

자율주행 차량 시스템의 사이버보안 위험 평가 방법론 비교 및 통합 방안

김주은 · 전상훈*

국민대학교 자동차IT융합학과

Comparative Analysis and Optimization Strategies of Risk Assessment Methodologies for Autonomous Vehicle Systems

Jueun Kim · Sanghoon Jeon*

Department of Automobile and IT Convergence, Kookmin University, Seoul 02707, Korea

(Received 18 July 2025 / Revised 22 August 2025 / Accepted 17 September 2025)

Abstract : As autonomous driving systems evolve, the number and variety of risk factors that targets these systems are also increasing. Particularly, manipulation of sensor data or attacks on communication technologies like V2X can severely impact the overall functionality of autonomous vehicles. To address such threats, the ISO/SAE 21434 standard introduces TARA, a process used in threat analysis and risk assessment. Based on TARA, various risk assessment methodologies have been proposed, with EVITA and HEAVENS among the most prominent. However, as these methodologies were originally designed for conventional vehicle systems, they are limited in terms of complex architecture and functionalities that are unique to autonomous driving systems. In this paper, we derived the risk levels of autonomous driving systems by using both EVITA and HEAVENS methodologies, analyzed their limitations, and proposed an integrated approach that combined their strengths so that they could provide a more suitable risk assessment framework for autonomous systems.

Key words : Risk assessment(위험 평가), TARA(위협 분석 및 위험 평가 프로세스), EVITA, HEAVENS, ISO/SAE 21434, SAE J3016, Automotive cybersecurity(차량 사이버보안)

1. 서론

최근 자율주행 시스템이 발달하면서 차량 사이버보안 분야에 대한 관심이 크게 증가하고 있다. 현대 사회는 완전 자율주행 시대를 지향하고 있기에, 자율주행 시스템에 대한 위협은 매우 치명적이다. 자율주행 시스템은 LiDAR, RADAR, Camera 등 다양한 센서로부터 수집한 데이터를 기반으로 판단 및 제어 기능을 수행하며, V2X와 같은 통신 기술을 통해 외부 환경에 대한 인지 기능을 보완한다. 그러나 이러한 센서 데이터가 조작되거나 통신 기술이 공격을 받을 경우, 차량의 의사 결정 및 제어 시스템 전반에 심각한 영향을 미칠 수 있다.¹⁾ 2021년 Komissarov와 Wool²⁾은 FMCW 레이더를 대상으로 가짜 장애물 생성 및 장애물 제거 공격을 수행하고 거리와 속도 데이터를 스푸핑하여 자율주행 차량의 충돌 방지 시스

템 오작동을 유발하였다. 또한 같은 해 Trkulja 등³⁾은 C-V2X 네트워크에서 협력형 DoS 공격을 통해 네트워크 성능이 악화되었음을 보였다. 2025년에는 Suzuki 등⁴⁾이 LiDAR 객체를 제거하고 가짜 장애물을 주입하는 공격을 실제 자율주행 자동차에 적용하여 96퍼센트 이상의 성공률을 보였다. 이처럼 자율주행 시스템에 대한 위협은 단순히 차량 내부 시스템에 대한 위협뿐만 아니라, V2X 등 차량 통신에 대한 위협으로까지 확장되고 있다. 이러한 위협 요소들에 효과적으로 대응하기 위해서는 자율주행 시스템을 대상으로 한 체계적인 위험 평가(Risk assessment) 방법론이 필요하다.

차량 사이버보안 강화를 목적으로 현재 국제적으로 채택되고 있는 표준은 ISO/SAE 21434⁵⁾이다. 해당 표준은 위험 감소, 차량 보안성 강화, 수명 주기 보장 등을 이점

*A part of this paper was presented at the KSAE 2025 Spring Conference

*Corresponding author, E-mail: sh.jeon@kookmin.ac.kr

¹⁾This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

이점으로 하며 특정 도구 또는 솔루션에 영향을 받지 않고 프로세스와 위협 관리에 중점을 둔다는 특징이 있다. ISO/SAE 21434에서는 TARA(Threat Analysis and Risk Assessment)⁶⁾라는 위협 분석 및 위협 관리 프로세스를 제안하고 있다. TARA 프로세스는 위협 식별(Threat identification), 위협 평가(Risk assessment), 위협 수준 분석(Risk analysis)의 세 절차를 거쳐 수행된다. TARA 수행 과정 중 위협 평가 단계에서 사용할 수 있는 방법론으로는 EVITA(E-safety Vehicle Intrusion proTected Applications)⁷⁾와 HEAVENS(HEALing Vulnerabilities to ENhance Software Security and Safety,⁸⁾ TVRA(Threat, Vulnerability And Risk Assessment)⁹⁾ 등이 있다. 이 중에서도 EVITA는 Attack Tree를 기반으로 위협 식별 및 위협 평가를 수행한다는 특징이 있으며, HEAVENS는 STRIDE¹⁰⁾라는 Threat modeling 기법을 이용하여 위협을 식별한다는 특징이 있다. 두 방법론의 서로 다른 위협 식별 방식 및 위협 레벨 도출 과정은 본 연구에서 하고자 하는 자율주행 시스템에 특화된 위협 평가 방법론 제시에서 다음과 같은 이점을 갖는다.

먼저, Attack Tree와 STRIDE 모델링은 특정 방법론에 국한되지 않고 널리 사용되는 기법이므로, 자율주행 시스템에 특화된 위협 평가의 위협 식별 단계를 구성하는데 유용하게 참고할 수 있다. 두 방법론은 동적 지표의 유무, 위협 레벨 평가 대상의 차이 등 서로 다른 특성을 지니고 있어, 실제 위협 평가 과정을 수행하며 비교 분석하기에 적합하다. 이에 본 논문에서는 이 두 가지 방법론을 적용하여 자율주행 시스템에 대한 위협 평가를 수행하고, 각 방법론의 한계점을 분석함과 동시에 장점들을 통합하여 자율주행 시스템에 적합한 새로운 위협 평가 방법론의 방향성을 제시하고자 한다. 본 연구의 기여점은 다음과 같다.

1) EVITA의 자산 중심 기법과 HEAVENS의 STRIDE 기반 시나리오 분석 기법을 비교 분석한다.

2) EVITA와 HEAVENS 방법론의 한계점을 보완하는, 자율주행 시스템에 최적화된 위협 평가 프레임워크를 제시하여 자율주행 차량의 보안성을 향상시키기 위한 기반을 마련한다. 특히 기존의 자산 중심 위협 식별 방식이 아닌, STRIDE 모델링을 도입하여 자율주행 차량 시스템의 구조를 보다 정확하게 파악할 수 있는 방안을 제시한다.

2. 배경 지식

자율주행 시스템의 위협 평가를 위해서는 해당 시스템에 적용 가능한 기존 위협 평가 방법론들에 대한 이해가 선행되어야 한다. 본 장에서는 본 연구에서 비교 대상으로 삼은 대표적인 두 가지 방법론인 EVITA와 HEAVENS에 대해 각각의 구조, 분석 절차, 평가 기준 등

을 중심으로 설명한다. 이를 통해 이후 장에서 수행될 비교 및 통합 방안의 이론적 기반을 마련하고자 한다.

2.1 EVITA

EVITA는 보안 요구 사항 분석, On-board 아키텍처의 안전, 보안 아키텍처 Prototyping, 검증 및 시연을 목적으로 유럽에서 진행된 프로젝트이다. EVITA 프로젝트에서는 Attack Tree를 기반으로 자산 중심의 위협을 식별하고 Attack probability와 Severity, Controllability를 통하여 Risk Level을 도출한다. Attack Tree란 자산 중심의 위협 식별을 위하여 사용하는 다이어그램으로, Top-down 방식으로 작성된다. 상위 노드부터 Attack goal, Attack objectives, Attack methods, Asset attacks의 순서로 나뉘는데, Attack goal이 Attack Tree의 최상위 노드가 되며 궁극적인 공격의 목표를 나타낸다.

