Light-weight ML Aided Autonomous IoT Networks

Sushmita Ghosh, Akash Kumar Mandal, Swades De, Shouri Chatterjee, and Marius Portmann

***Abstract*—A common challenge in Internet-of-Things (IoT) networks is managing and connecting a large number of field deployed IoT nodes. With the massive growth of IoT applications, any human intervention is even more impractical. Hence, node- and network-level coordination for low-cost sensing, energy sustainability, self-organized fault remediation capability, auto-calibration of field nodes are some of the desirable features. Aiming at autonomous IoT, this article first presents the recent progress on light-weight machine learning aided strategies for context-aware IoT applications, wherein, depending on the application context and adaptation requirement, the data-driven intelligence operates at the field nodes or at the nearby edge node or at the cloud storage. A few motivating results towards operational autonomy and network scalability on the chosen use cases are presented. Next, the requirements towards fully autonomous and self healing IoT networks are presented, highlighting several future research directions and challenges.**

***Index Terms*—Auto-reconfigurability; edge computing; energy autonomy; light weight machine learning; self healing; smart IoT**

I. Introduction

Internet-of-Things (IoT) devices, such as smart meters, wireless sensors, smart wearables, etc. are increasing the ease of living by providing advanced information along with easy communication. IoT devices have massive applications in multiple sectors, such as smart homes, environmental monitoring, grid monitoring, traffic management, smart parking, smart surveillance, autonomous vehicles, etc. Wireless sensor network (WSN) is one of the most widely used IoT networks, having multiple applications across various industries, such as healthcare, pollution monitoring, smart agriculture, and border surveillance. The WSNs consist of a large number of sensor nodes that are connected to a central entity (or edge node), with each sensor node consisting of multiple sensors to monitor various parameters of the system. In a network, the nodes can be connected to each other wirelessly, to share information which can further be connected to the central entity that controls the operations based on the sensed data. Managing a large scale IoT network poses numerous challenges in terms of uninterrupted operation, high volume data data communication, computation, storage, and data security. Although the energy and bandwidth constraints can be addressed by smart sensing or data pruning techniques [1], [2], human intervention-free adaptive network operation is of great importance from cost and scalability viewpoints.

This work was supported in part by the Science and Engineering Board, Department of Science and Technology (DST), Government of India, under Grant CRG/2019/002293; in part by the Indian National Academy of Engineering (INAE) through the Abdul Kalam Technology Innovation National Fellowship; and in part by the Prime Minister's Doctoral Research Fellowships.

S. Ghosh is with UQ-IITD Academy, IIT Delhi, New Delhi, India. A. K. Mandal, S. De, and S. Chatterjee are with the Dept. Electrical Eng. and Bharti School of Telecommun., IIT Delhi, New Delhi, India. M. Portmann is with the School of IT and Electrical Eng., Univ. Queensland, Brisbane, Australia.

A. Current Trends on Learning Aided Smart IoT

The field deployed wireless nodes often have energy constraints. Besides cost and convenience, frequent battery replacement of the field nodes disrupts the network operation. Two main energy consuming activities of WSNs are sensing and communication. Towards energy sustainability, energy efficient data acquisition optimization can be divided into two categories: $a)$ node level; $b)$ network level. Several *sensor data handling* schemes are proposed in literature. Learning aided real-time (RT) data pruning by dynamic prediction and non-RT (NRT) data reduction by compressive sampling [3] are used to minimize the communication energy and data footprint transmitted to the central entity. Further, due to more energy consumption by some sensors in sensing than in communication, optimal sampling interval is decided for each parameter without compromising the sensing quality [4].

In a dynamic stochastic environment, machine learning (ML) based *adaptive sensing* is employed at the sensor node, which can be a part of a large network or a distributed sensor node. In a multi-sensing node, if the parameters are cross-correlated, a few sensors are optimally activated in the next measurement cycle based on the cross-correlation and sensing energy consumption of the sensors, while the remaining sensors stay off [5]. The temporal correlation of each parameter of the active sensor set is also exploited to find the optimal sampling interval of that parameter. For the field nodes' energy sustainability, the edge nodes play a critical role in receiving the data, running the optimization algorithm, and communicating the active sensor set and their optimal sampling intervals back to the node. Thus, edge intelligence provides a global control on the IoT network, which can also detect and act on the battery status of node, making the system more independent from external intervention [6].

