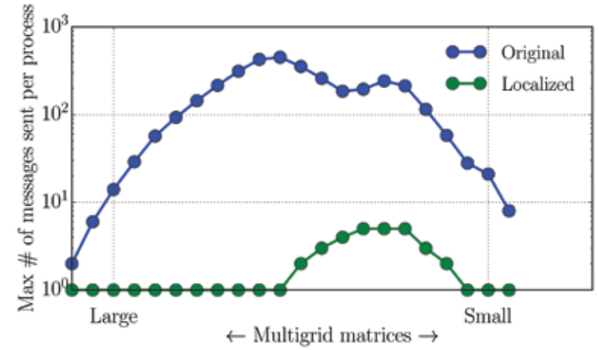

FIGURE 3: Effect of localized message communication on MPI messages.

it was observed that the standard approach to redistribution of data yielded a growth of $p^{1.1}$ in simulation time, while a redistribution designed to limit communication yields around $p^{0.17}$. This is a signficant saving, particularly as p approaches one million cores.

WHY BLUE WATERS

Blue Waters was key to this work as it provided access to the fast Cray Gemini interconnect as well as to the large core count needed for accurate scalability studies. Sparse matrix algorithms need to take advantage of multiple aspects of the compute architecture, and Blue Waters allowed for the development of both on-node features and long-distance communication decisions in the multigrid algorithm. As multigrid methods are expected to be a key solver technology on future systems, access to Blue Waters has been instrumental in advancing these methods.

NEXT GENERATION WORK

Increased concurrency will continue to challenge the scalability of sparse matrix computations. One goal is to capture the topology features of these systems more directly in the algorithms. This project provided several steps toward that goal. A future element of this work would be to add heterogeneous compute elements to the multigrid algorithm and to enhance the underlying performance models to effectively capture their use.

EVALUATING DATA-DRIVEN DETECTORS OF ELECTRICITY THEFT IN SMART GRIDS

Allocation: Illinois/50.0 knh
PI: William H. Sanders[1]
Co-PI: Varun Badrinath Krishna[1]
Collaborator: Juran Kirihara[1]

[1]University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

Electricity theft is a billion-dollar problem faced by utilities around the world and current measures are ineffective against sophisticated theft attacks that compromise the integrity of smart meter communications. We are devising algorithms that detect such theft attacks, and that are based on mathematical techniques in statistics and machine learning. The goal is to detect and mitigate theft by identifying anomalies in consumption patterns of electricity consumers. We used Blue Waters to evaluate the effectiveness of the autoregressive integrated moving average (ARIMA)-based approach to detect simulated anomalies in smart meter data. The best parameters for these algorithms need to be found using scanning techniques, and they need to account for a wide range of attack parameters that produce anomalies. Our evaluation is based on a large dataset obtained from a real smart meter deployment.

INTRODUCTION

Bloomberg News reported that electricity theft in India contributes to blackouts and costs $17 billion in lost revenue annually. According to the World Bank, electricity theft contributes to a loss in electricity delivery of over 25% of generated supply in India, 16% in Brazil, 6% in China and the U.S., and 5% in Australia. Theft in these countries is almost always achieved by tapping into electric distribution lines. To detect these thefts, utility companies such as BC Hydro have been convincing consumers to install smart meters. However, there has been some push-back as consumers have begun to realize that smart meters are vulnerable to cyber intrusions. In 2010, the Cyber Intelligence Section of the FBI reported that smart meter consumptions were being under-reported in Puerto Rico, leading to annual losses for the utility estimated at $400 million. In 2014, BBC News reported that smart meters in Spain were hacked to cut power bills. Given that smart meters can be compromised, the roll-out efforts of utilities such as BC Hydro may only increase the attack surface for cyber intrusion-based theft methods.

We identified seven classes of electricity attacks, some of which distribute the monetary loss to consumers, at no loss to the utility. Therefore, this problem is not only important to utilities, but also consumers around the world.

METHODS & RESULTS

The methods in this project are detailed in our earlier work [2], where we simulated electricity theft attacks on 500 consumers and tested our detector's false positive and false negative rates on them. The detector fitted an ARIMA model to the consumption data time series and then flagged outliers using a confidence interval created from the model.

We used methods in [3] to fit the ARIMA time series but learned from the larger simulation on Blue Waters that these methods do not scale well and are very sensitive to outliers. Also, we found unexpected results that used Python packages built by third parties. Upon further investigation, we identified errors in the algorithms coded in those Python packages (specifically the statsmodels.tsa.arima_model.ARIMA package). The algorithms simply do not implement ARIMA models correctly and use the differencing order term in the ARIMA model in a manner that is inconsistent with the theory.

WHY BLUE WATERS

Our earlier work published in [2] was performed at a much smaller scale (500 consumers) on a regular server rack, consuming inordinate amounts of processing time. We wanted to perform evaluation studies of our detector at a larger scale (2900 consumers). Also, we wanted to try out many parameters for our detector at that scale. Without Blue Waters, it would have taken years to complete these tasks.

NEXT GENERATION WORK

Our experience with Blue Waters helped us identify problems with third party software packages. We are redeveloping those packages ourselves and hope to complete them in time for the September 2016 call for proposals. With the correct algorithms, we hope to discover insights on our electricity theft detector from our use of Blue Waters.