

HAICon 2020 평가 척도 (TaPR) 설명

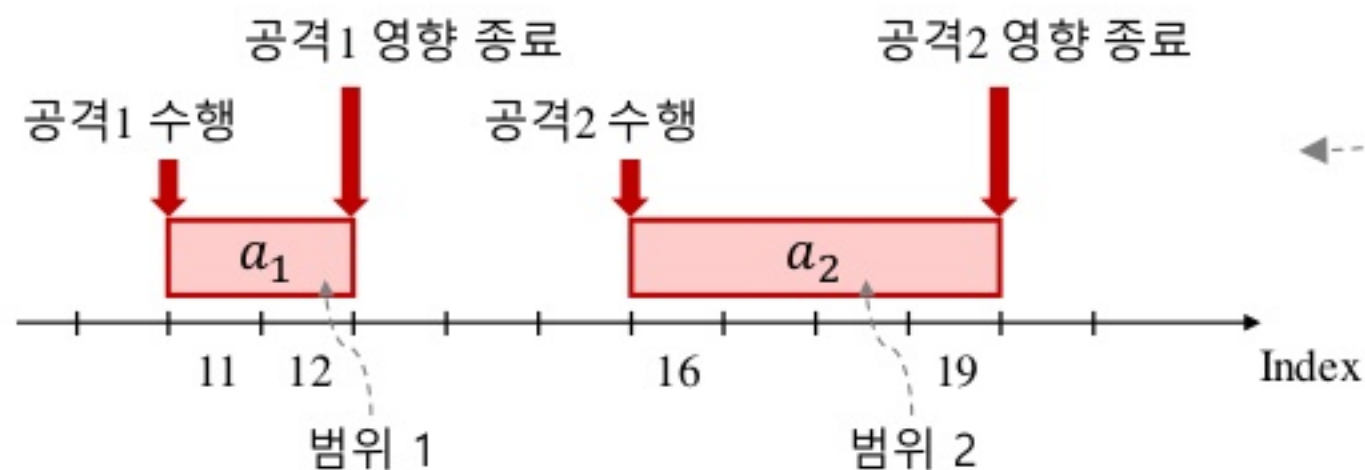
황원석

hws23@nsr.re.kr

*TaPR: W.-Hwang et al. "Time-Series Aware Precision and Recall for Anomaly Detection: Considering Variety of Detection Result and Addressing Ambiguous Labeling," In *Proc. of CIKM*, pp. 2241-2244, 2019.

배경 설명

- 탐지의 목표: 정상 중 공격의 영향을 받은 범위를 탐지
 - HAI 데이터는 오른쪽 표와 같이 구성됨
 - 공격인 구간에 해당하는 11~12, 16~19초 사이의 데이터를 탐지
 - 이 표는 아래와 같은 그림으로 표현됨
 - 각 범위는 서로 다른 공격의 영향



Index	Label
	⋮
10	0
11	1
12	1
13	0
14	0
15	0
16	1
17	1
18	1
19	1
20	0
21	0
	⋮

범위 1 (Range 1) covers indices 11 to 12.

범위 2 (Range 2) covers indices 16 to 19.

