SecureMCP

AI 데이터 접근 보안 프레임워크

MCP 기반 다계층 방어로 AI 에이전트의
RAG/RDBMS 데이터 접근을 안전하게 보호하는
정책 집행 미들웨어

Policy-Enforced
RBAC 기반의 강력한
접근 통제 정책

Fail-Closed
기본 차단 원칙의
안전한 설계

MCP-Native
표준 프로토콜 호환
확장성 보장

문제 정의

AI 에이전트 시대의 데이터 접근 보안 공백

MCP 급성장

LLM이 도구와 데이터베이스에 직접 연결되는 표준 프로토콜(MCP)이 급속도로 확산되고 있습니다.

Anthropic, OpenAI, Microsoft 등 주요 빅테크 기업들이 공식 채택하며 업계 표준으로 자리잡았습니다.

보안 공백

초기 MCP 서버 구현체들은 SQL 실행 기능은 제공하지만, **세밀한 역할 기반 접근 통제 (RBAC)**는 부재합니다.

이는 AI 에이전트가 과도한 권한을 갖게 되어 데이터 유출이나 시스템 손상 위험을 초래할 수 있습니다.

위협 스펙트럼: OWASP MCP Top 10

토큰 노출

자격 증명 및 API 키가 로그나 출력에 노출

권한 확대

느슨한 권한 설정으로 인한 비인가 행위 허용

도구 오염

도구 출력 변조를 통한 모델 행동 조작

인젝션 공격

프롬프트/명령 주입으로 시스템 명령 실행

감사 부재

비인가 행위 미탐지 및 사후 추적 불가

과공유

세션 간 민감 정보 누출 및 컨텍스트 오염

이중 경로 리스크 (Dual Path Risk)

RAG 경로 (비정형 데이터)

- 맥락 인젝션 (Context Injection)
- 문서 오염 (Document Poisoning)
- 검색 결과 조작

RDBMS 경로 (정형 데이터)

- SQL 인젝션 & NL2SQL 인젝션
- 권한 초과 데이터 조회
- 파괴적 DDL/DML 실행

솔루션 개요

SecureMCP의 핵심 가치

SecureMCP는 **AI 에이전트와 데이터 저장소 사이**에 위치하여 모든 데이터 접근 요청을 검증하는 정책 집행 미들웨어로서, 다음과 같은 핵심 가치를 제공합니다.

다계층 방어

5단계 심층 방어 체계

입력 정규화
↓
의도 분류
↓
접근 통제 (RBAC)
↓
이상 탐지 (ML)

정책 집행 미들웨어

Fail-Closed 원칙 적용

모든 질의는 기본적으로
차단되며,
명시적 정책을 통과해야만
실행 가능

보안 공백 최소화

MCP 표준 적합성

벤더 중립적 표준 준수

다양한 MCP 서버 및
AI 에이전트 도구와
호환되는 범용 보안 계층

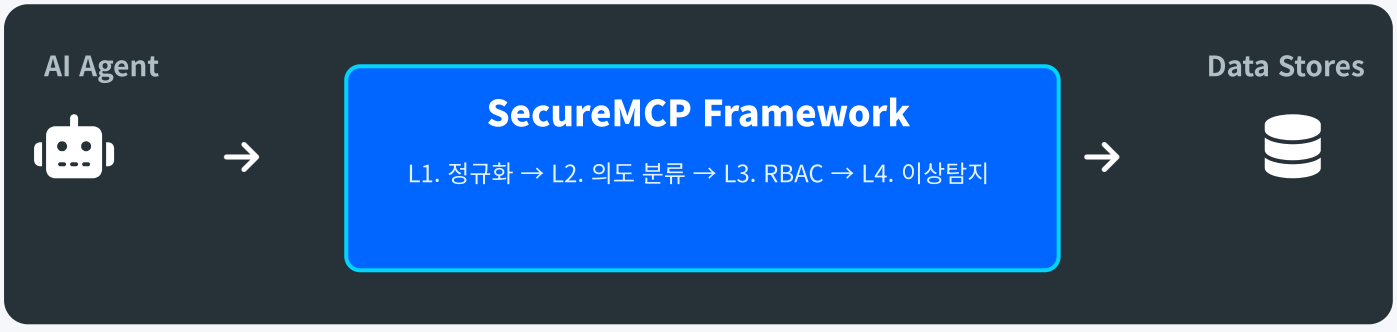
플러그 앤 플레이 방식

통합 보안 게이트웨이

RAG(비정형)와 RDBMS(정형)라는 이중 데이터 경로에 대해 일관된 보안 정책을 적용하는 단일 제어점(Single Point of Control) 제공

