כללי- עבדתי בBQ, העליתי את הCSV כטבלה אצלי בשם Veritas-Data-Set והSyntax בהתאם.

# שאלה 1:

## קוד:

SELECT `proto`, COUNT(id) AS NUM\_OF\_EVENTS\_PER\_PROTO # ספירת שורות לכל פרוטוקול

 FROM `redjony.MyDB.Veritas-Data-Set`

 GROUP BY `proto`#הספירה היא לכל פרוטוקול

order by NUM\_OF\_EVENTS\_PER\_PROTO desc # מיון כך שהערכים הגבוהים יהיו עליונים

## תשובה:

Tcp- 43095

Udp-29418

# שאלה 2:

## קוד:

SELECT `proto`, SUM(sbytes+dbytes) AS SUM\_OF\_BYTES # סכימת התקשורת

FROM `redjony.MyDB.Veritas-Data-Set`

GROUP BY `proto` `#הספירה היא לכל פרוטוקול

order by SUM\_OF\_BYTES desc #מיון כך שהערכים הגבוהים יהיו עליונים

## תשובה:

Tcp-1724299718

ניתן להסיק משאילתא זו ומהקודמת, מכך שהפרוטוקול הוא הכי נפוץ גם במספר השורות וגם בנפח התקשורת- שאין שונות רבה מאוד בין גודל הפקטות ללא תלות בפרוטוקול, אחרת היינו מצפים לראות שפרוטוקול אחר שבו פקטות גדולות יותר יעקוף אותו.

# שאלה 3:

## קוד:

#דומה לקודם, רק שכאן אני מגביל את התוצאות ל5 הכי גדולים

SELECT id, sbytes/(dbytes+1) AS ratio

FROM `redjony.MyDB.Veritas-Data-Set`

ORDER BY ratio desc

LIMIT 5

## תשובה:

|  |  |
| --- | --- |
| id | Ratio |
| 65563 | 33728 |
| 65560 | 33728 |
| 62059 | 29920 |
| 62115 | 29920 |
| 62064 | 29920 |

# שאלה 4:

## קוד:

SELECT `proto`, AVG(sbytes/(dbytes+1)) AS avg\_ratio #חישוב ממוצע היחס ע"פ הגדרה שניתנה

FROM `redjony.MyDB.Veritas-Data-Set`

where

`proto` IN ('udp','tcp') #הגבלה לפרוטקולים tcp,udp.

group by `proto` #הספירה היא לכל פרוטוקול

## תשובה:

הנתון שהתקבל-  
proto avg\_ratio

udp 207.17629063514309

tcp 13.426664880965713

יש לכך סיבה טכנית אפשרית, tcp הוא Connection oriented, מייצר חיבור ולכן יש לכל פקטה פקטת ack, היחס המקסימלי האפשרי ביניהם הרבה פעמים יהיה מוגבל ל1500 (תאורטית אפשר הרבה יותר, בד"כ מוגבל) מול Ack של נגיד כמה עשרות בייטים נניח 50, זה 1/30, הגיוני שלא כל פקטה תהיה בגודל מקסימלי,אז היחס של 1/13 נשמע הגיוני. udp הוא Stateless ולכן הוא די "חד כיווני" ניתן לומר, לכן היחס שונה לחלוטין.

# שאלה 5:

## קוד:

SELECT attack\_cat, AVG(spkts) AS avg\_s, AVG(dpkts) AS avg\_d #ממוצע הפקטות שנשלחו מכל צד

FROM `redjony.MyDB.Veritas-Data-Set`

group by attack\_cat #הספירה היא לכל קטגוריית תקיפה

## תשובה:

(ע"ב מיון בטבלה שהתקבלה ASC וDESC) מי ששולח הכי הרבה אלו Exploits, DOS, מי שמקבל הכי הרבה- Worms, DOS.

# שאלה 6:

## קוד:

SELECT service, count(id) AS cnt\_service #ספירת אירועים לכל שירות

FROM `redjony.MyDB.Veritas-Data-Set`

WHERE

label = 1 and

service != '-' #סינון של מצב בו לא התקבל השירות

group by service #הספירה היא לכל שירות

order by cnt\_service desc #סידור כשהערכים המקסימליים הם העליונים

LIMIT 5 #הגבלה ל5 רשומות ראשונות

## פתרון:

לת.ל. ההתייחסות לערך '-' היא כלא רלוונטי- אולי תעבורה שלא סווגה לשירות כלשהו, או מוצפנת.

הטבלה-

|  |  |
| --- | --- |
| service | cnt\_service |
| dns | 18299 |
| http | 4274 |
| smtp | 1216 |
| ftp | 794 |
| ftp-data | 447 |

# שאלת מחקר עצמאית:

האם יש פרוטוקולים שתקשורת בהם מהווה אינדיקציה לתקיפה גבוהה מבאחרים? ע"מ לברר זאת ננסה למצוא את כמות התקשורת של כל פרוטוקול מתוך תקשורת התקיפה, נשווה זאת לתקשורת שאינה תקשורת תקיפה, היחס ביניהם יענה על שאלה זו.

