협업 인식 시스템의 보안 강화를 위한 프레임워크 동향 분석

우겸혁, 이은규
인천대학교

{dnruagur1, eklee}@inu.ac.kr

Technical Advances in Security-Enhanced Frameworks for Collaborative Perception Systems

Kyeom-hyeok Woo and Eun-Kyu Lee
Incheon National Univ.

요 약

협업 인식 시스템은 협력형 자율주행의 안전성과 효율성을 높이는 핵심 기술이다. 그러나 공격자가 참여할 경우 전체 시스템의 성능을 저하시킬 수 있는 보안 취약점이 존재한다. 본 논문은 이러한 문제를 해결하기 위해 대표적인 협업 인식 프레임워크인 ROBOSAC 과 그 후속 연구들을 중심으로 기술 발전의 동향을 살펴본다.

Ⅰ. 서론

협력형 자율주행은 차량이 주변 차량이나 인프라와 정보를 공유하며 주행의 안전성과 효율성을 높이는 기술이다. 최근에는 한국도로교통공단에서 교통신호정보 KS 표준을 제정하면서 실시간 정보 교환의 중요성이 더욱 부각되고 있다. 특히 차량 간 공유되는 인식 정보는 자율주행의 핵심 요소로, 이를 기반으로 하는 협업 인식 시스템의 역할은 매우 중요하다.

협업 인식 시스템은 각 차량이 로컬 센서로 수집한 정보를 공유함으로써 인식 정확도를 높이는 방식이다. 하지만, 공격자가 협업 인식에 참여하여 전체 시스템의 성능을 저하시킬 수 있는 취약점이 존재한다. 이에 본 논문에서는 보안 취약성을 해결하기 위해 제안된 ROBOSAC 을 중심으로 다양한 기술적 개선 방안과 최근 연구 동향을 분석하고자 한다.

Ⅱ. 합의 기반의 신뢰성 있는 협업 인식 모델

ROBOSAC(ROBust cOllaborative SAmple Consensus)은 샘플링 기반 추정 알고리즘을 활용하여 협업 인식 시스템의 취약점을 해결하는 보안 프레임워크이다 [1]. 자아 차량은 해당 프레임워크를 통해 팀원들 중 협력자는 수용하고 공격자는 배제한다. 구체적으로, 자아 차량은 로컬 센서로 수집한 정보를 기반으로 개별 인식 결과를 생성한다. 이후 팀원들 중 일부를 샘플링하고 이들의 인식 정보를 수신한 뒤, 이를 융합하여 또 다른 인식 결과를 생성한다. 마지막으로 두 인식 결과 간의 일치 여부를 확인하여 합의를 검증한다. 합의가 성공하면 융합 결과를 최종 출력하고, 합의가 실패하면 공격자가 협업 인식에 참여했다고 간주하고 새로운 팀원들을 샘플링한다. 자아 차량은 이 과정을 주어진 예산 내에서 합의에 도달할 때까지 반복한다. 이는 그림 1 을 통해 확인 가능하다.

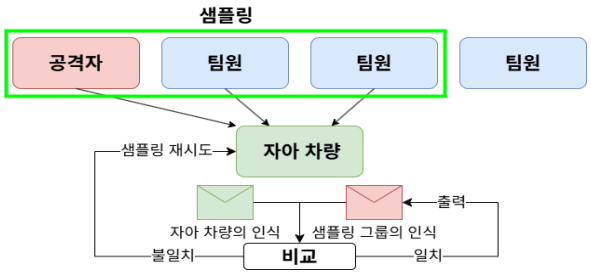

그림 1: ROBOSAC 구조도

ROBOSAC 은 두 가지 핵심 아이디어를 통해 신뢰 가능한 협력자를 파악한다. 샘플링 방식은 무작위 샘플 합의 기법을 사용한다. 이는 이상치가 많은 데이터셋에도 적용 가능한 추정 알고리즘이다. 또한, 환경마다 다른 공격자 비율을 추정하기 위해 공격적-보수적 프로빙(A2CP)을 사용한다. 이는 팀원의 인식 정보를 각각 분석하여 합의에 대한 기여도를 계산하고 팀원의 신뢰도를 측정한다.

적대적 공격에 강건한 ROBOSAC 도 한계가 존재한다. A2CP 로 추정한 공격자 비율은 고정적이지만, 실제 환경에서의 공격자 비율은 가변적이다. 또한, 이상 탐지 기반으로 동작하므로 우회 공격과 노이즈에 취약하다.

