

ARTA: Adaptive Redundancy-aware Telemetry node Activation in 6G Edge

Shatha Abbas, Nitinder Mohan, George Iosifidis, Fernando Kuipers

Abstract—Future 6G edge networks must support ultra-reliable and low-latency communication for applications such as autonomous driving and industrial automation, while operating under strict bandwidth and compute constraints. Achieving this balance requires telemetry systems that are adaptive, scalable, and resource-efficient. However, existing telemetry approaches either rely on static deployment or network-wide activation, leading to excessive overhead and redundant measurements in dense edge environments. This paper presents ARTA, an *Adaptive Redundancy-Aware Telemetry Node Activation* framework that formulates the telemetry activation problem as a redundancy-aware benefit–cost optimization. Each node is evaluated by anomaly likelihood, traffic importance, topological centrality, and context relevance, while redundancy, caused by nearby nodes reporting overlapping telemetry data, and resource limits are explicitly modeled. The resulting submodular objective enables a greedy algorithm with provable optimality guarantees. Simulations on edge–core topologies of up to 1,000 nodes show that ARTA achieves approximately 0.8 normalized accuracy (80% of the maximum achievable accuracy) while activating only 20% of nodes within a few seconds, demonstrating scalability and zero-touch observability.

Index Terms—6G networks, edge computing, adaptive node activation, submodular optimization, zero-touch orchestration.

I. INTRODUCTION

The evolution toward 6G networks introduces new challenges for network observability. Emerging applications such as autonomous driving, industrial automation, and extended reality depend on ultra-reliable and low-latency communication [1], requiring accurate, real-time telemetry across both the core and the edge. Unlike 5G, which mainly relied on centralized telemetry, 6G connects a large number of edge and far-edge nodes (e.g., vehicular units and lightweight servers), creating dense and dynamic environments that demand a rethinking of telemetry design and deployment [1], [2].

Existing telemetry approaches cannot meet these requirements. Current 5G and early In-band Network Telemetry (INT) systems rely on static placement or network-wide activation [3]–[5]. While offering fine-grained visibility, they incur excessive bandwidth, computation, and storage overhead in dense networks. This overhead is further exacerbated when adjacent telemetry nodes report correlated measurements, generating redundant data that burdens the network without improving observability [3], [5]. Telemetry must therefore become adaptive, activating only where information is most valuable while minimizing wasted resources [6].

All authors are with the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, 2628 CD Delft, The Netherlands. Emails: {s.o.m.abbas, n.mohan, g.iosifidis, f.a.kuipers}@tudelft.nl

Designing adaptive telemetry for 6G edge networks is challenging due to the large number of candidate nodes and the dependency between activation decisions, whereby activating one node affects the value of others, making the problem NP-hard. Prior approaches based on static probe placement [3]–[5] or topology heuristics [7] fail to scale because they overlook these interdependencies and redundancy effects, while activation decisions must still be computed rapidly under strict bandwidth and compute constraints.

To address this problem, we propose the **ARTA** framework that formulates telemetry activation as a redundancy-aware benefit–cost optimization problem. Each candidate node is evaluated by anomaly likelihood, traffic importance, topological centrality, and context relevance, while redundancy from overlapping telemetry and resource constraints is modeled.

We analyze the properties of the formulated optimization problem and show that it exhibits a submodular structure with cardinality constraints. Leveraging this property, ARTA employs a greedy algorithm that achieves provable optimality guarantees and runs in $\mathcal{O}(|V|K)$ time, where $|V|$ is the number of candidate telemetry nodes and K is the activation budget, ensuring scalability for large topologies [4], [5], [8].

ARTA models redundancy among nodes with overlapping monitoring coverage and enforces resource constraints to maintain efficient and nonoverlapping operation. This redundancy-aware formulation keeps telemetry adaptive and resource-efficient even in large 6G edge networks. Simulations on topologies of up to 1,000 nodes show that ARTA achieves high anomaly-detection accuracy while reducing cost and redundancy, providing a practical foundation for zero-touch and resource-aware observability in next-generation networks. This paper makes three main contributions:

- We define the telemetry node activation problem for dense 6G edge networks, highlighting the importance of balancing telemetry accuracy with resource efficiency.
- We formulate a redundancy-aware benefit–cost optimization problem and prove its submodularity, enabling scalable greedy algorithms with provable guarantees [8] (§III).
- We evaluate ARTA on network topologies with up to 1,000 nodes, achieving about 0.8 of the maximum accuracy while activating 20% of nodes with near-zero redundancy and improved scalability (§IV).

