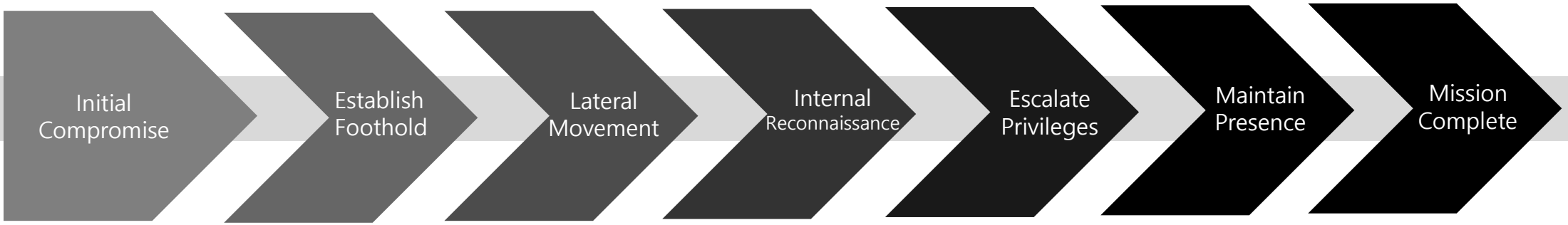
# Design of an eBPF and LLM assisted Detection Framework for APTs in Cloud Environments

Jongseop Kim[1], Changmin Son[2], Jinwoo Kim[3†]

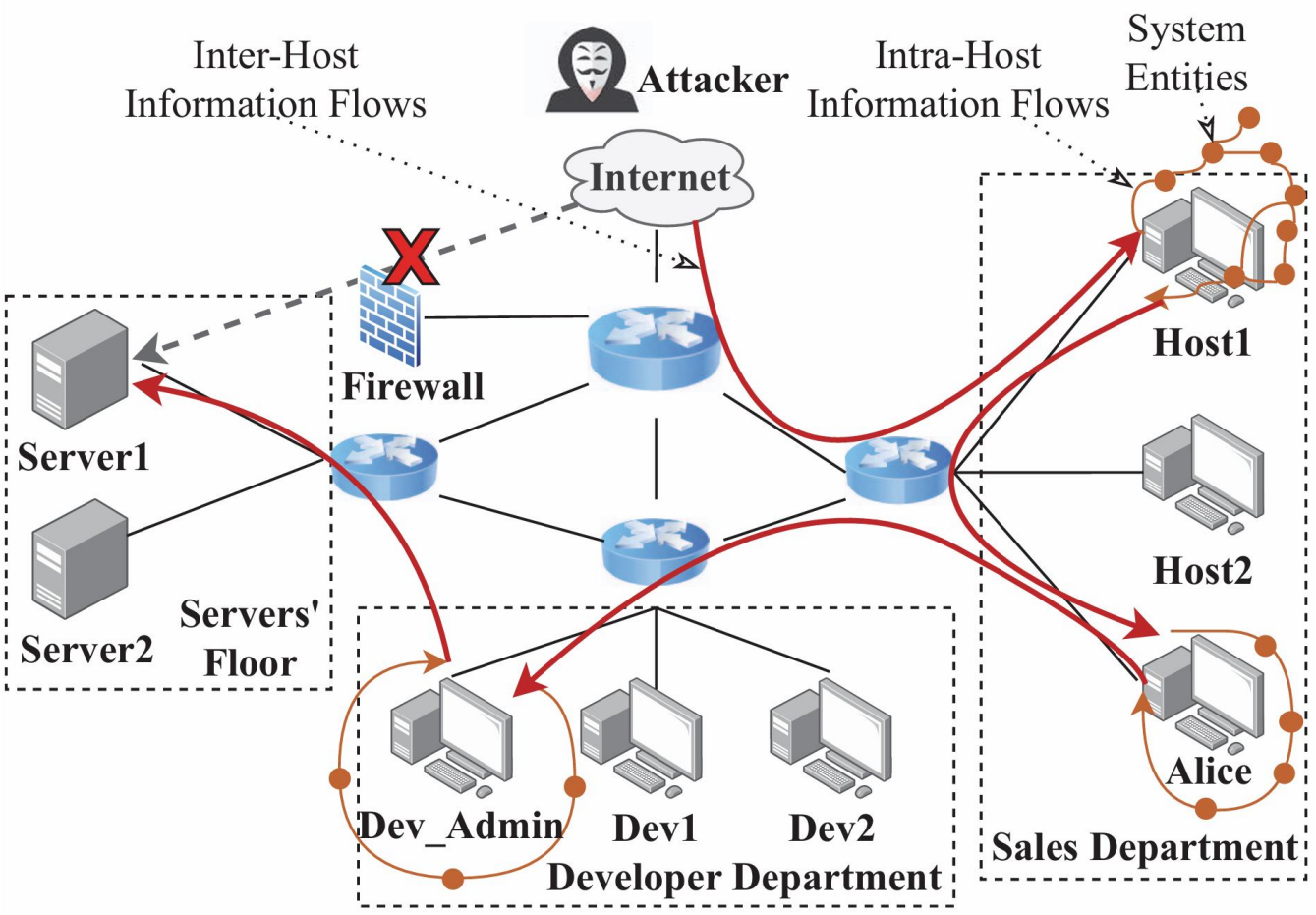
Kwangwoon University(Graduate student, Undergraduated student, Professor)

