תכנון והטמעת SIEM בארגון

מאת אריק יונאי

הקדמה

במאמר זה אסקור דרכים ואפשרויות להטמעת מערכת SIEM בארגון. מי שקרא את מאמרי הקודם, יודע שאני משתדל למעט בהגדרות המילוניות היבשות, אך אסקור בקצרה לריענון הזיכרון:

> מערכת SIEM היא מערכת המאפשרת שליטה ובקרה (שו"ב) על אירועי אבטחת מידע בארגון, ע"י איסוף התראות מרכיבים שונים, וע"י חיבור ותעדוף של ההתראות, מייצרת תמונת מצב כוללת של מצב אבטחת המידע בארגון (לרוב מצב אבטחת המידע בארגון, אך בהחלט ניתן להשתמש במערכת SIEM גם בכדי לתת מענה לצוותים אחרים בארגון).

הטרמינולוגיה משתנה מעט ממוצר SIEM אחד למשנהו, אך העיקרון זהה אצל כולם.

מערכת ה-SIEM תאסוף / תקבל לוגים מרכיבים שונים, כגון: Firewall, IPS, VPN, שרתי Windows, שרתי Web, DB, וכו', ותעביר את הלוגים תהליך (Correlation בין השאר) בכדי לקבוע בסופו של דבר מה לעשות עם הלוג, והאם להפוך אותו ל-Event.

Event יהיה אירוע (או התראה), שייווצר מלוג או ממספר לוגים שונים, ושיש לו חשיבות כלשהי ולרוב גם ידרוש התייחסות או תגובה.

ליוויתי מספר פרוייקטים של הטמעת SIEM בשלבים שונים, ואני חייב לציין שהטמעת מערכת SIEM לרוב היא הטמעה מורכבת וארוכה, הדורשת הרבה מבחינותרבות (כגון כסף, זמן, סבלנות, והמון ניסיון מקצועי). לעומת מוצרים רבים, בכדי להביא מערכת SIEM למצב של פעילות "איכותית" (התראות "אמיתיות" של אירועי אבטחת מידע הדורשים תגובה), יש צורך בהמון זמן הטמעה ותחזוקה, וצורך ביצרן איכותי + אינטגרטור מקצועי עם המון ניסיון במוצר הספציפי.

לאחר ליווי של מספר הטמעות SIEM, אני יכול לומר שרובן נכשלו (בפרספקטיבה של הדרישות המקוריות שהיו מהמוצר), ובסוף נשארה קופסה שחורה בחדר שרתים, בעל ערך מועט ביותר, ממספר סיבות שאפרט בהמשך. אנסה לומר זאת בעדינות, כנראה שאין פתרון SIEM אשר "מחברים והוא עובד", או כפי שאינטגרטורים מסויימים נוהגים לומר: "תן לי יומיים עבודה - תראה אילו התראות אני מוצא לך!".

אז זהו, שכנראה אין חיה כזאת בתחום ה-SIEM.

ועכשיו, נצלול פנימה ☺

סיבות להטמעת SIEM בארגון

אסקור את שתי הסיבות העיקריות להטמעת מערכת SIEM:

1) רגולציה - הסיבה הנפוצה להטמעת SIEM בארגון. אני מעריך שרוב הארגונים מתחילים את ההטמעה של מערכת SIEM מתוך דרישה של רגולציה כזו או אחרת, המחייבת את הארגון לאחסן ולנטר את הלוגים המכילים פעולות הנוגעות למידע רגיש, במקום מרכזי, ושתהיה לארגון אפשרות לקבל תמונת מצב כוללת על מצב אבטחת המידע בארגון, לרוב ע"י שליחת דו"חות מתוזמנים ממערכת ה-SIEM. לרוב, הטמעות מסוג זה הן פשוטות וקצרות יותר מהאפשרות שאציג מיד, ולרוב המטרה היא לסמן "V" על דרישות רגולציה. הערך של פתרון ה-SIEM בארגון בשיטה זו, מנקודת מבט של אבטחת מידע, יהיה מוגבל ברוב המקרים.

