סיכום מאמר, חלק א (Dry Part) פרויקט גמר רשתות תקשורת

מגישים: שירה צ'שלר 323825059,אוהד וולפמן 316552496

הדרישות בעבודה

לאחר קריאת המאמר נדרשנו להסביר במילים שלנו את הרעיון המרכזי של המאמר, ובפרט בהתייחסות לנקודות הבאות:

1. כיצד משיג התוקף ground truth על תעבורת הערוץ? (ground truth משמעותה בדיקת דיוק של תוצאות למידת מכונה מול העולם האמיתי)

2. כיצד התוקף מבצע האזנת סתר בתעבורת הרשת?

3. תאר בקצרה את המסקנות מטבלה II במאמר.

4. איור 8 במאמר.

הוסף לחלק היבש קישורים לפרויקט Github ולחשבונות LinkedIn שלך.

פתיחה

המאמר, "Practical Traffic Analysis Attacks on Secure Messaging Applications", נכתב על ידי אלירזה בהרמאלי, אמיר הומאנסאדר, רמין סולטני, דניס גוקל ודון טאוסלי מאמהרסט, מאוניברסיטת מסצ'וסטס אמהרסט. החוקרים בדקו כיצד ניתן להתקיף אפליקציות מאובטחות ופופולריות המיועדות להתכתבויות (IM=instant messaging) בהקשרים של מעקב וצנזורה. הרקע לכתיבת המאמר הוא העידן הנוכחי, בו אפליקציות התכתבויות מאובטחות כמו טלגרם, סיגנל וואטסאפ הפכו לחלק בלתי נפרד מהתקשורת היומיומית. פלטפורמות אלו, על אף שהן מספקות מרחב לתקשורת מיידית ומאובטחת (מבחינת תוכן ההתכתבויות), כפופות לסיכונים הנובעים ממעקב וצנזורה ממשלתיים. סיכונים אלו רלוונטיים במיוחד כאשר הפלטפורמות הללו משמשות מרחב להתכתבויות בנושאים רגישים חברתית ופוליטית.

הרעיון המרכזי

הרעיון המרכזי של המאמר הוא העלאת מודעות לגבי האבטחה של האפליקציות המדוברות. החוקרים מסבירים שאפילו הצפנה מתקדמת, המופעלת על ידי אפליקציות שונות, עשויה שלא להספיק כדי להגן באופן מלא על מידע משתמש רגיש מפני יריבים עם הכלים והטכניקות הנכונות – כאשר יריב מתואר כארגון מעקב (כממשלה או ארגון המשויך אליה) אשר מטרתו לזהות (על ידי זיהוי כתובות IP) של חברים או מנהלים של קבוצת התכתבויות מסוימת (כפי שצוין קודם לכן, מקור מטרה זו יכול להיות רצון למעקב אחרי מתנגדי משטר, צנזורת תכנים מסוימים וכדומה, כפי שניתן דוגמא במאמר לגבי איראן, רוסיה וסין).

על אף העובדה שתוכן התקשורת מוצפן בצורה מאובטחת באפליקציות הנ"ל, החוקרים הדגימו כי על ידי מעקב אחר התעבורה ברשת ניתן על ידי מודלים סטטיסטים לזהות דפוסים של זרימת התעבורה, שעל ידיהם הם יכולים לספק שפע של מידע, שהמשתמש ומפעיל הIM היו שמחים שלא יהיו חשופים ליריב/ תוקף. החוקרים גילו כי כל פעולה שמתבצעת בתוך יישומים אלה - כמו שליחת הודעה, העלאת קובץ או אפילו רק הקלדה - מייצרת זרימת נתוני תקשורת עם דפוסים ייחודיים. אם כן התוקף, על ידי ניטור וניתוח של דפוסים אלה בלבד, יכול לאסוף מידע רב ערך על משתמשים ופעילויותיהם. החוקרים מגדירים זאת כ'Traffic Analysis Attack'.

כיצד משיג התוקף ground truth על תעבורת הערוץ?