In a densely deployed WSN involving network level optimization, a large IoT network consisting of several sensor nodes is expected to use the spatio-temporal correlations among the environmental parameters to monitor the system with least redundancy. In this objective, the learning based adaptive sensing strategy activates an optimal set of nodes to sense the environment over a measurement cycle. Since such algorithms are computationally complex, they are executed at the central entity and the information is broadcast to all the sensor nodes [7], [8]. In contrast, in a relatively sparse WSN, as predominant in controlled deployments in a smart city context, the spatial data correlation is rather minimal; the redundancy can be only at the node-level due to temporal correlation in data for each monitored parameter and cross-correlation among multiple parameters. Thus, optimal sensing adaptation can be only at the node-level in case of sparsely deployed WSNs. However, in joint system monitoring setups,

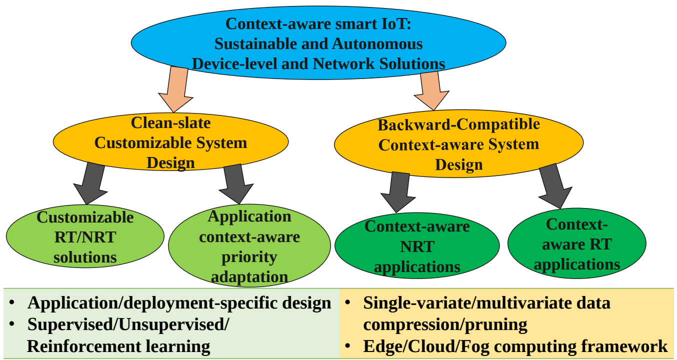

Fig. 1: LWML aided context aware IoT network solutions.

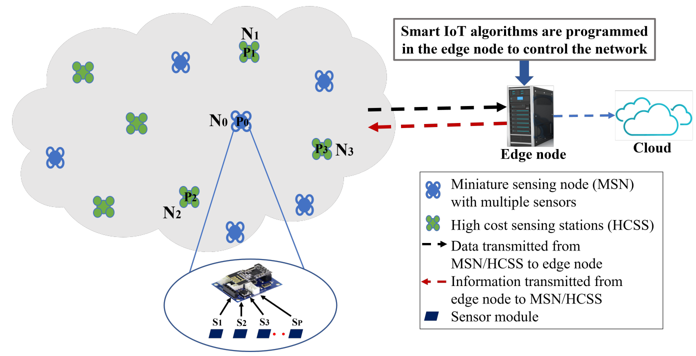

Fig. 2: Smart sensing system model.

where a minimum number of sensors are used to reduce installation and maintenance cost, node level data processing strategy is often not beneficial. This stems from the fact that, in creating an image of the system at the central entity, the data from all the sensors are used in predicting the state of the unmonitored system nodes. Thus, a system agnostic node-level data processing does not serve the purpose.

B. *Generalized Viewpoint on ML Aided Smart IoT*

Various adaptive sensing strategies, such as adaptive Nyquist sampling, auto-regressive models, ML based prediction models, etc., have been suggested for efficient data collection in IoT networks [4], [7]. Among others, light-weight ML (LWML) based adaptive sensing is proven to be quite energy efficient, with a comparatively low sensing error.

Fig. 1 depicts a generic representation of LWML aided context-aware smart IoT network solutions towards energy and operational autonomy. In developing sustainable and autonomous IoT solutions, context aware clean-slate system design for RT/NRT applications allows the implementation of smart sensing algorithms and priority adaptation at the node level as well as network level. In custom IoT applications without any control on sensing, such as in smart grid monitoring, smart energy metering, etc., for energy and bandwidth efficiency, backward-compatible context-aware dynamic data handling strategies are developed for RT/NRT solutions.

The following two sections present some representative LWML based advances towards autonomous and scalable IoT.

II. Smart Sensing in Customizable IoT Nodes

The energy consumption of the IoT nodes can be minimized by LWML techniques at the node-level, access-level, as well as network-level. Since a node typically monitors various parameters, the node's energy consumption can be reduced by optimally activating the different sensors. In case of densely deployed WSNs, the data collected form the various nodes contain redundancy. In a networked sensing, the nodes can be optimally activated to reduce the energy consumption. Some example approaches are discussed below.

A. *Customizable RT/NRT Design Solutions*

Owing to the continuously growing number of applications of IoT network, manually managing the nodes to collect data and reconfigure them is impractical and hence poses multiple challenges. Fig. 2 is an example of an IoT sensor network deployment architecture, consisting of a large number of field IoT sensor nodes equipped with multiple sensors to monitor the environmental phenomena. Each node is wirelessly connected to a central entity available in its vicinity. The network consists of both high cost sensing stations (HCSSs), equipped with good quality sensors and low cost miniature sensing nodes (MSNs), which need to be calibrated with respect to the HCSSs. Due to the high cost and large size, deploying HCSSs massively to monitor the entire geographical area is practically challenging. Thus, low cost MSNs can be deployed along with the HCSSs to make the network dense for fine granular monitoring of the environmental parameters. Since the edge node collects data from all the sensor nodes, the data collected from the HCSSs can be used to automatically calibrate/recalibrate the low cost MSNs data at the edge node.