또한 EVITA에서 최종 위협 레벨은 Risk Level로, Attack Probability, Severity, Controllability를 이용하여 도출한다. Attack Probability는 공격 가능성을 나타내는 지표로, 각 위협에 해당하는 공격을 수행하기 위해 필요한 요소의 점수를 결합하여 도출한 공격 잠재력(Attack Potential)의 범위에 따라 1부터 5까지 나뉜다. 숫자가 커질수록 공격 가능성이 높음을 의미한다. Severity는 각 위협이 네 가지 도메인(Safety, Privacy, Financial, Operational)에서 얼마나 영향을 미치는가를 S0부터 S4까지 총 다섯 단계로 나눈 지표이며, S4는 가장 심각한 영향을 미치는 단계를 의미한다.

Controllability는 운전자가 위협을 얼마나 완화할 수 있는지에 대한 가능성을 C1부터 C4까지의 레벨로 나타낸다. C4는 운전자의 제어로 위협을 완화할 수 없는 상태를, C1은 운전자의 제어로 위협 완화가 가능한 상태를 의미한다.

Controllability가 결정되면 이 값에 따라 Attack probability를 행으로, Severity를 열로 하는 Risk Level 도출 Matrix를 이용하여 최종 Risk Level을 결정한다.

EVITA는 Attack Tree로부터 식별된 공격 자산들이 결합된 위협의 레벨과 빈도에 따라 위협 대응 방안의 우선 순위를 나타내어 차량 보안성 향상에 기여한다는 평가를 받고 있다.⁷⁾

2.2 HEAVENS

HEAVENS는 자산 식별, 위협 모델링, 공격 가능성 및 영향 평가, Risk score 산출의 단계로 진행된다. 위협 모델링에는 STRIDE 기법을 사용하는데, 이는 Spoofing, Tampering, Repudiation, Information disclosure, Denial of Service(DoS), Elevation of privilege 총 여섯 종류의 공격 유

형에 따라 위협 시나리오를 나누어 매핑하는 기법이다. 이후 STRIDE 모델링을 통하여 여섯 가지 유형으로 분류된 위협 시나리오에 대해 Threat Level과 Impact Level을 도출하는 과정을 거친다. HEAVENS의 Threat Level은 EVITA에서의 Attack Probability와 유사한 개념으로, CC(Common Criteria), TVRA, ETSI, EVITA 등 타 위험 평가 방법론에서 Threat Level을 고려할 때 사용하는 파라미터 중 공통되는 부분들을 선정한다. 이를 0~3 범위 내에서 점수화한 값의 총합으로 Threat Level을 결정한다.

Threat Level 산정 후에는 Impact Level을 도출하는데, ISO 26262에 기반한 Safety 파라미터, BSI-Standard 100에 기반한 Financial 파라미터, 그 외 Operational과 Privacy, Legislation 파라미터를 고려한다. 각각의 파라미터에 점수를 부여한 후 합산하여 Impact Level을 결정한다. 이후, Threat Level과 Impact Level에 대해 Security Level을 나타내는 Matrix를 이용하여 최종 Security Level을 도출한다. HEAVENS는 기능 안전(HARA)와 사이버보안(TARA)을 상호 보완적으로 다루어 차량 외부 및 내부에 대한 위협들을 통합적으로 다룰 수 있다는 의의가 있다.⁸⁾

3. 관련 연구

차량에 대한 위험 평가 방법론의 필요성은 자율주행 차량 및 커넥티드 카의 발전과 확산에 따라 꾸준히 부각되고 있는 추세이다. 이에 따라 위험 평가 방법론을 비교 및 검증하려는 연구가 이어지고 있다.

2015년 Boudguiga 등¹¹⁾은 RACE(Risk Analysis for Cooperative Engines)라는 방법론을 제안하면서 EVITA, TVRA, RACE 위험 평가 방법론을 동일 시나리오에 적용하여 평가를 수행하였다. 해당 논문에서 RACE는 EVITA와 TVRA의 장점을 결합하여 만든 방법론으로, C-ITS(Cooperative Intelligent Transport Systems)를 위한 위험 분석 체계를 제공한다.

Macher 등¹²⁾은 2015년 차량에 대한 HARA (Hazard Analysis and Risk Assessment)와 STRIDE를 결합한 SAHARA라는 새로운 위험 평가 방법론을 제시하며, 실험을 통해 기존 HARA만 수행하였을 때와 대비하여 위험 상황이 34% 더 많이 식별되었다는 결과를 도출하여 SAHARA 방법론의 실효성을 입증하였다.

또한 Cui와 Zhang¹³⁾이 2020년 발표한 논문에서는 EVITA, RACE, CSRL과 차별화되는 VeRA(Vehicle Risk Analysis)라는 위험 평가 방법론을 제시하며 해당 방법론이 분석 시간 단축과 위험도 수준 측정의 정확도 면에서 자율주행 시스템에 특화된 위험 평가를 수행할 수 있다고 주장하였다. 해당 논문에서는 VeRA를 직접 수행하며 평가 결

과를 다른 방법론의 결과와 비교하여, 기존 자산 기반 식별 방식을 차용하여 일반화된 위험 평가 체계를 유지하고 있다. 그러나 자산 식별 방식의 경우 자율주행 시스템의 동적 운행 조건 및 통신 환경에 대한 복잡도 높은 위협을 식별하는 데 한계가 있을 수 있다.

Abouelnaga와 Jakobs¹⁴⁾의 논문에서는 TARA 프레임워크를 개괄적으로 설명한 후, 특정 케이스에 대해 EVITA와 HEAVENS 방법론을 비교하여 두 방법론이 실무 관점에서 어떠한 차이가 있는지 제시하고 있다.

앞서 언급한 연구들은 기존 방법론과 그를 보완한 개선된 방법론을 제시하면서, 어떤 측면에서 개선이 이루어졌는지를 시사하는 데 초점을 맞추고 있다. 그러나 이들 연구는 자율주행 시스템을 타겟으로 하여 EVITA와 HEAVENS 방법론대로 위험 평가를 수행하고 있지 않다. 본 논문에서는 위험 평가 대상이 자율주행 시스템임을 고려하여, 차량에 대한 대표적인 위험 평가 방법론인 EVITA와, 위험 분석 방식에서 EVITA와 가장 큰 차이를 보이는 HEAVENS를 비교 대상으로 선정하였다. HEAVENS는 STRIDE 모델링을 기반으로 위협을 식별하고, 정량적인 위험 평가를 수행함으로써, 자율주행 시스템의 복잡한 구조에 더욱 적합한 방법론으로 평가된다.⁸⁾ 이에 본 논문은 전통적인 차량 중심의 위험 평가를 넘어서, 자율주행 시스템에 특화된 위험 평가에는 어떤 접근 방식이 필요한지를 현실적으로 제시하고자 하며, 이를 통해 미래 자동차의 보안성 향상에 기여하고자 한다.

4. 위험 평가 수행 과정

본 연구는 자율주행 시스템에 특화된 위험 평가 방법론이 없다는 점에 주목하였다. 현재 선행 연구에서도 전통적인 ECU 기반의 차량 위험 평가 프레임워크를 다루는 부분이 여전히 많은 비중을 차지하고, 자율주행 시스템에 특화된다고 명확히 정의할 수 있는 위험 평가 방법론은 제시되지 않았다. 따라서 해당 한계를 개선하기 위해, 위험 식별 방식이 명확히 다른 두 방법론을 비교함으로써 더 나은 대안을 제시하고자 한다.

4.1 위험 평가 수행 주체 및 절차

본 논문에서 위험 평가 수행은 제1저자에 의하여 진행되었다. 위험 평가는 EVITA와 HEAVENS 방법론의 공식 절차에 따라 진행되었으며, 평가 방법론의 판단 지표는 Appendix A에서 확인할 수 있다.

먼저 EVITA 방법론의 공식 절차는 다음과 같다.

- 1) 자산 식별 및 Attack tree 구성
- 2) Attack potential을 Elapsed time, Expertise, System knowledge, Window of opportunity, Equipment에 따라 산출

- 3) Severity를 Safety, Financial, Operational, Privacy 도메인에 따라 산출
- 4) Controllability 및 최종 Risk Level 도출
다음으로 HEAVENS 방법론의 공식 절차는 다음과 같다.
 - 1) STRIDE 모델링을 통한 위협 시나리오 도출
 - 2) Threat Level을 Expertise, Knowledge of TOE, Window of opportunity, Equipment에 따라 산출
 - 3) Impact Level을 Safety, Financial, Operational, Privacy & Legislation 도메인에 따라 산출
 - 4) 최종 Security Level 도출
 본 연구에서는 사전 정의된 절차에 따라 위협 평가를 수행함으로써 평가자의 주관을 최소화하였다.