# APT Attack Overview

- Advanced Persistent Threat
  - **장기간에 걸쳐 은밀하게** 타겟에 접근하여 정보를 탈취하거나 시스템을 장악하는 공격



# APT Attack Overview

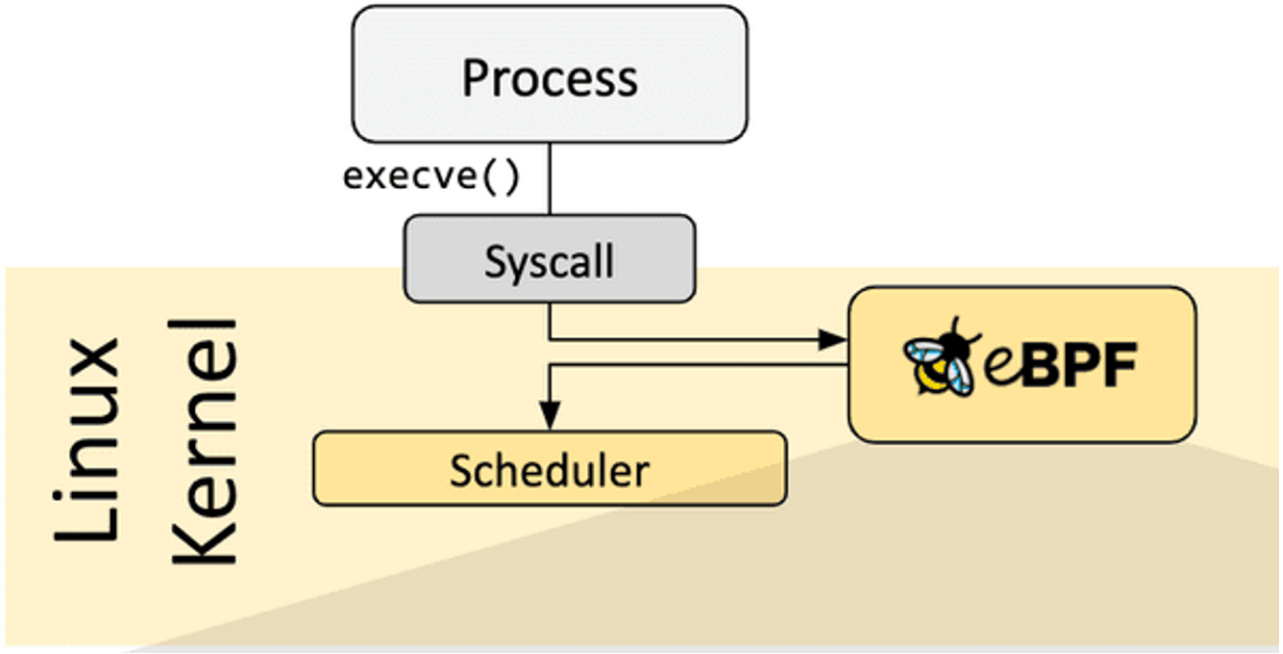





[1] J. Kang, H. Kim, H. Lee, "P4Control: Line-Rate Cross-Host Attack Prevention via In-Network Information Flow Control Enabled by Programmable Switches and eBPF," USENIX Security 2023.