Let, the distances between the HCSSs, such as N_1, N_2, N_3 deployed at positions P_1, P_2, P_3, respectively are on the order of kilometers, as shown in Fig. 2. However, deploying MSNs (N_0 at position P_0) makes the network more dense and improves the monitoring of environmental variations. The data collected at the nodes, based on sampling rate, are transmitted to the edge node. The edge node calibrates the MSNs data from the HCSSs data and applies the adaptive sensing algorithm on the calibrated data to find optimal system parameters, such as the number of active nodes, number of sensors at each active node to be activated in the next cycle, and the sampling interval of these sensors. The system takes decisions based on previous samples and experiences to optimize the energy efficiency. The edge nodes are connected to the base station or cloud for further data transmission and storage purposes. Thus, Fig. 2 depicts a self-managed IoT network.

1) Network level adaptive sensing : Fig. 2 depicts a multi-node IoT communication network, monitoring the variations of a system parameter over a geographical region. High sparsity in the spatio-temporal signature of the underlying parameter could be exploited to remove data redundancy in a centrally controlled setup [9]. Densely deployed WSNs present a practical use-case platform, where LWML aided adaptive sensing

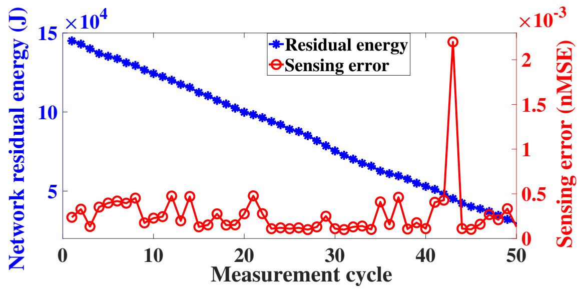

Fig. 3: Variation of sensing error and the residual energy of a network with measurement cycle [7].

strategies activate an optimal set of sensor nodes and sensing elements within them for subsequent measurement cycles, based on sparsity. These approaches increase the energy sustainability of the network. The sampling interval of the sensors in the active node can be decided based on the temporal variation of the signal [5]. In [7], an optimization function is defined to select a particular set of sensor nodes in a WSN for the next measurement cycle by jointly optimizing the trade-off between the sensing error and the sensing energy consumption. The optimization function minimizes the Bayesian Cramer-Rao bound (BCRB), which is the lower bound of mean squared error (MSE) of the estimated sparse signal to select the optimal active sensor set. BCRB is computed from the spatio-temporal variations of that specific signal. This method uses PCA based sparse signal representation and the signal is reconstructed using Sparse Bayesian learning (SBL) method. Fig. 3 describes the network level energy saving performance of the adaptive sensor node selection framework presented in [7], considering 32 sensor nodes in the WSN monitoring humidity in the environment. The PCA-SBL based framework is 50% more energy efficient compared to the other competitive methods without compromising on the sensing accuracy, as shown in Fig. 3. Thus, the network lifetime of the battery powered WSN has been increased by optimally activating the nodes by using LWML based methods without compromising the sensing quality. The central entity acts as an edge node that controls the activity of the sensor nodes connected wirelessly and makes the network self-manageable.

2) Node level adaptive sensing: In real-life scenarios, network level connectivity is not always possible in many applications, such as in home/office, medical health monitoring, etc., whereas node-level adaptation still can be used to make the IoT devices smart. Multiple works have been published on node-level energy efficient sensor data collection approaches using one-way ANOVA model, Nyquist criteria, Kalman Filter, etc., by considering a single parameter sensor node [4]. The work in [5] focuses on a learning based energy efficient sensor data collection mechanism by exploiting the temporal and cross-correlations of a multi-parameter sensor node.

As depicted in Fig 2, let us focus on node N_0 with multiple sensing elements for sensing various intended parameters, powered through a battery with limited capacity. In a real-life scenario, the sensed parameters exhibit cross-correlation. In such a case, the power hungry sensors can be judiciously turned off, with the corresponding parameters being predicted using the measurements of the lesser power hungry sensors. However, as the sensing signals are dynamic, the correlations vary with time. Hence, it is desirable to find an optimal sensor set for each measurement cycle, based on the correlations among the sensed parameters. In this objective, LWML based adaptive sensor selection algorithms employ reinforcement learning methods to determine the optimal sensor in a measurement cycle [5]. The monitored signals are used to predict the parameters of the 'inactive sensors' using efficient prediction models. While choosing an optimal set of sensors to turn on and collect samples from, the sampling interval of those sensors is decided by exploiting the temporal correlations of the sensing signals. Further, sensing and data reporting are dynamically adapted from non-time-critical ('good' state) to time-critical ('bad' state), depending on the values of some sensed parameters, which further optimizes the node-level multi-sensing and transmission. A sensing parameter is considered to be in a 'good' state if the signal remains within its satisfactory level (signal quality is satisfactory). Otherwise, the signal is considered to be in a 'bad' state.