Ⅲ. 공격 탐지 정밀도 향상

ROBOSAC 의 성능을 향상시키기 위해 다양한 후속 연구가 진행되었다. CoMamba 는 차량 간 공유되는 인식 정보를 효율적으로 통합하는 방법을 제안한다 [2]. 어텐션 비사용 상태 공간 모델에 기반한 CSS2D 및 GPM 모듈을 통해 실시간 협업 인지를 실현한다. 또한 기존 접근법들은 협력자 수가 증가하면 자원 소모가 기하급수적으로 증가했지만, CoMamba 는 선형적인 자원 소모를 통해 우수한 확장성을 보였다. 또한, MADE 는 다중 테스트를 통해 공격자를 탐지하고 제거하는 반응형 탐지 기반 방어 메커니즘이다 [3]. MADE 는 허위 탐지를 제어하기 위해 자아 차량과 협력자 간의 인식 결과 일관성을 평가한다. 이를 위해 제시된 매치 손실과 협업 재구성 손실은 각각 출력과 중간 특징 수준의 일관성을 평가한다. 이후, 평가 결과를 기반으로 교정된 p 를 계산하고 BH 검정에서 이를 조정하여 허위 탐지 발생 확률을 낮춘다. 이 과정은 그림 2 로 확인할 수 있다.

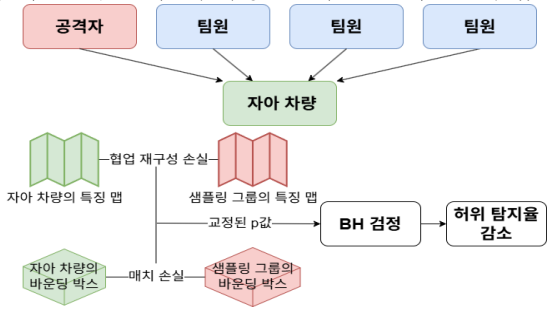

그림 2: MADE 구조도

한편, ROBOSAC 이 제시한 취약점을 보완하기 위한 연구도 존재한다. GCP 는 기존 가설-검증 기반 이상 탐지 방어 메커니즘을 우회하는 BAC 공격에 대응하기 위해 설계된 프레임워크이다[4]. GCP 는 신뢰도 기반 공간 일치 손실을 통해 공간적 정보를, LSTM-AE 기반의 시간적 조감도 흐름 재구성 기법을 통해 시간적 정보를 수집한다. 그리고 BH 검정을 적용하여 이들을 통합하고 분석한다. 실제 환경에서 발생하는 공격은 시간적 패턴을 가지기 때문에 시간적 정보를 활용하는 것은 방어에 중요한 요소로 작용한다. 또한, MRCNet 은 협업 인지 시스템에서 발생하는 노이즈 간섭을 완화하기 위해 제시된 통신 네트워크이다[5]. 자율주행 환경에서는 주로 차량이나 센서의 위치나 방향으로 인해 발생하는 자세 노이즈와 동적인 상황에서 발생하는 모션 블러가 인식 성능을 저하시킨다. 이를 해결하기 위해 제시된 다중 스케일 강건 융합은 다양한 스케일에서 특징을 융합하는 것으로 인식 성능을 개선하고 어텐션 기법을 통해 자세 노이즈의 영향을 줄인다. 그리고 모션 강화 기법은 과거의 프레임 정보와 함께 순환 유닛과 컨볼루션 레이어를 사용하여 모션 블러를 해결한다.

Ⅳ. 현실 적용 및 시뮬레이션

프레임워크를 실제로 적용하기 위해서는 모델 이질성을 해결해야 한다. 모델 이질성이란 협업 인식에 참여하는 차량 간의 입출력 모달리티, 센서 모달리티, 모델 아키텍처 등이 다른 것을 말한다. 이는 도메인 간 격차를 유발하여 협업 성능을 저하시킨다. HEAL 은 이를 해결하기 위해 통합된 특징 공간을 구축한다. 그리고 이질적인 차량이 협업에 참여할 때마다 역방향 정렬 메커니즘으로 로컬 학습을 수행시켜 통합된 특징 공간에 정렬시키는 방법으로 모델 이질성을 해결한다[6]. 또한, STAMP 는 가벼운 어댑터-리버터 쌍을 통해 이질적인 차량의 조감도 특징을 공통 프로토콜 조감도 도메인으로 변환한 뒤, 다른 차량들에게 전송한다. 이를 수신한 차량들은 해당 정보를 각자의 로컬 도메인으로 되돌려 이질성을 해결한다[7].