לתוקף ישנן מספר דרכים אופציונליות על מנת להשיג את דפוסי זרימת התעבורה של קבוצת התכתבויות/ הפצת מסרים מסוימת בהתאם לסוג קבוצת היעד:

א. אם קבוצת היעד הינה ציבורית (כלומר פתוחה לכל מי שרוצה להצטרף), התוקף יכול פשוט להצטרף לקבוצה כחבר. ברגע שהוא מצטרף, הוא יכול להקליט את ההודעות שנשלחו יחד עם הmeta data שלהן (כגון זמן וגודל ההודעות). סוג זה של נתונים יכול לתת ליריב תובנות לגבי מתי הקבוצה הכי פעילה, כמה חברים משתתפים ואולי אפילו באילו סוגי נושאים דנים (בהתבסס על גודל ותדירות ההודעות).
ב. אם קבוצת היעד אינה ציבורית אך התוקף כן הצליח להצטרף אליה והוא יכול לשלוח הודעות (ישנה הרשאה בקבוצה לכלל המשתתפים לשלוח הודעות או לחילופין שהתוקף הצליח לקבל הרשאה לשלוח הודעות על ידי קבלת הרשאות מנהל), הוא יכול לפרסם הודעות משלו עם דפוסי תעבורה ברורים (הידועים לו מראש). על ידי ניתוח האופן שבו חברים אחרים בערוץ מגיבים להודעות אלו, היריב יכול לקבל מידע נוסף על תנועת הערוץ.

ג. אם קבוצת היעד הינה פרטית, והתוקף אינו מצליח להצטרף אליה, אך הוא הצליח לזהות את כתובת הIP של אחד המשתתפים/ המנהלים בה, הוא יכול להאזין לתעבורת הרשת של המשתתף המזוהה. נתונים אלה יאפשרו לו לתעד את דפוסי התעבורה של המשתתף, מה שיכול לספק רמזים לגבי הפעילות בערוץ, גם אם התוקף לא יכול לראות את ההודעות שנשלחות בפועל.

<u>כיצד התוקף מבצע האזנת סתר בתעבורת הרשת?</u>

הדרך בה התוקף יכול להאזין לתעבורת הרשת של המשתתף המזוהה היא על ידי מספר דרכים:

א. על ידי האזנה לתעבורת הרשת של הISP (Internet Service Provider, ספק האינטרנט אשר דרך הנתבים שלו עוברת כל התעבורה)

ב. על ידי האזנה ל IXP (Internet eXchange Point, המיקום הפיזי דרכו חברות תשתיות אינטרנט כISP מתחברות לCDN-ים (Content Delivery Network, קבוצת שרתים מפוזרים המגבירים את מהירות הגעת מידע הרשת למשתמש ע"י הבאת המידע קרוב יותר אל מיקום המשתמש), בהנחה שהתוקף שולט בIXP/ISP.

ג. אחרת, התוקף יכול להאזין למשתמשים ספציפיים ע"י השגת צו האזנה נסתרת, למשל במידה והתוקף/ יריב הינו חברה ממשלתית ולכן יכול להשיג היתר משפטי שכזה.

על מנת להצליח ולדמות התקפה, החוקרים במאמר רצו למדל תקשורת SIM(Secured Instant Messaging). סוג ההתקפה במאמר הינו הדומה ביותר לסוג התקפה Flow correlation. בהתקפה זו התוקף מנסה לחבר בין זרימה ברשת לבין מאפייני התעבורה (זמנים של packets וגודליהן). כיוון שניתוח התעבורה מתמקד בגדלי ההודעות הנשלחות ובזמן השליחה שלהן, החוקרים מידלו דברים אלה. בטבלה II במאמר מוצגת טבלת ההתפלגות (סטטיסטיקות גודל ומספר) של 5 סוגי ההודעות הנפוצות ביותר – טקסט, תמונה, סרטון, קובץ וקובץ שמע:

TABLE II: Distribution of various message types

Type	Count	Volume (MB)	Size range	Avg. size
Text	12539 (29.4%)	3.85 (0.016%)	1B-4095B	306.61B
Photo	20471 (48%)	1869.57 (0.765%)	2.40Kb-378.68Kb	91.33KB
Video	6564 (15.4%)	232955.19 (95.3%)	10.16Kb-1.56Gb	35.49MB
File	903 (2.1%)	47.46 (0.019%)	2.54Kb-1.88Mg	52.56KB
Audio	2161 (5.1%)	9587.36 (3.92%)	2.83Kb-98.07Mg	4.44MB

המסקנות מטבלה זו:

א. ניתן לראות, לדוגמה כי בעוד שאחוז הסרטונים מתוך סך כל ההודעות הוא 15.4%, כמות המקום הנתפסת ע"י קבצים מסוג זה היא 95.3% (רוב נפח ההודעות בקבוצה).

