**Noname manuscript No.**
(will be inserted by the editor)

# Data-Driven Optimizations in IoT: A New Frontier of Challenges and Opportunities

**Sharda Tripathi · Swades De**

Received: date / Accepted: date

**Abstract** Internet of Things (IoT) has gained tremendous popularity with the recent fast-paced technological advances in embedded programmable electronic and electro-mechanical systems, miniaturization, and their networking ability. IoT is expected to change the way of human activities by extensively networked monitoring, automation, and control. However, widespread application of IoT is associated with numerous challenges on communication and storage requirements, energy sustainability, and security. Also, IoT data traffic as well as the service quality requirements are application-specific. Through a few practical example cases, this article presents IoT data driven unique communication approaches and optimization techniques to reduce the data handling footprint, leading to communication bandwidth, cloud storage, and energy saving, without compromising the service quality. Subsequently, it discusses newer challenges that are needed to be tackled, to make the IoT applications practically viable for their wide-ranging adoption.

**Keywords** Internet of Things (IoT) · Smart Grid · Smart Meters · Wireless Sensors · Data-driven Networking

This work has been partly supported by the Department of Telecommunication, Government of India, under the grant no. 4-23/5G test bed/2017-NT, for building end to end 5G test-bed.

Sharda Tripathi
E-mail: sharda.tripathi@ee.iitd.ac.in
Swades De
E-mail: swadesd@ee.iitd.ac.in

Department of Electrical Engineering
IIT Delhi, New Delhi, India

## 1 Introduction

The emerging Internet of Things (IoT) paradigm has brought together sensing, communication, big data, and artificial intelligence to achieve technological advancements for manifold benefits. The footprints of IoT applications is spread across all major industries and markets, as depicted in Fig.1. The key utilities include regulating energy consumption through applications like home automation and smart electric metering, streamlining operations in factories, improving social welfare through eHealthcare, smart education, transportation, and agriculture, process monitoring through sensor networks for power grid, military, and environment, and adding value to daily life with smart wristwear, clothes, and medical wearables. To support IoT applications for masses, studies on design of reliable and low-latency communication networks [1], centralized and distributed network architecture, access protocols at data link and network layers [2], and security features [3] have lately attained considerable research interest. However, in this article, we will focus on the area of *smart IoT communications*, which is relatively new and less explored.

Layout of this paper is as follows: Section 2 briefly motivates the context of data-driven framework in IoT, followed by detailed discussion of three IoT applications pertinent from data-driven perspective in Section 3. In Section 4, our findings from applying data-driven optimizations to the case studies of smart grid monitoring and smart metering are presented. A few pertinent open issues and challenges are discussed in Section 5. The paper is concluded in Section 6.

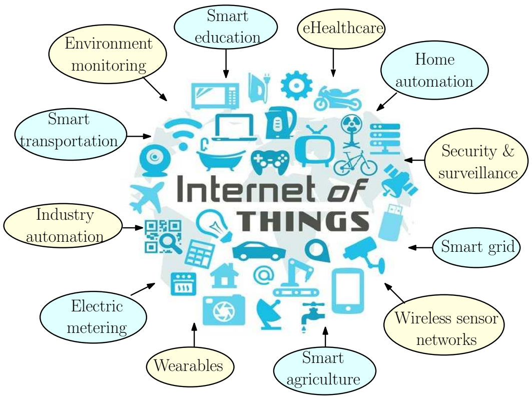


**Fig. 1** Various IoT applications.

## 2 Data-driven Framework in IoT

IoT serves as an adaptable platform for collection and exchange of sensor data in $5^{\text{th}}$ generation communication networks. Data acquisition and processing provides key insights into the dynamics of real-world systems, thereby enabling the development of effective control and automation techniques for improved efficiency and economic gain of the physical system. However, due to high sampling rates of IoT devices and near real-time data delivery, massive data is generated in the communication network. Consequently, efficient resource utilization for data transmission and archival is a major challenge with this rapidly growing technology.

Smart IoT communication is a data-driven framework, in which every IoT edge device is equipped with some intelligence based on the learning of underlying process dynamics. This intelligence enables to sample or communicate data judiciously without compromising on the information content for critical decision making purposes. The idea of augmenting node-level intelligence to the IoT sensors has dual benefits. Primarily, it makes them less communication resource hungry so as to reduce traffic load on communication network infrastructure. This is highly desirable in massive machine-type IoT communications context where huge number of edge devices simultaneously transmit their data over the communication network. Besides, it also minimizes the need for cloud data storage which rapidly improves the fetching and caching of content for learning algorithms interacting with this data.

In context of data analysis and pruning of field data in IoT it may be noted that, unlike the classical approaches of approximate modeling of traffic distribution, actual data-driven techniques are of contemporary interest for two reasons. One, in real-life applications the field data is unlikely to follow any fixed regular or stationary stochastic pattern; instead the data is highly dynamic, due to complex interplay of several system parameters and external influences. Therefore, any approximation in data pattern characterization is expected to highly deviate from the actual reality. The other reason is the current-day availability of fast processors and miniature embedded microcontrollers that can easily crunch high volume of data, *learn* from the dynamic data pattern, and take the needful predictive actions. Further, ability to execute the dynamic prediction task in a distributed manner aids in reducing the computation bottlenecks.