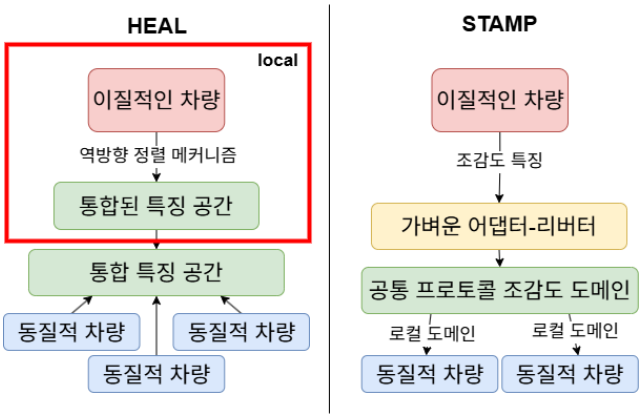

그림 3: HEAL 과 STAMP 의 구조도

한편, ROBOSAC 은 공격자에 대한 사전 확률이 중요한 파라미터이기 때문에 A2CP 로 이를 추정한다. 하지만, 실제 환경에서는 가변적이기 때문에 추정이 어렵다. CP-Guard 확률 비고려 샘플 합의 기법을 도입하여 공격자의 사전 확률 없이 협력자를 샘플링하게 된다[8]. 그리고, 자아 차량과 협력자 간의 불일치를 계산하기 위해 협업 일관성 손실을 사용한다. CP-Guard 는 두 개념을 통해서 '공격자의 사전 확률을 알고 있다.' 는 ROBOSAC 의 가정을 해소했다.
최근에는 성능 개선과 함께 실험을 위한 시뮬레이터가 중요해졌다. 때문에 CP-Guard+는 이중 중심 대조 손실로 정상 특징과 악성 특징 간의 차이를 키워 시스템 복잡도와 계산 비용을 절감하면서, 협업 인지 시스템에서 공격자 탐지 시뮬레이션을 위한 데이터셋 CP-GuardBench 을 제시한다[9]. 이를 통해서 강건한 프레임워크를 설계하는 것도 중요하지만, 적절한 데이터과 시뮬레이터를 사용하여 안전하게 프레임워크의 성능을 입증하는 것도 중요해졌다.

Ⅴ. 결론

협력형 자율주행의 상용화가 가속되면서, 협업 인식 시스템에 대한 필요성도 커지고 있다. 이러한 흐름에서 ROBOSAC 은 공격자 배제를 이용한 보안 프레임워크로 주목을 받았으나, 현실의 복잡한 환경을 충분히 반영하지 못한다. 이를 보완하기 위해서 다양한 후속 연구들이 실시간성, 허위 탐지 제어, 시간적 정보 활용, 노이즈 완화, 모델 이질성 해결, 사전 확률의 불확실성 해소 등 여러 측면에서 제안되었다. 최근에는 실제 상황에 대한 협업 인지의 적용 가능성을 향상시키기 위해서 시뮬레이션 데이터셋 구축도 활발히 이뤄지고 있으며, 이러한 연구들은 협력형 자율주행의 상용화를 한층 더 앞당기고 있다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 학석사연계 ICT 핵심인재양성사업의 연구결과로 수행되었음. (IITP-2025-RS-2024-00437024)

교신저자: 이은규 (eklee@inu.ac.kr)

참 고 문 헌

[1] Li, Yiming, et al. "Among us: Adversarially robust collaborative perception by consensus." *IEEE/CVF ICCV*, 2023.

[2] Li, Jinlong, et al. "Comamba: Real-time cooperative perception unlocked with state space models." *arXiv preprint arXiv:2409.10699* (2024).

[3] Zhao, Yangheng, et al. "MADE: Malicious Agent Detection for Robust Multi-Agent Collaborative Perception." *IEEE/RSJ IROS*, 2024.

[4] Tao, Yihang, et al. "GCP: Guarded Collaborative Perception with Spatial-Temporal Aware Malicious Agent Detection." *arXiv preprint arXiv:2501.02450* (2025).

[5] Hong, Shixin, et al. "Multi-agent collaborative perception via motion-aware robust communication network." *IEEE/CVF CVPR*, 2024.

[6] Lu, Yifan, et al. "An extensible framework for open heterogeneous collaborative perception." *arXiv preprint arXiv:2401.13964* (2024).

[7] Gao, Xiangbo, et al. "STAMP: Scalable Task And Model-agnostic Collaborative Perception." *arXiv preprint arXiv:2501.18616* (2025).

[8] Hu, Senkang, et al. "Cp-guard: Malicious agent detection and defense in collaborative bird's eye view perception." *AAAI*, Vol. 39. No. 22. 2025.

[9] Hu, Senkang, et al. "CP-Guard+: A New Paradigm for Malicious Agent Detection and Defense in Collaborative Perception." *arXiv preprint arXiv:2502.07807* (2025).