ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

NT230.P21.ANTT | G03 – S20

ADVANCED PERSISTENT THREAT ATTACK DETECTION

Giới thiệu **MAGIC -** mô hình **học máy tự giám sát** phát hiện APT ở nhiều mức độ. Nó học hành vi bình thường bằng cách **ẩn bớt thông tin** trong đồ thị, giúp trích xuất đặc trưng sâu và hiểu rõ cấu trúc hệ thống.

NỘI DUNG

1. Tổng quan đề tài
2. Kiến trúc & Phương pháp
3. Các thí nghiệm chính
4. Tổng kết & Định hướng

1. Tổng Quan

1.1. Tầm quan trọng

- APT là dạng tấn công tinh vi, lén lút và liên tục, rất khó phát hiện.
- Đồ thị nguồn gốc → phân tích quan hệ nhân quả → hỗ trợ phát hiện APT

1.2. Vấn đề cần giải quyết

Các phương pháp phát hiện APT dựa trên phân tích đồ thị nguồn gốc trước đây có hạn chế:

- Phụ thuộc dữ liệu tấn công, kiến thức tiên nghiệm -> khó áp dụng thực tế do thiếu dữ liệu và dễ bị tổn thương trước các mẫu tấn công mới
- Thiếu ngữ cảnh đồ thị → hiệu suất thấp, false positive cao
- Chi phí tính toán & bộ nhớ lớn → khó triển khai thực tế

1. Tổng Quan

1.3. Bản chất dự án

- **MAGIC**: Tự giám sát, không cần dữ liệu tấn công
- Dựa trên Masked Graph Learning + Anomaly Detection
- Tối ưu chi phí tính toán

2. Kiến trúc & Phương pháp

2.1. Dataset được sử dụng trong đề tài

MAGIC được đánh giá trên ba bộ dữ liệu công khai:

- **StreamSpot**: Bộ dữ liệu nhỏ, mô phỏng, chứa 600 lô nhật ký với 5 kịch bản lành tính, 1 kịch bản tấn công drive-by-download.
- **Unicorn Wget**: Bộ dữ liệu khó, 150 lô nhật ký với 25 cuộc tấn công chuỗi cung ứng.
- **DARPA E3**: Bộ dữ liệu lớn (51.69GB), chứa APT thực tế, chia thành các bộ dữ liệu con như Trace, THEIA, CADETS. MAGIC chỉ huấn luyện trên dữ liệu lành tính và kiểm tra trên cả dữ liệu lành tính và tấn công.