Data analytics as a subject area has been a major recent research interest in the Computer Science and Engineering community. However, data volume reduction coupled with network architecture-level solutions and context-aware caching are unique. A typical IoT network is shown in Fig.2. A lacuna observed in state-of-the-art is that due to conventionally low sampling rate and consequent unavailability of sufficient data in a short time frame at the edge node of IoT network, bandwidth saving between these nodes and data aggregator has not drawn much attention. In recent studies [4],[5], signal processing frameworks have been proposed for pruning of data volume in a typical IoT scenario. These methodologies, when integrated with scalability of the cloud, are shown to be highly effective for processing large amounts of data in near-real time. However, these schemes were developed with focus on big data streams generated at data aggregator or control center stage of the network architecture. With proliferating sampling rates and denser node deployments, and the applications becoming increasingly delay constrained, such approaches do not contribute to reduction in data volume from sensor to the aggregator. Thus, suitably leveraging of resources both at the edge and the core of the IoT network is essential.

Below, we discuss three pertinent example IoT applications from the perspective of smart IoT communications: wide area monitoring and control in smart grid, electric metering infrastructure in smart cities, and wireless sensor network for pollution monitoring.

## 3 Smart IoT Communication: Applications

There are wide-ranging IoT applications, ranging from human health care to smart manufacturing, and new applications have been constantly evolving. We elaborate studies on some key examples that are envisaged to have significant footprints from large-scale network resources viewpoints. These cases are smart grid monitoring and smart metering, respectively for power supply network and various user related optimizations, and environmental sensing, towards smart environments.

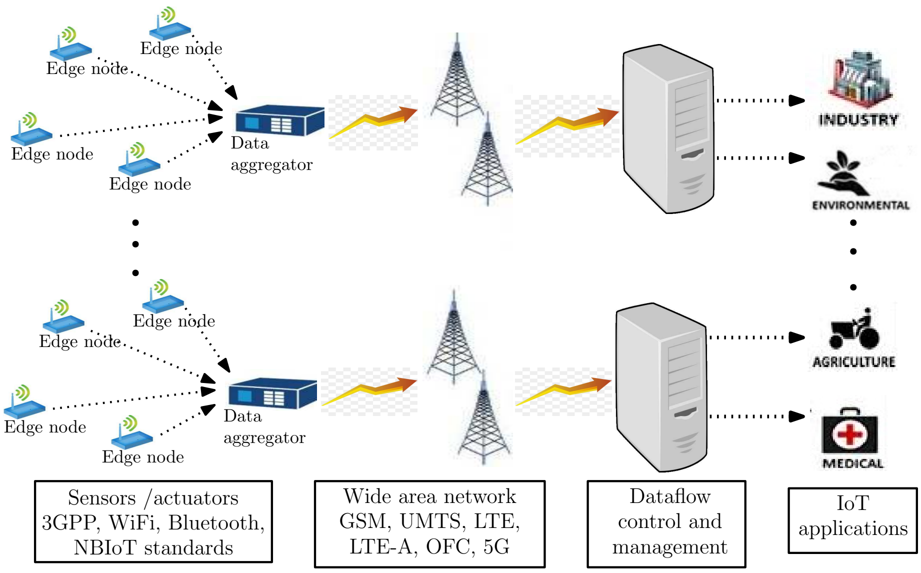


**Fig. 2** A typical IoT network.

### 3.1 Power Grid Health Monitoring

Health monitoring for wide area control in smart grid is an emerging IoT scenario. In a wide area smart grid network, the Phasor Measurement Units (PMUs) behave as edge devices or IoT nodes and generate loads of fine grained data from electricity distribution networks. This data is transmitted over communication network to a remotely located Phasor Data Concentrator (PDC) and further processed to facilitate development of effective monitoring and control techniques, thereby enhancing power grid reliability and meeting the quality of service requirements of electricity consumers.

Some unique challenges in smart grid networks are the following: (a) The smart grid health data is time-critical, which ensures that control actions from the PDC take place well in time in the event of any instability is signalled by the PMU. (b) The data is stochastically dynamic, influenced by complex factors. (c) There are network architecture and communication channel dependent packet delay, loss, and security vulnerabilities. This article discusses one key aspect, namely, communication and storage resource optimization by employing some processing intelligence at the 'edge', which requires to account for time-criticality of the traffic.