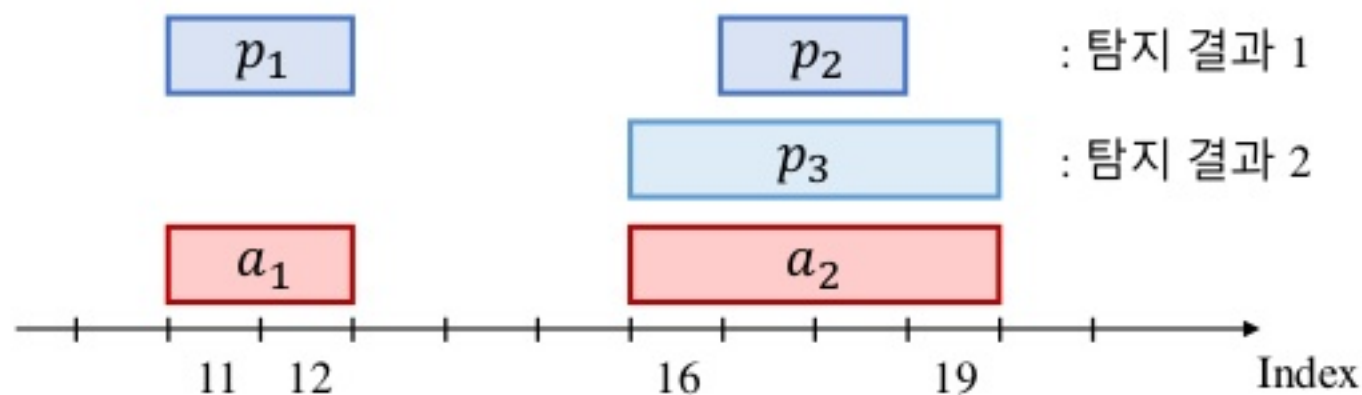
0: 정상 (Normal)
1: 공격 (Attack)

평가 목표

- 목표 1. 탐지된 공격의 다양성 평가

- 아래 그림의 예

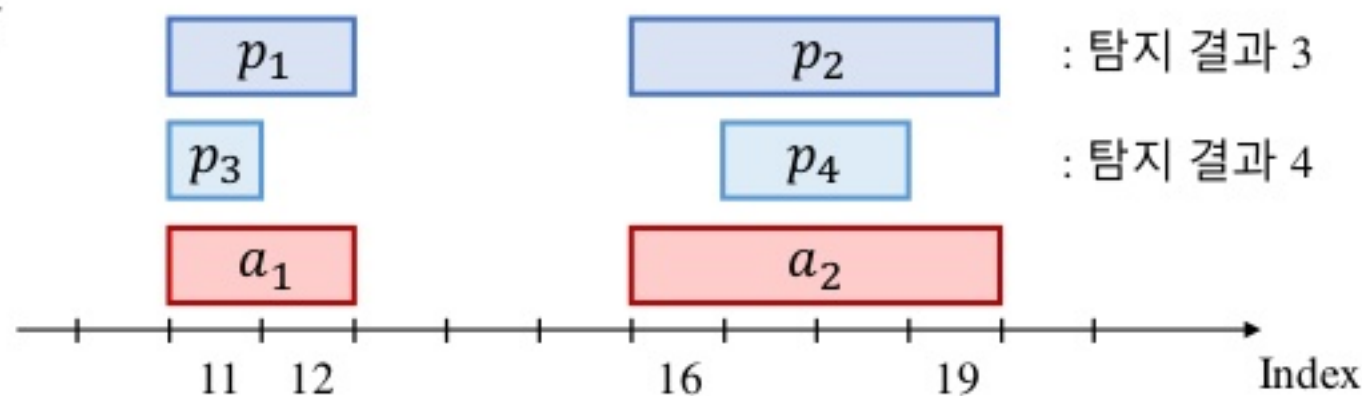
- 두 결과 모두 4초의 데이터를 탐지
- 결과 1을 더 높게 평가함
 - 결과 1은 두 가지 공격의 영향을 탐지
 - 결과 2는 한 가지 공격의 영향만을 탐지



평가 목표 (cont.)