2. Kiến trúc & Phương pháp

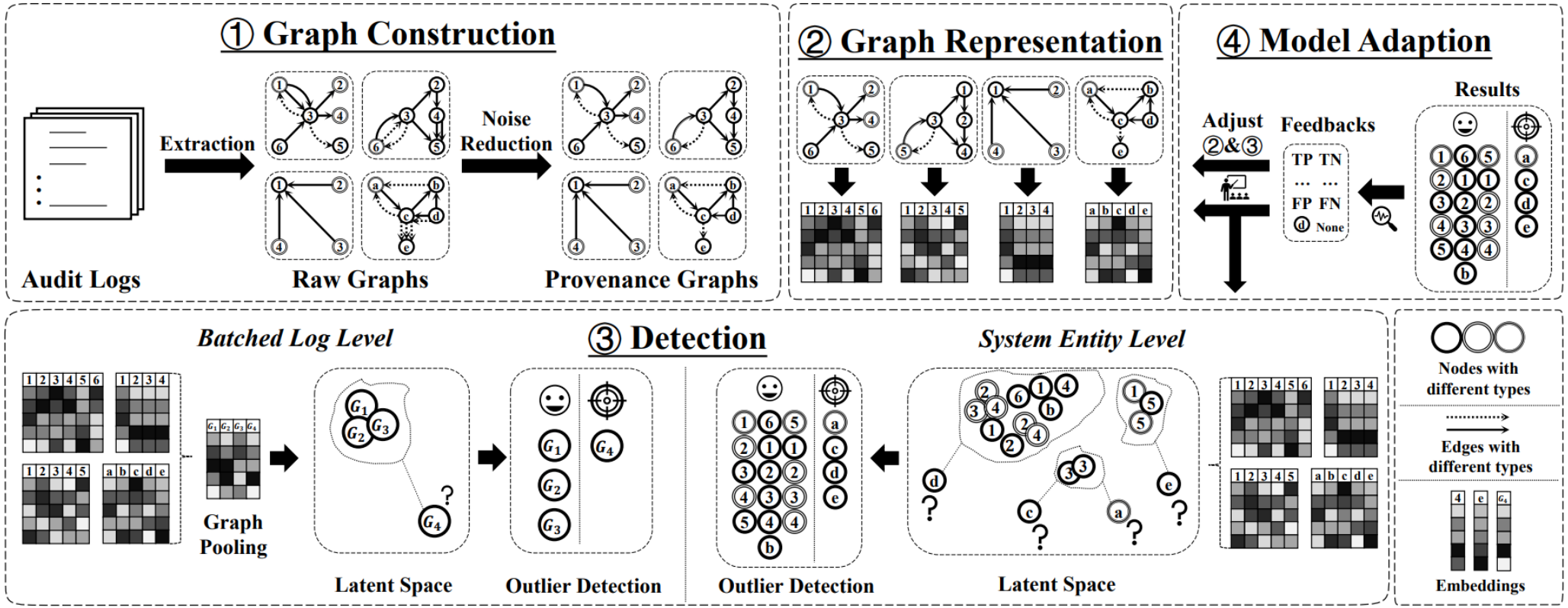
2.2. Tổng quan phương pháp

MAGIC gồm 3 mô-đun chính + 1 cơ chế tùy chọn:

1. Xây dựng đồ thị
2. Biểu diễn đồ thị
3. Phát hiện dị thường
4. Thích ứng theo phản hồi (tuỳ chọn)

2. Kiến trúc & Phương pháp

2.2. Tổng quan phương pháp



2. Kiến trúc & Phương pháp

2.3. Kiến trúc Mô hình MAGIC

- Đồ thị nguồn gốc:

+ Nút: tiến trình, file

+ Cạnh: thao tác hệ thống

+ Giảm nhiễu: gộp cạnh trùng (một tiến trình thực hiện đọc/ghi trên cùng 1 file => giảm nhiễu)

- Biểu diễn đồ thị:

+ Masking đặc trưng nút

+ Graph Attention Network (GAT)

+ Autoencoder tái cấu trúc

2. Kiến trúc & Phương pháp

2.2. Kiến trúc Mô hình MAGIC

- Phát hiện ngoại lệ:

+ K-D Tree

+ Khoảng cách nhúng → điểm bất thường

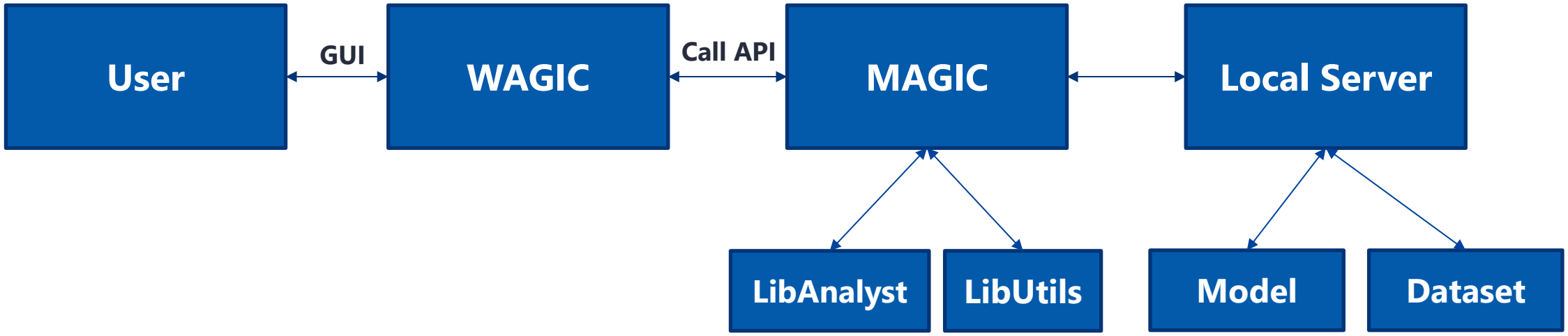
- Thích ứng tùy chọn:

+ Học từ phản hồi analyst

+ Giảm false positive, thích nghi hệ thống

3. Các thí nghiệm chính

Giới thiệu WAGIC: WPF for MAGIC



3. Các thí nghiệm chính

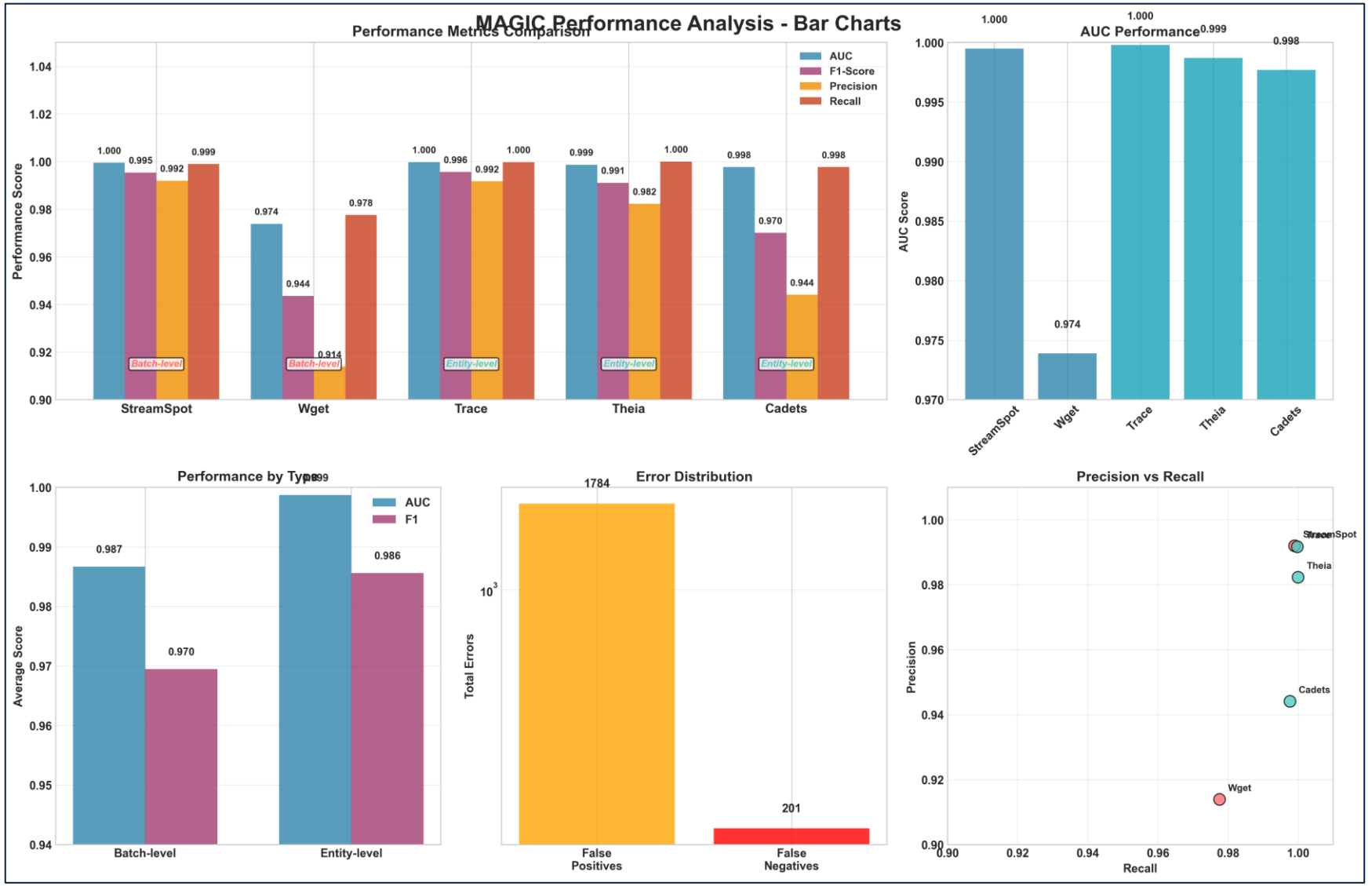
Kịch bản 1: Đánh giá mô hình được train trên tập Streamspot, Wget và DARPA