The remainder of this paper is organized as follows. Section II reviews background and related work, Section III presents the proposed redundancy-aware optimization frame-

work, Section IV evaluates ARTA, and Section V concludes the paper and outlines directions for future work.

II. BACKGROUND AND RELATED WORK

Network telemetry enables operators to monitor network performance such as latency, loss, and congestion in real time, supporting functions such as fault diagnosis and adaptive control. Telemetry systems can operate in two principal modes: *in-band*, where measurement metadata is embedded within live packets as they traverse the network, and *out-of-band*, where probes are generated separately and sent to collectors [9].

In 5G systems, most telemetry operations were concentrated at the network core. In contrast, 6G extends observability to a large number of edge and far-edge nodes, significantly increasing the scale and heterogeneity of telemetry data sources while intensifying resource constraints at the edge [1], [2]. Efficient and adaptive telemetry activation therefore becomes a fundamental design requirement for 6G.

A. Placement Strategies for Telemetry Nodes

Recent research on telemetry placement has evolved toward more flexible and context-aware strategies. Examples include INT-Label, which distributes labeling to reduce metadata [3], INTO, which applies topology heuristics [4], and DeltaINT, which reports only on local state changes [5]. Tomography-based selection [10] and other static designs perform well in stable networks but cannot adapt at runtime.

To move beyond static and topology-based placement, intent-driven approaches have emerged to make telemetry activation more adaptive and goal-oriented. Sel-INT enables programmable switches to insert telemetry selectively [11], ITU-T FG-AN promotes intent-based orchestration for objectives such as anomaly detection and SLA assurance [6], and Newton translates high-level intents into network-wide monitoring configurations [12]. While these designs improve flexibility, they typically assume a fixed set of active nodes, leaving open the question of where activation should occur. Prior work such as INT-balance [7] optimizes telemetry path planning to balance monitoring coverage and overhead, while NetView [13] orchestrates telemetry collection and visualization through controller coordination. These systems improve telemetry execution but not activation.

ARTA addresses the complementary problem of *where* telemetry should be activated, a challenging decision task due to the scale and coupling of 6G edge systems. It employs submodular optimization, which offers a principled basis for node selection via diminishing returns and greedy maximization. Prior work applied similar methods to sensor placement [8], online selection [14], and dynamical systems [15], but typically under static or centralized settings. ARTA generalizes these principles to redundancy-aware, adaptive activation with provable scalability guarantees.

B. Benefit–Cost and Design Factors in Telemetry

The usefulness of activating telemetry at a node depends on multiple factors. Prior studies emphasize individual dimensions: DeltaINT reports only when local states change beyond

a threshold, capturing anomaly likelihood [5]; AdapINT uses deep reinforcement learning to deploy probes under latency and bandwidth limits, highlighting traffic importance [16]; network tomography selects vantage points to maximize observability, reflecting topological centrality [10]; and intent-driven systems such as Newton and Sel-INT adapt telemetry to situational needs, emphasizing context relevance [11], [12]. These approaches optimize one metric or use heuristic learning rather than a unified optimization model. In contrast, ARTA jointly considers all four factors through a submodular optimization framework that captures their trade-offs and redundancy interactions.

Alongside these benefits, telemetry also incurs costs. Systems such as INTO and DeltaINT introduce extra probes and headers that consume bandwidth and may congest user traffic [4], [5]. Large-scale monitoring platforms like Pingmesh demonstrate how uncontrolled telemetry streams can overload data collectors [17]. On-device monitoring further stresses limited memory and compute resources, as shown in systems such as OpenSketch and lightweight telemetry designs for constrained edge environments [18], [19].

While a few studies combine cost and benefit factors, most address them separately. In contrast, ARTA provides a unified model that jointly optimizes benefit, cost, and redundancy, enabling adaptive and resource-efficient telemetry activation for large-scale 6G edge networks.

III. SYSTEM MODEL AND PROBLEM FORMULATION

This section presents the system model and optimization framework underlying ARTA. Telemetry-related signals used to derive benefit and cost are treated as configurable inputs inspired by existing monitoring systems, while the ARTA formulation, the explicit modeling of inter-node redundancy, and the resulting submodular optimization problem constitute the original contributions of this work. Telemetry activation is formulated as a decision problem in which a subset of candidate nodes is selected to maximize monitoring benefit under resource constraints. Figure 1a illustrates how ARTA favors high-benefit, moderate-cost nodes, achieving balanced and efficient activation across the network.