A case study of an edge intelligence based data-driven priority-aware sensing and communication framework is adopted to optimize the energy sustainability of a multi-parameter sensor node [5]. In this method, the edge node is programmed to exploit the spatiotemporal correlation among the sensing parameters from the data collected at the immediate past measurement cycle, to find the optimal active sensor set and their sampling intervals for the next measurement. An optimization function is developed using the discounted upper confidence bound (UCB) algorithm, which extracts the trade-off between the cross-correlation coefficient of the sensing parameters and the energy consumption. Due to the non-stationary distribution of the sensing parameters, UCB performs well in this case [10]. The optimal sampling interval of each active sensor is obtained by applying a user-defined temporal correlation threshold. Two different temporal correlation thresholds are decided for good state and bad state, as the sampling rate is higher in good state compared to that in bad state. Finally, the information containing the optimal active sensor set and optimal sampling intervals is fed back to the node. Accordingly, the node activates the sensors, collects data at their respective sampling interval, and sends all the data to the edge node at the end of the measurement cycle. The parameter values of the sleep sensor set are predicted from the parameter values of the active sensor set using a Gaussian process regressor model [11]. Thus, sensing and communication are dynamically varied from NRT ('good' state) to RT ('bad' state), depending on the value of the sensed parameters. The performance of this algorithm on an air quality monitoring dataset is presented in Fig. 4. It has been observed that the node lifetime can be increased significantly by applying the edge computing method to perform the optimal sensor selection. The GPR based framework is 41% energy efficient and 32% bandwidth efficient compared to the SBL based framework [7], while detecting the system states with 98% accuracy.

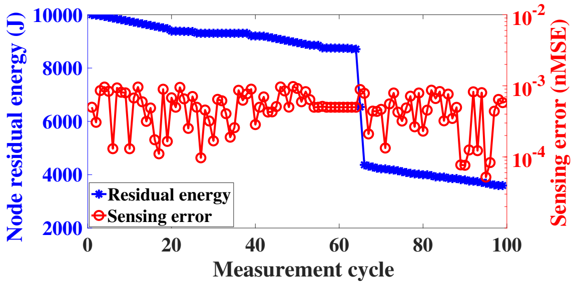

Fig. 4: Variation of sensing error and the residual energy of a multi-sensing node with measurement cycle [5].

B. Application Context Aware Priority Adaptation

Section II-A discussed the various optimization strategies applied at the application layer. In a customizable design, the IoT nodes need to be programmed for smart data handling both at the sensing stage as well as communication (networking) level. Dynamic transition of the system between NRT and RT creates a high variability of network resource requirements. Since the sampling rate of a node is higher in RT (critical/bad) state compared to that in NRT (non-critical/good) state, the bandwidth requirement is high in the critical state. Although efficient access control techniques, e.g., access class barring, exist in LTE (long term evolution), to handle a massive number of access requests, optimal scheduling, resource allocation, and routing methods are needed. The data transmitted from the node have different delay budgets based on its criticality level. Since the edge node controls the network, it can allocate resources dynamically based on the data priority. However, in massive machine type communications, the data packets are routed through multiple hops. Due to the delay associated with the routers, some packets which are exceeding the delay budgets are dropped even before reaching the data collector. To address this problem, a delay-aware priority access classification method is presented in [12]. This method dynamically assigns the access priority of the nodes having data to transmit, based on the packet delay due to access class barring.

III. ML BASED DATA OPTIMIZATION STRATEGIES FOR BACKWARD-COMPATIBLE IOT

A customizable IoT node design allows sampling the data at the required sampling interval, where smart sensing strategies can be implemented. However, this method cannot be applied to the already deployed IoT nodes where the sampling strategy is already fixed. Here, a backward-compatible smart processing can be added at the source IoT node, thus enabling to transmit only a fraction of useful samples required to reconstruct the signal. Based on the criticality of the data, backward-compatible data-driven pruning can have two major application contexts: a) NRT applications and b) RT applications.

A. Data Driven NRT Applications

Smart energy metering infrastructure is an example of NRT applications, where the energy meter collects the electrical parameters at a fixed sampling rate. While exploiting the temporal and cross-correlations among the various electrical parameters, such as energy, voltage, current, apparent power, frequency, etc., it has been observed that they possess a good degree of correlation. Thus, a dynamic data-driven resource optimization technique is employed to reduce the data volume, which is added backward with the source IoT node. At the sensor node, initially, the processing unit fetches the data in a batch of $x \times y$ matrix, where y is the number of parameters and x is the number of samples of each parameter collected by the energy meter. On the data matrix, Principal Component Analysis (PCA) is applied to reduce the sparsity and estimate the optimum number of columns (z) containing 99.9% variation of the data [13]. Thus, by pruning the less significant columns, a new data matrix of $x \times z$ is formed, where $z << y$. The transformed data matrix may have some sparsity in the time series samples (collected at a high sampling rate) even after applying PCA. To further reduce the data volume, compressive sampling is applied on each column of the transformed data matrix $x \times z$. With this, only the required samples are transmitted to the edge node, for fairly accurate reconstruction as opposed to Nyquist sampling.

