

# סקירת ספרות: זיהוי וסיווג מתקפות סייבר בסביבות ענן באמצעות למידת מכונה

ספיר שנהב ועופרי גרוס

## 1. אבטחת מידע בסביבות ענן

סביבות מחשוב ענן כגון AWS, Azure, ו-GCP הפכו בעשור האחרון לתשתית מרכזית וקריטית עבור ארגונים ומערכות מודרניות. המעבר לענן הביא עמו יתרונות רבים של גמישות וגידול, אך במקביל יצר מרחב תקיפה חדש ומורכב. סביבות אלו מייצרות לוגים מפורטים מאוד, כגון AWS CloudTrail המתעדים באופן רציף קריאות API, זהויות משתמשים (IAM Users), שירותים (כגון EC2, S3) אזורים גאוגרפיים וזמני ביצוע.

למרות עושר וזמינות המידע, נפח הנתונים העצום והמורכבות ההתנהגותית של משתמשים לגיטימיים מקשים מאוד על ניתוח ידני או מבוסס כללים פשוטים של הלוגים. כתוצאה מכך, זיהוי מתקפות בסביבות ענן הפך לאתגר מחקרי וטכנולוגי משמעותי. כפי שמציינים **Sharma et al. (2017)**, האתגר המרכזי בענן אינו רק איסוף הנתונים, אלא היכולת להבחין בין פעילות עסקית רגילה לבין דפוסי פעילות זדוניים המנסים להסתוות כפעילות תקינה. המחקר מדגיש כי מתקפות בענן הן לרוב תהליכים מתמשכים ולא אירועים נקודתיים, מה שדורש גישות ניתוח מתקדמות.

## 2. גישות לזיהוי מתקפות מבוססות לוגים

### 2.1 גישות מבוססות חוקים (Rule-Based)

מערכות אבטחה מסורתיות, כגון מערכות IDS (Intrusion Detection Systems) קלאסיות, נשענות על חוקים או חתימות ידועות מראש לצורך זיהוי פעילות חשודה. חוקים אלו עשויים לכלול זיהוי של גישה לפעולות רגישות במיוחד, ניסיונות הרשאה חריגים (Unauthorized access) או דפוסי שימוש לא שגרתיים במשאבי מחשב.

למרות פשטותן ויכולת ההסבר הגבוהה שלהן (שכן קל להבין מדוע חוק מסוים "קפץ"), גישות אלו סובלות ממגבלות קשות בסביבות ענן דינמיות. הן מתקשות להתמודד עם מתקפות "יום אפס" (Zero-day) או עם שינויים קלים בהתנהגות התוקף, ודורשות תחזוקה ידנית מתמדת של בסיס החוקים. בנוסף, כפי שמציין המאמר של **Sommer & Paxson (2010)**, מערכות מבוססות חוקים מתקשות לזהות מתקפות מורכבות המורכבות מרצפים של פעולות שנראות לגיטימיות כשלעצמן (Living off the Land) אך יחד הן מהוות וקטור תקיפה.

### 2.2 למידת מכונה (Machine Learning) לזיהוי מתקפות

כדי לגשר על פערי הגישות המסורתיות, המחקר האקדמי והתעשייה פנו לשימוש באלגוריתמים של למידת מכונה לניתוח לוגים. לפי **Kim et al. (2018)**, שיטות אלו מאפשרות ללמוד את ההתנהגות הנורמלית של המערכת ולזהות חריגות (Anomalies) באופן אוטונומי. הגישות נחלקות לשתי קטגוריות:

- **למידה לא־מונחית (Unsupervised):** שמטרתה לזהות חריגות ללא ידע מוקדם או לייבלים, בדרך כלל באמצעות Clustering או Outlier Detection.
- **למידה מונחית (Supervised):** שבה נעשה שימוש בנתונים היסטוריים הכוללים לייבלים המסמנים פעילות תקינה לעומת זדונית.

בלמידה מונחית, המודל לומד לקשר בין מאפיינים סטטיסטיים (Features) לבין סיווג האירוע. זוהי הגישה שנבחרה לפרויקט זה, שכן היא מאפשרת דיוק גבוה בזיהוי סוגי מתקפות מוכרים ומתבססת על ידע מוקדם של דפוסי תקיפה בענף.

### 3. ניתוח התנהגותי מבוסס חלונות זמן

אחד התובנות המרכזיות במחקר של Lopez et al. (2019) היא שמתקפות ענף אינן מתרחשות לרוב כאירוע בודד. תוקף המבצע סריקה (Reconnaissance) או תנועה רוחבית (Lateral Movement) מבצע סדרה של פעולות לאורך זמן. לכן, ניתוח של שורת לוג בודדת אינו מספק מספיק הקשר (Context) כדי לקבוע אם מדובר בתקיפה.