A. Network Model

The network is represented as a graph $G = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} denotes the set of candidate telemetry nodes and \mathcal{E} the set of communication links among them. The node set \mathcal{V} may include backbone, edge, and far-edge elements of the 6G system, while \mathcal{E} includes all real or virtual links over which telemetry information is transmitted between nodes.

Each node $i \in \mathcal{V}$ is associated with a *benefit score* B_i , capturing its monitoring value, and an *activation cost* C_i , representing resource consumption. Both parameters are derived from telemetry-related metrics obtained from passive measurements or historical statistics, consistent with prior work [3]–[5], [17], [18], [20]. These parameters are treated as inputs to ARTA. Table I summarizes the main notation used in the formulation.

The main decision variables of ARTA indicate whether telemetry is activated at each candidate node. Formally,

Decision variable:

$$x_i = \begin{cases} 1, & \text{if telemetry is activated at node } i, \\ 0, & \text{otherwise.} \end{cases}$$

The binary vector $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$ represents the activation state of all candidate telemetry nodes. The parameters and activation decisions can be revisited periodically, depending on network dynamics and controller policies, and ARTA operates efficiently enough to update these decisions even at fine time scales. Activation decisions are computed centrally by a network controller and applied across distributed edge and core nodes.

Constraint: Activating telemetry at many nodes can overwhelm network and device resources [3], [5], [17], [21]. To prevent such overload, we impose a cardinality budget on the number of simultaneously active telemetry nodes:

$$\sum_{i \in \mathcal{V}} x_i \leq K, \quad (1)$$

where K is the maximum number of simultaneously activated nodes (e.g., 20% of edge devices). Such constraints are common in network utility maximization, network design, and sensor placement to prevent overload while preserving tractability [22]–[24].

B. Benefit and Redundancy-Adjusted Benefit Function

As defined in the network model, each node i is associated with a benefit score B_i that quantifies the monitoring value of activating it. The benefit is expressed as four normalized components capturing different aspects of telemetry value:

$$B_i = \sum_{k=1}^4 \lambda_k s_{i,k}, \quad (2)$$

where $s_{i,k} \in [0, 1]$ are feature scores and λ_k are their non-negative weights (defaulting to 0.25 each):

- Anomaly likelihood: Probability that a node experiences or is about to experience faults, inferred from standard counters such as queue occupancy or packet drops, without requiring prior telemetry [5], [21], [25].
- Traffic importance: Criticality of flows handled by the node, derived from QoS or SLA descriptors.
- Topological centrality: Structural importance, measured by path coverage or connectivity, indicating its effectiveness as a monitoring point.
- Context relevance: Situational factors that increase monitoring value, e.g., V2X segments during rush hours or boundary nodes in cross-domain settings.

In dense 6G edge networks, activating telemetry on nearby nodes often yields correlated measurements, such as nearly identical latency or queue statistics across connected switches [3], [5], [21]. These redundant probes consume bandwidth and collector resources without improving visibility.

TABLE I
SUMMARY OF NOTATION

Symbol	Definition
\mathcal{V}	Set of candidate telemetry nodes
$S \subseteq \mathcal{V}$	Activated subset of nodes
B_i	Baseline benefit score of node i
C_i	Activation cost of node i
$J_i(\mathbf{x})$	Redundancy-adjusted benefit of node i
$J(S)$	Set-level objective function for selected nodes S
w_{ij}	Redundancy weight (similarity) between nodes i and j
$\mathcal{N}(i)$	Set of nodes within a limited hop distance from node i
x_i	Activation decision for node i
λ_k	Weight of benefit component k
$s_{i,k}$	Normalized score of benefit component k for node i
γ_c	Weight of cost component c
$q_{i,c}$	Normalized score of cost component c for node i
β	Redundancy sensitivity parameter
α	Weight for benefit–cost trade-off
K	Activation budget (maximum nodes that can be active)

To address this inefficiency, ARTA models redundancy by assigning weights $w_{ij} \in [0, 1]$ for all $j \in \mathcal{N}(i)$, where $\mathcal{N}(i)$ denotes the set of nodes within a limited hop distance from node i . The weight w_{ij} decreases with hop distance, reflecting weaker redundancy between more distant nodes. Such correlation weights have also been adopted in prior INT and network monitoring studies [3], [4], where w_{ij} can be estimated from shared traffic flows, overlapping path segments, or statistical similarity among monitored data streams.

