סיכום פגישה בצ'קמרקס, ינואר 2019

בפגישה הציגו דניאלה (ראש קבוצה וחוקרת) ומתי (CTO) את הבעיה, וכן בעיות קרובות שכרגע בשלבי מחקר ופיתוח בצ'קמרקס.

דניאלה הציגה מספר שיטות למציאת דמיון בין קטעי קוד, והשימוש בהן לחיפוש קטע קוד דומה לקטע קוד עם בעיה. הכיוון שמתבצע בתוך צ'קמרקס כיום מתבסס על שבירת הקוד ל-statements, שבכל אחד מהם יש מספר tokens. משווים פונקציות שלמות, עם בעיות אבטחת מידע ידועות, לפונקציות שיש לנו בקוד, או בקוד פתוח, ומחפשים דברים מספיק דומים. מה שמוצאים מצביע על פוטנציאל גבוה לבעיית אבטחת מידע.

מתי הציג את הפוטנציאל בפרויקט שאנחנו מתחילים – כל המקרים בהם חריגה קטנה מהקוד יכולה לתת רמז לבעיית אבטחת מידע.

הקשר בין הפרויקט שלנו לפרויקט הפנימי בחברה היא שימוש פוטנציאלי באותן פונקציות מרחק שחקרו והשוו, וימשיכו לחקור בחודשיים הקרובים. אין סיבה שלא ניקח את זה כנקודת פתיחה (לפחות) עבור הפרויקט שלנו.

ההבדל המהותי הוא מה עושים עם פונקציות הדמיון. במוקם להשוות בין קוד פגיע לקוד שלנו, וחיפוש דמיון, מה שאנחנו צפוים לעשות הוא השוואות הרבה יותר מאסיביות, על מנת לבנות clusters של קטעי קוד דומים – פונקציות שלמות או חלקים מהן – ומציאת החריגים בכל cluster. ואז לנסות לראות אם החריגות האלה הן בעיות אבטחת מידע, או אחרות.