2) רצון לשפר את אבטחת המידע בארגון - במצב זה, בניית פתרון ה-SIEM יבוצע בצורה "נכונה" יותר ברוב המקרים, אך ההטמעה תהיה מורכבת יותר. ארגונים המעוניינים לשפר את מערך אבטחת מידע בוחרים פעמים רבות להטמיע מערכת SIEM, כפתרון "שליטה ובקרה" (שו"ב), המאפשר ניהול אירועים מנקודה מרכזית. כמו כן, מערכת SIEM גם מקנה יכולות תחקור (Forensic), כך שבמקרה אירוע ניתן לחפש מידע לאחור.

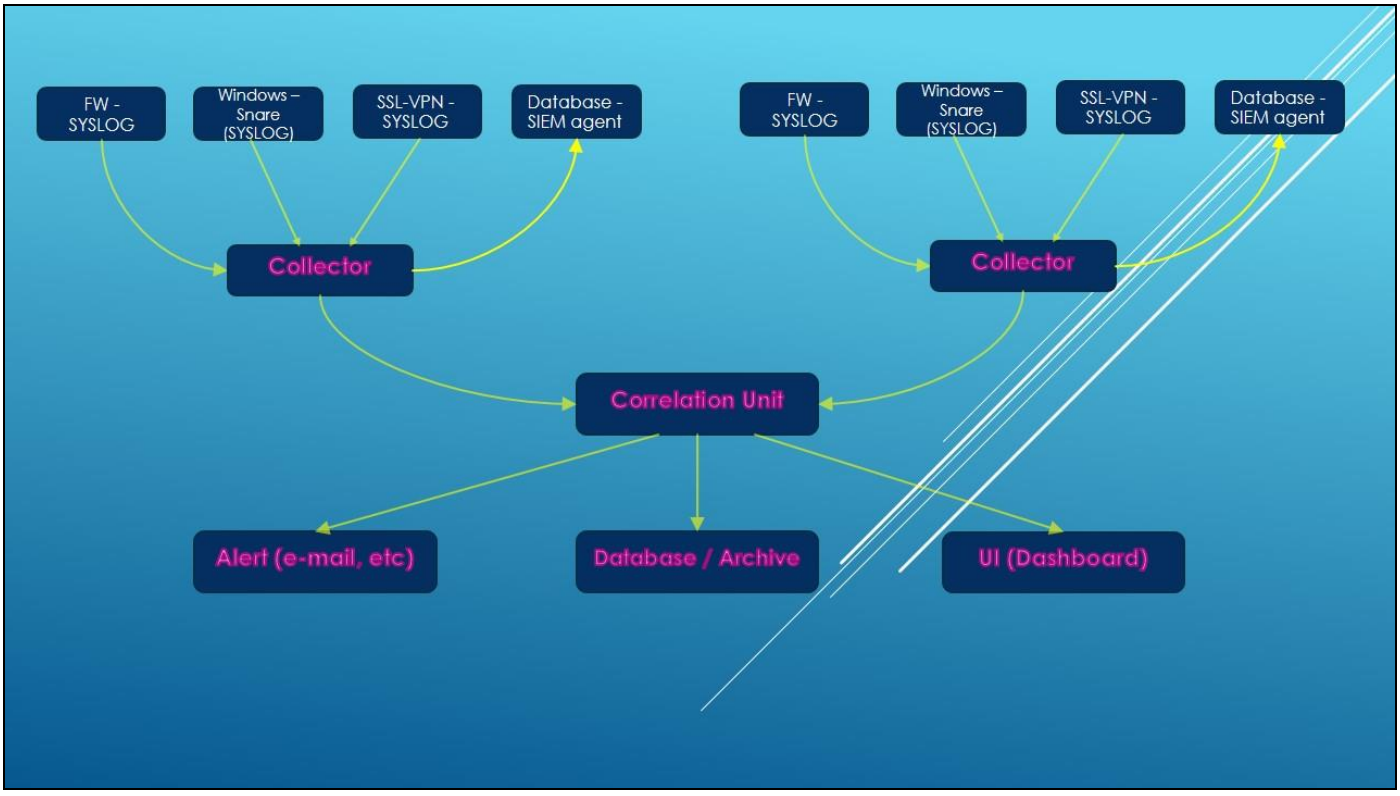
הבעיה והפתרון

כמובן שאין באפשרותנו לעבור על כל הלוגים הנכתבים במערכות הארגון, ובוודאי שלא בו-זמנית מרכיבים שונים, בזמן אמת. מערכת SIEM מאפשרת לקבל במקום מרכזי אחד, עם Dashboard אחד, את הלוגים מרכיבים שונים ברשת, ועליהם יבוצעו פעולות שיאפשרו בסופו של תהליך לקבוע האם הלוג יהפוך לאירוע (Event), והאם יש לטפל בו.

מבנה ורכיבי המערכת

ליצרני SIEM שונים יש שמות שונים לרכיבים שונים, אך אפרט מהם הרכיבים הכלליים המשותפים למערכות SIEM (כאמור השמות יכולים להיות שונים מיצרן ליצרן).

מערכת SIEM היא תוכנה המוכרבת ממספר רכיבים, לעיתים תהיה מותקנת על Appliance ייעודי, ולפעמים על Open server או שרתי VM. חלק מהפתרונות מבוססי Linux, וחלקם מבוססי Windows. אצל רוב היצרנים ניתן יהיה להפריד את הרכיבים השונים לשרתים נפרדים במקומות שונים ברשת.



[מבנה מערך SIEM סטנדרטי]

למערכת ה-SIEM יש כמה רכיבים:

1) **Collector** - רכיב אשר אמון על איסוף / קבלת הלוגים מהרכיבים השונים. רכיב זה לרוב יודע לקבל / לאסוף לוגים במספר פרוטוקולים, כאשר בדר"כ הנפוץ מביניהם יהיה SYSLOG.
בחלק מההתקנים ברשת נגדיר שליחת SYSLOG לרכיב (שרת) ה-Collector (כאשר החיבור יפתח מהתקני הרשת, לכיוון שרת ה-Collector).