# eBPF

- Extended Berkeley Packet Filter
  - 리눅스 커널 상, 샌드박싱된 환경에서 **커널의 관찰성**을 보여주는 도구




Hook

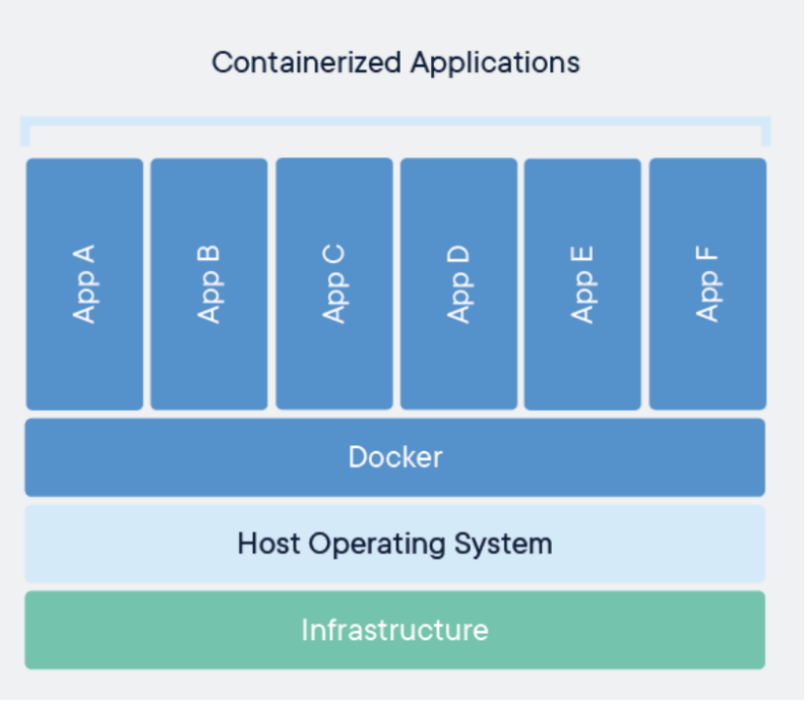
- System call
- Function
- Network
- …



[2] eBPF Project, “What is eBPF?” Available: https://ebpf.io/what-is-ebpf/

# Motivation

- 클라우드 컨테이너 환경의 확산
  - Docker, Kubernetes 기반 서비스들이 증가하며 **하나의 호스트 커널을 공유**하는 구조에서의 공격 표면의 증가[3]






[3] N. Papastergiou, K. Mallas, and V. P. Kemerlis, *"Cross-Container Attacks: The Bewildered eBPF on the Hunt for the Vile Namespace,"* USENIX Security 2022

# Motivation

- 높아진 국내 APT 공격 빈도[4]
  - 2024년 대비 2025년 APT 활동이 56% 증가


| Category | Format | Type | Percentage |
|---|---|---|---|
| SpearPhishing | LNK | Type A | 17.50% |
| | | Type B | 7.50% |
| | | Type C | 13.75% |
| | | Type D | 15.00% |
| | | Type E | 3.75% |
| | | Type F | 8.75% |
| | JSE | Type G | 33.75% |



[4] Ahnlab, “2025년 10월 APT 공격 동향 보고서”, https://asec.ahnlab.com/ko/91047/

# Related Work

- APT(Advanced Persistent Threat) 탐지 방식의 한계
  - 여러 로그를 단편적으로 모니터링하는 기존의 APT 탐지 방식의 공격 맥락 및 흐름 정보 부재[5]

- 의미와 맥락을 파악할 수 있는 LLM의 등장
  - **여러 이벤트의 결합**으로 이루어진 APT 공격에 대해 맥락을 파악할 수 있는 LLM의 등장[6]