ב. בנוסף, ניתן לראות כי שונות גודל הסרטון הינה גדולה מאוד (טווח גודל 10.16Kb-1.56Gb), בעוד שונות גודל הודעה, למשל, קטן מאוד (טווח גודל 1B-4095B). מלבד זאת, ניתן לראות כי אחוז הקבצים וקבצי השמע יחדיו אינו רב (פחות מ10%: 7.2%).

ג. כמו כן נוכל לשים לב כי מעל 77% מהתעבורה היא על ידי הודעות טקסט או תמונות, כלומר ניתן לחשוד שמדובר בקבוצה פעילה הן מבחינת השיח רב המשתתפים והן מבחינת שיתוף תמונות ועדכונים, פרטיים או עדכונים של תמונות בזמן אמת כמו חדשות וכד'.

לאחר המידול, החוקרים מציעים שני אלגוריתמי גילוי לפיהם ניתן לעקוב אחר תעבורה של קבוצה ולהחליט האם משתמש מסוים אכן חבר בקבוצה:

א. האלגוריתם הראשון הוא 'Event-Based Detector' (מזהה מבוסס אירועים).
Event (אירוע) יקרה כאשר תבוצע שליחה שתגרום לפרץ של packets בגודל מקסימלי - MTU (Maximum Transmission Unit, גודל המנה המקסימלי ששכבה נתונה בפרוטוקול מסוים יכולה להעביר). ואז בהינתן שלתוקף ישנה גישה לתעבורת הקבוצה הרצויה, ותעבורות משתמשים נוספים, באמצעות פונקציית קורלציה – ניתן לבדוק האם משתמש שייך לקבוצה או לא (אם הקורלציה בין התעבורה שלו בעלת קורלציה גבוהה מספיק, מעל threshold מסוים, לזו של הקבוצה לה מאזין התוקף).

ב. האלגוריתם השני הוא Shape-based Detector (אלגוריתם מבוסס-צורה), המתאם את צורות זרימת תעבורת ה-SIM על מנת לשייך את המשתמשים לערוצי היעד. האלגוריתם מבוסס צורות הוא איטי יותר אך מציע ביצועי זיהוי מדויקים יותר מאלגוריתם מבוסס אירועים (רק 15 דקות של תעבורת טלגרם מספיקות לגלאי מבוסס הצורה כדי לזהות את המנהל של ערוץ SIM מסוים עם דיוק של 94%).
'צורות תעבורה', כפי שכינו אותה החוקרים, הכוונה לוקטור של אורך הפקטות כתלות בזמן, ומכיל 4 שלבים:

1) חילוץ אירוע, כמו באלגוריתם הראשון – על ידי זיהוי של פרצים של השימוש ברוחב הפס.
2) נרמול צורות תעבורה - מסירה את ההשפעה של רוחב הפס של המשתמש, הנרמול מתבצע על ידי החלפת כל אירוע (כלומר, כל פרץ) בפס תנועה שרוחבו הוא te 2, כאשר te הוא הסף

המשמש במהלך חילוץ אירוע. גובה כל bar בגרף נבחר כך שהשטח מתחת לbar יהיה שווה לגודל האירוע. לבסוף אנו מחלקים כל bar לbins קטנים יותר ברוחב הts, ועם גובה השווה לגובה הפס המקביל של הקבוצה הנבקדת. לכן כל bar מורכב ממספר bins בעלי רוחב וגובה שווים. צורת התנועה החדשה תהיה הווקטור של גבהים של bins לאורך זמן.

3) קורלציה של צורות התעבורה המנורמלות לזו של הקבוצה הנבדקת.

4) השוואה ביחס לtreshold מוגדר שתביא לחיזוי ביחס לכתובת המודגמת.

(נעיר כי בפועל, התוקף יכול לשלב את שני האלגוריתמים כדי לאזן את הtrade-off בין ייעול עלות החישוב (והמדרגיות) לעומת ביצועי (אחוז) הזיהוי)

באיור 8 במאמר ניתן לראות תרשימים של גודל הpacket בקרב שני משתמשים – אחד שלא שייך לקבוצה (שעל כן הקורלציה בינו ובין התעבורה בקבוצה אינו גבוה, וגרף אחד של משתמש שאכן שייך לקבוצה ובעל קורלציה גבוהה לתעבורה של הקבוצה אל מול ציר זמן.