The redundancy-aware benefit of node i under an activation vector \mathbf{x} is defined as

$$J_i(\mathbf{x}) = B_i - \beta \log\left(1 + \sum_{j \in \mathcal{N}(i)} w_{ij} x_j\right), \quad (3)$$

where $\beta \in [0, 1]$ controls redundancy sensitivity. The logarithmic term captures diminishing returns from correlated activations, balancing observability and scalability in dense 6G edge networks.

C. Cost Function

Complementing the benefit score, the cost C_i is defined as the resource burden of enabling telemetry at node i . It is expressed as a weighted sum of three normalized components:

$$C_i = \sum_{c=1}^3 \gamma_c q_{i,c}, \quad (4)$$

where $q_{i,c} \in [0, 1]$ are normalized scores for each cost component, and γ_c are non-negative weights (defaulting to equal values). The three representative components are:

- Bandwidth Cost: additional link capacity consumed by probes or metadata.
- Collector/Processing Cost: load imposed on back-end collectors and analytics systems.
- On-Device Resource Cost: memory and compute usage at telemetry-capable nodes.

Higher cost values indicate limited resource availability or temporary resource stress at a node, which naturally reduces the likelihood of activating telemetry there. These dimensions

capture the main sources of overhead while remaining flexible, allowing additional factors (e.g., energy consumption) to be incorporated as needed.

D. Optimization problem

The telemetry activation problem is formalized by combining the redundancy-aware benefit and the activation cost into a single optimization objective:

$$\begin{aligned} \max_{\mathbf{x} \in \{0,1\}^{|\mathcal{V}|}} \quad & \sum_{i \in \mathcal{V}} J_i(\mathbf{x}) x_i - \alpha \sum_{i \in \mathcal{V}} C_i x_i, \\ \text{s.t.} \quad & \sum_{i \in \mathcal{V}} x_i \leq K, x_i \in \{0,1\}, \forall i \in \mathcal{V}. \end{aligned} \quad (5)$$

This formulation integrates monitoring utility, redundancy, and resource constraints, framing the problem as a submodular maximization under a cardinality constraint. The greedy algorithm provides the classical $(1 - 1/e)$ guarantee [26], offering a tractable and theoretically grounded approach for selective telemetry activation in 6G edge networks. This is a challenging problem, since each activation decision $x_i = 1$ depends on its benefit $J_i(x)$, which in turn varies with other activated nodes. Next, we analyze its properties and develop a scalable solution.

E. Complexity, Submodularity, and Greedy Solution

The computational and structural properties of the telemetry activation problem are analyzed under the logarithmic redundancy penalty. To examine complexity, the objective is rewritten in equivalent set-function form [8], following standard submodular optimization notation. Let $S \subseteq \mathcal{V}$ denote the subset of activated nodes corresponding to the binary vector \mathbf{x} , where $x_i = 1$ if and only if $i \in S$. The benefit term is expressed as $J(S)$, while the complete objective also includes the cost and the cardinality constraint $|S| \leq K$. This function is submodular, defined as follows:

Definition 1 (Submodularity). *A set function $J : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ is submodular if for all $A \subseteq B \subseteq \mathcal{V}$ and $x \in \mathcal{V} \setminus B$,*

$$J(A \cup \{x\}) - J(A) \geq J(B \cup \{x\}) - J(B).$$

The redundancy-adjusted benefit can then be written as:

$$\max_{S \subseteq \mathcal{V}} J(S) = \sum_{i \in S} B_i - \beta \sum_{i \in S} \log \left(1 + \sum_{j \in S \setminus \{i\}} w_{ij} \right),$$

subject to a cardinality constraint $|S| \leq K$.