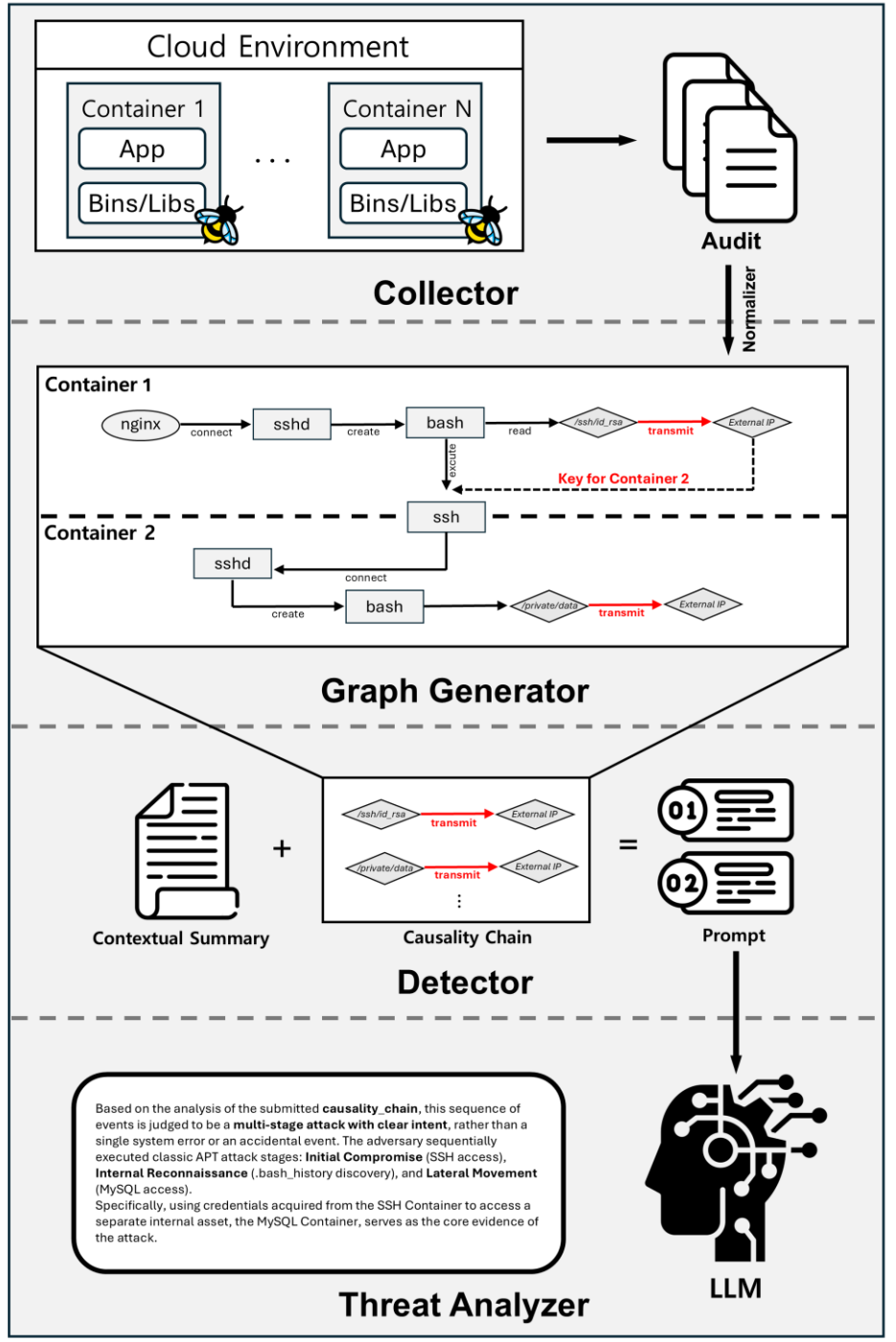
[5] S. M. Milajerdi, R. Gjomemo, B. Eshete, R. Sekar, and V. N. Venkatakrishnan,
*"HOLMES: Real-time APT detection through correlation of suspicious information flows,"* IEEE S&P 2019.
[6]S. Park, S. Lee, and T. Kim, "SHIELD: Large Language Model-Augmented Provenance Analysis for APT Detection," 2024.