=> MAGIC trên 5 dataset gốc (Streamspot, Wget, D.Trace, D.Theia, D.Cadets) đều đạt hiệu suất rất cao với AUC > 0.97, trung bình ~0.99; F1-score dao động 0.94–0.99; Precision & Recall luôn trên 0.91. Điều này khẳng định khả năng phân biệt bất thường của MAGIC trong cả kịch bản batch-level và entity-level là rất tốt.

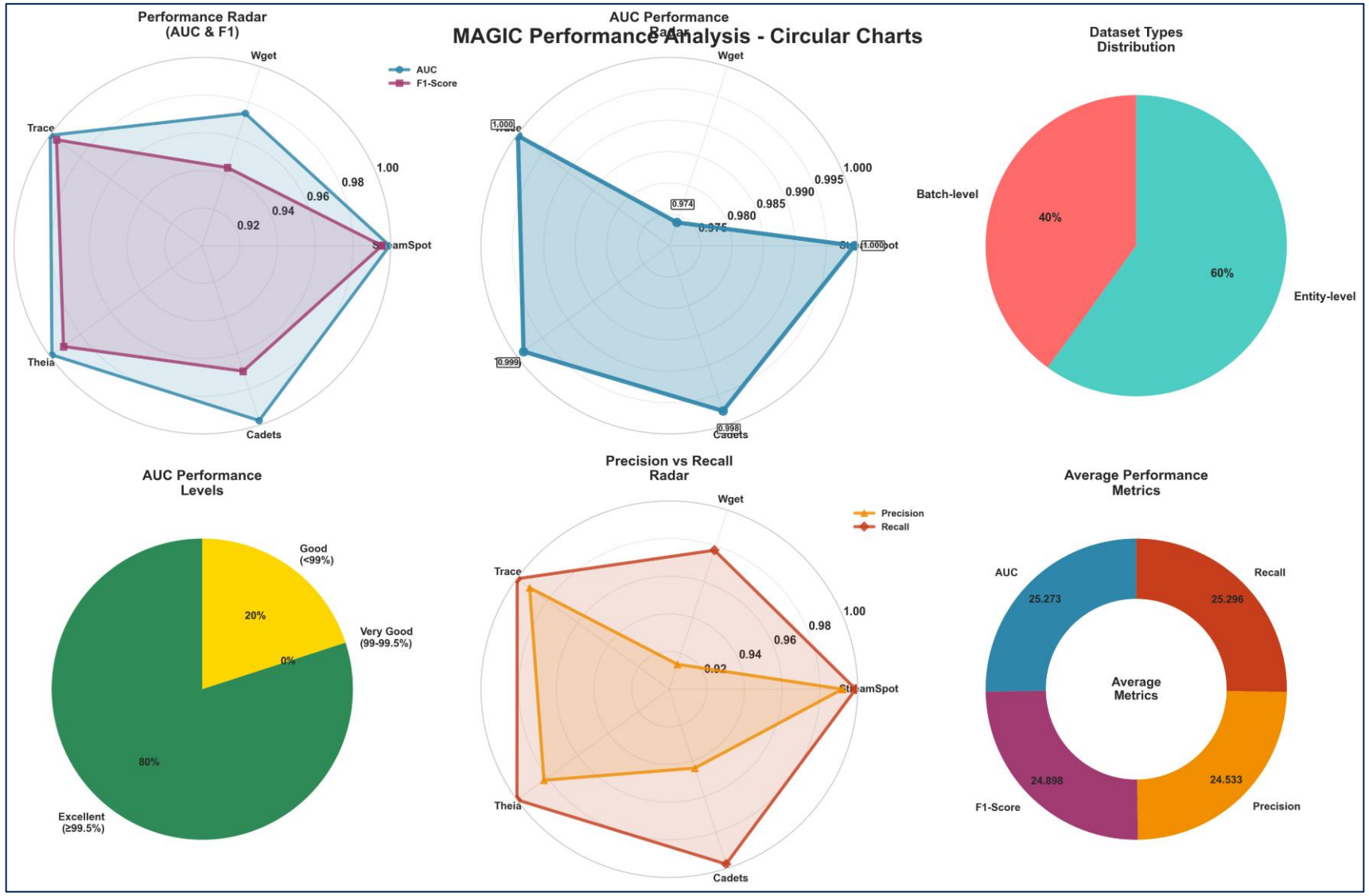
Kịch bản 2: Đánh giá mô hình được train trên tập Five Directions (Nhóm tìm thêm)