Though acquisition and analysis of this massive data imparts intelligence to the analytical framework to adapt to the dynamics of real world systems, its efficient communication and storage remains a challenge [6]. In the existing studies, a few works have addressed the issue of data reduction in wide area measurement system. These include autoregressive modelling of PMU data sequence [7], short-term frequency prediction using state-space approach [8], dimensionality reduction of PMU data using linear principal component analysis [9], [10], signal processing algorithms based on compressive sampling [11], wavelet packet decomposition [4], [12], lossless encoding [13] and a fuzzy-based paradigm for efficient processing and compression of smart grid data [14]. It may be noted that most of the algorithms proposed in literature [7], [8], [9], [10],[4], [12], [13], [14] investigate data reduction at the PDC with the perspective of designing efficient state estimation techniques and cut down on the storage requirements. Although the objective in [11] has been communication bandwidth reduction, it does not deal with nonstationary nature of PMU data. Therefore, in absence of continuous learning and adaptation, quality of compression and hence the quality of power system health monitoring is expected to degrade over time. As an advance, in [15,16] data-driven optimization study for wide area monitoring and control in smart grid has been developed. Here we study how the data collected by PMUs in wide area monitoring is exploited in [15,16] to intelligently learn the process and optimally prune redundant content in accordance with the process dynamics.

Presently, standard reporting rate from PMU to the control center is fixed at 25 and 30 samples/s, respectively for 50 Hz and 60 Hz systems [17]. Since the transient occurrences in the power grid are sporadic and PMU data is highly redundant, for optimal use of communication bandwidth, fixed-rate data transmission at all times is not required. The proposed framework is based on $\epsilon$-Support Vector Regression (SVR) learning to predict the data at the PDC, thereby intelligently pruning the transmission of redundant data.

The application-specific data is noted to be stochastically dynamic. Thus, the process is modelled from the individual time series of sensed data at the IoT node. Due to inherent non-stationary nature of PMU data, the hyper-parameters of the learning model are dynamically recomputed as necessary, thereby maintaining the accuracy of prediction and robustness of the algorithm. Performance of the proposed algorithm is evaluated via large scale simulations using powerline frequency data. A trade-off between prediction quality and runtime of the algorithm is observed, which is addressed by suitable selection of hyper-parameters.

### 3.2 Smart Metering

Similar to the PMU data, smart metering is another source of data which is expected to increase the volume of network traffic simultaneously. While there have been several motivations on collecting power consumption data at a much finer granularity, ranging up to sampling at one second interval, investigating extraction of the needful data and judiciously deciding on the optimum locations of incorporating the intelligence is essential. Intuitively, unlike the smart grid monitoring data, smart metering does not require strictly real-time delivery guarantee. Study on the smart meter data patterns suggest that, it does not possess the same nature of dynamics as the smart grid data. Also, the vulnerability of smart metering data include privacy concerns, more *ad hoc* network infrastructure that carries the data, and easy commercial exploitation by third party. Hence it necessitates a different approach to securing and compression of smart meter data. In our subsequent discussion, data compression aspect is considered.

Strategies proposed in literature for the reduction of smart meter data are based on singular value decomposition [5], generalized extreme value characterization [18], dictionary learning and sparse encoding [19], and burrow-wheeler transform with entropy encoding [20]. The resolution of data considered in these studies are on the order of one sample per several minutes. Modern-day smart metering framework is capable of supporting capture of energy consumption data at a rate as high as 1 sample per second. To handle compression of high granularity data at the meter level, a lossy compression method based on piece-wise approximation of original data [21] and loss-less compression algorithms based on differential coding [22],[23] are proposed. Algorithms proposed to operate on low resolution data at the aggregation points [5], [18], [19], [20] have access to large data chunks from several smart meters, thus identifying daily, weekly, seasonal, or behavioural patterns in the data, and exploiting them to achieve data compression becomes relatively easy, since aggregated data from several smart meters could be huge. However, for high resolution smart meter data, load patterns are more erratic and they also vary considerably even for a single user over a time frame. This reduces the data compressibility and hence more resources are required for its transmission and storage. Algorithms proposed to work on high resolution data at the smart meter [21], [22], [23] are sensitive to small consecutive value differences in smart meter data and their compression performance degrades with increasing sampling interval and presence of corrupted samples in data transmission/collection process. Besides, they work fairly well for appliance level data, but with coarse granularity these tend to become less effective.

In view of limitations in existing studies, a novel characterization of smart meter data based on Gaussian mixture (GM) model is presented in [24]. It is shown that compared to the existing characterization models, the proposed GM model provides a significantly better fit for smart meter data. Further, at each smart meter, sparsity of data is exploited to devise an adaptive data reduction algorithm using compressive sampling technique such that the bandwidth requirement for smart meter data transmission is reduced with minimum loss of information. Specifically, an adaptive data reduction scheme using compressive sampling is devised to operate at the smart meter which achieves about 40% bandwidth saving in data transmission to the nearest collection center.