בתחתית האיור ניתן לראות events שקורים במהלך פרק הזמן בו מתבוננים על תעבורת שני המשתמשים (שליחת תמונה, שליחת סרטון, שליחת קובץ אודיו, ושליחת קובץ), כאשר עבור המשתמש אשר זוהה כשייך לקבוצה – ניתן לראות שבכל שליחה שכזו יש מספר MTU (לפחות 1, בדיוק בזמן הevent), בעוד שעבור המשתמש שלא זוהה כשייך לקבוצה ישנם MTU בזמנים שונים, לא בהכרח בזמן event.

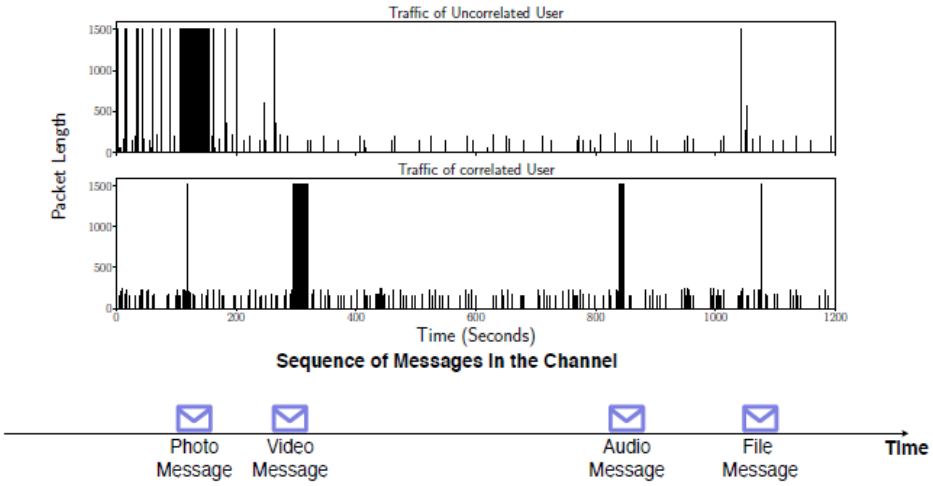

Fig. 8: Event extraction: IM Messages sent/received by a target user create bursts of (encrypted) packets; the adversary can extract events from packet bursts.

לאחר מכן, בודקים החוקרים את המודלים שלהם על טלגרם, ווטסאפ וסיגנל. הם מראים שהפעלת האלגוריתמים שלהם נותנים תוצאות זיהוי טובות. דבר זה מהווה איום משמעותי על המשתמשים, לאור הניסיונות ההולכים וגדלים של ממשלות מדכאות לפצח את הערוצים השנויים במחלוקת בפלטפורמות

הללו. המחקר מהווה קריאת השכמה הן למשתמשים והן לספקים של שירותי הודעות כאלה. המשתמשים צריכים להיות מודעים לסיכונים ולהתאים את דפוסי השימוש שלהם בהתאם. במקביל, זה גם מדגיש את הצורך של ספקי שירותים לשלב אמצעי נגד אפקטיביים של ערפול תעבורה במערכות שלהם, מעבר להצפנה בלבד, כדי להבטיח את הפרטיות והבטיחות של המשתמשים שלהם.

לבסוף, המחברים מציגים מערכת אמצעי נגד הנקרא IMProxy . מערכת זו נועדה להגן מפני ההתקפות מסוג זה בדיוק. החוקרים גילו כי ביצוע של tunnelling (מנהור) של תעבורת SIM דרך VPN וערבוב שלה עם תעבורת גלישה באינטרנט מפחיתה את דיוק ההתקפה באמצעות שני האלגוריתמים שלהם מ-93% ל-70%, והוספת תעבורת כיסוי (cover traffic) עם תקורה של 17% מורידה את הדיוק ל-62%.

על כן כפי שהצגנו, החוקרים פרסמו מערכת אמצעי נגד, הזמינה לציבור בקוד פתוח, הנקראת IMProxy, שיכולה לשמש לקוחות IM ללא צורך בתמיכה כלשהי מספקי IM, אשר מביאה לתוצאות טובות אשר הוכחו גם בניסויים.

<u>קישורים</u>

לינקדאין –

1. https://www.linkedin.com/in/shira-chesler-4438b5222/
2. https://www.linkedin.com/in/ohad-wolfman/

גיטהאב –

https://github.com/ohadwolfman/Networks_Final_Project