=> MAGIC trên dataset tự thêm (Five Directions) cho thấy AUC ~ 0.75 nhưng Recall = 0 ⇒ mô hình chưa bắt được node độc hại nào, mặc dù Precision = 1.0. Kết quả này chỉ ra sự khác biệt lớn về phân phối đặc trưng của FiveDirections so với các dataset trong bài báo, và nhấn mạnh nhu cầu tinh chỉnh masking strategy hoặc thêm bước self-supervision phù hợp để mô hình có thể generalize tốt hơn trên data "thô" thực tế.

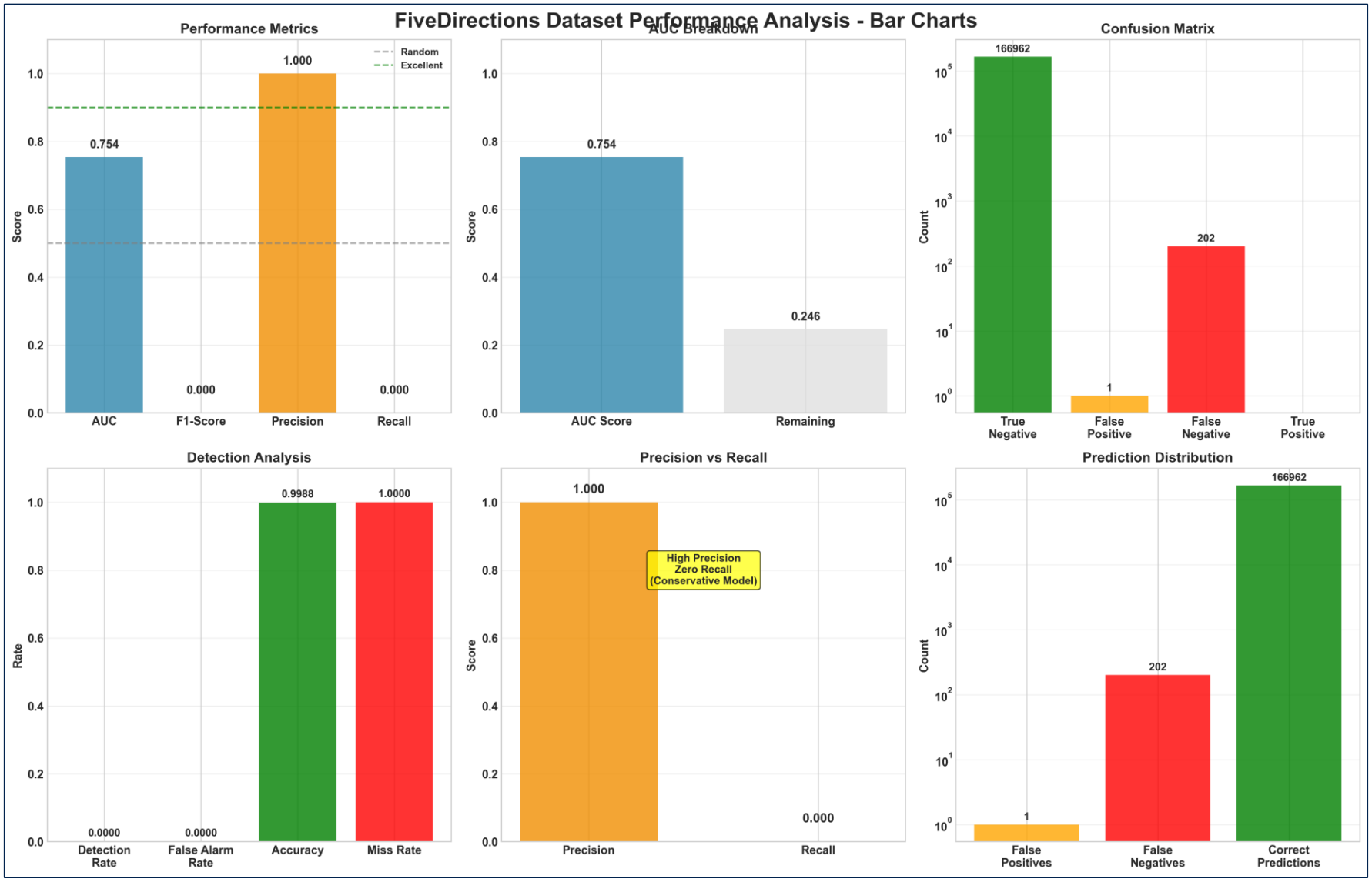
3. Các thí nghiệm chính – Kịch bản 1



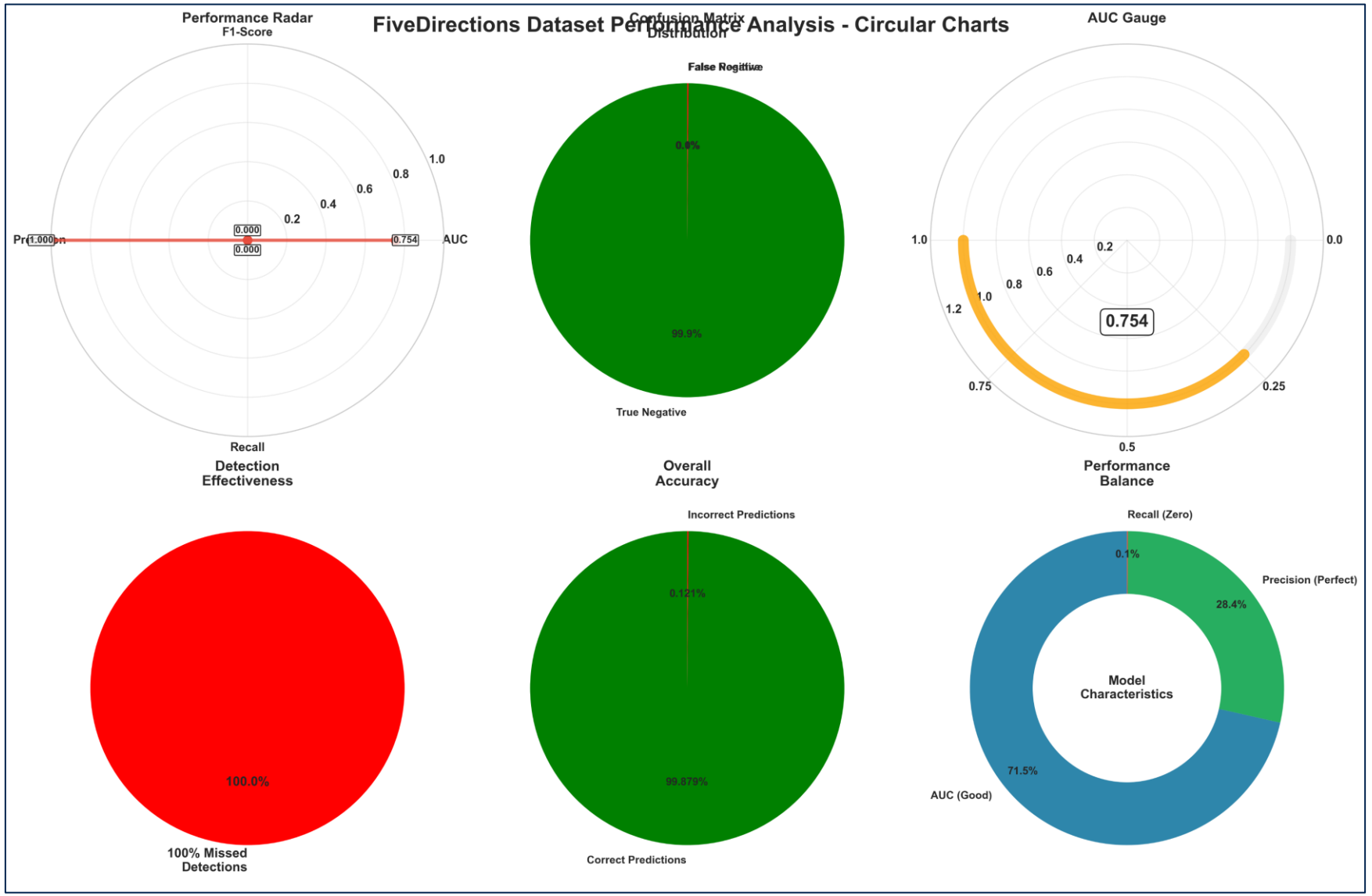
3. Các thí nghiệm chính – Kịch bản 1



3. Các thí nghiệm chính – Kịch bản 2



3. Các thí nghiệm chính – Kịch bản 2