# Key Idea

- 컨테이너 기반 행동 수집(eBPF)
  - eBPF(extended Berkeley Packet Filter) 기반 **컨테이너 별** Kernel-Level tracing
- 의미적 인과관계 구조화
  - 수집된 Provenance를 **PROV-O 기반 그래프로 변환**하여, 컨테이너 간 인과관계 표현
- LLM 기반 공격 단계 해석
  - 구조화된 그래프를 LLM에 입력하여 **공격 의도** 및 단계에 대한 MITRE ATT&CK TTP 매핑

# System Overview

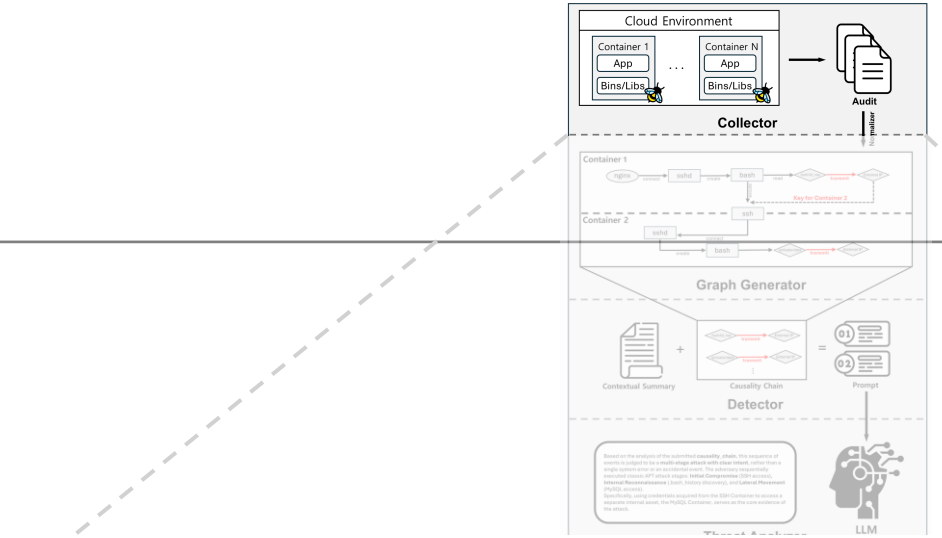
- Collector
  - eBPF 이벤트 수집기
- Graph Generator
  - PROV-O 기반 그래프 생성기
- Detector
  - LLM 입력 프롬프트 생성기
- Threat Analyzer
  - LLM기반 위협 해석기