4.2 Architecture 선정

본 논문에서는 Limbasiya 등¹⁹⁾의 논문에서 제시하고 있는 자율주행 시스템의 구조와 그에 따른 Attack surface를 EVITA와 HEAVENS 방법론에 따라 위협 평가를 수행하기 위한 기준 Architecture로 선정하였다. 해당 Architecture는 자율주행 차량의 내부 구조에 따라 가능한 Attack method를 나타내고 있기에, 위협 평가를 수행한 후 위협이 적절하게 식별되었는지 비교할 수 있다. Figure 1에 나타난 것처럼 자율주행 자동차의 내부 시스템은 크게 다음과 같이 6가지로 구분된다. 6가지의 항목에는 Figure 1의 회색 항목들과 동일한 번호를 부여하였다.

- 1) Wireless interfaces: Wi-Fi, LTE, Bluetooth 등 무선 인터페이스로 구성된 영역으로 무선 통신 기능을 수행한다. 해당 영역에서는 Impersonation, DoS, Sybil, Modification, Replay, Injection, Side-Channel 공격이 가능하다.
- 2) Physical ports: USB 또는 OBD 단자 등으로 구성되어 있는 물리적 포트로, 차량 외부의 하드웨어 장치를 차량에 연결하는 기능을 수행한다. Injection, Side-Channel, Impersonation 등의 공격이 가능하다.

- 3) In-Vehicle network: 차량 ECU와 통신 프로토콜이 이루어지는 CAN Bus가 속하는 영역이며 Fuzzing, Bus-off, Modification, Injection, Impersonation 공격이 가능하다.
- 4) Keyless entry system: Remote Key 등 차량의 물리적 잠금 해제를 실제 키 없이 수행하는 시스템이다. 해당 시스템에서는 Impersonation, Replay 공격이 가능하다.
- 5) Infotainment system: OTA 펌웨어 업데이트를 통해 차량 계기판 및 라디오 등 인포테인먼트 기능을 수행하는 시스템 영역이다. 이 시스템을 대상으로 가능한 공격은 Impersonation, Modification이 있다.
- 6) Perception sensors: Radar, LiDAR, Camera 등 외부 물체 인식을 담당하는 센서들이 포함되어 있다. 해당 센서들을 대상으로 Modification, DoS, Remote sensor 공격을 수행할 수 있다.

이 구조는 자율주행 차량에 탑재되어 있는 시스템을 기능에 따라 6가지로 구분함으로써 각 항목에 대해 적절히 위협 모델링을 수행할 수 있다. 본 논문에서는 해당 Architecture를 기준으로 하여 EVITA와 HEAVENS 방법론에 따라 위협 평가를 수행하였다.

4.3 Attack Tree 기반의 EVITA 위협 평가 방법론 수행

Figure 1을 바탕으로, EVITA와 HEAVENS 방법론에 대해 Architecture 기반 위협 식별 및 위협 평가를 진행하고 자 한다.

EVITA 방법론에서는 Attack Tree를 통하여 공격 자산을 분석한다. Figure 2는 Figure 1에 나타난 각 Architecture의 각 항목을 Attack goal로 설정하여 Attack Tree를 작성한 모습이다. Figure 1의 (a)-(f)에는 총 6가지의 세부항목이 명시되어 있는데, 각각의 항목에 대한 공격이 Figure 2에 나타난 Attack Tree의 최상위 노드에 해당하는 Attack goal이다. Attack objective에는 Attack goal을 달성하기 위한 공격

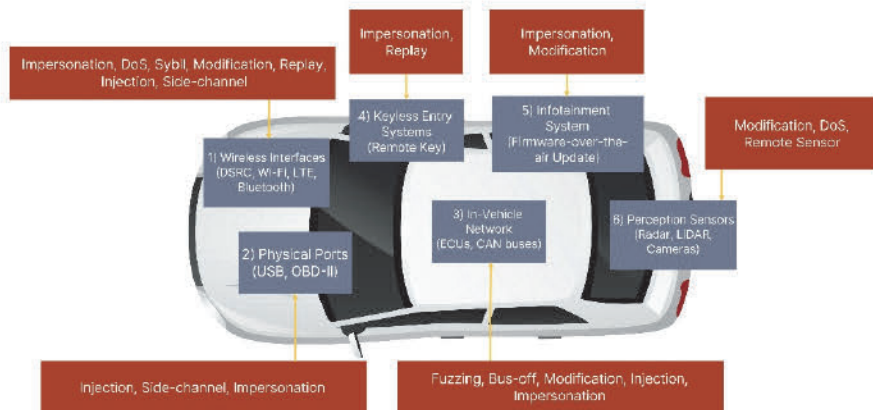
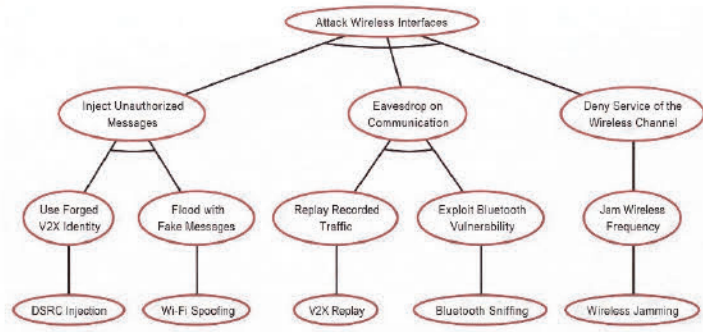
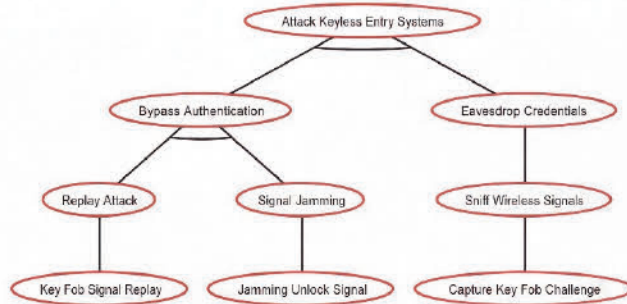


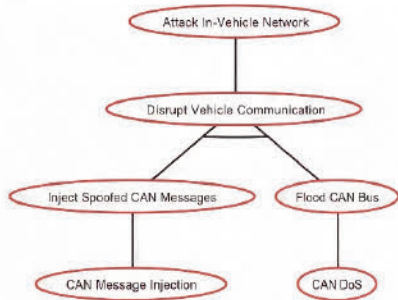
Fig. 1 The attack surface model for connected and automotive driving system operations



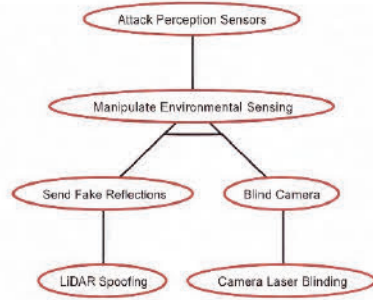
(a) Attack tree_wireless interfaces



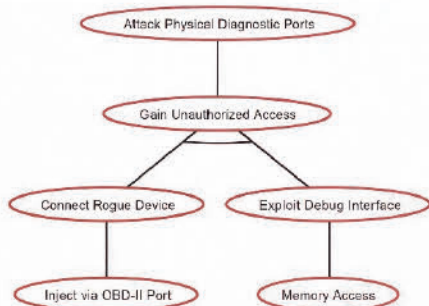
(b) Attack tree_keyless entry system



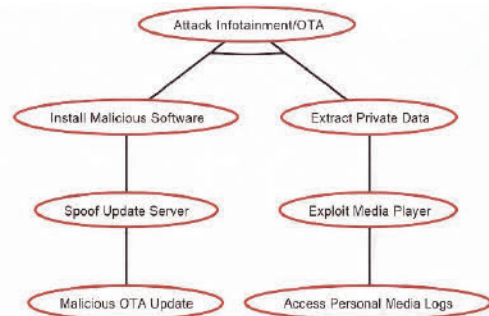
(c) Attack tree_in-vehicle network



(d) Attack tree_perception sensors



(e) Attack tree_physical diagnostic ports



(f) Attack tree_infotainment/ota

Fig. 2 Attack trees of automotive driving system

목표를 작성하였고, 하위 노드인 Attack method에는 Replay, Spoofing 등 실질적인 공격 방법을 작성하였다. 마지막으로 Asset attack에 해당하는 최하위 노드에는 각 공격 방법들이 어떤 자산으로부터 시작되는지 분석하여 공격 자산을 기록하였다. 예를 들어, Inject spoofed CAN messages라는 CAN 메시지 주입 공격은 CAN message injection이라는 자산으로부터 달성할 수 있는 공격 방법 이므로, Attack tree 상에서 Inject spoofed CAN messages 노드의 하위에 CAN message injection 노드를 작성하였다. Attack tree를 작성한 이후, 이 내용을 기반으로 Attack Probability를 도출하였다.