בהתקנים אחרים ברשת (שרתי Database לצורך העניין), יתכן ויהיה צורך להגדיר Authentication ברכיב ה-Collector, אשר **יפנה לשרת** ה-Database **וימשוך את המידע הרלוונטי אליו.**

התקנים אחרים ישלחו מידע ע"י Agent אשר יותקן על אותו רכיב (לצורך העניין שרתי Windows), וה-Agent הוא זה שישלח את הלוגים הרלוונטים לרכיב ה-Collector בפרוטוקול כזה או אחר.

2) **Parsing** - רכיב אשר בדר"כ יהיה חלק מרכיב אחר.
תפקידו של רכיב זה הוא לפרק את הלוגים "הגולמיים" (לא מחולקים לשדות, ועל כן חסרי משמעות למערכת ה-SIEM), לשדות מוגדרים אשר יוכלו להיכנס לטבלאות קבועות ב-DB, בכדי שניתן יהיה לתת להם משמעות ולהגדיר על-פיהם תסריטים (יפורט בהמשך).

הלוג הבא (במצב "נקי", ללא Parsing), מדגים כיצד ה-Collector מקבל לוג מ-Firewall ב-SYSLOG:

```
"14:55:20 accept gw.foobar.com >eth1 product VPN-1 & Firewall-1 src
10.5.5.1 s_port 4523 dst 10.10.10.2"
```

ובכן, כנראה שלרובנו יהיה קל יחסית להבין מה כתוב פה פחות או יותר, כי ראינו כאלה ואנחנו מסוגלים להשתמש בהיגיון בריא בכדי לנתח את המידע. ובכן, ל- SIEM, כך מסתבר, לא יהיה קל לנתח את המידע.

מערכת ה-SIEM זקוקה ל-Parsers, מערך של Regular expressions, שיעניק משמעות לחלקי הלוג השונים. לכל פתרון SIEM קיים מערך Parsers, אשר מתרגם את הלוג הגולמי, ללוג בעל משמעות.
בדוגמא זו, הלוג (ללא Parsing):

```
"14:55:20 accept fw-1.test.com >eth1 product VPN-1 & Firewall-1 src
10.5.5.1 s_port 4523 dst 10.10.10.2"
```

לאחר Parsing יראה פחות או יותר כך:

Time of event	Action	Firewall IP	Interface	Source	Source port	Destination
14:55:20	accept	fw-1.test.com	eth1	10.5.5.1	4523	10.10.10.2

אתייחס לבעיות צפויות בתחום ה-Parsing בהמשך המאמר.