Consider a vector $X = (X_1, X_2, \ldots, X_n)$, consisting of n time series samples collected by the IoT node at the Nyquist sampling rate. Thus, X can be represented as, $X = \psi F$, where ψ is the $n \times n$ sparse basis matrix and F is the transformed sparse vector of coefficients corresponding to ψ. F consists of few non-zero samples that denote the optimum number of samples (comprising 99.99% energy) required to reconstruct the entire signal. Hence, only k samples ($k << n$) are randomly chosen from X for transmission, which can be represented as, $K = \phi X = \phi\psi F$. Here, ϕ is the sensing matrix of size $k \times n$. From k random samples, the original signal X is accurately reconstructed at the receiver end by using the Subspace pursuit algorithm, which solves an underdetermined system of linear equations. In this case, Discrete Fourier Transform (DFT) and Random Gaussian Matrix are chosen as sparse basis matrix and sensing matrix, respectively to satisfy the restricted isometry and incoherence property [3].

The data compression algorithm reduces the data volume to be transmitted to the edge node, thereby reducing the storage memory requirement at the edge node. It also reduces the communication energy consumption, while losing some energy in processing the algorithm at the IoT node. Thus, the complexity of the compression algorithm should be as low as possible to reduce the processing energy consumption and achieve maximum gain in terms of total energy saving. This method achieves a bandwidth saving up to 81% in 3.6 ms with an average reconstruction error of $< 10^{-3}$ when implemented in a practical smart metering infrastructure. This concept can be easily extended to similar IoT devices generating time-series data consisting of single or multiple parameters.

Smart data pruning provides inherent data security by transforming data into feature space through the parameter k. Since the original data is converted to another domain, the pruning-induced 'encryption' at the transmitter end can only be 'decrypted' at the receiver having an accurate information of the system variables k and z, by employing appropriate decompression algorithm. These parameters are estimated for

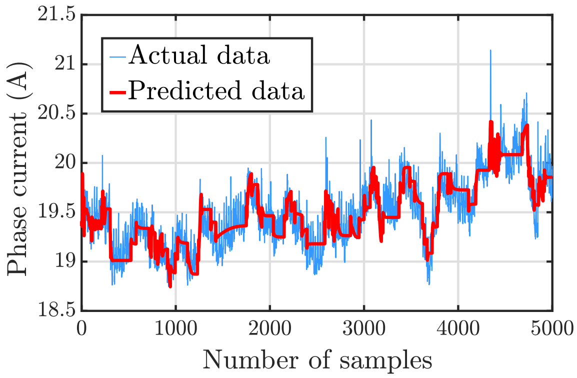

Fig. 5: Current attribute reconstruction.

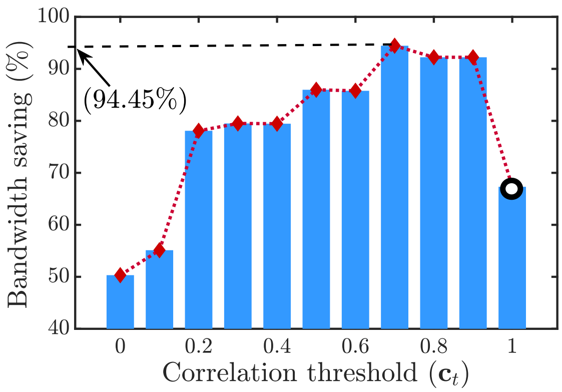

Fig. 6: Optimum bandwidth saving.

every batch depending on the sparsity of the data, thus making it difficult to maliciously estimate these parameters.

B. Data Driven RT Applications

Most of the critical IoT applications cater to the monitoring task within a defined delay budget. As the budget becomes strict, the sensing and control actions have to be accurate and fast-acting. However, with a rapid sensing requirement (higher sampling rate), the data processing exceeds the delay budget, owing to huge data volumes. This elongates the action time through the control centre. Thus, a data processing technique for such low latency systems is required, before the data analysis is done. One such RT IoT device is the phasor measurement unit (PMU), capturing the dynamic state vectors of the electrical power grids. These IoT devices report data to the edge node served by a phasor data concentrator (PDC). Therefore, a smart PMU-PDC communication strategy is required, employing RT data pruning at the IoT node with an ambient reconstruction at the edge node.