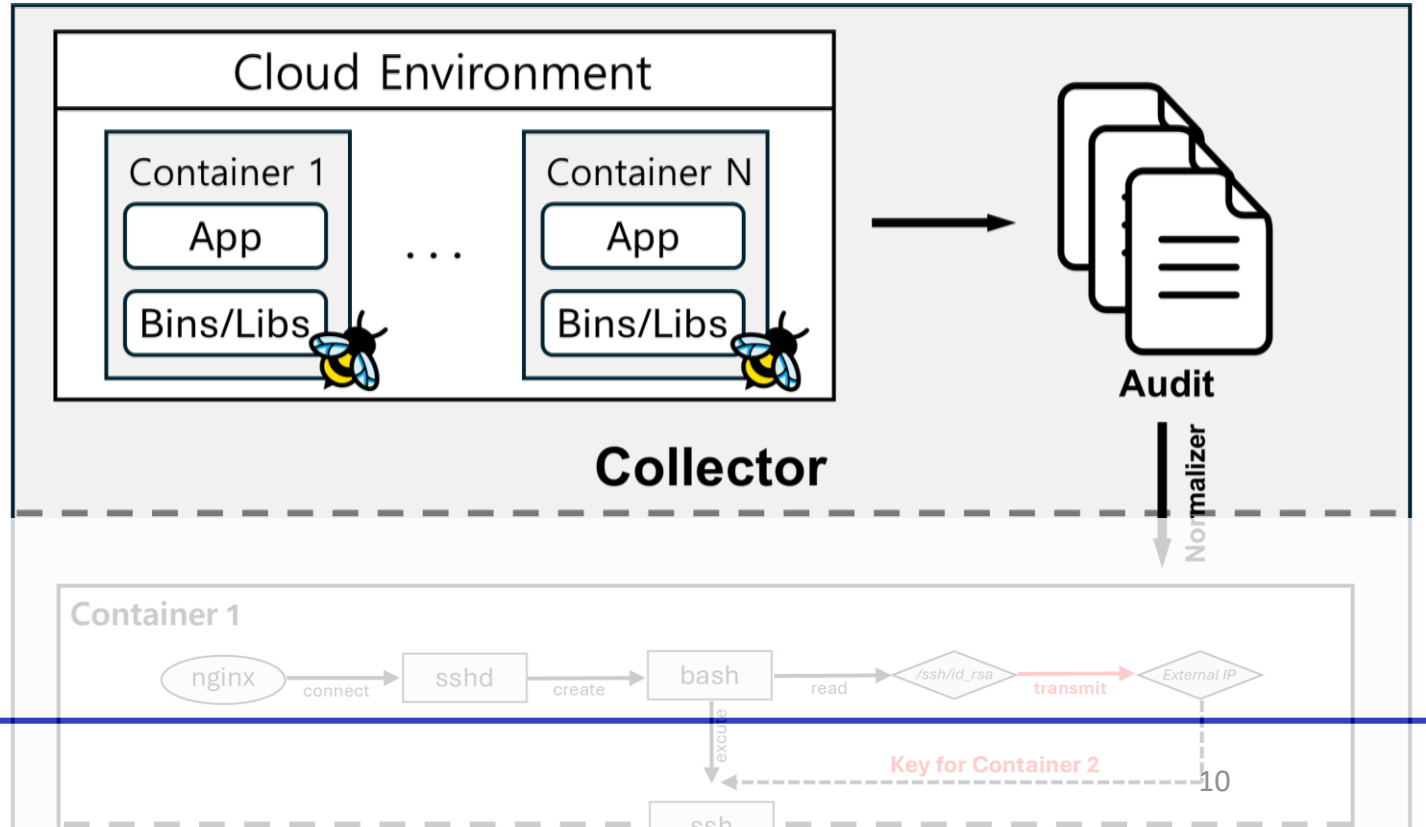


# Collector

- eBPF 기반 이벤트 수집기

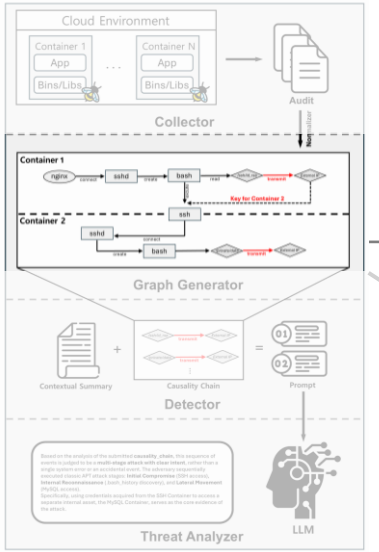

```
{
  "ts": "2025-11-18T21:32:16",
  "host": "jsebpf",
  "container_cgid": "22941",
  "container_name": "ssh-server",
  "uid": 0,
  "pid": 118308,
  "proc": "nc",
  "syscall": "connect",
  "path": "127.0.0.11:53",
  "op": "connect"
}
```