가시성과 책임성

모든 요청·응답·차단 이벤트를 100% 감사 로그로 기록하여, 규제 준수(Compliance) 대응 및 사후 보안 감사(Audit) 지원



기술 아키텍처

5계층 방어 시스템 및 이중 경로 라우팅

주요 기능

5개 핵심 방어 모듈 상세

SecureMCP 프레임워크는 다음 5가지 특화된 방어 모듈을 통해 다양한 보안 위협을 심층적으로 차단합니다.

1. check_policy (RBAC 엔진)

SQL AST(Abstract Syntax Tree) 파싱을 통해 질의가 접근하려는 테이블, 컬럼, 연산 유형을 추출하고, 사용자 역할별 권한 매트릭스와 대조하여 허용 여부를 결정합니다.

대응 위협: 권한 확대(Privilege Escalation), 민감 정보 노출

2. explain_gate (비용 검증)

실제 실행 전 `EXPLAIN` 명령을 통해 예상 조회 행 수를 확인합니다. 설정된 임계값(예: 500,000행)을 초과하는 고비용 질의를 사전에 차단하여 DB 과부하를 방지합니다.

대응 위협: 서비스 거부(DoS), 비용 폭증(Cost Explosion)

3. SQL Interceptor (패턴 차단)

정규식(Regex)을 이용해 위험한 SQL 패턴(UNION SELECT, LOAD_FILE, BENCHMARK, SLEEP 등)과 과도하게 긴 쿼리(2,000자 이상)를 즉시 차단합니다.

대응 위협: SQL 인젝션, 시스템 파일 접근, 타이밍 공격

4. Risk Level Filter (위험 등급 필터)

쿼리를 위험도에 따라 4단계(LOW, MEDIUM, HIGH, CRITICAL)로 분류하고, 오직 LOW(단순 조회) 등급만 허용합니다. DDL, 데이터 변경, 다중 구문 실행을 원천 봉쇄합니다.

대응 위협: 데이터 파괴, 멀티 스테이트먼트 인젝션

5. DB Isolation (데이터베이스 격리)

Strict 모드를 통해 지정된 데이터베이스 이외의 접근을 차단합니다. `information_schema`, `mysql.user` 등 시스템 테이블에 대한 접근 시도를 방어하여 스키마 유출을 막습니다.

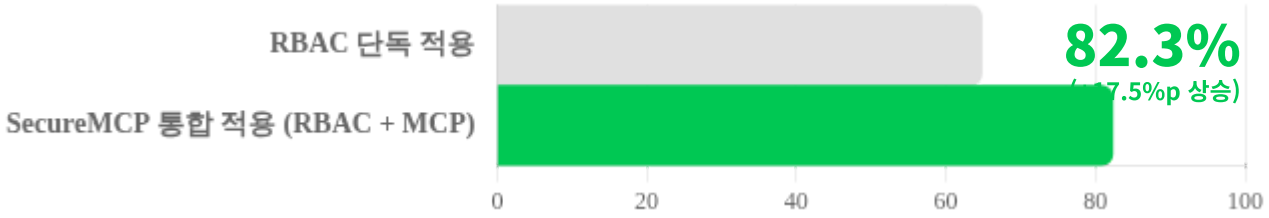
대응 위협: 스키마 유출(Schema Exfiltration), 섀도 서버