Attack Probability는 각각의 Attack tree 노드 중 Asset attack에서 도출한다. Attack Probability를 도출하기 위해서는 Elapsed time(공격에 걸리는 시간), Expertise(공격자의 능력), System knowledge(요구되는 시스템 지식 수준), Window of opportunity(공격 기회), Equipment(공격에 필요한 장비) 총 6가지의 Factor를 고려한다. Elapsed time은 공격에 걸리는 시간을 의미하며, 1일, 1주, 1달, 3달, 6달 단위로 구분하여 차등 점수를 부여한다. Expertise는 공격자의 경험치를 나타내는 지표로 Layman, Proficient, Expert, Multiple experts로 구분하여 점수를 매긴다. System knowledge는 공격에 요구되는 시스템에 대한 지식 수준이 얼마나 제한적인지를 나타내며, Public, Restricted, Sensitive, Critical로 나누어 점수를 부여한다. Window of opportunity는 공격에 필요한 타겟 수와 공격 윈도우에 접근하기까지의 시간을 Unnecessary/unlimited, Easy, Moderate, Difficult, None으로 구분하여 차등적으로 점수를 부여한다. 또한 Equipment는 공격을 수행하기 위한 장비의 수준을 나타내며 Standard, Specialized, Bespoke, Multiple bespoke의 단계로 나누어 각 항목에 점수를 매긴다. 해당 점수 기준표는 Appendix A.1에서 확인할 수 있다.

이렇게 도출된 점수를 합산하여 Attack Potential을 도출한 후, Attack Potential의 범위에 따라 Attack Probability의 점수를 매긴다. Attack Potential은 공격을 위한 요소들의 합산을 나타내는 지표이므로 해당 항목과 Attack

Probability는 반비례한다. Table 1은 Attack Potential의 범위에 따라 Attack Probability를 도출하는 기준을 나타낸 표이고, Table 2에서 Attack Tree의 Asset attack에 대해 Attack Potential과 Attack Probability를 도출한 결과를 볼 수 있다. Table 1의 표를 보면 Attack Potential과 Attack Probability는 반비례하므로, Table 2에서 Attack Potential이 최고점인 LiDAR spoofing은 Attack Probability가 2로 가장 낮게 도출되었고, Attack Potential이 최저점인 Key fob signal replay와 Capture key fob challenge는 Attack Probability가 4로 가장 높게 도출되었다.

다음으로 Attack objective에 대해 공격의 심각도, 즉 Severity를 도출하였다. Table 3은 EVITA 공식 문서에서 제공하는 Severity 기준 표이다. Severity에서 고려하는 것은 Safety, Privacy, Financial, Operational 네 가지 도메인이다. Safety는 위협이 발생했을 경우 운전자의 생명과 차량 시스템에 어느 정도 영향을 주는지를 나타내는 도메인이고, Privacy는 위협 발생 시 운전자나 차량의 데이터에 접근할 수 있는 가능성을 나타내는 도메인이다. 또한 Financial은 위협 발생 시 금전적 손해를 나타내는 도메인이며, Operational은 위협 발생 시 운용에 영향을 주는 정도를 나타내는 도메인이다.

해당 표를 기반으로 Table 4와 같이 4가지 Factor에 대하여 점수를 매겨 각각의 Severity를 도출하고, 최종 Severity의 도메인을 결정하였다. EVITA에서 Severity는 Safety가 0이 아닐 때 거의 모든 경우에서 메인 도메인으로 Safety를 사용한다. Safety가 0일 경우는 Privacy, Financial, Operational 중 가장 점수가 높은 도메인을 메인 도메인으로 한다. 예를 들어, Table 4에서 볼 수 있듯이 Attack objective가 Eavesdrop on communication인 항목에서는 Safety 도메인의 값이 0이기에 가장 점수가 높은 Privacy 도메인으로 Severity를 도출하였다. 또한 Bypass authentication과 Deny service of the wireless channel 항목과 같이 Safety 도메인에 비해 타 도메인의 점수가 상대적으로 높을 경우 Privacy, Operational 등 타 도메인이 Severity 도출에 고려되었음을 볼 수 있다.

Table 1 Attack potential and attack probability classification

Attack potential		Attack probability	
Rating	Description	Likelihood	Ranking
0-9	Basic	Highly likely	5
10-13	Enhanced basic	Likely	4
14-19	Moderate	Possible	3
20-24	High	Unlikely	2
>=25	Beyond High	Remote	1

Table 2 EVITA-Based attack probability for attack methods

Asset attack	Elapsed time	Expertise	System knowledge	Window of opportunity	Equipment	Attack potential	Attack probability
CAN message injection	2	3	2	3	3	13	4
CAN DoS	2	3	2	2	3	12	4
Key fob signal replay	1	3	2	2	2	10	4
Jamming unlock signal	2	3	2	2	2	11	4
Capture key fob challenge	1	3	2	2	2	10	4
LiDAR spoofing	3	6	4	3	4	20	2
Camera laser blinding	3	4	3	2	4	16	3
Inject via OBD-II Port	2	3	3	2	2	12	4
Memory access	3	6	4	2	3	18	3
DSRC injection	2	4	3	3	3	15	3
Wi-Fi flooding	1	3	2	3	2	11	4
V2X replay	1	3	3	2	2	11	4
Bluetooth sniffing	1	3	2	2	2	10	4
Wireless jamming	1	3	2	3	2	11	4
Malicious OTA update	3	6	3	3	4	19	3
Access personal media logs	1	3	2	2	2	10	4

Table 3 Severity classification

Class	Safety	Privacy	Financial	Operational
S0	No injuries.	No data access.	No financial loss.	No impact on operation.
S1	Light/moderate injuries.	Anonymous data only (no specific user or vehicle data).	Low level loss (\approx €10).	Impact not discernible to driver.
S2	Severe injuries (survival probable). Moderate injuries for multiple units.	Vehicle specific data (vehicle or model). Anonymous data for multiple units.	Moderate loss (\approx €100). Low losses for multiple units.	Driver aware. Not discernible in multiple units.
S3	Life threatening or fatal injuries. Severe injuries for multiple units.	Driver identity compromised. Vehicle data for multiple units.	Heavy loss (\approx €1000). Multiple moderate loss.	Significant impact. Multiple units with driver aware.
S4	Fatal for multiple vehicles.	Driver identity access for multiple units.	Multiple heavy losses.	Significant impact for multiple units.

Table 4 Derived severity scores per EVITA attack objective

Attack objective	S_safety	S_privacy	S_financial	S_operational	Severity
Disrupt vehicle communication	3	1	1	3	Safety (3)
Bypass authentication	1	2	2	1	Privacy (2)
Eavesdrop credentials	0	3	2	1	Privacy (3)
Manipulate environmental sensing	4	1	1	3	Safety (4)
Gain unauthorized access	2	2	2	2	Safety (2)
Inject unauthorized messages	3	1	1	2	Safety (3)
Eavesdrop on communication	0	3	2	1	Privacy (3)
Deny service of the wireless channel	1	1	1	2	Operational (2)
Install malicious software	4	2	2	3	Safety (4)
Extract private data	0	3	2	1	Privacy (3)

Table 5 Controllability classification

Class	Meaning
C1	Despite operational limitations, avoidance of an accident is normally possible with a normal human response.
C2	Avoidance of an accident is difficult, but usually possible with a sensible human response.
C3	Avoidance of an accident is very difficult, but under favorable circumstances some control can be maintained with an experienced human response.
C4	Situation cannot be influenced by a human response.

