Adaptive User Profiling with Online Incremental Machine Learning for Security Information and Event Management

Dilli P. Sharma*, Barjinder Kaur*, Farzaneh Shoeleh*, Masoud Erfani*, Duc-Phong Le[ɟ],
Arash Habibi Lashkari*, Ali A. Ghorbani*
*Canadian Institute for Cybersecurity, University of New Brunswick, Canada
[ɟ] Bank of Canada, Ottawa, Canada

Presenter: Dilli P. Sharma
Email: dilli.sharma@unb.ca

November 14-18, 2021
Athens, Greece

About myself

- **Personal Details:**
  - Name: Dr. Dilli P. Sharma
  - Work: Postdoctoral Research Fellow at University of New Brunswick, Canada.
  - Study: Ph. D. (Computer Science), University of Canterbury, NZ
  - Email: dilli.sharma@unb.ca
- **Research Interest:**
  - Cybersecurity Analysis
  - Intrusion Detections
  - Moving Target Defense Techniques
  - Security Metrics
  - IoT Security Analysis
  - Applications of Machine Learning and Deep Learning in Cybersecurity
- **Selected Publications:**
  - "Dynamic Security Metrics for Software-Defined Network-based Moving Target Defense", *Journal of Network and Computer Applications*, 2020.
  - "Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense", *IEEE Communications Surveys and Tutorials*, vol. 22, no. 1, pp. 709-745, Firstquarter 2020.
  - "FRVM: Flexible Random Virtual IP Multiplexing in Software-Defined Networks," (*TrustCom-2018)*

Introduction

CIC

- Security-centric user profiling:
  - Monitors and analyzes the user activities
  - It also helps to identify the malicious user behavior
- However, the user behavior is unpredictable:
  - Change over time
  - Dynamic nature of the user
- State-of-the-art user profiling models lack to capture the dynamic user behavior as they are static
- **Our proposed approach**:
  - An anomaly detection-based adaptive user profiling model that dynamically learns user behavior from the user activities and updates the model over time