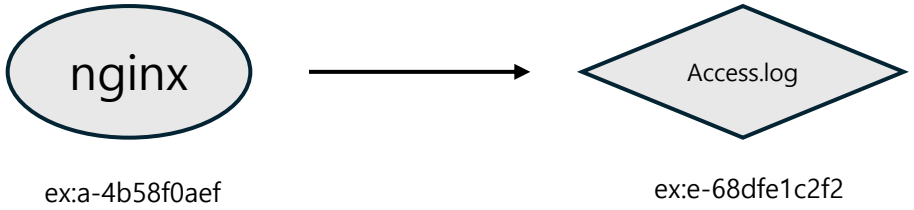



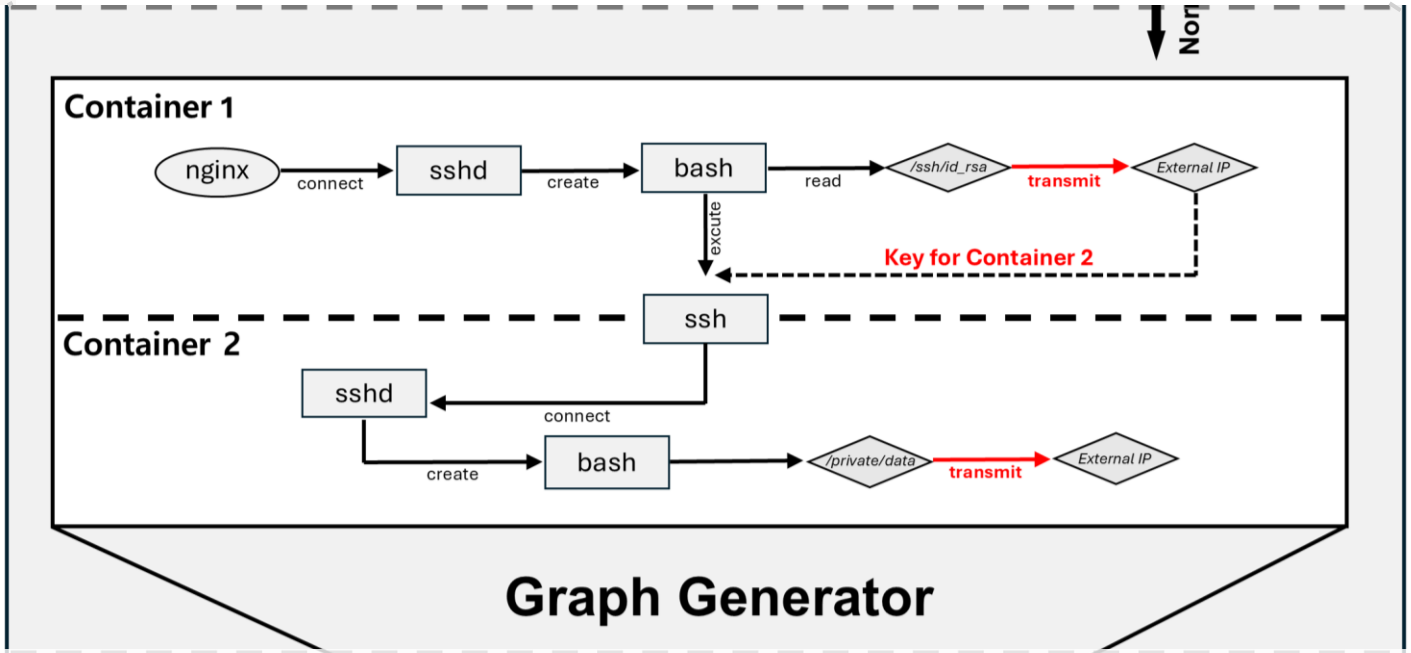
# Graph Generator

- Prov-O 기반 그래프 생성기
  - (Entity, Activity, Agent)




```
ex:act-000001-ec754f48f2 a prov:Activity ;
  prov:label "write" ;
  prov:generated ex:e-68dfe1c2f2 ;
  prov:wasAssociatedWith ex:a-4b58f0aef, ex:a-76c5572beb .
```