Table 6 Controllability classification of attack objectives

Attack objective	Controllability
Disrupt vehicle communication	C3
Manipulate environmental sensing	C4
Inject unauthorized messages	C4
Install malicious software	C4

EVITA에서 고려하는 마지막 요소는 Controllability이다. Controllability는 위협이 위협으로 나타났을 때, 운전자가 대응할 가능성을 나타내는 지표이다. Controllability는 Attack objective로부터 도출한 Severity 중, Safety domain을 Main domain으로 하는 Objective에 대해서만 4단계로 값을 매기는데, 공격에 대해 운전자의 대응이 가능한 레벨은 C1부터 C3까지이고 C4는 인간의 대응으로 상황이 바뀌지 않는 레벨을 의미한다.

Table 5는 EVITA에서 명시하고 있는 Controllability 산정 표이고, Table 6은 이에 기반하여 Safety domain일 경우 Controllability를 측정하는 표이다.

마지막으로, 지금까지 도출한 Attack Probability와 Severity, Controllability를 사용하여 최종 Risk Level을 도출한다.

Risk Level은 Controllability 지표의 값에 따라 총 4개의 다른 Matrix를 사용한다. Matrix는 Appendix A.1에서 확인할 수 있다. 이때 Attack Probability와 Severity를 Matrix에 매핑하여 최종 Risk Level을 결정하는데, Table 7에서 최종 Risk Level을 확인할 수 있다. Table 7을 보았을 때 Inject Spoofed CAN Messages, Send Fake Reflections 등 Controllability가 C3 이상의 등급일 경우 Risk Level이 R6 이상인 반면, Signal Jamming과 Exploit Debug Interface 등 Controllability가 C1일 경우 Risk Level이 R2, R3로 도출된다.

4.4 STRIDE 모델링 기반의 HEAVENS 방법론을 사용한 위협 시나리오 분석

HEAVENS에서는 EVITA와 달리 STRIDE 모델링을 사용한다.⁹⁾ 본 논문에서는 Table 8과 같이 Figure 1의 Architecture 상에서 발생 가능한 위협에 대해 STRIDE 모

델링을 수행한 후, 식별된 위협에 대하여 Threat ID를 할당하였다.

HEAVENS 방법론을 수행한 과정의 기준 표는 Appendix A.2에 제시되어 있다. Table A.0은 Threat Level 산정 기준, Table A.1은 Impact Level 산정 기준, Table A.2는 Risk Score 산정 Matrix를 나타낸다.

각 Threat ID별로 Threat Level(TL)과 Impact Level(IL)을 도출할 수 있다. Threat Level은 Expertise, Knowledge of TOE(Target of Evaluation), Window of Opportunity, Equipment 총 4가지 파라미터의 점수를 합산하여 도출한다. 해당 파라미터는 EVITA에서 Attack Potential을 구하기 위해 사용하는 요소들과 유사한 의미를 가지지만, 부여하는 점수를 0부터 3 사이로 제한하여 각 파라미터별로 정량적이고 일관적인 결과를 도출하도록 하였다. Table 9는 HEAVENS 2.0에서 제시하고 있는 기준에 따라 TL을 도출한 것이다. Table 9의 Threat ID는 STRIDE 모델링을 사용하여 식별한 위협에 대해 각각 ID를 할당한 것이며, 각 요소들의 합에 따라 Threat Level 구간을 분류하여 점수를 도출하였다.

Impact Level은 EVITA의 Severity와 유사하게 Safety, Financial, Operational, Privacy와 Legislation을 이용하여 도출한다. 이때 EVITA는 개인 정보 보호 측면에서 Privacy만을 고려하지만, HEAVENS에서는 Legislation, 즉 위협이 발생 시 법률에 위배되는 상황이 유발되는지 여부도 함께 판단한다.

또한 EVITA가 S0부터 S3으로 Severity class를 구분하는 반면, HEAVENS는 0,1,10,100의 값을 부여하고 총 점수를 합산하여 Impact Level을 도출한다. Table 10은 Threat ID별로 Impact Level을 도출한 것이다.

Table 7 Final risk levels based on EVITA

Objective	Attack method	Attack probability	Severity	Controllability	Risk
Disrupt vehicle communication	Inject spoofed CAN messages	4	3	C3	R6
Disrupt vehicle communication	Flood CAN bus	4	3	C1	R4
Bypass authentication	Relay attack	4	2	C1	R3
Bypass authentication	Signal jamming	4	2	C1	R3
Eavesdrop credentials	Sniff wireless signals	4	2	C1	R3
Manipulate environmental sensing	Send fake reflections	2	4	C4	R6
Manipulate environmental sensing	Blind camera	3	4	C1	R4
Gain unauthorized access	Connect rogue device	4	2	C1	R3
Gain unauthorized access	Exploit debug interface	3	2	C1	R2
Inject unauthorized messages	Use forged V2X identity	3	3	C4	R6
Inject unauthorized messages	Flood with fake messages	4	3	C1	R4
Eavesdrop on communication	Replay recorded traffic	4	2	C1	R3
Eavesdrop on communication	Exploit bluetooth vulnerability	4	2	C1	R3
Deny service of the wireless channel	Jam wireless frequency	4	2	C1	R3
Install malicious software	Spoof update server	3	4	C4	R7
Extract private data	Exploit media player	4	3	C1	R4

Table 8 STRIDE-Based threat modeling of automotive driving system

Threat ID	Component	Threat	STRIDE category
T1	In-Vehicle network	Bus-off	Denial of service
T2	In-Vehicle Network	Impersonation	Spoofing
T3	In-Vehicle network	Fuzzing	Tampering
T4	In-Vehicle network	Injection	Tampering
T5	In-Vehicle network	Modification	Tampering
T6	Infotainment system	Impersonation	Spoofing
T7	Infotainment system	Modification	Tampering
T8	Keyless entry	Replay	Repudiation
T9	Keyless entry	Impersonation	Spoofing
T10	Perception sensors	DoS	Denial of service
T11	Perception sensors	Modification	Tampering
T12	Physical ports	Side-channel	Information disclosure
T13	Physical ports	Impersonation	Spoofing
T14	Physical ports	Injection	Tampering
T15	Wireless interfaces	DoS	Denial of service
T16	Wireless interfaces	Side-channel	Information disclosure
T17	Wireless interfaces	Replay	Repudiation
T18	Wireless interfaces	Impersonation	Spoofing
T19	Wireless interfaces	Sybil	Spoofing
T20	Wireless interfaces	Modification	Tampering
T21	Wireless interfaces	Injection	Tampering

Table 9 Threat level per threat

Threat ID	Expertise	Knowledge of TOE	Window of opportunity	Equipment	Threat level (TL)
T1	2	2	2	2	1
T2	2	2	2	2	1
T3	3	3	2	2	2
T4	2	2	2	3	1
T5	2	2	2	2	1
T6	2	2	2	3	1
T7	2	3	2	2	1
T8	2	2	2	2	1
T9	2	2	2	2	1
T10	3	3	2	2	2
T11	3	3	3	2	2
T12	4	4	4	3	4
T13	3	2	3	2	2
T14	4	3	4	3	4
T15	2	2	2	2	1
T16	2	2	2	2	1
T17	2	2	2	2	1
T18	2	2	2	2	1
T19	1	1	2	2	2
T20	1	1	1	1	1
T21	1	1	1	1	1