Proposed Framework for User Profiling

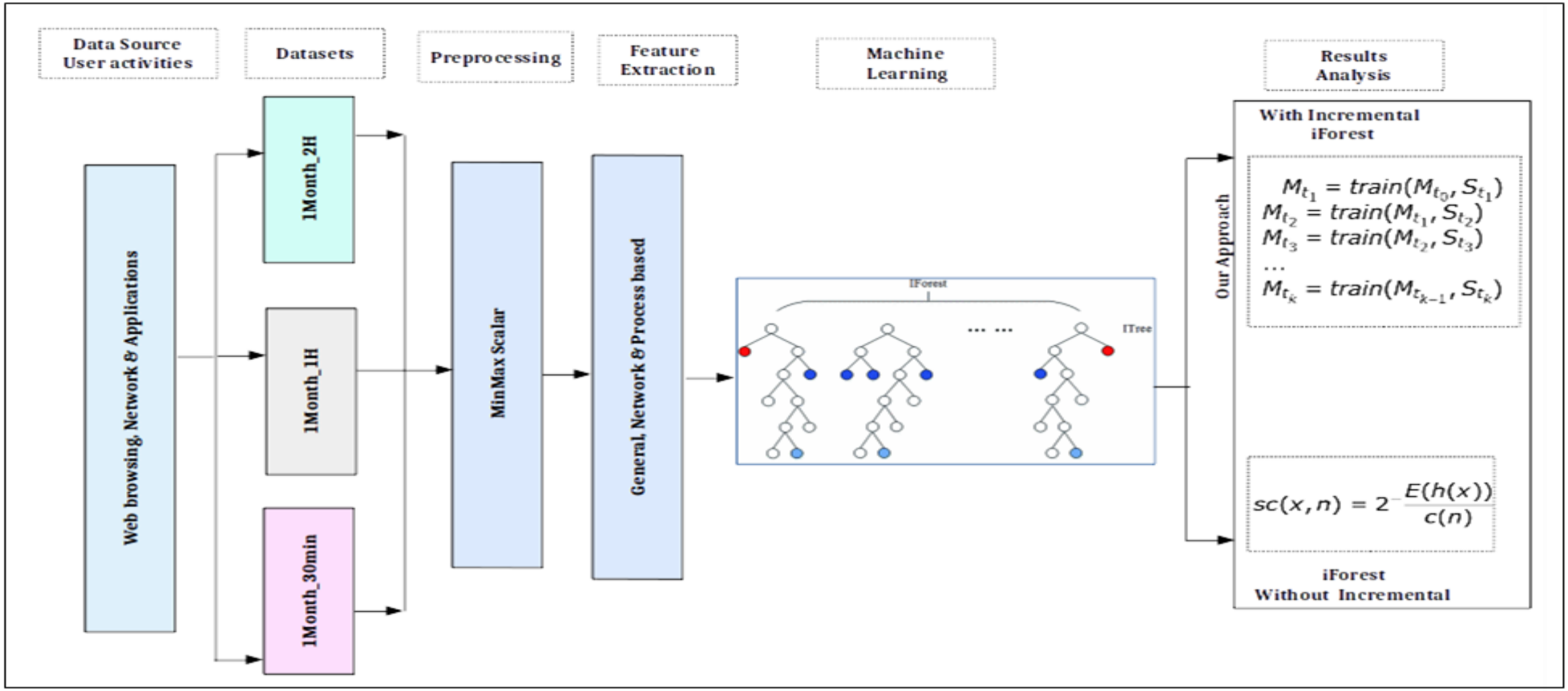

Fig 1: User profiling framework with the proposed incremental approach

Proposed Online Incremental Machine Learning Model

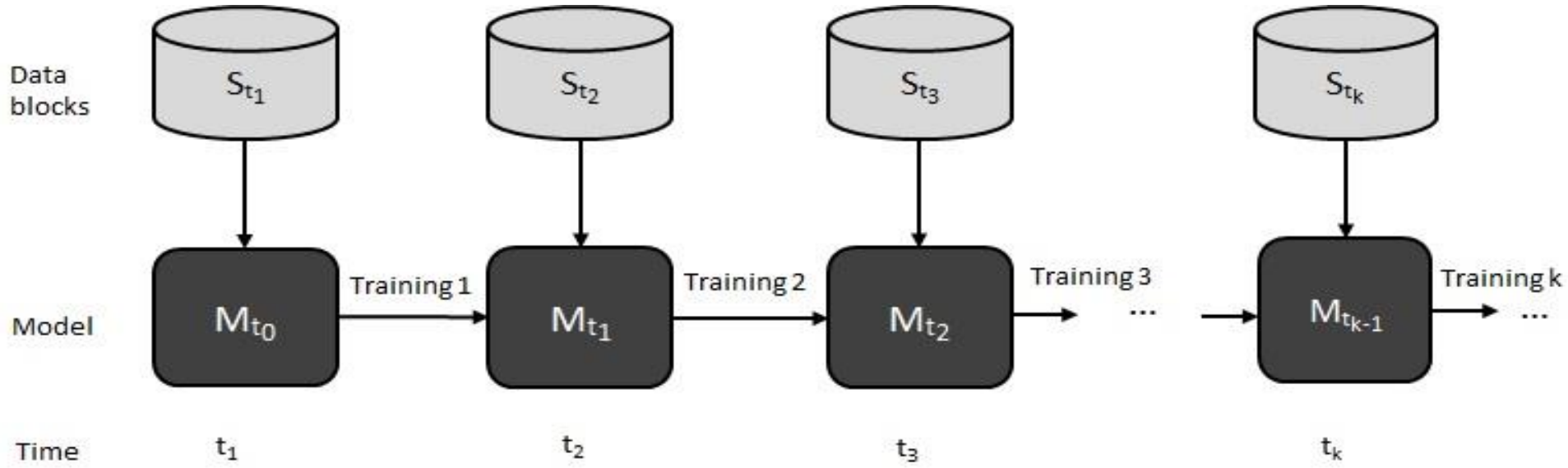

Fig 2: Evolving a model (machine) with online incremental learning

$$
\begin{aligned}
M_{t_1} &= train(M_{t_0}, S_{t_1}) \\
M_{t_2} &= train(M_{t_1}, S_{t_2}) \\
M_{t_3} &= train(M_{t_2}, S_{t_3}) \\
&\dots \\
M_{t_k} &= train(M_{t_{k-1}}, S_{t_k})
\end{aligned}
$$

Dataset Description

- **Datasets:**
  - 1Month_30min
  - 1Month_1H
  - 1Month_2H
- **User activities:**
  - Network
  - Web
  - Application
- All the user activities recorded between 8:00 AM to 5:00PM

TABLE I: SUMMARY OF THE DATASETS

Name	Time period	Session duration	Number of instances	Features
1Month_30min	One month	30 minutes	2280	44
1Month_1H	One month	1 hour	1116	44
1Month_2H	One month	2 hours	560	44

Experimental Setup

CIC

- We used sklearn[1] API for the experimentation
- Baseline model: **Isolation Forest Classifier**
- For each dataset:
  - **90**% data used for training and **10**% used for testing
  - Training data is further divided into **10** blocks (**3**-day time window)
- Trained both ***non-incremental*** and ***our proposed incremental*** models
- **Incremental Training and Updating the model**
  - Incremental training adds **more estimators** in every training and **updated the model**
- **Number of anomalies detections** after each training are recorded