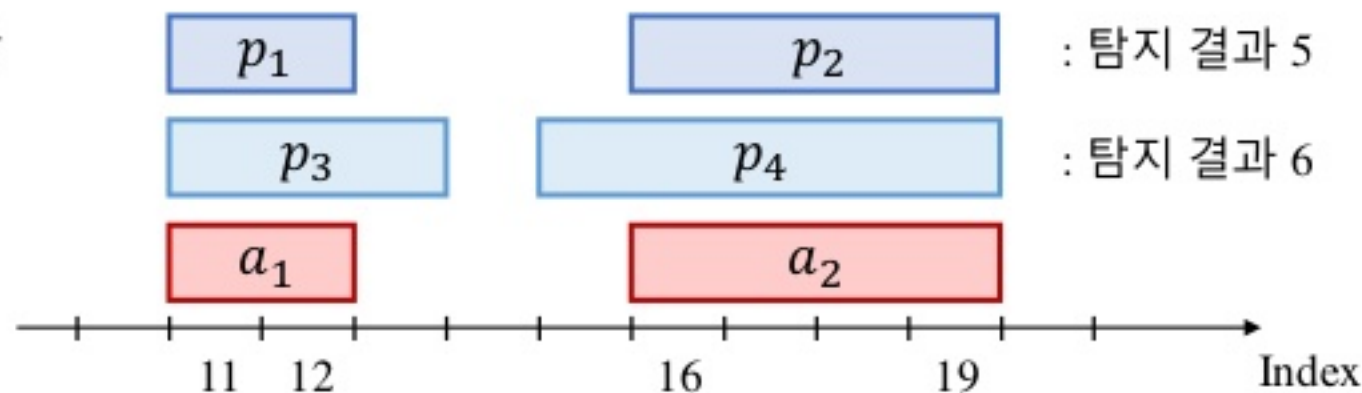
- 목표 2. 탐지의 정확성

- 범위 1과 2에 해당하는 시각(6초)을 모두 탐지할수록 높은 점수 부여
- 오른쪽 그림에서 결과 3이 4보다 높게 평가



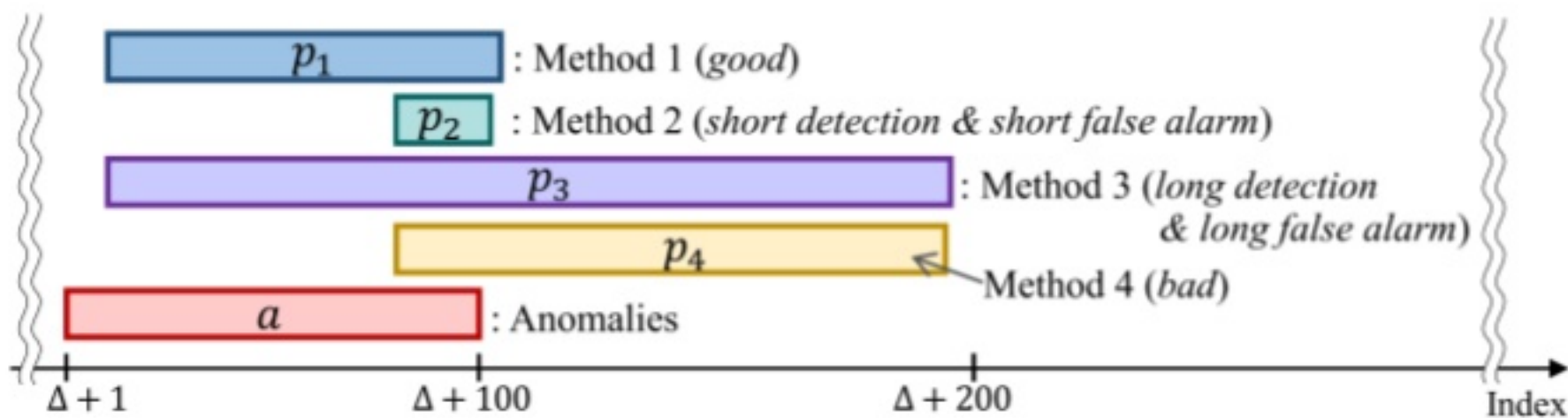
- 목표 3. 낮은 오탐

- 정답 외의 시각을 탐지하지 않을수록 높은 점수 부여
- 오른쪽 그림에서 결과 5가 6보다 높게 평가



TaPR

- TaPR은 상기 목표들을 만족하는 평가를 제공
 - TaP와 TaR로 구성
 - TaP: 예측 결과가 **오탐 없이** 이상 징후를 찾아내는가? (점수: 0 ~ 1)
 - TaR: 얼마나 **다양한 공격 범위**를 찾아내는가? (점수: 0 ~ 1)
 - 각 점수에 대해서는 아래 그림과 표 참고
- 본 대회에서는 TaP와 TaR의 F1 점수로 최종 평가



