

# Optimization Techniques for Minimizing Energy Consumption in IoT Devices at the Edge

Burla Srinivas<sup>1</sup> and Dr. Pawan Kumar<sup>2</sup>

Research Scholar, Department of Computer Science & Engineering<sup>1</sup>

Research Guide, Department of Computer Science & Engineering<sup>2</sup>

NIILM University, Kaithal, India

**Abstract:** *The rapid proliferation of Internet of Things (IoT) devices has led to an unprecedented surge in energy consumption, posing significant challenges to sustainability and operational efficiency. This abstract explores optimization techniques aimed at minimizing energy consumption in IoT devices at the edge, where data is processed closer to the source rather than relying solely on centralized cloud resources. The paramount importance of energy efficiency in IoT devices stems from their resource-constrained nature, making them susceptible to premature battery depletion and environmental impact. This paper investigates a range of optimization strategies, including low-power hardware design, energy-aware algorithms, and adaptive power management schemes. Leveraging edge computing capabilities, these techniques aim to strike a balance between computation and energy efficiency by offloading processing tasks to edge nodes. Furthermore, the abstract delves into the significance of machine learning algorithms for predicting and optimizing energy consumption patterns in real-time. The findings presented here contribute to the ongoing discourse on sustainable IoT ecosystems, shedding light on practical approaches to mitigate energy challenges and enhance the longevity and reliability of IoT devices at the edge.*

**Keywords:** IoT devices, Edge computing, Energy optimization

## REFERENCES

- [1]. H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the internet of things with edge computing," *IEEE Network*, vol. 32, no. 1, pp. 96–101, 2018.
- [2]. Y. You, Z. Zhang, C. Hsieh, J. Demmel, and K. Keutzer, "Fast deep neural network training on distributed systems and cloud TPUs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 11, pp. 2449–2462, 2019.
- [3]. C. Hu, W. Bao, D. Wang, and F. Liu, "Dynamic adaptive DNN surgery for inference acceleration on the edge," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2019.
- [4]. Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," in the *22nd International Conference on Architectural Support for Programming Languages and Operating Systems*, 2017.
- [5]. H. Badri, T. Bahreini, D. Grosu, and K. Yang, "Energy-aware application placement in mobile edge computing: A stochastic optimization approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 4, pp. 909–922, 2020.
- [6]. B. Lin, Y. Huang, J. Zhang, J. Hu, X. Chen, and J. Li, "Cost-driven off-loading for DNN-based applications over cloud, edge, and end devices," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5456–5466, 2020.
- [7]. M. Chen, S. Guo, K. Liu, X. Liao, and B. Xiao, "Robust computation offloading and resource scheduling in cloudlet-based mobile cloud computing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 2025–2040, 2021.
- [8]. M. Altamimi, A. Abdrabou, K. Naik, and A. Nayak, "Energy cost models of smartphones for task offloading to the cloud," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp. 384–398, 2015.