Multiple power system attributes, such as 3-phase complex voltage, 3-phase complex current, positive sequence voltage and current, frequency, rate of change of frequency (ROCOF) etc., are sensed by the IoT node. Owing to the high sampling rate used by the IoT node, the generated multi-attribute data set has a very high dimensionality. Therefore, a learning based pruning algorithm is used to compress the multi-attribute data in real time. ϵ-SVR aids this purpose by mapping the input features to a higher-dimensional space and performing regressions to find the best fit on the given data. It creates an ϵ-tube around the data which is useful when the error in prediction has to be contained within a particular range.

Two groups $\mathbf{G1}$ and $\mathbf{G2}$ are formed to segregate the base attributes in $\mathbf{G1}$ and the non-base attributes in $\mathbf{G2}$. Initially, all attributes are kept in $\mathbf{G2}$ with two parameters ρ_i^ν and δ_i^ν associated with them, where ρ_i^ν denotes the value of maximum correlation of the ν^{th} element in $\mathbf{G2}$ with the elements in group $\mathbf{G1}$ during i^{th} sample estimation, and δ_i^ν captures the position index of that attribute in group $\mathbf{G1}$. These parameters are initially set to 0 and -1, since $\mathbf{G1}$ is empty. A cross-correlation threshold $\mathbf{c}_t$ is defined using the Pearson's correlation coefficient to generate the couple attributes with highest correlation, as it is known to provide best correlation results for true experiments involving associative or causal hypothesis. Using this threshold, we iterate through $\mathbf{G2}$ and find $d_i^\nu = \rho_i^\nu - \mathbf{c}_t$, defining the distance of ρ^ν for i^{th} attribute from $\mathbf{c}_t$. Attributes having negative distance are the candidate attributes for shifting to $\mathbf{G1}$. A transfer score κ^ν for the ν^{th} attribute is calculated as $\kappa_i^\nu = \sum_{A_j \in \mathbf{G2}} ||\mathbf{m}_{kj} - \rho_j||$, where $\mathbf{m}_{kj}$ is an element of the matrix $\mathbf{M}_{l \times l}$. Finally, the one with the maximum score is shifted to $\mathbf{G1}$ and the values of ρ and ν_M are updated for all the attributes in $\mathbf{G2}$. This process is repeated until all the attributes in $\mathbf{G2}$ validate $\rho_i^\nu > \mathbf{c}_t$.

Two SVR models utilizing the radial basis function, namely, base SVR model and non-base SVR model are formed respectively at the source IoT node and the edge node. The base SVR model uses auto-regression for predicting its own subsequent samples, which are used by the non-base SVR model(s) for predicting their values. One flag each is defined for capturing run-time prediction errors in base and non-base attributes. If such a flag is raised, based on the model the flag is raised for, either a retraining or remodeling is in order. This helps to save the communication bandwidth by transmitting data only at the beginning, or during the retraining instants. The proposed method is compared with the N-single variate data pruning algorithm, which corresponds to $c_t = 1$. Thus harnessing no cross-correlation among the attributes. From Fig. 5, depicting the reconstruction graph for the current attribute, a close correspondence between the actual and reconstructed graphs was observed with a RMSE of 4.34×10^{-5}, with the maximum computational footprint upper bounded by 0.04 s. Further, from Fig. 6 we observe that by setting an optimum value of $c_t = 0.7$ with the SVR algorithm, maximum bandwidth saving ($\approx 94\%$) is attained. Also, for $c_t = 1$, a reduced bandwidth saving of 70% is achieved, which is $\approx 40\%$ lesser than the optimum achieved using the proposed strategy.

A few future research directions on fully autonomous IoT networks are outlined below.

IV. Towards Fully Autonomous IoT Networks

High energy consumption of field deployed IoT nodes is one of the most challenging problems, as the nodes are mostly powered by batteries with limited capacity. Section II discussed various learning based optimization strategies, which are implemented to reduce the energy consumption at the sensor nodes and related communication modules. Another major challenge is to accommodate the high bandwidth requirement of the already deployed non-customizable IoT devices in transmitting data sampled at high rate to the receiver node. In this context, Section III discussed the various data compression methods to prune/compress the data at the transmitter side and decompress at the edge node/receiver side. Though the work undertaken provides potential solutions in a lot of relevant use-cases, a few of the challenging aspects remain to be addressed. As we look into the future, dynamism in the IoT devices and environment pave their way through the existing infrastructure. The problem becomes further pertinent as we move towards auto-reconfigurable and self-healing networks. The work presented thus far helps to motivate such a case.

A. Energy Autonomy

Energy sustainability and green sensing/communication form the base for all future smart IoT networks. Since the battery capacity of the field deployed sensor nodes are limited, efficient energy harvesting techniques need to be incorporated to make the nodes energy autonomous. The currently available harvesting methods, such as solar, RF (radio frequency), piezoelectric, etc. generate a limited amount of energy, as opposed to the high energy requirement of the sensor nodes having multiple good quality sensors. Thus, the design of a completely self-healing IoT network involving good quality sensors is still an unresolved challenge with an abundance of future scope.