### 3.3 Environmental Sensing using Wireless Sensors

Another IoT application of interest is monitoring of environmental parameters, such as industrial or city air quality, using a network of miniature wireless sensors. The network comprises of several spatially distributed autonomous sensors to measure physical or environmental parameters. They routinely transfer their data, to be delivered to a central aggregator. Wireless sensor nodes are microelectronic devices which can only be equipped with a limited power resource. However, the sensors typically used for pollution monitoring work by the principle of electro-chemical reaction or consist of a resistor heater to facilitate sensing, which result in high energy consumption [25]. Also, a field sensor node is normally equipped with a set of sensing elements mounted on a hub, for monitoring multiple environmental parameters [26]. The consequence of periodic multi-sensing is that, the battery for powering the sensor node is rapidly exhausted. Thus, besides resource saving for transmission and cloud storage of sensor data, as in case of smart grid and smart metering, energy efficiency for sensing is also a motivating factor in sensor networks using smart IoT [27].

Recent studies on air pollution monitoring have addressed forecasting of pollutants like PM2.5 [28], ground level ozone, nitrogen dioxide, and sulphur dioxide using advanced machine learning models. Towards energy-efficient sensing, in [28], cluster based hybrid model that utilizes an autoregressive integrated moving average and neural network autoregression model is used to forecast hourly PM2.5 levels from its historical values for different nodes in a cluster. Performance of the prediction model is compared for two types of clustering techniques, based on geographical separation of the sensing nodes and coefficients obtained from wavelet decomposition of time series of PM2.5 values from individual nodes and it is found that the latter model has high prediction accuracy and low computation time. Similarly, study in [29] predicts the concentration of pollutants using support vector machines, M5P model trees and artificial neural network. For each machine learning technique, a univariate and well as multivariate modelling is considered. For different prediction horizons, M5P method to found to have better prediction accuracy and powerful generalization ability. In another direction on networked sensing of spatio-temporally varying signal, aiming at extended network lifetime, the authors in [30] presented a sparse Baysian learning based adaptive sensor selection strategy that trades between sensing quality and energy efficiency.

Thus, in context of wireless sensors, besides the communication and storage issues, the data-driven optimization framework looks for critical energy efficiency and energy-sustainable networking. The learning frameworks here additionally require to deal with the other node-level parameters, such as, the nature of sensing elements, data collection network architecture and communication protocol used, and recharging resources.

## 4 Data-driven Optimization Case Studies

The techniques used for evaluating the performance of any physical system can be broadly grouped into three categories, namely, analytical, simulation based, and data driven. The analytical technique refers to developing mathematical model of the physical system. Parameter values measuring the underlying process vary randomly with time, which are assumed to follow certain probability distributions. While this technique is widely accepted by the scientific community due to inclusion of elaborate mathematical proofs and reproducibility of the claims, it requires a deep system level understanding and substantial mathematical skills to come up with a practical model. Also, for tractability

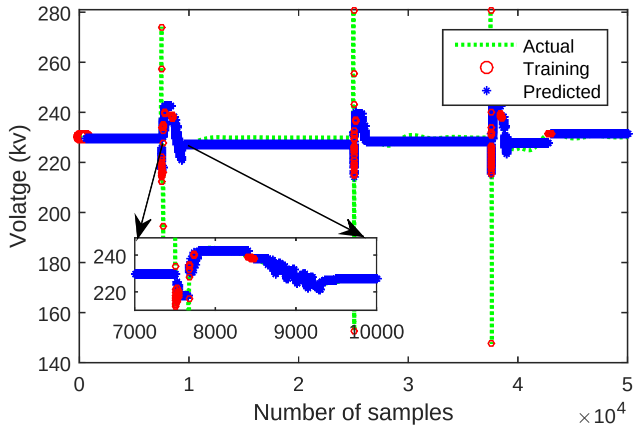


**Fig. 3** Dynamic prediction model performance on simulated phase voltage data.

of the solution, unrealistic assumptions are required to be made which may not be true to real-world application. Consequently, through analytical technique, only an approximate representation of the actual physical process is obtained. By simulation based studies, it is possible to execute high control over system parameters for a flexible model of the physical system. Thus, it is likely to be more accurate in comparison to the analytical technique. However, its performance is limited by the simulator design and programming know-how of the system designer. More often, analytical modelling is used for the validation of simulation results. On the contrary, data-driven optimization studies use observations of the randomly varying parameters obtained from an actual system implementation as direct input to mathematical programming problems. In this manner, behaviour of the physical system and its evolution over time is *learned* by the optimizer. Though this approach is expensive in terms of hardware cost involved in the system implementation, it is adaptive to the dynamics of real world system, thus most accurate and pertinent in IoT context among all three techniques.

Below, we discuss our findings on two data-driven optimization case studies, namely, smart grid monitoring and smart metering.

### 4.1 Smart Grid Monitoring

As discussed in Section 3.1, owing to high sampling rates and rapid deployment of PMUs, huge volume of data is generated in the wide area smart grid network, most of which is redundant. Here we apply dynamic prediction algorithm [15,16] on transmission phase voltage data measured by the PMU and evaluate its performance.