Lemma 1 (Computational Hardness). *The telemetry activation problem is NP-hard.*

Proof. Maximizing a nonnegative submodular function under a cardinality constraint is NP-hard [26]. The logarithmic redundancy term couples node selections, making the objective non-separable and combinatorial. Therefore, solving it exactly requires exploring an exponential number of activation subsets, and the problem is NP-hard. \square

Lemma 2 (Submodularity of J). *The utility function $J(S)$ is submodular. Consequently, the greedy algorithm used in ARTA*

Algorithm 1 Redundancy-Aware Greedy Algorithm

```

1:  $S_t \leftarrow \emptyset, U \leftarrow \mathcal{V}$   $\triangleright$  initialize selected and candidate sets
2: while  $|S_t| < K$  and  $U \neq \emptyset$  do
3:   for all  $i \in U$  do
4:      $\Delta_i(S_t) \leftarrow B_i - \beta \log \left( 1 + \sum_{j \in S_t} w_{ij} \right)$ 
5:      $\text{UTIL}(i) \leftarrow \Delta_i(S_t) - \alpha C_i$ 
6:    $p \leftarrow \arg \max_{i \in U} \text{UTIL}(i)$ 
7:   if  $\text{UTIL}(p) > 0$  then
8:      $S_t \leftarrow S_t \cup \{p\}$   $\triangleright$  add the most valuable node
9:    $U \leftarrow U \setminus \{p\}$ 
10: return  $S_t$   $\triangleright$  final set of activated telemetry nodes
```

achieves the classical $(1 - 1/e)$ approximation ratio to the optimal solution under the cardinality constraint [26].

Proof. For $S \subseteq \mathcal{V}$ and $x \notin S$, the redundancy penalty of adding x is

$$\Delta_S(x) = B_x - \beta \log \left(1 + \sum_{j \in S} w_{xj} \right).$$

Since $\log(1 + z)$ is concave and $\sum_{j \in S} w_{xj}$ is modular, their composition is concave-over-modular, which implies submodularity [8], [27]. \square

Proposition 1 (Sufficient Monotonicity). *If for every $x \in \mathcal{V}$ and $S \subseteq \mathcal{V} \setminus \{x\}$ it holds that*

$$B_x \geq \beta \log \left(1 + \sum_{j \in S} w_{xj} \right), \quad \text{then } J \text{ is monotone.}$$

Since the problem is NP-hard but submodular, ARTA uses a greedy selection that activates the node with the highest marginal utility, defined as redundancy-penalized benefit minus weighted cost. Greedy algorithms yield near-optimal solutions for submodular maximization under cardinality constraints with linear complexity per iteration [26].

The method is summarized in Algorithm 1. It begins with an empty selected set S_t and a candidate pool U (Line 1). While the number of selected nodes is below the upper limit K and candidates remain (Line 2), the algorithm computes for each candidate $i \in U$ its redundancy-adjusted gain $\Delta_i(S_t)$ and marginal utility $\text{UTIL}(i) = \Delta_i(S_t) - \alpha C_i$ (Lines 3–4). The node p with the highest utility is identified (Line 5) and added to S_t if it provides a positive improvement (Lines 6–7). The process repeats, removing each chosen node from the pool until the budget K is reached (Lines 8–9). The final output is the selected set S_t of active telemetry nodes (Line 10).

IV. EVALUATION

ARTA is evaluated on hierarchical edge-core topologies generated with NetworkX, where dense edge clusters connect to few core nodes, reflecting typical 6G architectures [1], [2]. Candidate nodes represent telemetry-capable devices and lightweight edge servers. Each experiment is repeated with different random seeds, and results are averaged across runs.