הפתרון המקובל הוא איחוד אירועים בודדים לתוך **חלונות זמן קבועים** (Time Windows). הניתוח מתבצע ברמת החלון, כאשר לכל חלון מופקים מאפיינים מצטברים כגון:

- מספר האירועים הכולל בחלון הזמן.
- מספר פעולות API ייחודיות (מגוון הפעולות של המשתמש).
- מספר השירותים והאזורים (Regions) השונים אליהם פנה המשתמש.
- מספר כתובות ה-IP השונות ששימשו לביצוע הפעולות.
- נוכחות של פעולות ניהול (Management events) או פעולות כתיבה/מחיקה רגישות.

גישה זו מאפשרת למודלים של למידת מכונה לזהות דפוסים מתמשכים ולהפחית רעש הנובע מפעולות בודדות שאינן מעידות על כוונה זדונית. בפרויקט הנוכחי יושמה גישה זו באמצעות חלונות זמן באורך של 20 דקות, שהוכחו במחקרים קודמים כאיזון נכון בין זיהוי מהיר לבין צבירת מספיק נתונים לאפיון התנהגות.

### 4. מודלים לסיווג מתקפות – הרחבה והשוואה

במסגרת המחקר, התמקדנו בהשוואה בין שני מודלים מרכזיים המייצגים גישות שונות בלמידה מונחית:

#### 4.1 רגרסיה לוגיסטית (Logistic Regression)

רגרסיה לוגיסטית היא מודל בסיס (Baseline) נפוץ מאוד בתחום אבטחת המידע בזכות פשטותו, מהירות האימון שלו ויכולת ההסבר (Interpretability). המודל מניח קשר לינארי בין המאפיינים (כמו מספר ה-IPs) לבין ההסתברות שהחלון הוא "תקיפה". בפרויקט זה, הרגרסיה הלוגיסטית שימשה כקו בסיס. למרות שהשיגה

תוצאות סבירות, היא מוגבלת ביכולתה לזהות קשרים מורכבים – למשל, מצב שבו מספר פעולות הוא תקין באזור אחד אך נחשב לתקיפה באזור אחר.

## Random Forest 4.2

מודל ה-Random Forest הוא אלגוריתם מסוג Ensemble המשלב מספר רב של עצי החלטה. לפי הספרות, מודלים מבוססי עצים מתאימים במיוחד לנתוני לוגים של ענן מכמה סיבות:

1. **התמודדות עם נתונים הטרוגניים:** לוגים מכילים ערכים מספריים וקטגוריאליים גם יחד.
2. **עמידות לחוסר איזון (Data Imbalance):** בעיית יסוד בזהוי מתקפות היא שרוב הדאטה הוא "תקין". Random Forest יודע להתמודד טוב יותר עם מחלקות מיעוט.
3. **זיהוי אינטראקציות לא-ליניאריות:** המודל מסוגל לזהות שילובים מורכבים של מאפיינים המעידים על תקיפה. בפרויקט זה, ה-Random Forest נבחר כמודל המרכזי לאחר שהציג ביצועים עדיפים משמעותית, במיוחד במדד ה-F1 המאזן בין דיוק (Precision) לרגישות (Recall).

## 5. סיווג רב-מחלקתי (Multi-class Classification) ואתגריו

מעבר לשאלה הבינארית "האם יש תקיפה?", ארגונים זקוקים לידע על סוג התקיפה כדי להגיב נכון. מחקרים מתקדמים מחלקים את התקיפות לפי מסגרות עבודה כגון MITRE ATT&CK. סוגי התקיפות שנבחנו כוללים:

- **Reconnaissance** - סריקת הרשאות ומשאבים.
- **Cryptojacking** - ניצול כוח עיבוד של EC2 לכריית מטבעות.
- **Lateral Movement** - ניסיון לעבור בין זהויות ושירותים.
- **Billing Attacks** - יצירת משאבים יקרים כדי לגרום נזק כלכלי.

הסיווג הרב-מחלקתי מציב אתגרים טכניים משמעותיים: חוסר איזון קיצוני (סוגי תקיפה נדירים מאוד) וחפיפה התנהגותית. בפרויקט זה בוצע עיבוד מקדים שכלל Label Encoding וסינון רעשים (הסרת תקיפות עם פחות מ-2 מופעים). המודלים הושו באמצעות **Weighted F1 Score** כאשר ה-Random Forest הוביל עם ציון של 0.88 לעומת 0.82 של הרגרסיה הלוגיסטית, מה שמוכיח את עליונותו בטיפול בדאטה מורכב ולא מאוזן.