One approach in achieving this milestone would be to have the sensor nodes transmit battery status along with data to the edge node. If the energy level falls below a threshold, the edge node or the network controller can take the decision to substitute that sensor node with another energy-surplus node, while the former switches to replenishing mode by recharging from ambient or via on-demand wireless energy transfer. This leads to another direction of research, towards the development of specialized sensor-to-edge communication standards. Since edge intelligence is one of the important features in 3GPP standards for the future generation networks, standard methods need to be developed for automated actuation and control of the sensor nodes. If the edge processing capability is placed in a mobile robot used to collect data from the static/mobile sensor nodes, the edge node can be equipped with a charging unit, such as RF transmitter, while the sensor nodes are equipped with RF energy harvester. Optimum methods need to be developed at the edge node for recharging the sensor node and collecting data from the node simultaneously.

B. Auto-Reconfigurability

Dynamic adaptability at the node and network levels represent the next most important aspect of autonomy in future IoT networks. Mobile sensor nodes provide great avenues for robust and low-cost system monitoring. However, the management in such an auto-reconfigurable setting is a complex task. Beyond efficient data collection and processing at the edge node, energy management in such node-level mobility poses a potential research perspective in re-configurable autonomous IoT networks. The entire system under test could be divided into multiple small zones, each having a local data collector, which are jointly monitored by a central edge node. An efficient data offloading policy through a structured handshaking and hand-off would be of interest. Furthermore, this provides an alternate vantage point in self-healing scenarios, where an energy-deficient node can be turned off, while being replaced by an energy-surplus sensor node.

Network-level auto-reconfigurability is a step further towards autonomous IoT networks. One of the major challenges in this task is an energy efficient autonomous system reorganization, thus suggesting the need for inter-node coordination in a massive IoT node deployment scenario. This not only necessitates the use of LWML techniques at the node level, but a network level coordinated learning also becomes pertinent. One of the crucial addressed phenomena in self-healing network aspect is allowing redundancy through network-level mobility by intelligently awakening a group of sensors. This takes to the more demanding, yet appealing task of distributed analysis and joint data processing, thereby constructing an entire situational image of the system using a limited deployment of IoT sensor nodes.

C. Secure Autonomous IoT Networks

Security makes the backbone of the future autonomous IoT networks, as mass-scale adaptability of such network cannot happen without data integrity [14]. With a higher order of dynamism, higher node deployment, and sparse data offloading, the security and reliability of IoT networks becomes a challenge. As the handling of network and node level auto-reconfigurability seeks distributed processing of sensor data, correct data interpretation at the edge node and false data identification and rejection becomes increasingly important in establishing a secure and reliable IoT network. Though computation offloading at multiple levels (edge/cloud/fog) enhances the IoT node longevity, it amplifies the data access points for a possible breach. This seeks a robust data processing formulation, thereby segregating false data from the true one. Furthermore, a specialized public key infrastructure could be devised for such a distributed autonomous IoT framework.

It is notable that, as distributed data processing requires aggregate system visualization, optimum control action for maintaining or restoring system stability holds crucial importance for the IoT infrastructure. Non-ideality in the control channel must be appropriately modeled and combated to ensure reliable network functioning. Further, as the data processing becomes distributed, tail data mis-classification gains prominence, thus leading to ‘attack of tails’. Such attacks can be defended using a central observer that ensures that the mis-classification is handled as an outlier. Further, from graph-theoretic viewpoint, when a malicious user launches such attacks, the graph can have a small *quotient-cut* that disconnects multiple graph nodes. Thus, the malicious attacker can be identified and removed using a ’SybilGuard’ approach. An allied concern stems from the data reconstruction purview under channel noise. Channel adaptive transmission protocols needs to be built to facilitate robust and secure data delivery.

D. Human-Machine Interface

The factor of fully human intervention free operation has proven to be a demerit of many artificial intelligence enabled autonomous systems. The initiative by Defense Advanced Research Agency over the years has shown the pitfall in the autonomous systems owing to the human intervention-free nature of ML based solutions [15]. The mid-2015 cyber-attack drills suggested an important strategy in false event management, namely rapid attack detection, isolation, and characterization. This posits a major lacuna arising from fully autonomous networks. A potential direction of research lies in the monitoring system smartly seeking human intervention over precarious events. This could secure robust control in many false attack events, as a human perspective of system and surroundings provides a different viewpoint in event characterization. Furthermore, the idea of human machine interaction strengthens the sensor life-span by providing preemptive maintenance of the nodes beyond the network's self healing capability limit. Thus, a controlled human moderation could prove valuable in strengthening the security and reliability of the IoT network, thereby enhancing the benefits of network-level autonomy while mitigating the pitfalls of full autonomy.

