

Edge AI를 위한 데이터 분산 서비스(DDS) 아키텍처 분석 연구

김재우, 이재민, 김동성*

ICT융합특성화연구센터, *금오공과대학교

{jaewookim, ljmpaul, dskim*}@kumoh.ac.kr

An Analytical Study of Data Distribution Service (DDS) Architectures for Edge AI

Jae Woo Kim, Jae Min Lee, *Dong Seong Kim

ICT convergence Research Center, * Kumoh National Institute of Technology

요약

본 논문은 실시간 데이터 공유 표준인 DDS(Data Distribution Service)를 Edge AI 환경에 적용하기 위한 최적의 아키텍처를 제안한다. 표준 DDS, ROS2의 RMW 계층, 그리고 자원 제약 기기를 위한 DDS-XRCE의 구조적 장단점을 비교 분석하였다. 연구 결과, Edge AI의 가변적인 네트워크 환경과 자원 제약을 극복하기 위해 계층형 멀티 티어 구조와 적응형 QoS(Quality of Service) 로직을 결합한 'HE-DDS(Hybrid Edge-DDS)' 구조를 설계하였다. 제안된 구조는 발견(Discovery) 트래픽을 90% 이상 절감하며, 제로 카피(Zero-Copy) 기술을 통해 AI 추론 데이터의 전송 지연을 최소화한다.

1. 서론

현대 분산 시스템은 데이터 중심(Data-Centric) 패러다임으로 전환되고 있다. DDS는 OMG(Object Management Group)에서 제정한 통신 표준으로 기기 간 실시간 데이터 교환을 보장한다. DDS는 RTPS 프로토콜을 통해 브로커 없는 P2P 통신, 동적 discovery, 다양한 QoS 정책(신뢰성, 우선순위, 대역폭 등)을 제공함으로써 저지연 통신을 가능케 한다[1]. 이러한 장점으로 DDS는 다양한 애플리케이션에서 활용되어 왔으며, 로봇틱스 및 국방 플랫폼과 같은 서로 다른 도메인의 요구사항을 반영하여 표준 DDS, ROS2(Robot Operating System 2) DDS, DDS-XRCE 등으로 확장·특화되어 발전해 왔다[2][3]. 엣지 AI는 클라우드 전송 없이 로컬에서 즉각적인 추론을 수행해야 하므로 50~200ms의 네트워크 지연을 제거하는 것이 필수적이다. 그러나 엣지 기기는 전력, 메모리, 컴퓨팅 성능이 제한적이다. 따라서 미들웨어는 초저지연 성능을 유지하면서도 낮은 자원 점유율(Memory Footprint)을 가져야 하며, 네트워크의 가변적인 품질에 대응할 수 있는 신뢰성 정책을 지원해야 한다. 본 논문에서는 표준 DDS와 ROS2에 구현된 DDS, 그리고 DDS-XRCE 아키텍처의 구조적 원리와 장단점을 심층적으로 분석한다. 이를 바탕으로 엣지 AI 환경에서 요구되는 DDS의 핵심 요구사항을 도출하고, 가변적인 네트워크 품질, 모델 최적화에 따른 데이터 특성의 변화, 그리고 실시간 결정성을 동시에 만족할 수 있는 새로운 DDS 아키텍처 설계안을 제안한다.

II. 본론

1. 기존 DDS 아키텍처 분석

표준 DDS는 데이터 중심 발행-구독(Publish-Subscribe)모델을 통해 응용 프로그램 간의 결합도를 낮추고 데이터 공유 효율성을 극대화한다. DDS는 도메인(Domain) 개념을 통해 통신 범위를 논리적으로 구분하고, 각 도메인 내에서 자동 디스커버리(discovery) 기능으로 참가자와 토픽을 탐지한다. DDS의 이점은 통신에 참여하는 응용 프로그램들이 서로 존재 여부를 알 필요가 없는 공간적, 시간적 분리에 있다. 미들웨어가 메시지 전달, 수신자 위치 확인, 전달 실패 시의 처리 등을 투명하게 관리하므로 개발자의 구현 효율성을 제공한다. 그림1은 표준 DDS 라이브러리 아키텍처와 ROS2의 아키텍처다. ROS 2는 로봇 애플리케이션을 위한 프레임워크로, DDS를 통신 엔진으로 활용하고 RMW(ROS Middleware)라는 추상화 계층을 제공한다. ROS 2의 노드, 토픽, 서비스 등은 내부적으로