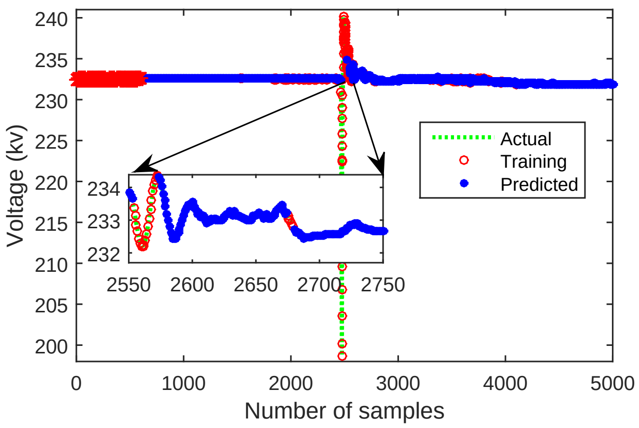


**Fig. 4** Dynamic prediction model performance on field phase voltage data.

The algorithm operates at the PMU to identify and eliminate the redundant samples before transmission using $\epsilon$-support vector regression model, while its counterpart simultaneously operates at the PDC to estimate missing samples within the prescribed latency bound. It may be noted that due to non-stationarity of PMU data, retraining of the regression model is required once the predicted sample deviates from actual value by a margin greater than predefined threshold $\epsilon$. Since the behaviour of transmission phase voltage and powerline frequency is similar in steady state and disturbed state of the power system, we choose values of hyperparameters optimum training length, lag, and $\epsilon$ as proposed in [16] to be respectively, 600 samples, 5 samples and 0.01. Remaining hyperparameters $C$ and $\gamma$ are computed on the fly during the optimization using cross-validation.

In Fig. 3, performance of dynamic prediction algorithm is shown on time series of transmission phase voltage values obtained by simulating a two-area power system via online simulink model. It may be noted that simulated data do not exactly mimic the actual field data as measured by hardware PMUs. This is because it is difficult to incorporate external factors like precision and compatibility of the measuring devices from different manufacturers and measurement noises. To demonstrate the accuracy of prediction with the field data, in Fig. 4, performance of dynamic prediction algorithm is shown on an instance of transmission phase voltage data logged in by PMU located in Vindhyachal region during the tripping incident of Rihand thermal power station unit in India on June 1, 2010.

Performance of the algorithm is measured in terms of bandwidth saving (BWS), retraining count per sample (RC), disturbance identification index (DI) and root mean square error (RMSE). BWS is the percentage of redundant samples eliminated at the PMU and successfully predicted within the predefined threshold $\epsilon$ at the PDC. For a given threshold $\epsilon$, time complexity of dynamic prediction algorithm is measured in terms of RC. It is defined as the number of retraining instances required to make successful predictions over length of time-series considered in dynamic prediction algorithm implementation. It can be observed from Table 1, that due to measurement noises in the field data, more retrainings are required in dynamic prediction implementation on data from Rihand tripping. As a result, bandwidth saving in this case in reduced to 81% compared to 94% as obtained from simulated data. To identify the samples belonging to disturbed state, it is checked that the undervoltage trigger is set at 85% of the normal operating voltage for a duration of 5 seconds [4]. Over a large interval $\Delta$, let $l_{dist}$ and $\hat{l}_{dist}$ be respectively the actual and the estimated number of frequency samples designated to be in disturbed states. Then, $DI = \lim_{\Delta\to\infty}(\hat{l}_{dist}/l_{dist})$. For both simulated and field data, RMSE is well below the acceptable upper threshold limit $\epsilon$ and all disturbance instances are correctly identified.

**Table 1** Performance indices for dynamic prediction algorithm on simulated data and field data implementation.

| Performance Index | Simulated data | Field data |
|---|---|---|
| BWS (%) | 94.65 | 81 |
| RC | 0.007 | 0.096 |
| DI | 1.07 | 1 |
| RMSE | 0.004 | 0.007 |

*Thus, for the same order of prediction accuracy, time complexity of the dynamic prediction algorithm is higher with actual field data and its bandwidth saving is reduced by* 13% *compared to that obtained using simulated data. This demonstrates the impact of measurement noise in actual system. Nevertheless, the amount of bandwidth saving is above 80%, which is significant.*

## 4.2 Smart Metering

Real smart meters sense multiple variables, such as power consumption, current, voltage, frequency, energy, and meter-health related parameters for their installation premises. Unlike their analog counterparts, they follow a rapid and automated data logging approach to generate a massive amount of multivariate data. This data is then transmitted to the data aggregator through a communication network which could be wired or wireless. High resolution smart meter data is used by near real time applications, like energy feedback, demand response, dynamic pricing, load monitoring, and short-term load forecasting. However, to address efficient com-