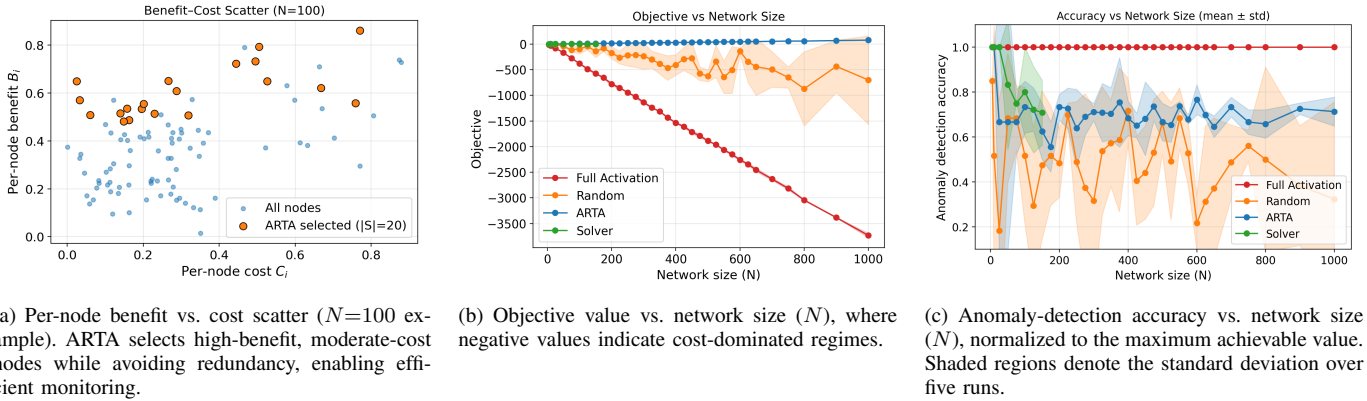


Fig. 1. ARTA performance compared with baselines across objective, accuracy, and selection.

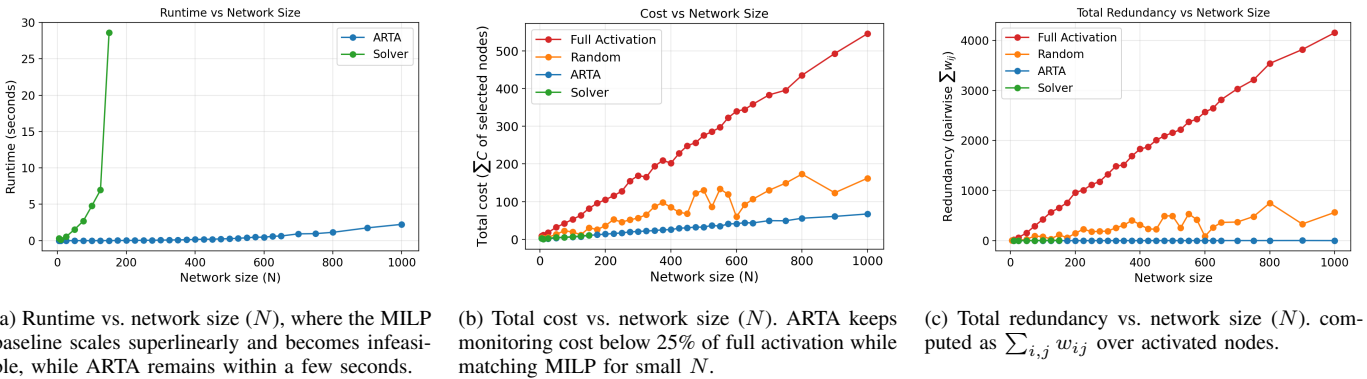


Fig. 2. ARTA compared with baseline methods across runtime, cost, and redundancy metrics.

Randomized parameters and multiple seeds are intentionally used to avoid bias toward specific topologies or traffic patterns and to evaluate ARTA under diverse and representative operating conditions.

Redundancy weights w_{ij} are randomly generated in the range $[0, 1]$, where higher values represent stronger correlation between nearby nodes. ARTA selects nodes by their marginal benefit-to-cost ratio while accounting for redundancy. The benefit combines four normalized components: anomaly likelihood, traffic importance, topological centrality, and context relevance, each with equal weights ($\lambda_k = 0.25$). The activation cost aggregates bandwidth, collector, and on-device usage with uniform weights ($\gamma_c = 1/3$). Equal weights ensuring neutral contribution among factors for both benefit and cost [4], [5].

The default parameters are set to $\alpha = 0.7$, $\beta = 0.2$, and $K = 0.2|\mathcal{V}|$. These values were empirically chosen to balance accuracy, cost, and redundancy across network sizes. We also verified that small variations ($\pm 20\%$) in α or β did not affect the overall performance trends, confirming that ARTA’s results are robust to parameter changes. Evaluation metrics include accuracy, cost, redundancy, and runtime scalability.

The evaluation compares ARTA with three baselines: (i) *Full activation*, which enables telemetry at all nodes and achieves maximum coverage at prohibitive cost [3]–[5]; (ii)

Random selection, which activates K nodes uniformly at random and serves as a standard benchmark in sensor placement studies [8]; and (iii) *Solver-based optimization (MILP BASELINE)*, implemented in CVXPY [28] with the CBC solver, computes the exact optimum of the benefit–cost formulation for small topologies due to its high complexity.

Figure 1b reports the objective value as network size increases. ARTA matches the MILP baseline on small topologies and maintains near-optimal performance as the network scales, while random selection exhibits high variance and full activation degrades due to accumulated cost and redundancy. Shaded areas denote the standard deviation over five runs.