DDS의 도메인으로 매핑되지만, RMW 계층이 이를 자동 설정 및 관리하여 사용자는 DDS의 복잡한 설정(도메인 ID, 파티션, IDL 메시지 생성 등)을 직접 다루지 않고도 통신을 구현할 수 있다. 이러한 추상화로 개발 편의성은 높아졌으나 추상화 계층으로 인해 발생하는 오버헤드로 표준 DDS 통신에 비해 약 50%의 지연 시간이 발생한다. 오버헤드의 주요 원인은 Rclcpp 알림 지연(Notification Delay)으로, 메시지가 수신인지 시점부터 실제 응용 프로그램이 데이터를 가져가기까지 걸리는 시간이다.

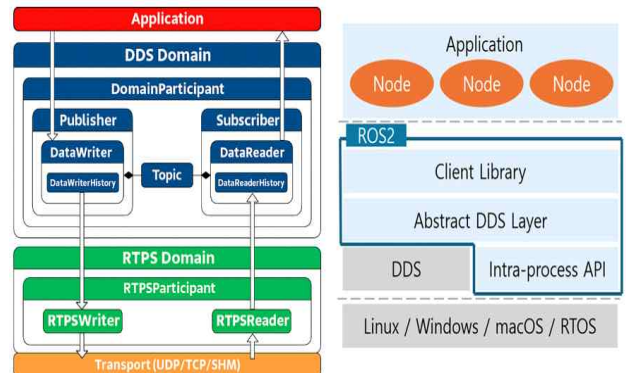


그림 1 표준 DDS와 ROS2 DDS 구조

또한, 고해상도 페이로드와 같은 대용량 데이터 전송 시 단편화 문제로 성능 저하도 중요한 이슈이다. 페이로드가 단편화 임계치를 초과할 때 지연 시간이 급격히 증가하며, 이는 무선 링크 기반의 모바일 로봇이나 드론 시스템에서 통신 실패의 원인이 되기도 한다. 한편 Edge AI 시스템은 고성능 서버뿐만 아니라 센서 데이터를 수집하고 말단 제어를 수행하는 저전력 마이크로컨트롤러(MCU)를 포함한다. 일반적인 DDS는 수 MB 이상의 메모리를 요구하므로 수 KB의 RAM을 가진 MCU에는 적합하지 않다. 이를 해결하기 위해 OMG는 DDS-XRCE(DDS for eXtremely Resource Constrained Environments) 표준을 정의했다[3]. Micro-DDS는 메모리·연산 자원이 극도로 제한된 임베디드 기기에서도 DDS 네트워크에 참여할 수 있도록 고안되었다. 표준 DDS가 P2P 방식인 것과 달리, DDS-XRCE는 클라이언트-서버 (Client/Agent) 아키텍처를 취한다. 즉, 마이크로 디바이스 상에서는 경량 XRCE 클라이언트가 동작하며, 이 클라이언트는 네트워크 상의 XRCE 에이전트(Agent) 서버와 통신하면서 DDS 도메인에 간접 참여한다. 이 구조의 핵심 장점은 클라이언트가 상대

정보를 유지할 필요가 없는 상태 비저장(Stateless) 시스템으로 운영될 수 있다는 점이다. 이를 통해 MCU 기기는 통신이 없는 시간 동안 슬립 모드로 진입하여 전력을 절감할 수 있고, 에이전트가 브릿지 역할을 수행하므로 연결이 끊어져도 DDS 네트워크 상에서의 존재는 유지된다.