4. Tổng kết & Định hướng

4.1. Ưu điểm

- Tự giám sát, không cần dữ liệu tấn công
- Trích xuất đặc trưng sâu: Masked graph nắm bắt ngữ cảnh phong phú
- Chi phí tính toán thấp
- Phát hiện đa cấp độ: batch-level & entity-level
- Thích ứng liên tục: Học từ phản hồi để giảm dương tính giả và đối phó drift

4. Tổng kết & Định hướng

4.2. Hạn chế

- Khó phát hiện thực thể thụ động: Node “im lặng” (ví dụ file, thư viện) dễ bị bỏ sót
- Với APT cực kỳ tinh vi: Những mẫu quá mới, quá mờ nhạt có thể không hiện rõ trên đồ thị
- Tinh chỉnh siêu tham số: Cần điều chỉnh tỷ lệ masking, ngưỡng phát hiện cho từng môi trường

4. Tổng kết & Định hướng

4.3. Định hướng Tương Lai

- Kết hợp dữ liệu đa nguồn (mạng, người dùng, intel)
- Nâng cấp anomaly detection với học sâu không giám sát hoặc reinforcement learning
- Tự động cập nhật thời gian thực qua cơ chế online learning
- Mở rộng sang tấn công mới (fileless, supply-chain…) để đảm bảo tính tổng quát

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

NT230.P21.ANTT | G03 – S20

ADVANCED PERSISTENT THREAT ATTACK DETECTION

Giới thiệu **MAGIC -** mô hình **học máy tự giám sát** phát hiện APT ở nhiều mức độ. Nó học hành vi bình thường bằng cách **ẩn bớt thông tin** trong đồ thị, giúp trích xuất đặc trưng sâu và hiểu rõ cấu trúc hệ thống.