Table 10 Impact level per threat

Threat ID	Safety	Financial	Operational	Privacy and legislation	IL total	Impact level (IL)
T1	10	100	10	10	130	3
T2	100	100	100	1	301	3
T3	10	10	10	10	40	2
T4	100	100	10	10	220	3
T5	10	10	1	1	22	2
T6	100	100	10	10	220	3
T7	10	10	1	100	121	3
T8	100	100	10	1	211	3
T9	10	10	1	1	22	2
T10	10	100	1	1	112	3
T11	10	100	1	1	112	3
T12	1000	100	100	10	1210	4
T13	100	10	100	1	211	3
T14	100	100	100	10	310	3
T15	100	100	100	10	310	3
T16	100	100	100	10	310	3
T17	100	100	100	10	310	3
T18	100	100	100	10	310	3
T19	1000	10	10	1	1021	4
T20	10	10	1	100	121	3
T21	10	10	10	1	31	2

Table 11 Final security levels based on HEAVENS

Threat ID	Threat	Threat level (TL)	Impact level (IL)	Security level (SL)
T1	Bus-off	1	3	Low
T2	Impersonation	1	3	Low
T3	Fuzzing	2	2	Medium
T4	Injection	1	3	Low
T5	Modification	1	2	Low
T6	Impersonation	1	3	Low
T7	Modification	1	3	Low
T8	Replay	1	3	Low
T9	Impersonation	1	2	Low
T10	DoS	2	3	Medium
T11	Modification	2	3	Medium
T12	Side-channel	4	4	Critical
T13	Impersonation	2	3	Medium
T14	Injection	4	3	High
T15	DoS	1	3	Low
T16	Side-channel	1	3	Low
T17	Replay	1	3	Low
T18	Impersonation	1	3	Low
T19	Sybil	2	4	High
T20	Modification	1	3	Low
T21	Injection	1	2	Low

HEAVENS에서는 Threat Level과 Impact Level을 고려하여 최종 Security Level을 도출하는 표를 사용한다. HEAVENS는 EVITA와 다르게 Controllability를 고려하지 않기 때문에, 하나의 Matrix로 최종 결과를 도출할 수 있다. Table 11은 Threat Level과 Impact Level에 따른 Matrix를 참고하여, 최종 Security Level을 도출한 표이다. Side-channel과 같은 위협은 Threat Level과 Impact Level이 모두 최고점으로, Security Level도 최고 등급인 Critical로 도출되었다. 반면 Modification과 Impersonation 등의 위협은 Threat Level이 1, Impact Level이 2로 Security Level이 최저 등급인 Low로 도출되었음을 알 수 있다.

5. 비교 분석 및 결과

이와 같이 EVITA와 HEAVENS 방법론에 따라 자율주행 시스템에 대한 위험 평가를 수행하였다. 두 방법론 모두 자율주행 시스템의 위험을 평가하는 데 있어서 일부 유의미한 결과를 도출하였지만, 각 방법론은 위협 식별 및 평가 지표 측면에서 뚜렷한 차이를 보였다. 해당 차이점은 Table 12에 나타나 있다. 본 논문에서 ‘위험 식별’은 위험 평가 수행 과정 중 각 위협 항목들이 독립적으로 구분되었음을 의미한다. Table 12의 식별 결과를 보았을 때,

위험 식별 과정에서 HEAVENS는 Figure 1의 Architecture에 나타나 있는 위협들을 모두 식별하였으며 EVITA에서는 일부 위협들이 누락되었다. Table 2와 Figure 1을 비교해 보았을 때 대표적으로 누락된 위협은 Bus off(In-Vehicle Network), Side Channel Attack(Physical Ports), Sybil(Wireless Interfaces)인데, 이 세 위협 중 특히 Sybil과 Side Channel Attack은 HEAVENS 방법론에 따라 평가하였을 때 Security Level이 각각 High와 Critical 수준으로 평가되는 위험도가 매우 높은 위협이다. 자율주행 시스템에서 Sybil Attack은 가짜 노드를 생성하여 트래픽의 방향을 임의로 전환할 수 있는 치명적인 공격¹⁶⁾이기에 반드시 식별되어야 한다. 또한 Side Channel Attack의 경우 차량의 제어 소프트웨어가 캐시에 접근하는 패턴을 이용하여 자율주행 차량의 위치를 실시간으로 추적할 수 있으므로, 심각한 프라이버시 침해 및 차량 경로 조작을 유발할 수 있는 공격¹⁷⁾이기에 Sybil Attack과 함께 반드시 식별되어야 하는 공격 중 하나이다.

EVITA가 이와 같은 치명적인 공격들을 식별하지 못한 이유는 EVITA가 자산 중심의 논리적 공격 탐지에 초점이 맞추어져 있기 때문이다. Side Channel Attack은 하드웨어 레벨의 누출 정보를 분석하는 과정이 포함되어 있는

Table 12 Threat identification – EVITA vs HEAVENS

Threat (from Fig.1)	Component(s) in Fig.1	Identified by EVITA	Identified by HEAVENS
Impersonation	Wireless interfaces, Physical ports, In-Vehicle network, Keyless entry systems, Infotainment system	Yes	Yes
DoS	Wireless interfaces, Perception sensors	Yes	Yes
Sybil	Wireless interfaces	No	Yes
Modification	Wireless interfaces, In-Vehicle network, Infotainment system, Perception sensors	Yes	Yes
Replay	Wireless interfaces, Keyless entry systems	Yes	Yes
Injection	Wireless interfaces, Physical ports, In-Vehicle network	Yes	Yes
Side-Channel	Wireless interfaces, Physical ports	No	Yes
Fuzzing	In-Vehicle network	Yes	Yes
Bus-off	In-Vehicle network	No	Yes
Remote sensor	Perception sensors	Yes	Yes

데, EVITA에서는 하드웨어 기반 공격을 고려하지 않는다.⁷⁾ 또한 Sybil Attack은 네트워크 통신의 위조 공격, 즉 Spoofing 공격으로 분류되는데 EVITA에서는 이러한 Spoofing과 같은 복잡한 네트워크 위협을 탐지하지 못한다.⁸⁾ 반면 HEAVENS는 STRIDE 모델링을 기반으로 자산 중심이 아닌 위협 중심의 분류를 수행하며, Spoofing, Information disclosure와 같이 위협의 카테고리를 체계적으로 분류하기 때문에 Table 8에 표시된 것처럼 Sybil Attack과 Side Channel Attack을 식별해낼 수 있었다. 따라서 EVITA의 자산 중심 분석보다는 HEAVENS와 같이 위협의 유형을 세분화하여 식별하는 것이 더 세밀한 위협을 찾아낼 수 있다.

다음으로 HEAVENS에서 가장 뚜렷하게 드러나는 단점은 EVITA의 Controllability와 같은 동적 지표가 없어, 실제 상황에 적용되는 평가 기준이 될 수 없다는 점이다. 이는 실제 운행 중 발생할 수 있는 다양한 환경 변화 및 공격 시도에 적절하게 대응하지 못하는 한계로 이어질 수 있다.

앞서 분석했듯이, EVITA에는 HEAVENS와 다르게 동적 지표를 반영할 수 있는 평가 기준인 Controllability가 있다. 그러나 Controllability 또한 동적 상황을 완벽히 반영하지는 못한다.

해당 지표는 ISO 26262의 사고 위협 모델이 반영된 것인데, 이것은 운전자가 차량에 탑승해 있고 주행 상황 제

어가 가능하다는 전제를 기반으로 하고 있으므로 자율주행 레벨 4 이상의 시스템에 적용하기 어렵다. 예를 들어, Table 6에서 Attack Objective가 Disrupt Vehicle Communication 인 부분을 보자. Controllability가 C3로 제어가 특정 상황에서 가능하다는 결과가 나타나 있다. 그러나 자율주행 레벨 4 이상의 환경에서 운전자가 탑승하지 않았을 경우, 차량 간 통신이 무력화되면 운전자의 제어는 불가능하다. 또한 EVITA에서는 Controllability를 기업별 시나리오 등을 고려하여 상향 조정할 수 있다고 명시하고 있지만,⁹⁾ 결국 상향 조정하는 것도 객관적인 지표가 존재하지 않아 정량적 평가가 불가능하다는 한계가 있다. 따라서 EVITA의 Controllability를 실제 상황에 사용하기 위해서는 객관적인 기준과 함께 정량적 평가가 가능하도록 보완해야 한다.