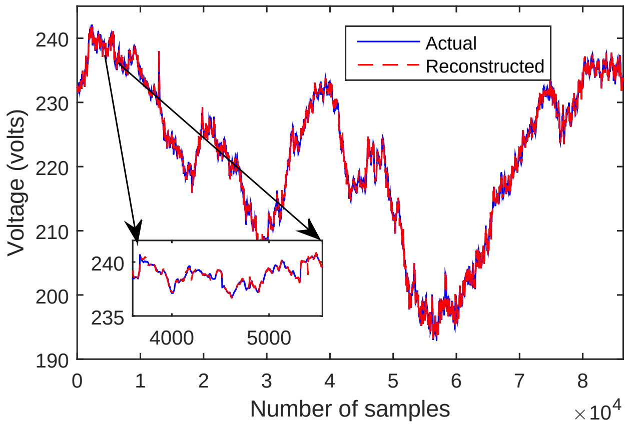


**Fig. 5** Reconstructed versus actual values for voltage variable sampled at 1 sec.

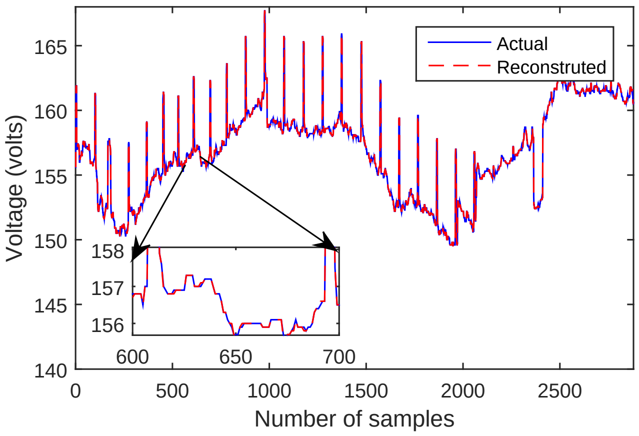


**Fig. 6** Reconstructed versus actual values for voltage variable sampled at 30 sec.

**Table 2** Performance indices for adaptive compressive sampling algorithm on smart meter data sampled at interval of 1 Hz and 30 Hz.

| Performance Index | 1 Hz | 30 Hz |
|---|---|---|
| Optimum batch size | 600 samples | 2 samples |
| Bandwidth saving (%) | 48.99 | 41.32 |
| RMSE | 0.038 | 0.046 |
| nRMSE | $1.22X10^{-6}$ | $7.17X10^{-7}$ |

munication and storage of data in this IoT scenario, reduction of data at granular level of network architecture is essential.

Here, we investigate the performance of adaptive compressive sampling algorithm as proposed in [24] on voltage variable of 2 high resolution smart meter datasets sampled at 1 Hz and $\frac{1}{30}$ Hz. The 1 Hz dataset is Indian dataset for Ambient Water and Energy (iAWE) [31]. It is openly available online and captures electricity consumption monitoring variables from a residential setup in New Delhi, India. The dataset captured at 30 Hz frequency is from smart metering framework deployed at IIT Delhi campus.

The algorithm operates at the smart meter on an optimum batch size of the targeted variable. It may be noted that for varying sampling intervals keeping the optimum batch-size fixed for compression does not lead to maximum bandwidth savings. This is because the correlation in consecutive data samples with 1 second sampling interval is much higher than that in 30 seconds sampling interval. Consequently, if batch size remains the same, bandwidth saving reduces for 30 seconds sampling interval, since owing to lower autocorrelation, much data cannot be discarded. Here we adopt a data dependent optimum batch size selection approach, wherein, before configuring the proposed adaptive compressive sampling algorithm on the metering device, the data variability pattern and the sampling frequency are studied to investigate the optimum batch size from reconstruction accuracy and bandwidth saving trade-off. As proposed in [24], optimum batchsize for data sampled at 1 Hz is taken as 600 samples. From a similar study, we found that the optimum batch size for dataset sampled 30 sec is 2 samples. Unlike conventional compressive sampling, sparsity of the data is evaluated for every batch from number of DFT coefficient containing 99.99% energy of samples in the data window and the compression is performed accordingly. It helps to capture the rapidly varying behavior of smart meter data and is essential in reducing the count of transmitted samples without compromising on the reconstruction accuracy in a dynamic environment. Performance of adaptive compressing sampling algorithm is shown in Fig. 5 and Fig. 6 for 1 day of data sampled at 1 sec and 30 seconds respectively. It can be observed that for both cases, reconstructed data and actual data closely match with each other. Corresponding bandwidth saving and accuracy metrices- RMSE and normalized RMSE (nRMSE) are presented in Table 2.

*It is found that by using data driven optimization in compression of high frequency smart data, over* 40% *bandwidth saving can be obtained. Also, for a given order of reconstruction accuracy, bandwidth saving for data sampled at 1 Hz is* 7% *higher than that at 30 Hz.*

## 5 Open Issues and Challenges

From the discussions so far, it is apparent that data-driven IoT framework for a physical system facilitates development of a cognitive application aware platform. We will now highlight a few challenges that lie ahead in practical implementation of smart IoT infrastructure.