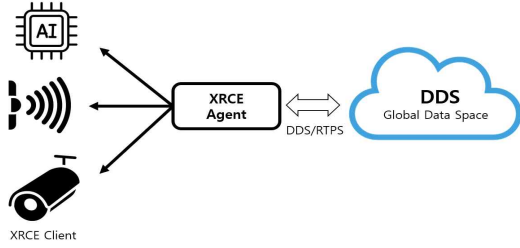


그림 2 DDS-XRES 운용 구조

2. DDS 기반 엣지 AI의 요구 조건

엣지 AI 시스템에서 DDS 미들웨어가 충족해야 할 요구 사항은 다음과 같다.

- (1) 저지연 실시간 통신 : 엣지 AI의 추론 결과나 제어 명령은 실시간 응답성이 요구되므로, 네트워크 왕복 시간을 최소화하며 근거리 통신과 지역 처리를 활용하여 지연을 줄이는 아키텍처가 필요하다.
- (2) 경량 메모리/CPU 사용 : 엣지 디바이스들은 제한된 자원으로 동작하므로, 미들웨어 자체의 메모리 Footprint와 연산 오버헤드를 낮춰야 한다.
- (3) 통신 신뢰성과 품질 : 분산 AI 모델 동기화나 로봇 제어와 같이 중요한 데이터는 손실 없이 전달되어야 하므로 신뢰성 QoS(Reliable)와 QoS 프로파일 튜닝이 필수적이다. 또한, 대역폭에 따라 일부 덜 중요한 스트림은 베스트 에포트로 전송하여 하는 등 QoS를 조절할 수 있어야 한다.
- (4) 확장성과 Discovery 효율 : 엣지에는 수십~수백개의 센서/기기가 참여할 수 있는데, DDS의 디스커버리 트래픽이 증가하면 네트워크 과부하와 연결 지연을 초래한다. 따라서 노드 수 증가에 따라 효율적인 discovery 메커니즘이 요구된다. 계층적 discovery나 필터링 기법 등 노드 탐색 범위를 제한하거나 트래픽을 감축해야 한다.
- (5) 계층 구조 및 로컬 처리 : 엣지 연산의 이점은 데이터를 발생한 근처에서 처리함으로써 응답 시간을 줄이고 백홀 트래픽을 감소시키는 데 있다. 이를 지원하려면 미들웨어가 단일 평면 구조보다는 말단(End)-엣지-클라우드 구조로 계층화되어 데이터 흐름을 분리하는 편이 유리하다.

3. 제안하는 Edge AI 최적화 DDS 구조(HE-DDS)

본 연구에서는 엣지 컴퓨팅의 효율성을 극대화하기 위해 다음과 같은 하이브리드 엣지 DDS(HE-DDS) 구조를 설계한다. 제안하는 HE-DDS 구조는 Edge AI 최적화를 위해 3가지 특징을 가진다.

- (1) 계층형 멀티티어 구조 : 제안하는 최적화 구조는 시스템의 자원량에 따라 End, Edge, Cloud 계층으로 통신 모델을 분리한다. End-Tier는 센서/액추에이터로 구성되며 DDS-XRES 클라이언트를 사용하여 전력 소모를 최소화한다. 엣지 티어 (Edge-Tier)는 현장에 배치된 엣지 서버, 엣지 게이트웨이, 또는 MEC 서버들로 구성되며, End 디바이스와 Cloud를 연결하는 중간 허브 역할을 한다., 다른 엣지 노드와는 표준 P2P DDS로 통신하여 실시간성을 확보한다. 클라우드 티어 (Cloud-Tier): 대규모 모델 업데이트 및 전역 관리를 담당한다.