이러한 분석 결과에 따라, 본 연구에서는 Table 13과 같이 해당 지표를 SAE J3016에서 정의하는 자율주행 레벨¹⁰⁾을 결합할 것을 제안한다. SAE J3016에서는 자율주행 레벨을 0부터 5까지 총 6단계로 나누며, 그중 0단계부터 3단계까지의 레벨에서 운전자의 개입을 의무화한다. 따라서 본 연구는 자율주행 레벨 0부터 3까지는 기존 EVITA에서 사용하는 것과 같이 운전자의 Controllability를 평가하고, 운전자가 없을 경우에는 일관적으로 C4로 두어 운전자의 부재를 전제하는 프레임워크를 제안한다. 이를

Table 13 Proposed controllability to SAE J3016 driving automation level mapping

SAE Level	Label	Driver engagement	Controllability evaluation (Proposed)	Notes / Rationale
Level 0	No driving automation	Required (Driver must intervene)	Evaluate by driver controllability (C1-C4, Scenario-dependent)	Driver performs all DDT and monitoring.
Level 1	Driver assistance	Required (Driver must intervene)	Evaluate by driver controllability (C1-C4, Scenario-dependent)	System assists steering or acceleration/deceleration; Driver monitors.
Level 2	Partial driving automation	Required (Driver must intervene)	Evaluate by driver controllability (C1-C4, Scenario-dependent)	System assists both steering and acceleration/deceleration; Driver monitors.
Level 3	Conditional driving automation	Required (Driver must intervene)	Evaluate by driver controllability (C1-C4, Scenario-dependent)	System performs DDT; Driver must take over upon request.
Level 4	High driving automation	Not required (Driver absent assumed)	Fix at C4 (Driver absent)	System performs DDT, OEDR, and fallback in defined ODD; No driver assumed.
Level 5	Full driving automation	Not required (Driver absent assumed)	Fix at C4 (Driver absent)	All conditions; Fully autonomous; No human driver concept.

Table 14 Risk under standard vs SAE J3016 automation level controllability

Attack method	Controllability	Risk_E	Risk_C4	(Risk_C4) – (Risk_E)
Inject spoofed CAN messages	C3	R6	R7	1
Flood CAN bus	C1	R4	R7	3
Relay attack	C1	R3	R6	3
Signal jamming	C1	R3	R6	3
Sniff wireless signals	C1	R3	R6	3
Send fake reflections	C4	R6	R6	0
Blind camera	C1	R4	R7	3
Connect rogue device	C1	R3	R6	3
Exploit debug interface	C1	R2	R5	3
Use forged V2X identity	C4	R6	R6	0
Flood with fake messages	C1	R4	R7	3
Replay recorded traffic	C1	R3	R6	3
Exploit bluetooth vulnerability	C1	R3	R6	3
Jam wireless frequency	C1	R3	R6	3
Spoof update server	C4	R7	R7	0
Exploit media player	C1	R4	R7	3

통해 운전자의 제어 가능성을 객관적인 지표에 따라 보다 정량적으로 평가할 수 있다.

결과적으로 EVITA는 위협 식별과 위험 평가 면에서 각각의 장점을 가지고 있다. 특히 위협 식별 면에서는 STRIDE 모델링을 수행하며 위협 시나리오별로 위협을 식별해내는 HEAVENS가 자율주행 시스템의 복잡한 구조의 위험을 평가하기에 적절하다. 또한 동적 상황 평가에는 Controllability를 고려하는 EVITA의 위험 평가 방식

이 자율주행 시스템에서 발생할 수 있는 다양하고 복잡한 위험을 평가하기에 상대적으로 유리한 구조를 갖추고 있다. 이러한 두 가지 특성에 의해 기존의 EVITA나 HEAVENS 방법론 중 한 가지만을 사용해서는 자율주행 시스템에 특화된 위험 평가를 진행할 수 없다.

따라서 자율주행 시스템에 특화된 방법론을 제시하기 위해서는 HEAVENS의 위협 식별 모델과 EVITA의 동적 위험 평가 모델 및 SAE J3016의 자율주행 레벨을 결합해

야 한다.²⁰⁾

본 논문에서는 해당 제안의 기대 효과를 Table 14의 결과로 검증하였다. Table 14에서는 자율주행 Level 4,5일 때 Controllability를 C4로 고정한 경우의 Risk Level(Risk_C4)과, 자율주행 Level을 고려하지 않았을 때 Controllability에 따른 EVITA의 Risk Level(Risk_E)을 비교하고 있다. 표의 (Risk_C4)-(Risk_E) 항목을 보았을 때 Controllability가 C4인 경우에는 차이가 없고, 모든 Attack Method에 대해 Risk Level이 하나 이상 증가했음을 확인할 수 있다. 이는 자율주행 Level 4,5의 상황에서의 Controllability 도입이 상황에 따른 민감도를 부여함과 동시에 과도한 보정을 유발하지 않는다는 점을 시사한다.

6. 결론

본 연구는 자율주행 시스템에 대한 체계적이고 정밀한 위험 평가 방법론의 부재라는 문제의식에서 출발하였다. 기존의 대표적 방법론인 EVITA와 HEAVENS를 자율주행 시스템 아키텍처에 적용하여 각각의 한계와 장점을 분석하였고, 이를 통해 자율주행 시스템의 복잡성과 동적 특성을 반영할 수 있는 평가 요소를 도출하였다.

분석 결과, EVITA는 자산 중심의 위협 식별에 강점을 가지며 Controllability와 같은 동적 지표를 포함하고 있어 실제 운행 중 발생할 수 있는 위험 상황에 대한 평가에 효과적이다. 반면 HEAVENS는 STRIDE 기반의 위협 중심 모델링을 통해 보다 정교한 위협 식별이 가능하며, 자율주행 시스템 특유의 다양한 위험 시나리오를 포괄하는데 적합한 구조를 갖추고 있다. 그러나 각각의 방법론은 동적 상황 반영의 미흡함(EVITA) 또는 Controllability 부재(HEAVENS)와 같은 보완이 필요한 부분이 존재함을 확인하였다.

이에 본 연구는 향후 위험 평가 체계 개선을 위한 방향으로, 자율주행 레벨에 기반한 동적 위험 평가 지표의 도입을 제안하였다. 특히, SAE J3016에 정의된 자율주행 레벨을 기준으로 Controllability의 적용 범위를 차별화할 수 있음을 제안하였다. 이를 통해 자율주행 시스템의 구조적 특성과 운용 환경을 반영한 정밀한 위험 분석이 가능하다.

이후 후속 연구로 STRIDE 모델링과 자율주행 레벨을 도입한 Controllability를 정량화하고, 이를 결합한 방법론 정립 및 관련 파라미터 조정을 수행하여 제안한 내용을 구체화할 수 있을 것으로 기대된다.

References

1) S. Kwon and J. H. Lee, "Security Threats and

Technological Trends in Autonomous Vehicles," Review of KIISC, Vol.30, No.2, pp.31-39, 2020.