For accuracy evaluation, 5–10% of nodes are randomly marked as anomalous, and a node is considered correctly detected if it belongs to the activated set covering an anomalous node. Figure 1c reports the normalized accuracy, where 1 denotes the maximum achievable value. ARTA achieves 0.8 accuracy (80% of the maximum), remaining stable across network sizes, while random selection fluctuates due to uncoordinated activation. Full activation yields perfect accuracy, and the MILP baseline matches ARTA on small networks but becomes infeasible as N grows.

The MILP baseline uses a piecewise-linear approximation of the logarithmic redundancy penalty. Figure 2a shows that

its runtime grows superlinearly with N , becoming infeasible beyond 200 nodes, while ARTA remains within a few seconds at 1,000 nodes, confirming its scalability.

Figure 2b and Fig. 2c report the monitoring cost and total redundancy, respectively. As network size increases, full activation quickly becomes inefficient, while random selection incurs high cost due to uncoordinated activation. ARTA maintains low cost by selecting informative nodes, matching the MILP baseline on small topologies before the latter becomes infeasible. Redundancy exhibits the same qualitative behavior across methods.

Increasing the activation budget K from 10–30% improves accuracy as expected, without changing the relative performance trends, demonstrating ARTA’s robustness to parameter choice. At $N = 1,000$ nodes, ARTA achieves about 0.78 accuracy with less than 25% of the cost of full activation and near-zero redundancy, confirming its efficiency at scale.

V. CONCLUSION AND FUTURE WORK

This paper presented ARTA, an Adaptive Redundancy-Aware Telemetry Node Activation framework for efficient observability in 6G edge networks. By formulating telemetry activation as a redundancy-aware benefit–cost optimization, ARTA balances monitoring accuracy with resource efficiency and scales effectively to large topologies. Simulations show that ARTA achieves ~ 0.8 normalized accuracy (80% of the maximum achievable accuracy) while activating only 20% of nodes and maintaining minimal redundancy. These results demonstrate the practicality of selective activation for resource-efficient telemetry in next-generation networks. Future work will extend ARTA toward online and dynamic settings, including highly mobile vehicular and IoT scenarios, and integrate adaptive learning and intent-driven orchestration to enable self-healing and autonomous telemetry.

ACKNOWLEDGMENT

This research was supported by the National Growth Fund through the Dutch 6G flagship project *Future Network Services*, and by the European Commission through Grant No. 101139270 (ORIGAMI) and Grant No. 101192462 (FLECON-6G).