3) **Correlation Unit** - הרכיב מבצע ניתוח של הלוגים (לאחר שלב ה-Parsing), ומחליט האם ליצור התראה (Event). ה-Correlation Unit **מורכב מתסריטים (Scenarios)**, שהם לצורך העניין ליבת מערכת ה-SIEM.

תסריט הוא קבוצה של **חוקים**, המגדירים ל-Correlation Unit כיצד להתייחס ללוגים השונים.

לצורך הדוגמא, **אציג תסריט פשוט, המוגדר מחוק אחד בלבד, כאשר מטרתו היא להתריע כאשר נוצר Domain Admin חדש ב-Active Directory.**

החוק אומר כך:

אם:

א. המקור הוא Active Directory (ניתן להגדיר את החוק ע"פ סוג הרכיב, כתובת IP, וכו').

ב. ומגיע לוג המכיל פעולה X (יצירת Domain Admin חדש).

ג. הפעולה התרחשה בין השעות 21:00 ל-7:00 (החלטתי שאלו שעות חשודות שבהן אינני מצפה לפעילות כזו, לצורך הדוגמא).

אז: שלח התראה (דוא"ל לצורך הדוגמא) למנהל הרשת עם פירוט הפעילות (מי יצר, מתי, וכו'), ברמת דחיפות Medium.

דוגמא ל**תסריט נוסף, המורכב משני חוקים (ששניהם חייבים להתקיים יחדיו). המהות של התסריט, היא למנוע מצב שבו אדם יגנוב זהות של עובד ויצליח להתחבר ב-VPN מחוץ למשרד, בעוד שהעובד נמצא בתוך המשרד.**

חוק א':

אם:

א. המקור הוא Active Directory.

ב. מגיע לוג המאשר שמשתמש X ביצע פעילות בדומיין, **מתוך רשת המשרד**.

חוק ב':

אם:

א. תנאי: חוק ב' יתקיים **רק אם חוק א' התקיים בחצי שעה האחרונה** (אם חוק א' התקיים לפני יומיים, חוק ב' יהיה חסר משמעות, ועל כן כל התסריט לא יופעל).

ב. המקור הוא SSL-VPN.

ג. משתמש X הזדהה בהצלחה מול ה-VPN.

אז: אם חוק א' + חוק ב' התקיימו ביחד, אז שלח התראה למנהל אבטחת מידע בארגון + המשתמש שביצע את הפעולה (בכדי לקבל מהמשתמש תגובה לגבי הפעילות: האם הוא זה שביצע את הפעולה, או זהותו נגנבה?), עם פירוט הפעילות, בדרגת דחיפות Critical.

ודוגמא אחרונה לתסריט קצר המורכב מחוק אחד (תלוי במוצר SIEM), שמטרתו היא **להתריע בפני Brute force על SSL-VPN מכתובת מקור אחת**.

החוק:

אם:

א. המקור הוא SSL-VPN.

ב. מגיע לוג עם הודעת "Failed login" מכתובת Public IP כלשהי.

אז: שלח התראה למנהל אבטחת המידע בארגון, **אך התנאי הוא שהחוק יופעל כ-30 פעמים בטווח זמן של חמש דקות** (אחרת החוק והתסריט לא יופעלו).

מערכת SIEM תכיל לרוב **מספר תסריטים שונים**, שכל אחד מהם **יכיל חוק אחד או כמה חוקים.**

ע"י מערך של תסריטים, יכול הארגון לקבל התראות ולהגיב במהירות לתסריטים אותם הוא הגדיר כקריטיים להמשך פעילות תקינה של הארגון.

