External Consumer/enterprise: Diurnal workloads
  - User-responsive, Geographically constrained
  - Metrics: Response Time [ms], Throughput [M reqs/sec], Yield [fraction served] (availability)

**Analytics Workloads – Online to Offline**

- Real-time, streaming (clickstream)
- Compute Inverse Search map
- Analyze inventory/logistics
- Recommendations, pricing
- Optimize siting/logistics/vendors
- Longer-term (IoT): Precision agriculture, urban planning, research

- Analytics: Services, Optimization
  - Deadline driven, limited by Data logistics
  - Metrics: Cost, Quality, Ontime % (Deadline)
New Problems (Opportunities!)

- Prediction
- Classical Problems Revisited
- New Algorithms and Protocols
- Design of Resource Properties (and Mix)

Classical Distributed Systems Problems Revisited

- Leader election
- Consensus (Paxos, Spanner)
- Termination
- Snapshots

- Message, Time Complexity? Improved bounds
- Impact of high outage rates?
- Value of Notice?
- Value of Prediction?
Traditional Resources

- Capacity: requests/second, total bytes
- Latency: synchronous, asynchronous, milliseconds (L)
- Reliability: yes (how many 9’s), no (P2P)

Intermittent Resources

- Traditional dimensions plus...
  - Duty “Factor” (i.e. Availability, 63%)
  - Availability Trace
  - Arrival Process, Departure Process (pdf)
  - Notice [Going offline NOW, in 5 minutes]
Predictable, Intermittent Resources

- All of above plus...
- Intermittence Prediction based on external information
  - Confidence intervals for "event times", "resource capacity"

New Algorithms and Protocols

- Algorithms for X (graphs, ML, optimization, ...)
- Protocols: Data replication strategies, Consistency
Design Resource Properties

- Theory insights to “useful” resource behaviors
  - What behaviors enable much better solution to classical problems?
- Examples of what’s possible:
  - Notice: “perfect” zero-latency failure detector
  - Warning: 10-second warning for intermittence (energy storage, ?)
  - Prediction: open problems for intermittence, return to availability
- Gradual:
  - Pairing with reliable and intermittent resources
  - Variations of storage, networking across multiple levels

Design of Mixed Resource Collections

- Traditional and Intermittent Resources
- Intermittent A (Wind)+ Intermittent B (Solar)
- Covariation and Countervariation
- Location-dependence
- Cost
Summary

• Drive to renewable power gives rise to growing quantities of stranded power
• Zero-carbon Cloud is an ambitious attempt to exploit this growing waste for Intermittent computing
• The advent of such resources gives rise to a host of new research questions
  • Prediction
  • Revisit classical problems
  • New optimization problems
  • Design of Intermittent resources

QUESTIONS?
More about ZCCloud