- [9]. K. Elgazzar, P. Martin, and H. S. Hassanein, "Cloud-assisted computation offloading to support mobile services," *IEEE Transactions on Cloud Computing*, vol. 4, no. 3, pp. 279–292, 2016.
- [10]. Z. Fang, T. Yu, O. J. Mengshoel, and R. K. Gupta, "Qos-aware scheduling of heterogeneous servers for inference in deep neural networks," in *International Conference on Information and Knowledge Management (ICIKM)*, 2017.
- [11]. K. Kumar and Y. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *Computer*, vol. 43, no. 4, pp. 51–56, 2010.
- [12]. H. J. Jeong, "Lightweight offloading system for mobile edge computing," in *IEEE International Conference on Pervasive Computing and Communications Workshops (ICPCCW)*, 2019.
- [13]. M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 587–597, 2018.
- [14]. X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.
- [15]. Z. Ali, L. Jiao, T. Baker, G. Abbas, Z. H. Abbas, and S. Khaf, "A deep learning approach for energy efficient computational offloading in mobile edge computing," *IEEE Access*, vol. 7, pp. 149 623–149 633, 2019.
- [16]. C. Lo, Y. Y. Su, C. Y. Lee, and S. C. Chang, "A dynamic deep neural network design for efficient workload allocation in edge computing," in *IEEE International Conference on Computer Design (ICCD)*, 2017.
- [17]. H. J. Jeong, I. Jeong, H. J. Lee, and S. M. Moon, "Computation offloading for machine learning web apps in the edge server environment," in *International Conference on Distributed Computing Systems (ICDCS)*, 2018.
- [18]. L. Yang, J. Cao, S. Tang, D. Han, and N. Suri, "Run time application repartitioning in dynamic mobile cloud environments," *IEEE Transactions on Cloud Computing*, vol. 4, no. 3, pp. 336–348, 2016.
- [19]. T. Mohammed, C. Joe-Wong, R. Babbar, and M. D. Francesco, "Distributed inference acceleration with adaptive DNN partitioning and offloading," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2020.
- [20]. H. Wu, W. J. Knottenbelt, and K. Wolter, "An efficient application partitioning algorithm in mobile environments," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 7, pp. 1464–1480, 2019.
- [21]. S. Teerapittayanon, B. McDanel, and H. Kung, "Distributed deep neural networks over the cloud, the edge and end devices," in *International Conference on Distributed Computing Systems (ICDCS)*, 2017.
- [22]. H. Lin, S. Zeadally, Z. Chen, H. Labiod, and L. Wang, "A survey on computation offloading modeling for edge computing," *Journal of Network and Computer Applications*, vol. 169, pp. 1–25, 2020.
- [23]. C. Hu, W. Bao, D. Wang, and F. Liu, "Dynamic adaptive DNN surgery for inference acceleration on the edge," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2019.
- [24]. N. Zhao, X. Liu, F. R. Yu, M. Li, and V. C. M. Leung, "Communications, caching, and computing oriented small cell networks with interference alignment," *IEEE Communications Magazine*, vol. 54, pp. 29–35, 2016.
- [25]. G. Xie, G. Zeng, X. Xiao, R. Li, and K. Li, "Energy-efficient scheduling algorithms for real-time parallel applications on heterogeneous distributed embedded systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 12, pp. 3426–3442, 2017.
- [26]. C. Jiang, Z. Chen, R. Su, and Y. C. Soh, "Group greedy method for sensor placement," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2249–2262, 2019.
- [27]. D. S. Hochba, "Approximation algorithms for NP-hard problems," *SIGACT News*, vol. 28, pp. 40–52, 1997.
- [28]. B. Lin, F. Zhu, J. Zhang, J. Chen, X. Chen, N. N. Xiong, and J. Lloret Mauri, "A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4254–4265, 2019.
- [29]. L. Cui, J. Zhang, L. Yue, Y. Shi, H. Li, and D. Yuan, "A genetic algorithm based data replica placement strategy for scientific applications in clouds," *IEEE Transactions on Services Computing*, vol. 11, no. 4, pp. 727–739, 2018.
- [30]. J. S. Su, W. Z. Guo, C. L. Yu, and G. L. Chen, "Fault-tolerance clustering algorithm with load-balance aware in wireless sensor network," *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 37, pp. 445–456, 2014.

- [31]. M. A. Rodriguez and R. Buyya, "Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds," IEEE Transactions on Cloud Computing, vol. 2, no. 2, pp. 222–235, 2014.
- [32]. H. Li, D. Yang, W. Su, J. LÄij, and X. Yu, "An overall distribution particle swarm optimization MPPT algorithm for photovoltaic system under partial shading," IEEE Transactions on Industrial Electronics, vol. 66, no. 1, pp. 265–275, 2019.
- [33]. D. O'Neill, A. Lensen, B. Xue, and M. Zhang, "Particle swarm optimisation for feature selection and weighting in high-dimensional clustering," in IEEE Congress on Evolutionary Computation (CEC), 2018.
- [34]. M. Masdari, F. Salehi, M. Jalali, and M. Bidaki, "A survey of PSO-based scheduling algorithms in cloud computing," Journal of Network and Systems Management, vol. 25, pp. 122–158, 2017.
- [35]. T. R. Chavan and A. V. Nandedkar, "A hybrid deep neural network for online learning," in the 9th International Conference on Advances in Pattern Recognition (ICAPR), 2017.
- [36]. X. Chen, M. Li, H. Zhong, Y. Ma, and C.-H. Hsu, "DNNOff: Offloading dnn-based intelligent iot applications in mobile edge computing," IEEE Transactions on Industrial Informatics, pp. 1–1, 2021.
- [37]. H. Topcuoglu, S. Hariri, and Min-You Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing," IEEE Transactions on Parallel and Distributed Systems, vol. 13, no. 3, pp. 260–274, 2002.
- [38]. Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in IEEE International Conference on Evolutionary Computation Proceedings (ICECP), 1998.