[1]https://scikit-learn.org/stable/index.html

Results & Analysis

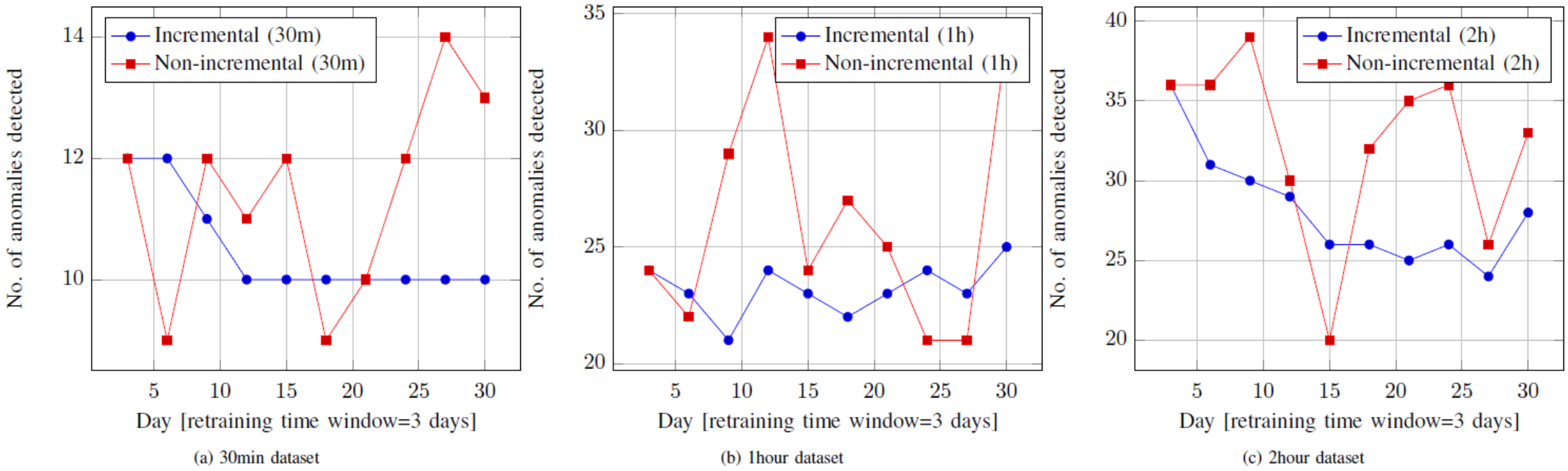

Fig. 3: Comparison of results of our proposed incremental model with a non-incremental model.

Results & Analysis (cont.)

TABLE II: SUMMARY OF THE COMPARATIVE PERFORMANCE ANALYSIS OF RESULTS

Model	Datasets	# of anomalies, # of new anomalies after each (re)training									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Our Incremental using iForest Model	1month_30min	12	12 (0) -	11 (0) ↓	10 (0) ↓	10 (0) -	10 (0) -	10 (0) -	10 (0) -	10 (0) -	10 (0) -
	1month_1h	24	23 (2) ↓	21 (0) ↓	24 (4) ↑	23 (0) ↓	22 (0) ↓	23 (1) ↑	24 (2) ↑	23 (0) ↓	25 (2) ↑
	1month_2h	36	31 (1) ↓	30 (1) ↓	29 (1) ↓	26 (0) ↓	26 (0) -	25 (0) ↓	26 (1) ↑	24 (0) ↓	28 (4) ↑
Non-Incremental using iForest Model	1month_30min	12	9 (0) ↓	12 (3) ↑	11 (0) ↓	12 (2) ↑	9 (0) ↓	10 (1) ↑	12 (2) ↑	14 (2) ↑	13 (0) ↓
	1month_1h	24	22 (2) ↓	29 (11) ↑	34 (12) ↑	24 (1) ↓	27 (7) ↑	25 (3) ↓	21 (2) ↓	21 (6) -	34 (16) ↑
	1month_2h	36	26 (1) ↓	39 (16) ↑	30 (3) ↓	20 (1) ↓	32 (17) ↑	35 (12) ↑	36 (7) ↑	26 (2) ↓	33 (9) ↑

Conclusion

CIC

- We proposed **an online incremental anomaly detection-based user profiling model** for Security Information and Event Management (SIEM) systems.
  - The proposed model **dynamically learns from the user activities and updates the model incrementally over time**.
- We validated the performance of the proposed incremental approach against the non-incremental baseline model **in terms of adaptability of user activities** for **3-different datasets.**
- The experimental results **proved that our proposed incremental model outperformed its baseline counterpart model.**
- Our findings suggest that the proposed model **should be applied more opportunistically to profile users as SIEM system component**.

Thank you !