REFERENCES

- [1] Hexa-X Project Consortium, “Hexa-x deliverable d1.1: 6g vision, use cases and key societal values,” Tech. Rep., 2021.
- [2] NGMN Alliance, “6g requirements and design considerations,” Tech. Rep., 2022.
- [3] E. Song, T. Pan, C. Jia, W. Cao, J. Zhang, T. Huang, and Y. Liu, “Int-label: Lightweight in-band network-wide telemetry via interval-based distributed labelling,” in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021.
- [4] J. A. Marques, M. C. Luizelli, R. I. T. da Costa Filho, and L. P. Gaspar, “An optimization-based approach for efficient network monitoring using in-band network telemetry,” *Journal of Internet Services and Applications*, vol. 10, pp. 1 – 20, 2019.
- [5] J. Sheng, Q. Huang, and P. P. C. Lee, “Deltaint: Toward general in-band network telemetry with extremely low bandwidth overhead,” in *2021 IEEE 29th ICNP*, 2021.
- [6] ITU-T Focus Group on Autonomous Networks (FG-AN), “Artificial intelligence for autonomous networks,” International Telecommunication Union, Tech. Rep., 2022.
- [7] Y. Zhang, T. Pan, Y. Zheng, E. Song, J. Liu, T. Huang, and Y. Liu, “Int-balance: In-band network-wide telemetry with balanced monitoring path planning,” in *ICC 2023 - IEEE*, 2023.
- [8] A. Krause, A. Singh, and C. Guestrin, “Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies,” *J. Mach. Learn. Res.*, vol. 9, Jun. 2008.
- [9] L. Tan, W. Su, W. Zhang, J. Lv, Z. Zhang, J. Miao, X. Liu, and N. Li, “In-band network telemetry: A survey,” *Computer Networks*, vol. 186, p. 107763, 2021.
- [10] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, “Network tomography: Recent developments,” *Statistical Science*, vol. 19, no. 3, Aug 2004.
- [11] S. Tang, D. Li, B. Niu, J. Peng, and Z. Zhu, “Sel-int: A runtime-programmable selective in-band network telemetry system,” *IEEE TNSM*, vol. 17, no. 2, 2020.
- [12] Z. Xi, Y. Zhou, D. Zhang, K. Gao, C. Sun, J. Cao, Y. Wang, M. Xu, and J. Wu, “Newton: Intent-driven network traffic monitoring,” *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, 2022.
- [13] Y. Lin, Y. Zhou, Z. Liu, K. Liu, Y. Wang, M. Xu, J. Bi, Y. Liu, and J. Wu, “Netview: Towards on-demand network-wide telemetry in the data center,” in *ICC 2020 - 2020 IEEE ICC*, 2020.
- [14] D. Golovin, M. Faulkner, and A. Krause, “Online distributed sensor selection,” in *Proceedings of the 9th ACM/IEEE IPSN*, ser. IPSN ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 220–231.
- [15] T. H. Summers, F. L. Cortesi, and J. Lygeros, “On submodularity and controllability in complex dynamical networks,” *IEEE Transactions on Control of Network Systems*, vol. 3, no. 1, pp. 91–101, 2016.
- [16] P. Zhang, H. Zhang, Y. Pi, Z. Cao, J. Wang, and J. Liao, “Adapint: A flexible and adaptive in-band network telemetry system based on deep reinforcement learning,” *IEEE TNSM*, vol. 21, no. 5, 2024.
- [17] C. Guo, L. Yuan, D. Xiang, Y. Dang, R. Huang, D. Maltz, Z. Liu, V. Wang, B. Pang, H. Chen, Z.-W. Lin, and V. Kurien, “Pingmesh: A large-scale system for data center network latency measurement and analysis,” vol. 45, no. 4. Association for Computing Machinery, Aug. 2015.
- [18] M. Yu, L. Jose, and R. Miao, “Software defined traffic measurement with opensketch,” in *Proceedings of the 10th USENIX Conference on NSDI*, ser. nsdi’13. USENIX Association, 2013.
- [19] M. Otero, J. M. García, and P. Fernandez, “An extensible lightweight framework for distributed telemetry of microservices,” *Sustainable Computing: Informatics and Systems*, vol. 46, 2025.
- [20] M. Otero, J. M. Garcia, and P. Fernandez, “Towards a lightweight distributed telemetry for microservices,” in *2024 IEEE 44th ICDCSW*, 2024.
- [21] Q. Huang, H. Sun, P. P. C. Lee, W. Bai, F. Zhu, and Y. Bao, “Omnimon: Re-architecting network telemetry with resource efficiency and full accuracy,” in *Proc. of ACM SIGCOMM*, ser. SIGCOMM ’20. Association for Computing Machinery, 2020, p. 404–421.
- [22] Y. Bi, C. W. Tan, and A. Tang, “Network utility maximization with path cardinality constraints,” in *IEEE INFOCOM 2016*, 2016.
- [23] C. Sun, R. Dai, and M. Mesbahi, “Weighted network design with cardinality constraints via alternating direction method of multipliers,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, 2018.
- [24] N. Karnik, M. G. Abdo, C. E. Estrada-Perez, J. S. Yoo, J. J. Cogliati, R. S. Skifton, P. Calderoni, S. L. Brunton, and K. Manohar, “Constrained optimization of sensor placement for nuclear digital twins,” *IEEE Sensors Journal*, vol. 24, no. 9, 2024.
- [25] H. Sun, Q. Huang, P. P. C. Lee, W. Bai, F. Zhu, and Y. Bao, “Distributed network telemetry with resource efficiency and full accuracy,” *IEEE/ACM Transactions on Networking*, vol. 32, no. 3, 2024.
- [26] G. Nemhauser, L. Wolsey, and M. Fisher, “An analysis of approximations for maximizing submodular set functions—i,” *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 12 1978.
- [27] R. Iyer and J. Bilmes, “Concave aspects of submodular functions,” in *2020 IEEE ISIT*, 2020.
- [28] S. Diamond and S. Boyd, “Cvxpy: a python-embedded modeling language for convex optimization,” *J. Mach. Learn. Res.*, vol. 17, no. 1, Jan. 2016.