- 2) R. Komissarov and A. Wool, "Spoofing Attacks Against Vehicular FMCW Radar," Proceedings of the 5th Workshop on Attacks and Solutions in Hardware Security, pp.91-97, 2021.
- 3) N. Trkulja, D. Starobinski and R. A. Berry, "Denial-of-Service Attacks on C-V2X Networks," arXiv preprint arXiv:2010.13725, 2020.
- 4) R. Suzuki, T. Sato, Y. Hayakawa, K. Ikeda, O. Sako, R. Nagata, Q. A. Chen and K. Yoshioka, "Real World LiDAR Spoofing Attacks on Driving Vehicles at Cruising Speeds," NDSS Symposium, 2025.
- 5) International Organization for Standardization, ISO/SAE 21434:2021 – Road vehicles — Cybersecurity engineering, International Organization for Standardization, <https://www.iso.org/standard/70918.html>, 2021.
- 6) G. Macher, E. Armengaud, E. Brenner and C. Kreiner, "A Review of Threat Analysis and Risk Assessment Methods in the Automotive Context," SAFECOMP 2016: Computer Safety, Reliability, and Security, pp.130-141, 2016.
- 7) A. Ruddle, "Security Risk Analysis Approach for On-Board Vehicle Networks," The Fully Networked Car Workshop at the Geneva International Motor Show, Geneva, Switzerland, <http://evita-project.org/Publications/Rud10.pdf>, 2010.
- 8) A. Lautenbach and M. Islam, "HEAVENS—Healing Vulnerabilities to Enhance Software Security and Safety," The HEAVENS Consortium, Borås, Sweden, 2016.
- 9) European Telecommunications Standards Institute, "Methods for Testing and Specification (MTS); Security; Part 1: Threat, Vulnerability, Risk Analysis (TVRA)," ETSI TS 102 165-1 V4.2.3, https://www.etsi.org/deliver/etsi_ts/102100_102199/10216501/04.02.03_60/ts_10216501v040203p.pdf, 2011.
- 10) A. Shostack, Threat Modeling: Designing for Security, New York, John Wiley & Sons, 2014.
- 11) A. Boudguiga, A. Boulanger, P. Chiron, W. Kludel, H. Labiod and J.-C. Seguy, "RACE: Risk Analysis for Cooperative Engines," Proceedings of the 2015 7th International Conference on New Technologies, Mobility and Security (NTMS), pp.1-5, 2015.
- 12) G. Macher, H. Sporer, R. Berlach, E. Armengaud and C. Kreiner, "SAHARA: A Security-Aware Hazard and Risk Analysis Method," 2015 Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, France, pp.621-624, 2015.
- 13) J. Cui and B. Zhang, "VERA: A Simplified Security Risk Analysis Method for Autonomous Vehicles," IEEE

- Transactions on Vehicular Technology, Vol.69, No.10, pp.10494-10505, 2020.
- 14) M. Abouelnaga and C. Jakobs, "Security Risk Analysis Methodologies for Automotive Systems," arXiv preprint arXiv:2307.02261, 2023.
 - 15) T. Limbasiya, K. Z. Teng, S. Chattopadhyay and J. Zhou, "A Systematic Survey of Attack Detection and Prevention in Connected and Autonomous Vehicles," Vehicular Communications, Vol.37, Paper No.100515, 2022.
 - 16) S. Gupta, C. Maple and R. Passerone, "An Investigation of Cyberattacks and Security Mechanisms for Connected and Autonomous Vehicles," IEEE Access, Vol.11, pp.90641-90669, 2023.
 - 17) M. Luo, A. C. Myers and G. E. Suh, "Stealthy Tracking of Autonomous Vehicles with Cache Side Channels," 29th USENIX Security Symposium (USENIX Security 20), pp.859-876, 2020.
 - 18) SAE International, SAE Updates J3016 Automated Driving Graphic, <https://www.sae.org>, 2019.
 - 19) EVITA Project, Deliverables, <https://www.evita-project.org/deliverables.html>, 2025.
 - 20) A. Lautenbach and M. Islam, HEAVENS Deliverable D2: Security Models (Version 2.0), https://autosec.se/wp-content/uploads/2018/03/HEAVENS_D2_v2.0.pdf, 2025.
 - 21) J. Kim and S. Jeon, "Towards Improved Risk Assessment for Autonomous Driving Systems: A Comparative Study," KSAE Spring Conference Proceedings, pp.1973-1974, 2025.

Appendix A. Scoring Rules for Table 1-11

본 Appendix에서는 본문의 Table 1 ~ 11의 Score 및 Level 산정 규칙을 EVITA/HEAVENS 공식 문서^{19,20)}에 따라 제시한다. 본문에서는 Appendix의 동일 규칙을 일관적으로 적용하였다.

A.1 EVITA

Table A.0 EVITA attack potential (Summarized and adapted)

Factor	Level	Comment	Value
Elapsed time	≤1 day		0
	≤1 week		1
	≤1 month		4
	≤3 months		10
	≤6 months		17
	>6 months		19
	Not practical	The Attack cannot be measured in a timescale	∞
Expertise	Layman	People with relatively little knowledge	0
	Proficient	People familiar with security system	3
	Expert	People familiar with algorithms and disciplines used in security	6
	Multiple experts	People who are experts in various fields related to attacks	8
Knowledge of system	Public	Can be obtained through the Internet	0
	Restricted	Some internal details known	3
	Sensitive	Constrained only to group members	7
	Critical	Known by a few individuals	11
Window of opportunity	Unnecessary/unlimited	Does not need any kind of opportunity	0
	Easy	Access is required for ≤1 day, number of targets ≤10	1
	Moderate	Access is required for ≤1 month, number of targets ≤100	4
	Difficult	Access is required for >1 month, number of targets >100	10
	None	Access time is too short or number of targets is too small	∞
Equipment	Standard	Easily available to the attacker	0
	Specialised	Attacker can get without undue effort	4
	Bespoke	Distribution restricted or very expensive	7
	Multiple bespoke	Different types of bespoke equipment are required	9

Table A.1 EVITA risk matrix (Summarized and adapted)

Controllability	Severity	Attack probability				
		1	2	3	4	5
1	1	R0	R0	R1	R2	R3
	2	R0	R1	R2	R3	R4
	3	R1	R2	R3	R4	R5
	4	R2	R3	R4	R5	R6
2	1	R0	R1	R2	R3	R4
	2	R1	R2	R3	R4	R5
	3	R2	R3	R4	R5	R6
	4	R3	R4	R5	R6	R7
3	1	R1	R2	R3	R4	R5
	2	R2	R3	R4	R5	R6
	3	R3	R4	R5	R6	R7
	4	R4	R5	R6	R7	R7+
4	1	R2	R3	R4	R5	R6
	2	R3	R4	R5	R6	R7
	3	R4	R5	R6	R7	R7+
	4	R5	R6	R7	R7+	R7+

A.2 HEAVENS

Table A.2 HEAVENS Threat Level (TL) parameter (Summarized and adapted)

Parameter	Grade	Value
Expertise	Layman (non-Expert)	0
	Proficient (can perform common attacks)	1
	Expert (can define new attack)	2
	Multiple experts (professional in each attack stage)	3
Knowledge about TOE	Public (public knowledge)	0
	Restricted (limited sharing with external parties)	1
	Sensitive (limited sharing with internal team)	2
	Critical (extremely limited)	3
Window of opportunity	Critical (always accessible)	0
	High (accessible within a limited time)	1
	Medium (restricted physical/logical access)	2
	Low (very low availability)	3
Equipment	Standard (easily available)	0
	Specialized (available with little cost or effort)	1
	Bespoke (expensive or professional equipment)	2
	Multiple bespokes (different equipment required for each attack stage)	3

Table A.3 HEAVENS threat level (TL) score (Summarized and adapted)

Sum of the Values of TL parameters	Threat level	Threat level value
>9	None	0
7-9	Low	1
4-6	Medium	2
2-3	High	3
0-1	Critical	4

Table A.4 HEAVENS Impact Level (IL) parameter (Summarized and adapted)

Parameter	Explanation	Value
Safety	No injury	0
	Light/moderate injuries	10
	Life-threatening but survival probable injuries	100
	Life-threatening and survival uncertain injuries	1000
Financial	No impact	0
	The damage remains most of the organization	10
	The damage leads to financial losses but is not critical	100
	The damage threatens the existence of the organization	1000
Operational	No effect	0
	Emergence items or audio noise	1
	Disruption of convenience functions	10
	Disruption of safety-related functions	100
Privacy & Legislation	No impact	0
	Violation of privacy/legislations but may not lead to abuses	1
	Violation of privacy/legislations leading to abuses	10
	Violation of privacy/legislations leading to abuses and causing significant consequences	100

Table A.5 HEAVENS Impact Level (IL) score (Summarized and adapted)

Sum of the values of IL parameters	Impact level	Impact level value
0	No impact	0
1-19	Low	1
20-99	Medium	2
100-999	High	3
≥1000	Critical	4

Table A.6 HEAVENS Security Level (SL) matrix (Summarized and adapted)

Security Level (SL)	Impact level (IL)					
		0	1	2	3	4
Threat level (TL)	0	QM	QM	QM	QM	Low
	1	QM	Low	Low	Low	Medium
	2	QM	Low	Medium	Medium	High
	3	QM	Low	Medium	High	High
	4	Low	Medium	High	High	Critical