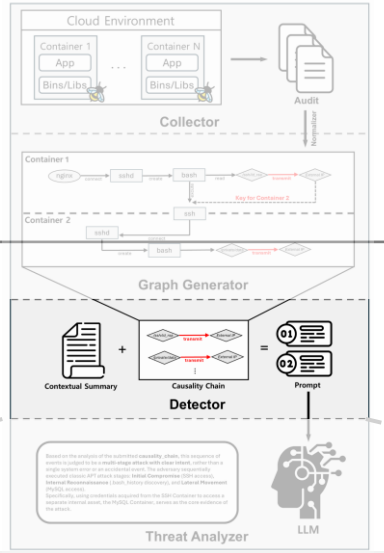


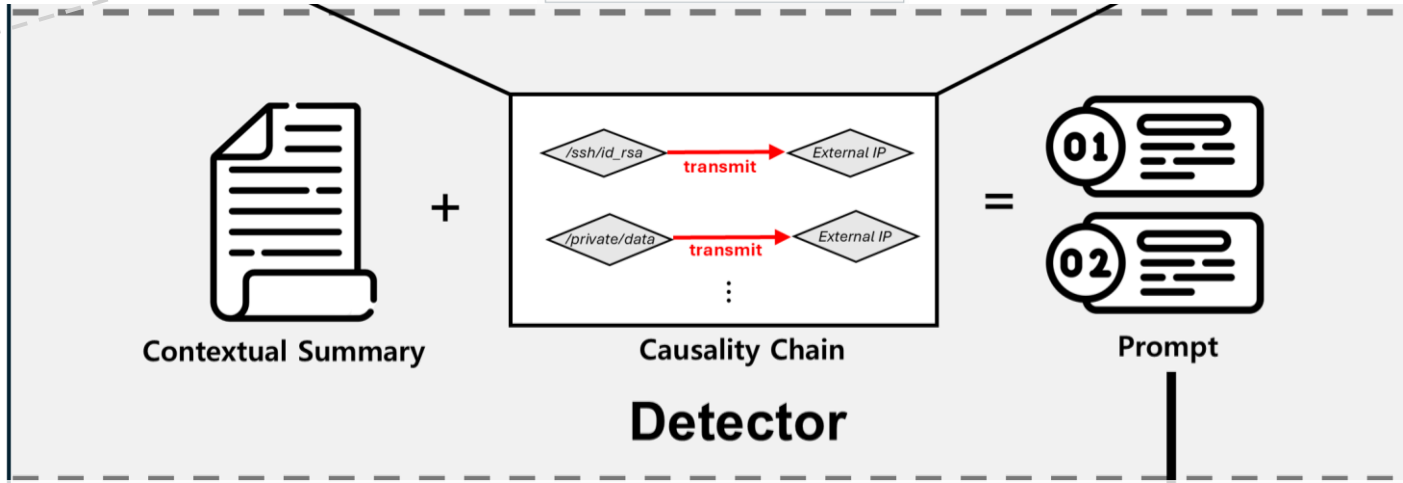



# Detector

- LLM 입력 프롬프트 생성기
  - Contextual Summary + Causality Chain

You are a security analyst.
Based on the contextual summary, identify:
1. Attack stages (MITRE ATT&CK)
2. Probable attacker intent
3. Cross-container movement
4. Final impact




Based on the analysis of the submitted **causality_chain**, this sequence of events is judged to be a **multi-stage attack with clear intent**, rather than a single system error or an accidental event. The adversary sequentially executed classic APT attack stages: **Initial Compromise** (SSH access), **Internal Reconnaissance** (.bash_history discovery), and **Lateral Movement** (MySQL access).
Specifically, using credentials acquired from the SSH Container to access a separate internal asset, the MySQL Container, serves as the core evidence of the attack.

# Threat Analyzer

- Output

Based on the analysis of the submitted causality_chain, this sequence of events is judged to be **a multi-stage attack with clear intent**, rather than a single system error or an accidental event.The adversary sequentially executed classic APT attack stages: **Initial Compromise (SSH access), Internal Reconnaissance (.bash_history discovery), and Lateral Movement (MySQL access)**.
Specifically, using credentials acquired from the SSH Container to access a separate internal asset, the MySQL Container, serves as the core evidence of the attack.

# Future Work

- 공격 시나리오 확장 및 수집
- 외부 APT 데이터셋(e.g. DARPA TC) 기반 모델 학습 가능성
- LLM 프롬프트 엔지니어링 개선 및 최적화
- eBPF 이벤트 소스 확장(추가적인 syscall 및 feature 활용)

# Expected Contribution

- 컨테이너 기반 APT 탐지를 위한 eBPF 기반 저수준 수집 구조 제안
- PROV-O 기반의 공격 행위 그래프 생성
- LLM을 활용한 APT 공격 단계 해석 프레임워크 제안
- 기존 Rule 기반 탐지를 보완하는 APT 분석 모델 제공

들어주셔서
감사합니다