## 6. ייחוס תוקפים (Attribution) וניתוח פוסט-אנליטי

זיהוי של חלון זמן זדוני הוא רק תחילת האירוע. שלב קריטי נוסף, המודגש במחקרים של Zuech et al. (2015) הוא ייחוס התוקף – (Attribution) היכולת לקשר בין אוסף אירועים לבין זהות ספציפית או קמפיין. בפרויקט זה יושמה שכבת ניתוח פוסט-אנליטית המבצעת איגום (Aggregation) של תוצאות המודל. במקום להסתכל על כל חלון בנפרד, הסתכלנו על רצף החלונות המשויכים לכל משתמש. גישה זו אפשרה לנו:

- לזהות את התוקפים המרכזיים במערכת.
- להבחין בין תוקפים "רועשים" (שמבצעים פעולות רבות בזמן קצר) לבין תוקפים "שקטים" שפועלים לאורך זמן. הממצאים הראו כי מתוך 11 זהויות שזוהו כחשודות, רק 5 היו אחראיות לרוב המכריע של הפעילות הזדונית. ממצא זה מאשש את ההנחה המחקרית כי מספר קטן של תוקפים מיומנים מוביל קמפיינים משמעותיים במערכות ענן.

## 7. סיכום והקשר לפרויקט הנוכחי

הפרויקט הנוכחי מיישם הלכה למעשה את התובנות העולות מן הספרות האקדמית. שילבנו גישה מונחית המבוססת על חלונות זמן (20 דקות) עם מודלים של למידת מכונה (RF ו-LR). העבודה כללה Pipeline מלא: החל מעיבוד מקדים (StandardScaler) דרך אופטימיזציית Threshold להתמודדות עם חוסר איזון, ועד לשכבת הייחוס הפוסט-אנליטית. התוצאות, שהראו עדיפות ברורה ל-Random Forest מדגימות את החשיבות של בחירת מודל המסוגל להתמודד עם המורכבות הלא-ליניארית של לוגי ה-CloudTrail. שילוב זה מאפשר מעבר מזיהוי נקודתי של אירועים בודדים להבנה מערכתית ומושכלת של איומי סייבר בסביבת הענן.

מסקנה מההשוואה	#	תוצאה (Score)	מדד הערכה	קבוצת נתונים	מודל	שלב הניתוח
פער קטן מאוד בין אימון לולידציה מעיד על מודל יציב.	0.91	F1-Score	אימון (Train)		Random Forest	סיווג בינארי
	0.883	F1-Score	ולידציה (Val)			
תוצאות גבוהות בשני השלבים, אך המודל פשוט מדי לדפוסים מורכבים.	0.98	Accuracy	אימון (Train)		Logistic Regression	
	0.976	Accuracy	ולידציה (Val)			
המודל מצליח להכליל טוב גם כשיש הרבה סוגי תקיפה.	0.92	Weighted F1	אימון (Train)		Random Forest	סיווג רב-מחלקתי
	0.883	Weighted F1	ולידציה (Val)			
ידידה בביצועים במעבר לרב-מחלקתי עקב ליניאריות המודל.	0.84	Weighted F1	אימון (Train)		Logistic Regression	
	0.822	Weighted F1	ולידציה (Val)			

**Al-Zewairi, M., et al. (2020).** "Deep Learning for Cloud Intrusion Detection Systems: A Survey."

**AWS CloudTrail Documentation.** Official technical guide.  
<https://docs.aws.amazon.com/cloudtrail/>

**Buczak, A. L., & Guven, E. (2016).** "A Survey of Data Mining and Machine Learning Methods for Cybersecurity Intrusion Detection."

**Chawla, N. V. (2009).** "Data Mining for Imbalanced Datasets: An Overview."

**Injadat, M., et al. (2020).** "Multi-stage Optimized Machine Learning Framework for Intrusion Detection."

**Kim, J., et al. (2018).** "Detecting malicious activities in cloud environments using machine learning." *Computers & Security*.

**Lopez, M., et al. (2019).** "Behavior-based anomaly detection in cloud systems." *Future Generation Computer Systems (FGCS)*.

**MITRE ATT&CK Framework.** Cloud Matrix documentation. <https://attack.mitre.org/>

**Sharma, A., et al. (2017).** "Cloud security issues and challenges: A survey." *IEEE International Conference on Computing, Communication and Automation*.

**Sommer, R., & Paxson, V. (2010).** "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection." *IEEE Symposium on Security and Privacy*.

**Zuech, R., et al. (2015).** "Intrusion detection and Big Data Analytics - A Survey." *Journal of Big Data*.