הנחות עבודה:

1. הdata מהווה מדגם מייצג
2. סיווג הפרוטוקולים הוא אמין
3. אין כפילות בפרוטוקולים שבנויים אחד על השני (לדוגמא- http לרוב מתקשר על tcp, אז תקשורת http תסווג כhttp ולא תופיע פעם כhttp ופעם כtcp)

## קוד

#יצירת סכימה של תקשורת תקיפה לפי פרוטוקול

With attack\_com AS

(SELECT `proto`, COUNT(id) AS SUM\_OF\_ROWS # סכימת התקשורת

FROM `redjony.MyDB.Veritas-Data-Set`

WHERE

  label = 1 #תקשורת תקיפה

GROUP BY `proto`), #הספירה היא לכל פרוטוקול

#יצירת סכימה של תקשורת שאינה תקיפה לפי פרוטוקול

non\_attack\_com AS

(SELECT `proto`, COUNT(id) AS SUM\_OF\_ROWS # סכימת התקשורת

FROM `redjony.MyDB.Veritas-Data-Set`

WHERE

  label = 0 #תקשורת שאינה תקיפה

GROUP BY `proto`), #הספירה היא לכל פרוטוקול

#חלוקת תקשורת התקיפה לאחוזים

attack\_com\_per AS

(SELECT `proto`, SUM\_OF\_ROWS/(SUM(SUM\_OF\_ROWS) OVER())\*100 AS proto\_percentage #אחוז התקשורת בפרוטוקןל מסויים מתוך כלל התקשורת, מתוקננת לאחוזים

FROM attack\_com),

#חלוקת התקשורת שאינה תקיפה לאחוזים

non\_attack\_com\_per AS

(SELECT `proto`, SUM\_OF\_ROWS/(SUM(SUM\_OF\_ROWS) OVER())\*100 AS proto\_percentage #אחוז התקשורת בפרוטוקול מסויים מתוך כלל התקשורת, מתוקננת לאחוזים

FROM non\_attack\_com)

#ביצוע השאילתא- באלו פרוטוקולים האחוז בטבלת התקיפה היה גבוה באופן חריג?

SELECT

  COALESCE(a.proto, b.proto) AS `proto`, a.proto\_percentage AS attack\_per, b.proto\_percentage AS non\_attack\_per,

  CASE

    WHEN COALESCE(a.proto\_percentage, 0) > COALESCE(b.proto\_percentage, 0)

      THEN 'Greater'

    ELSE

      'Non Greater'

    END AS is\_greater #בדיקה האם אחוז התקשורת בפרוטוקול מסויים בתקשורת התקיפה גדול מהאחוז בתקשורת שאינה תקיפה

FROM

  attack\_com\_per AS a FULL JOIN non\_attack\_com\_per AS b

  ON a.proto = b.proto; #השוואה בין שתי הטבלאות בהן האחוזים

## תשובה

יש שני פרוטוקולים בהן התקבלה התשובה כי אחוז תקשורת התקיפה גבוה מאחוזה בתקשורת שאינה כזו- udp, ospf.

מצ"ב הטבלה במקום בו הייתה השוואה (לשני הפרוטוקולים התקבלו נתונים):

|  |  |  |  |
| --- | --- | --- | --- |
| Proto | attack\_per | non\_attack\_per | is\_greater |
| Tcp | 33.63407747 | 75.26486486 | Non Greater |
| Udp | 47.03300097 | 21.88378378 | Greater |
| Ospf | 1.407394335 | 0.1027027027 | Greater |

ניתן להניח בזהירות כי אחוז הudp פחות משנה, מאחר ומדובר בפרוטוקול מאוד נפוץ ובסיסי, והיחס בין האחוזים הוא בערך 1/2, ניתן להניח בזהירות שאין פה משמעות להצלחת התקיפה. ניתן לראות גם שבtcp היחס הפוך בערך.

לגבי ospf המצב שונה, היחס גדול בהרבה ביניהם, כמו כן מאחר ומדובר בפרוטוקול המשמש לניהול טופולוגיית רשת ניתן להשתמש בו לטובת spoofing ו/או הפניה של נקודות קצה לשרת זדוני, לכן ניתן להניח בזהירות שכמות חריגה של תקשורת ospf תהווה אינדיקציה לפעילות תוקף ברשת כלשהי.

כמו כן ניתן להסיק שבתקשורת תקיפה אחוז הפרוטוקולים המאוד נפוצים (udp, tcp) נמוך מאשר בתקשורת סטנדרטית, מה שיכול להוות אינדיקציה קלה לזיהוי ללא צורך בניתוח תעבורת התקיפה עצמה- אלא ממבט על התקשורת השגרתית ברשת.

תובנות מעשיות-

1. בראיית התוקף, מאחר ובתקשורת תקיפה בסה"כ יש הרבה פחות תקשורת של הפרוטוקולים הנפוצים ביותר (tcp,udp), 96% מול 80% לכן ברגע שתקיפה הצליחה מומלץ יהיה להשתמש רק בפרוטוקולים אלו כדי לא לייצר חריגות ברשת.
2. בראיית המגן, אם אחוז התקשורת בפרוטוקולים האלו יורד מתך לרף כלשהו (נניח 90%) ייתכן ומדובר באינדיקציה לפעילות חריגה.

# תרגיל Tableau:

https://public.tableau.com/app/profile/jony.klein/viz/VeritasDashboard/Dashboard1