예측 결과	TaP	TaR
p_1	High	High
p_2	High	Low
p_3	Low	High
p_4	Low	Low

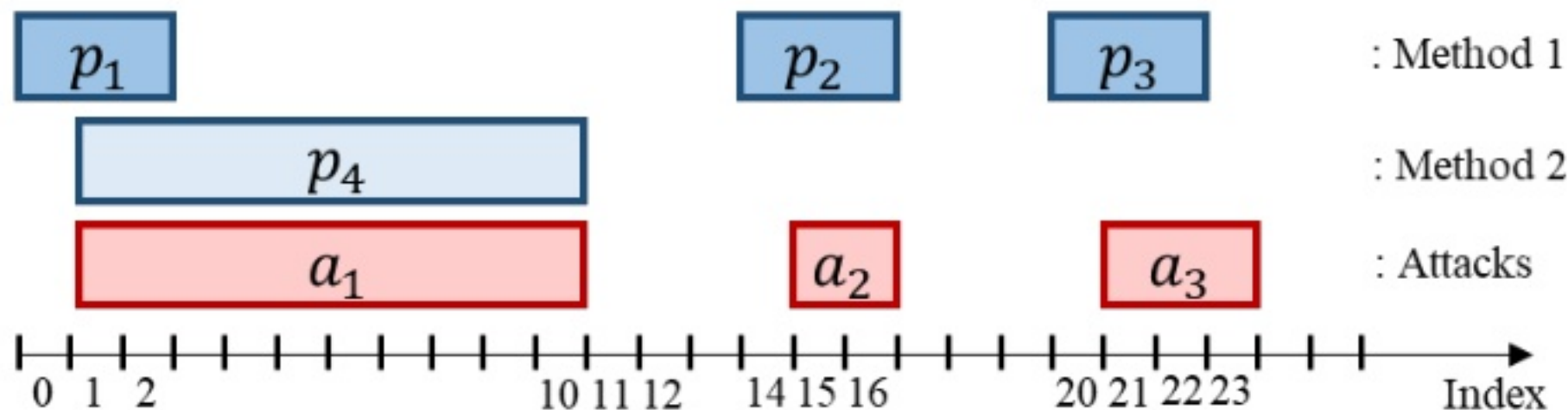
기타

- 평가 척도의 자세한 내용은 다음 논문 참고
 - W.-Hwang et al. “Time-Series Aware Precision and Recall for Anomaly Detection: Considering Variety of Detection Result and Addressing Ambiguous Labeling,” In *Proc. of CIKM*, pp. 2241-2244, 2019.
- HAICon 2020에 사용하는 TaPR은 상기 논문에서 제안한 방법을 **개선한 버전**
 - **대회 홈페이지에서 제공하는 버전만을 사용!**
 - 제공되는 whl파일을 설치 (예제 참고)
 - 그 외의 경로에서 찾은 TaPR은 다른 결과를 도출할 수 있음

추가: Precision 및 Recall을 사용하지 않는 이유

- Precision 과 recall은 널리 사용되는 평가 방법 중 하나
- Precision 과 recall은 다양한 공격을 탐지 하였는지 평가하는데 부적합
 - 방법 2는 가장 긴 시간 동안 영향을 준 공격 a_1 만을 탐지하였으나, 높은 점수를 받음

Method	Metric	
	Precision	Recall
1	0.67	0.40
2	1.00	0.67



추가: TaPR의 파라미터

- TaPR의 파라미터는 제공된 whl 파일에 설정하여 제공
 - 원본 TaPR에는 세팅할 파라미터들이 존재
 - 파라미터는 대회 목적에 맞도록 설정
 - 공격 범위 중 **일부만 찾아도 높은 점수** 부여 (TaR 점수는 후하게 측정)
 - 오탐은 **조금만 발생해도 큰 감점** (TaP 점수는 박하게 측정)
 - 그 결과, 아래 그림의 방법 2도 높은 점수를 받을 수 있음