מחזור חיים סטנדרטי של התראה:

- ☒ הלוגים נשלחים / נאספים ע"י ה-Collectors מהרכיבים השונים הרשת.
- ☒ הלוגים עוברים תהליך של Parsing (ע"פ Regular expressions), ומוכנסים לטבלאות הרלוונטיות (Date, IP addresses, actions, etc.).
- ☒ הלוגים מועברים (לאחר ה-Parsing) ל-Correlation Unit, **ונבדקת התאמה בין הלוגים לתסריטים השונים.**
- ☒ במידה וקיימת התאמה בין לוגים לתסריט (במידה והלוגים עונים על הדרישות של החוקים באחד התסריטים), מתבצעת פעולה, לרוב שליחת התראה למשתמש (קיימים מוצרי SIEM אשר יודעים "להגיב אוטומטית" לאירועים, כגון להריץ סקריפטים, וכו').
- ☒ לאחר מכן הלוגים / ההתראות ישלחו ל-Database (יפורט בהמשך), לצורך אפשרויות תחקור או Archive.

4) מאפיינים נוספים למערכות SIEM:

במערכות SIEM רבות יופיעו מספר רכיבים נוספים, אשר נותנים ערך מוסף חשוב ולפעמים קריטי בבחירת המוצר המתאים:

א. **Compression** - קיימים שני סוגים של דחיסה: הראשון - דחיסה של התעבורה, והשני - דחיסה של המידע המאוחסן (Database, logs, etc). כאשר עוברים לוגים רבים על גבי קווי WAN, דחיסת המידע יכולה להיות קריטית, במקרים מסויימים משהו שלא ניתן להסתדר בלעדיו.
בדקו היטב מה יכולות המוצר בנוגע לדחיסת המידע שעובר על גבי הרשת, במידה וקווי ה-WAN יקרים ללבכם.

הסוג השני - דחיסה של המידע על גבי Storage. מבלי להיכנס עמוק מדי לסוגיה זו, מערכת SIEM גם מקנה יכולות תחקור, מאחר וניתן לחפש לאחור לוגים / אירועים. לדחיסת המידע באחסון (ולא רק על גבי רשת התקשורת), יש משמעות גדולה, כי היא מגדילה את הזמן שניתן לתחקר לאחור אירועים.

ב. **תחקור (Forensic)** - לכל מערכת SIEM יש יכולת חיפוש / תחקור כזו או אחרת, אך איכותה הוא נושא חשוב מאוד. מה ניתן לחפש ואיך, והכי חשוב: **באיזו מהירות יזרמו התוצאות.** בזמן אירוע, נהיה מעוניינים לחפש כתובת IP, שם משתמש, וכו'. אני יכול לומר באחריות שקיימות מערכות SIEM בהן **תוצאות חיפוש של 24 שעות אחורה (על מיליוני שורות לוג), יגיעו רק לאחר שעות ארוכות ואף ימים (!)**, ואילו קיימות מערכות SIEM **אשר יהיה ניתן לחפש חודשים אחורה על כמות אדירה של שורות לוג, ולקבל את התוצאות תוך דקות בודדות (ואף בפחות).**

קחו בחשבון את נושא מהירות החיפוש במערכת השיקולים בבחירת מוצר SIEM (זה נושא יותר משמעותי ממה שהוא נראה על פני השטח, מאחר וחיפוש איטי לאחור יכול לפגוע משמעותית ביכולות התחקור).

ג. **Dynamic C&C database** (או IP Reputation) - רכיב אשר מכיל רשימה של כתובות IP ודומיינים זדוניים, ומתעדכן בתדירות גבוהה.
ניתן להגדיר תסריטים אשר יתריעו בפני תעבורה הנכנסת / יוצאת מהרשת לכתובות זדוניות, ובכך לזהות Malwares שונים (ראו מאמר בנושא "Malwares 2.0, ודרכי התמודדות בארגון": http://www.digitalwhisper.co.il/files/Zines/0x2A/DigitalWhisper42.pdf).
רכיב יעיל מאוד כנגד סוסים טרויאנים ותולעים.

ד. **Compliance Reports** - במידה והארגון נדרש לעמוד ברגולציה (או ידרש לעמוד באחת כזו בעתיד), כדאי לוודא שקיימים דו"חות Out-of-the-box לרגולציות שונות (רוב המוצרים תומכים בדו"חות מסוג זה).