**Latency:** Presently, a majority of IoT traffic rely on existing communication architectures like wireless LAN or cellular networks. However, existing system architectures were not designed with billions of IoT devices in mind. With traffic from a massively large number of IoT devices, meeting stringent latency constraints for real-time applications, such as in smart grid monitoring, autonomous vehicles, tele-surgery, will be challenging. While application context aware data-driven optimization schemes are expected to aid in reduced traffic handling over the network, a trade-off exists between learning accuracy and processing time of these algorithms, which must be carefully addressed for IoT devices to work efficiently.

**Network architecture:** An IoT network is supposed to handle a variety of devices with diverse power requirements, computational capabilities, and uses. To this end, multi-layered network architecture comprising of four basic layers has evolved. These are: (a) sensing layer, comprising of physical devices for data collection, object identification, and physical connectivity; (b) networking layer, wired or wireless infrastructure based on 4G, LTE, LTE-A, 5G, for transportation of data to higher layers; (c) management layer, for data processing, analytics, and decision making; (d) application layer, for ubiquitous cloud computing, intelligent processing, mega database handling. However, with induction of new features such as edge computing and smart IoT communications, source-level data processing, optimum distribution of computational load at the edge nodes, and intelligent device management are required. Besides, accommodating massive machine-to-machine communications and providing channel access to millions of IoT devices simultaneously will also require major architectural and protocol-level advancements.

**Power and energy efficiency:** As the Information and Communications community is warming up to the IoT needs, an extensive range of networking-capable devices and sensors are continuing to evolve. In most cases, these are powered using batteries alone. It may be difficult or even impossible to replenish these batteries. Besides, incorporating data-driven algorithms in sensor nodes to make them smart will incur additional computational cost that is expected to further drain out the batteries. Thus, utility of the sensing device is limited by its battery life. This has opened up new research avenues to improve the lifetime of batteries in the directions of solar / radio frequency (RF) / unmanned ariel vehicle (UAV) aided charging of sensor nodes, energy harvesting techniques for wireless sensor networks, and optimal policies for sleep-wakeup schedule of the sensor nodes. However, most of these techniques are still in investigative stage. More devices- and circuits-level technological maturity and systems-level innovative protocol optimization solutions are required to move towards green and sustainable IoT solutions.

**Security and privacy:** With rapidly-increasing number of IoT devices being connected together, more decentralized entry points, vulnerable to security attacks, are created. As a natural tradeoff with low-cost and energy-efficient sensors with small footprint, these resource-constrained devices do not incorporate strong security measures, which may lead to lucid tampering and security breaches. Additionally, once the devices are deployed on a large scale in the field, security patches are barely updated. Another primary security concern is in interfacing legacy devices that are not inherently designed for IoT connectivity, and consequently they do not have any inbuilt-mechanism against modern threats. To address these risks, it essential to devise an industry-accepted IoT security framework with policy-driven approaches to enhance system protection and ensure secure interoperability among the IoT nodes.

**System-level integration:** Seamless integration of multiple platforms for sensing, computation, communication, and intelligence, with protocols at different architectural layers and large number of application programming interfaces is cost-prohibitive and incurs technological risks. Also, rapid evolution of IoT features and lack of existing standards have further complicated the structural model. Thus, commercial success of IoT framework is currently limited by lack of expertise in system understanding, unanticipated resource requirements, and budget overruns.

## 6 Concluding Remarks

To summarize, IoT has unfolded as a promising technology for process monitoring, automation, and control in recent times. In this article, we have identified novel methodologies for exploiting data-driven IoT framework towards optimized resource utilization and development of context-aware cognitive applications in a massive machine type communication setting. Performance of data-driven optimization are analyzed based on dynamic prediction using $\epsilon$-support vector regression and adaptive compressive sampling for two germane IoT applications, namely, smart grid monitoring and smart metering. Around 80% reduction in bandwidth resource requirement is observed in transmission of PMU data, while for smart meter traffic 48% and 41% bandwidth saving are achieved for data sampled respectively at 1 Hz and 30 Hz. Thus, data-driven optimization studies have proved to be very useful in handling the stochasticity of real-world applications for which exact mathematical models do not exist.

While newer optimization approaches are required to be evolved for different application and performance context specific resource optimization, newer challenges and hence enormous research opportunities await on cost-effective, scalable, and secure solutions for massive-scale of IoT deployment.