적용 효과

검증된 보안성과 안정성

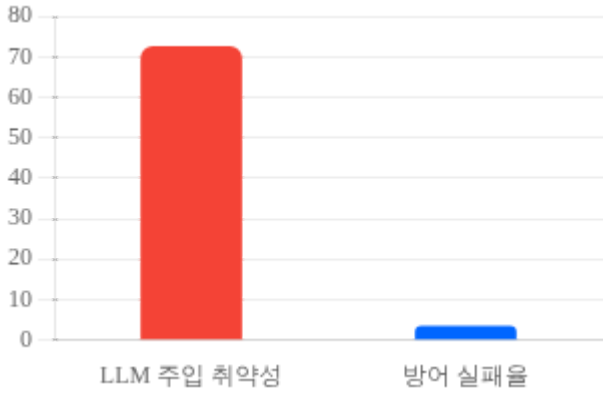
정책 준수율 (Policy Compliance)

적대적 주입(Injection) 공격에 대해 올바르게 차단(Block)한 비율



방어 효율성 및 성능 보존

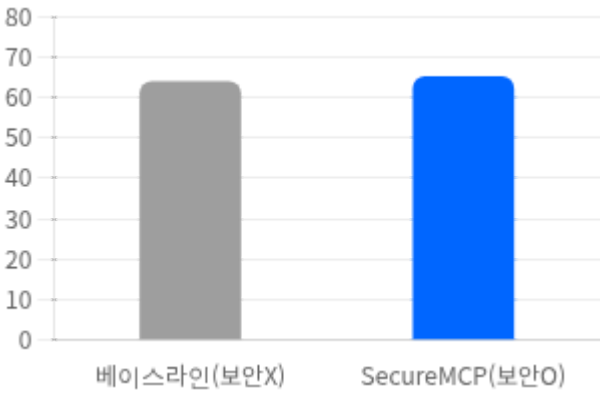
위협 심각성 vs 방어 성공



LLM 자체는 취약함 (72.5%):
외부 보안 없이는 10번 중 7번 이상 공격 성공.

SecureMCP 방어 (실패율 3.4%):
외부 정책 집행을 통해 실패율을 극소화.

정상 질의 실행 정확도 (EX-in-ALLOW)



성능 저하 없음:
보안 계층을 추가했음에도 불구하고, 허용된 정상 질의의 실행 정확도는 베이스라인(63.8%)보다 소폭 상승(65.1%)하거나 유지됨.
→ **안전하면서도 유용한 시스템**

결론: 다계층 통합 방어는 단일 기법 대비 유의미한 보안성 향상(+17.5%p)과 정확도 보존을 동시에 달성함을 실증적으로 입증하였습니다.

적용 분야

공공·AIoT·GovTech 활용 사례

SecureMCP는 공공 데이터의 투명한 관리부터 스마트 시티 인프라 보호까지,
다양한 도메인에서 **안전한 AI 도입을 가속화**합니다.

공공 복지 (Public Welfare)

시민 | 복지 플래너 | 감사자

- **역할 기반 데이터 격리:** 시민은 본인 정보만, 플래너는 담당 구역만 조회 (최소 권한).
- **규제 준수 감사:** 모든 조회 이력을 100% 로깅하여 민감 정보 오남용 사후 추적.
- **접근성 향상:** 복잡한 신청 서식 대신 자연어 질의로 복지 혜택 조회 및 신청 자동화.

AIoT & 스마트빌딩 (Smart Infrastructure)

네트워크 로그 | 센서 데이터 | 이상 탐지

- **운영 안정성 보장:** 대용량 로그 조회 시 비용 폭증(Cost Explosion) 질의 사전 차단.
- **인프라 보호:** 스키마 유출 방지 및 파괴적 명령(DROP/DELETE) 실행 원천 봉쇄.
- **현장 업무 효율화:** SQL을 모르는 시설 관리자도 자연어로 센서 통계 및 상태 점검.

GovTech (Government Technology)

MCP 표준 | 데이터 자산화 | 행정 혁신

- **빠른 통합과 확장:** MCP 표준 기반으로 기존 행정 데이터베이스와 즉시 연동 가능.
- **일관된 보안 거버넌스:** 부서별로 파편화된 데이터 접근 정책을 중앙에서 통합 관리.
- **데이터 기반 의사결정:** 보안이 확보된 상태에서 대규모 행정 데이터를 AI로 분석.

SecureMCP는 안전한 연결을 통해 데이터의 가치를 극대화하고, 디지털 혁신의 기반을 마련합니다.