בעיות צפויות ונקודות קריטיות ל-POC

העצה הטובה ביותר שקיבלתי לפני הטמעת SIEM, היה לעשות PoC מעמיק וארוך! אם יש מוצר שהערך של POC בו הוא הגדול ביותר, לדעתי זה מוצר SIEM.

הגדירו יעדים ומדדים ברורים ומדוייקים להצלחה!

1) אתחיל דווקא מהסוף: אולי אחד הדברים החשובים ב-POC / לאחריו, הוא כתיבת מסמך Incident response.

הגעתם ליעד המבוקש, דאגתם להגדיר את ה-SIEM היטב, תיקנתם, שיפצתם, קניתם, והמוצר עובד פיקס. מה עכשיו? בסופו של יום, יגיעו התראות מה-SIEM, ומישהו יצטרך לטפל בהן. המלצתי החמה היא ליצור מסמך שירכז כמה עשרות אירועי אבטחת מידע (נפוצים וחריגים), ושרשרת הפעולות שיש לבצע לאחר קבלת דיווח על אירוע מה-SIEM.

הקמת צוות SOC (או לפחות אדם שיודע שתפקידו הוא לטפל באירועי אבטחת מידע שיגיעו מה-SIEM), היא פעילות מורכבת ממה שנדמה, ומומלץ מאוד להגדיר פעולות מסודרות לפני שיגיעו

האירועים (מה אני עושה כשמתקבלת התראה על Brute force? מה אני עושה כשמשתמש מתחבר ב-VPN, למרות שהוא במשרד? כיצד מגיבים לאירוע DDoS וכו').

2) בדקו היטב את מודל הרישוי של המוצר, וסכמו לפני הרכישה על אפשרות גדילה. רוב מוצרי ה-SIEM נמדדים ב-EPS (Event per second), שהמשמעות היא כמה לוגים המוצר מסוגל לעבד בשניה. חלק מהמוצרים דווקא לא מגבילים כמות EPS ע"י רישוי (אלא מוגבלים למגבלות חומרה), אך הרישוי מגביל כמות Collectors, או רכיבים אחרים. סכמו מראש על מחירי שדרוג של חומרה / תוכנה / רישוי, במידה ויוצר הצורך בעתיד.

3) נושא שבוודאות ידרוש מכם להשקיע הרבה אנרגיה יהיה תחום ה-Parsers. רוב ה-Parsers, מגיעים מיצרן ה-SIEM. יחד עם זאת, Parsers נכתבים למוצר ספציפי, ולרוב גם לגרסה ספציפית. חשוב מאוד להבין את המשמעויות של העניין.

נניח כי יש לי בארגון Check Point Firewall, ומוצר ה-SIEM תומך ב-Check Point זה נהדר, אבל זה ממש לא מחייב שה-SIEM ידע לבצע Parsing נכון של הלוגים. יכול להיות שה-Parsers במוצר ה-SIEM נכתבו ל-Check Point R70, ואילו בארגון יש לי Check Point R75.40 (לדוגמא), ומבנה הלוגים הוא שונה לחלוטין.

המשמעות במקרה זה (או במקרים של תוכנה שאיננה "תוכנת מדף", אלא תוכנה שפותחה בתוך הארגון), **היא שתצטרכו ליצור את ה-Parsers בעצמכם**. לחלק ממוצרי ה-SIEM יש כלים חצי אוטומטיים לכתיבת ה-Parsers, אך בהחלט זהו נושא שחובה להתנסות בו בעצמכם במהלך ה-POC (מומחה Regular expressions מאוד יכולה לסייע), ולהבין האם כתיבת Parsers חדשים היא משימה לא מורכבת באופן יחסי, או משימה מאוד מאוד קשה.

אני מדגיש את החשיבות של ה-Parsers, מאחר וראיתי במו עיניי מקרים שבהם יצרן X כתב רשימה באורך הגלות של מוצרים נתמכים, ואילו בפועל, רק כ-30% מהמוצרים עברו Parsing כמו שצריך. המשמעות היא שבכדי לתקן את זה, צריך לכתוב עשרות אלפי Regular expressions, ובפועל, ניתן לזרוק את המוצר לפח.