## References


1. S. Lien, S. Hung, D. Deng, and Y. J. Wang, "Efficient ultra-reliable and low latency communications and massive machine-type communications in 5g new radio," in *Proc. IEEE GLOBECOM*, Dec 2017, pp. 1–7.
2. A. Aijaz and A. H. Aghvami, "Cognitive machine-to-machine communications for internet-of-things: A protocol stack perspective," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 103–112, April 2015.
3. I. Tomić and J. A. McCann, "A survey of potential security issues in existing wireless sensor network protocols," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1910–1923, Dec 2017.
4. P. H. Gadde, M. Biswal, S. Brahma, and H. Cao, "Efficient compression of PMU data in WAMS," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2406–2413, Sep. 2016.
5. J. C. S. de Souza, T. M. L. Assis, and B. C. Pal, "Data compression in smart distribution systems via singular value decomposition," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 275–284, Jan. 2017.
6. V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards," *IEEE Trans. Ind. Informat.*, vol. 7, no. 4, pp. 529–539, Nov. 2011.
7. E. Cotilla-Sanchez, P. D. H. Hines, and C. M. Danforth, "Predicting critical transitions from time series synchrophasor data," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1832–1840, Dec. 2012.
8. J. Dong, X. Ma, S. M. Djouadi, H. Li, and Y. Liu, "Frequency prediction of power systems in FNET based on state-space approach and uncertain basis functions," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2602–2612, Nov. 2014.
9. L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.
10. Y. Ge, A. J. Flueck, D. K. Kim, J. B. Ahn, J. D. Lee, and D. Y. Kwon, "Power system real-time event detection and associated data archival reduction based on synchrophasors," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2088–2097, Jul. 2015.
11. S. Das and T. S. Sidhu, "Application of compressive sampling in synchrophasor data communication in WAMS," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 450–460, Feb. 2014.
12. J. Khan, S. Bhuiyan, G. Murphy, and J. Williams, "Data denoising and compression for smart grid communication," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 200–214, Jun. 2016.
13. J. E. Tate, "Preprocessing and golomb -rice encoding for lossless compression of phasor angle data," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 718–729, Mar. 2016.
14. V. Loia, S. Tomasiello, and A. Vaccaro, "Fuzzy transform based compression of electric signal waveforms for smart grids," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 1, pp. 121–132, Jan. 2017.
15. S. Tripathi and S. De, "Assessment of power system stability using reduced-rate synchrophasor data," in *Proc. IEEE ANTS*, Kolkata, India, Dec. 2015, pp. 1–5.
16. ——, "Dynamic prediction of powerline frequency for wide area monitoring and control," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 2837–2846, July 2018.
17. "IEEE standard for synchrophasor data transfer for power systems," *IEEE Std C37.118.2-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1–53, Dec. 2011.
18. X. Tong, C. Kang, and Q. Xia, "Smart metering load data compression based on load feature identification," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2414–2422, Sept. 2016.
19. Y. Wang, Q. Chen, C. Kang, Q. Xia, and M. Luo, "Sparse and redundant representation-based smart meter data compression and pattern extraction," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2142–2151, May 2017.
20. A. Abuadbba, I. Khalil, and X. Yu, "Gaussian approximation based lossless compression of smart meter readings," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
21. F. Eichinger, P. Efros, S. Karnouskos, and K. Böhm, "A time-series compression technique and its application to the smart grid," *The VLDB Journal*, vol. 24, no. 2, pp. 193–218, 2015.
22. M. Ringwelski, C. Renner, A. Reinhardt, A. Weigel, and V. Turau, "The hitchhiker's guide to choosing the compression algorithm for your smart meter data," in *IEEE International Energy Conference and Exhibition (ENERGYCON)*, Sept. 2012, pp. 935–940.
23. A. Unterweger and D. Engel, "Resumable load data compression in smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 919–929, Mar. 2015.
24. S. Tripathi and S. De, "An efficient data characterization and reduction scheme for smart metering infrastructure," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4300–4308, Oct 2018.
25. P. Gupta, K. Kaushik, S. De, and S. Jana, "Feasibility analysis on integrated recharging and data collection in pollution sensor networks," in *Proc. Nat. Conf. Commun*, Feb. 2013, pp. 1–6.
26. M. Roychowdhury, N. K. Shukla, S. De, and R. Biswas, "Energy-efficient air pollution monitoring with optimum duty-cycling on a sensor hub," in *Proc. Nat. Conf. Commun.*, Mar. 2018, pp. 1–6.
27. S. De and R. Singhal, "Toward uninterrupted operation of wireless sensor networks," *IEEE Computer Mag.*, vol. 45, no. 9, pp. 24–30, Sep. 2012.
28. S. Mahajan, H. Liu, T. Tsai, and L. Chen, "Improving the accuracy and efficiency of pm2.5 forecast service using cluster-based hybrid neural network model," *IEEE Access*, vol. 6, pp. 19 193–19 204, 2018.
29. K. B. Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models," *IEEE Sensors J.*, vol. 16, no. 8, pp. 2598–2606, April 2016.
30. V. Gupta and S. De, "SBL-based adaptive sensing framework for WSN-assisted IoT applications," *IEEE Internet of Things J.*, vol. 99, pp. 1–15, Sep. 2018.
31. N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, "It's different: Insights into home energy consumption in India," in *Proc. ACM Wksp Embedded Systems For Energy-Efficient Buildings*, Nov. 2013, pp. 3:1–3:8.