V. Concluding Remarks

This article has presented strategies for realizing autonomous IoT networks through various LWML aided node- and network-level optimizations. Through some implementation case studies it has shown how some of the autonomous features can be fulfilled by energy and data footprint reduction. These include context-aware smart sensing and smart data pruning, and further edge-intelligence based network-level coordination, aiding scalability and energy autonomy. The glimpses of learning-aided auto-calibration feature of the IoT sensor nodes discussed in the article has demonstrated the possibility of operational autonomy. To achieve fully autonomous IoT networks and to work around the pitfalls of full autonomy, more research and innovations are necessary. Some of the exciting future works and challenges have been highlighted, which include smarter network towards self-healing capability, further advanced techniques on deployment agnostic energy autonomy, and on-demand wireless energy replenishability. A pitfall of learning-aided fully machine-type/autonomous system is the potential lack of human control on security, privacy, and actuation/control of system. Some of these challenges and research directions and the need for human machine interaction in smart IoT networks are also highlighted.

References

[1] H. Hu and L. Tang, "Edge intelligence for real-time data analytics in an IoT-based smart metering system," *IEEE Network*, vol. 34, no. 5, pp. 68–74, 2020.

[2] V. Hayyolalam, M. Aloqaily, O. Ozkasap, and M. Guizani, "Edge intelligence for empowering IoT-based healthcare systems," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 6–14, 2021.

[3] 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, 2018.

[4] H. Harb and A. Makhoul, "Energy-efficient sensor data collection approach for industrial process monitoring," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 661–672, 2017.

[5] S. Ghosh, S. De, S. Chatterjee, and M. Portmann, "Edge intelligence framework for data-driven dynamic priority sensing and transmission," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 376–390, 2022.

[6] M. Liyanage, Q.-V. Pham, K. Dev, S. Bhattacharya, P. K. R. Maddikunta, T. R. Gadekallu, and G. Yenduri, "A survey on zero touch network and service (ZSM) management for 5G and beyond networks," *Journal of Network and Computer Applications*, p. 103362, 2022.

[7] V. Gupta and S. De, "SBL-based adaptive sensing framework for WSN-assisted IoT applications," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4598–4612, 2018.

[8] S. Hwang, R. Ran, J. Yang, and D. K. Kim, "Multivariated Bayesian compressive sensing in wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2196–2206, 2015.

[9] J. Wang, S. Tang, B. Yin, and X.-Y. Li, "Data gathering in wireless sensor networks through intelligent compressive sensing," in *Proc. IEEE INFOCOM, Orlando, FL, USA*, 2012, pp. 603–611.

[10] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[11] C. Rasmussen and C. Williams, *Gaussian processes for machine learning*. MIT Press: Cambridge, MA, 2006.

[12] M. R. Chowdhury and S. De, "Delay-aware priority access classification for massive machine-type communication," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 238–13 254, 2021.

[13] M. R. Chowdhury, S. Tripathi, and S. De, "Adaptive multivariate data compression in smart metering Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1287–1297, 2021.

[14] T. H. Szymanski, "Security and privacy for a green Internet of Things," *IT Professional*, vol. 19, no. 5, pp. 34–41, 2017.

[15] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," *IEEE commun. Surveys Tuts.*, vol. 16, no. 1, pp. 266–282, 2013.

Sushmita Ghosh (Graduate Student Member, IEEE) is pursuing the PhD degree in the UQ-IITD Academy of Research, IIT Delhi, and is a Prime Minister's Research Fellow. Her current research interest includes design of energy-efficient IoT communication systems and networks.

Akash Kumar Mandal (Graduate Student Member, IEEE) is pursuing his PhD in the Department of Electrical Engineering, IIT Delhi. He is a Prime Minister's Research Fellow. His research interests broadly lies in smart grid communication with emphasis on channel modeling, study of system dynamics, and stability analysis.

Swades De (Senior Member, IEEE) is a Professor in the Department of Electrical Engineering, IIT Delhi. His research interests are broadly in communication networks, with emphasis on performance modeling and analysis. Current directions include energy harvesting IoT communications, broadband wireless access and routing, spectrum coexistence, and smart grid networks.

Shouri Chatterjee (Senior Member, IEEE) is a Professor in the Department of Electrical Engineering, IIT Delhi. His primary area of work is in integrated circuit design, circuits, and systems. His research interests include filter design, ultra-low power circuits for energy scavenging, and wireless front-ends for millimeter-wave radios.

Marius Portmann (Senior Member, IEEE) is an Associate Professor with the School of Information Technology and Electrical Engineering, The University of Queensland, Australia. His research interests include general networking, pervasive computing, the IoT, and cyber security.