4) במהלך ה-POC למדו כיצד לכתוב תסריטים, ובדקו אותם. לדוגמא, אם כתבתם תסריט שמתריע מפני Brute force על ה-VPN, בדקו בפועל שאכן ה-SIEM ידווח לכם על האירוע.

5) תנו ל-SIEM לאסוף מידע במהלך ה-POC מהרכיבים אותם תרצו לחבר לרשת, במשך כמה שבועות. לאחר מכן, בצעו חיפוש על פרמטר כלשהו (כתובת IP, שם משתמש, וכו'). ראו תוך כמה זמן אתם מקבלים את תוצאות החיפוש.

6) כתבו מסמך תסריטים מפורט ומסודר. במסמך, פרטו מהו התסריט (דיווח על התחברות משתמש דרך VPN, בעוד המשתמש נמצא ברשת המשרד, לדוגמא), אילו רכיבים יש לחבר ל-SIEM לצורך יישום תסריט זה (Active Directory + SSL-VPN), ואת כל שאר הפרמטרים הדרושים (לאחר כמה לוגים ה-SIEM ישלח את ההתראה, באילו שעות, באיזו עדיפות, וכו').

7) כאמור מערכת SIEM היא מערכת שמיועדת לצורכי אבטחת מידע, אך בהחלט ניתן להשתמש בה גם לשימושים נוספים, לצורך העניין לשימושי מחלקת ה-IT: קבלת התרעות על נפילות של שרתים, מעקב אחר פעולות System מסויימות, נפילת Services, ועוד. שימוש מערכת ה-SIEM גם ע"י צוותים נוספים בארגון, יכול להעלות את ערכו וחשיבותו של מוצר ה-SIEM בארגון.

לאמיצים בלבד - Open Source

לגיבורים שבנינו, קיימים פתרונות Open source חינמיים לחלוטין, או מאוד מאוד זולים (יכולים לעלות אחוזים בודדים ממחיר פתרון SIEM של יצרן), אך לרוב הם דורשים ידע מעמיק ביותר ב-Linux (לדעת לכתוב ifconfig ו-reboot לא יספיקו במקרה זה ☺).

בפתרונות אלו, בדר"כ לא יהיו את כל הרכיבים הקיימים בפתרונות בתשלום, אך יש להגדיר היטב את צרכי הארגון ולהשוואתם ליכולות המוצרים, בהחלט יתכן ויהיו פתרונות Open source אשר יתאימו לצרכי הארגון.

לסיכום

כמובן שיש עוד מידע רב על התחום, אך במאמר זה ניסיתי לתת את אבני היסוד להטמעת SIEM, בכדי להימנע מהטעויות הקרדינליות שרבים מאיתנו עשו, בעיקר בבחירת המוצר והאינטגרטור.

אני לא חושב שיש הטמעת SIEM קלה ומהירה (לא רק לחבר את ה-SIEM לחשמל, אלא SIEM שבאמת יעשה עבודה), ודורש תקופה ארוכה מאוד של Fine tuning, אבל בהחלט כדאי לבדוק האם SIEM יכול לסייע לכם.

בתחילת הדרך, כנראה ויגיעו אליכם יותר מדי התראות מהמוצר, ואז לאחר מכן, מדי התראות, ואז עוד פעם יותר מדי, עד שתגיעו למצב שבו אתם מקבלים כמות התראות שניתן להתמודד איתה. מטבעם של דברים, רוב ההתראות שתקבלו יהיו "False-positive", ולא באמת בעלות משמעות - היו חזקים, והמשיכו לבצע Fine tuning על התסריטים שהגדרתם עד שתגידו לתוצאה הרצויה! ואדגיש פעם נוספת את **חשיבות איכותו של האינטגרטור והניסיון המקצועי שלו!**

המצב הגרוע ביותר הוא שתקבלו אלפי התראות ביום, ואף אחד כבר לא יפתח אותן בכדי לבדוק מה יש בפנים.

תודה על תשומת הלב, אריק יונאי.