(2) 적응형 QoS 로직 : HE-DDS 구조는 적응형 QoS 로직(Adaptive QoS Logic, AQL) 엔진을 도입한다. 이는 고정된 QoS 대신 데이터의 긴급도와 네트워크 상태를 반영하기 위함이다. 우선순위 산출은 긴급도 (Urgency_Score)를 계산하여 장애물 알람 등은 'Reliable' 모드로, 단순 로그는 'Best-Effort' 모드로 동적 전환하도록 한다.

- (3) 데이터 평면 제로 카피 (Zero-Copy) 및 또한 동일 노드 내 AI 추론

파이프라인(예: 카메라 노드 → 객체 인식 노드) 간에는 공유 메모리 기반의 제로 카피 기술을 적용하여 지연 시간을 최대 98%까지 단축한다[4]. Zero-Copy는 데이터 전달 과정에서 발생하는 다중 메모리 복사를 제거하거나 최소화함으로써, 대용량 센서 데이터 기반의 Edge AI 파이프라인에서 지연·CPU 부하·메모리 대역폭 사용을 크게 줄이는 핵심 기법이다. 특히 동일 호스트 내에서 공유메모리 기반의 inter-process zero-copy를 적용하면 영상·LiDAR와 같은 대용량 데이터의 직렬화/복사 비용을 줄여 전체 지연과 tail latency를 완화할 수 있다.

- (4) Discovery 성능 최적화 : HE-DDS는 발견(Discovery) 및 메모리 최적화 구조를 가진다. 알려진 노드의 엔드포인트 데이터를 사전에 로딩하여 초기 부팅 시 발생하는 발견 패킷 폭증을 90% 이상 억제하고 MAD-DDS 기법을 적용하여 네트워크 상에 교환되는 발견 정보를 ब्लूप 필터로 요약, 메모리 소비를 절감한다[5].

III. 결론

본 논문에서는 표준 DDS 및 ROS2의 DDS, 그리고 DDS-XRES의 특징을 분석하고 고성능 제어 능력과 자원 효율성을 결합한 Edge AI 위한 DDS 통신 구조를 제안하였다. 분석 결과, ROS2 RMW의 추상화 오버헤드를 줄이기 위해서는 AI 추론 데이터에 특화된 제로 카피 메커니즘이 필수적이며, 자원 제약 기기를 DDS 생태계에 통합하기 위해 클라이언트-에이전트 모델이 가장 유효함을 확인하였다. 제안된 HE-DDS 구조는 가변적인 엣지 환경에서 안정적인 데이터 전송과 신속한 AI 판단을 지원함으로써, 국방 Edge AI 플랫폼 및 스마트 제조 분야의 신뢰성을 높일 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2026년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업과 기초연구사업과 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행된 연구임 (2018R1A6A1A03024003, IITP-2025-RS-2024-00438430)

참 고 문 헌

- [1] Pardo-Castellote, G. "OMG Data-Distribution Service (DDS): Architectural Overview". Proceedings of the 23rd IEEE International Conference on Distributed Computing Systems Workshops, pp 200-206, 2003.
- [2] Macenski, S., Foote, T., Gerkey, B., Lalancette, C., & Woodall, W. "Robot Operating System 2: Design, architecture, and uses in the wild". Science Robotics, Vol. 7, No. 66, eabm6074, 2022.
- [3] Object Management Group. "DDS For Extremely Resource Constrained Environments (DDS-XRCE) 1.0". OMG Formal Specification, 2020.
- [4] He, Yuankai, and Weisong Shi. "A Faster and More Reliable Middleware for Autonomous Driving Systems." arXiv preprint, pp 2510.11448, 2025.
- [5] Nwadiugwu, W. P., Kim, D. S., Ejaz, W., & Anpalagan, A. "MAD-DDS: Memory efficient automatic discovery data distribution service for large scale distributed control network". IET Communications, Vol. 17, No. 12, pp 1432-1446, 2023