Traffic Analysis attacks on IM service users project

שמות המגישים:
יונתן בוריצקי - 207254194 קישור לפוסט בלינקדין
עידן ינאי - 214293300 קישור לפוסט בלינקדין

קישור לגיט: https://github.com/Jon400/Network-Final-Project.git

סקירת מאמר Practical Traffic Analysis Attacks on Secure Messaging Applications:

הרעיון המרכזי של המאמר מדבר על פרצה בפרטיות המשתמשים באפליקציות מסוג Instant Messaging כמו למשל Whatsapp או Telegram. המאמר עוסק בפירצה המשתמשת ב-Traffic Analysis בהקשר של IM services, הסכנות אליהן המשתמשים נחשפים מהפירצה והפתרונות האפשריים עבורה.

מטרת המאמר היא להראות שקיימת דרך שבה ניתן יהיה להסיק את זהות משתמשים הנמצאים בקבוצה, על ידי שימוש ב-Traffic analysis, בנוסף מציעה פתרונות אשר מאפשרים להתמודד איתה.

רוב שירותי ה-IM service בנויים על שיטה של Client-Server שבו נעשה שימוש בשרת מרכזי אליו נשלחות הודעות ממשתמש מסוים, והשרת אחראי על העברת ההודעות למשתמשים אחרים. גישה אחרת שקיימת היא גישת הP2P ובגישה זו נעשה שימוש בשרת רק לצורך יצירת קשר, תהליך זה נקרא handshaking. בגישה הראשונה של שרת-לקוח, ניתן להשיג כתובת ה-IP של השרת המרכזי ובכך קל יותר לתוקף לבצע סינון בעת הסנפת החבילות, בעוד שבגישת הP2P, ה-IP של משתמשי הקצה מועבר באופן מוצפן ובכך קשה הרבה יותר לסנן חבילות.

שירותי ה-IM service משתמשים לרוב בפרוטוקולי תקשורת מוצפנים ברמת האפליקציה כגון TLS, המידע המוצפן מכיל גם את תוכן ההודעה וגם את ה-metadata המכיל בין היתר מידע לגבי השולח או המקבל. לכן יש קושי להסיק את זהות המשתמשים על ידי חילוץ ישיר של המידע מהחבילות. המאמר מציע אפשרות להסיק את זהות על ידי ניטור תעבורת הרשת - Traffic Analysis.

בשיטה זו התוקף מאזין אחר הצד המותקף, דרך אחת לעשות זאת זה על ידי שליטה בספקי האינטרנט (לדוגמה מדינה שיש לה שליטה על שירותי התקשורת), ובכך יש לו אפשרות לבצע ניטור אחר החבילות שנשלחות אל שרת של ה-IM Service ובמקביל להכיר את זהות המשתמש לפי כתובת ה-IP שלו. התוקף משתמש ב-Wiretaping, כלומר הוא מתחבר אל השכבה הפיזית שבה עוברת התעבורה בין המשתמש המותקף אל השרת, ומבצע העתקה של החבילות מבלי לפגוע בחבילות המקוריות הנשלחות מהמשתמש המותקף אל השרת ובכיוון ההפוך. לדוגמה, חדירת ה-LAN של המשתמש באמצעות חדירה לרשת ה-WIFI שבה המשתמש המותקף עושה שימוש. במידה והתוקף הוא מנהל ה-ISP של המשתמש המותקף, אז יש לו יכולת להאזין ללקוח בחלק משלבי הניתוב של הלקוח אל השרת.

בנוסף התוקף מאזין לתעבורה העוברת באפליקציה של ה-IM service וזה נעשה בשלוש דרכים גישות עיקריות:

- התוקף מתחבר כמשתמש בקבוצה או ערוץ של IM Service ובכך משיג מידע אודות המידע שנשלח בקבוצה. יתרון של שיטה זו היא בכך שהתוקף מקבל מידע נקי ומדויק אודות התעבורה של הערוץ מבלי להיות חשוף לרעשים. החיסרון בגישה זו היא במידה והמשתמש אינו מכיר את מאפייני התעבורה בקבוצה, יהיה לו קשה יותר לזהות דפוסי התאמה בין התעבורה שעוברת בקבוצה לבין התעבורה שעוברת אצל המשתמש המותקף.
- התוקף מתחבר כמשתמש בעל הרשאות פרסום בקבוצה של IM. יתרונותיה של הגישה היא בכך שיש לתוקף יכולת לשלוח מידע מלאכותי שייצר מאפייני תעבורה ייחודיים שאותם יהיה קל לתוקף לזהות. התוקף יכול לייצר תעבורה מלאכותית ולקבוע עבורה: Inter message delays ו-Message sizes לפי התפלגות מסוימת, באמצעות האלגוריתם שמסופק במאמר - "Algorithm 1 Algorithm for Generating Synthetic IM Traffic" ולזהות התאמה לתעבורה שעוברת אצל המשתמש המותקף.
- בגישה זו מניחים שהתוקף מכיר את אחד מחברי הקבוצה. התוקף מבצע Wiretapping על אותו חבר קבוצה ומנתח את התעבורה שלו על ידי הסנפת חבילות. היתרון בגישה זו היא בכך שאין צורך למדל את מאפייני התעבורה בקבוצה, אך יש צורך לבצע השוואה בין תעבורת המשתמש המותקף לתעבורה של חבר הקבוצה המזוהה. החסרונות של גישה זו היא שהיא חשופה יותר לרעשים. כמו כן, קשה לדעת את גודל ההודעות וזמני שליחתן, ומכך קשה יותר לבצע התאמה.

ישנם שלושה מאפייני תעבורה עיקריים באפליקציות IM:

Inter message delays - מדד המתאר את זמני השהיות בין חבילות הנשלחות בעת שליחת קובץ גדול כגון תמונה או הקלטה. ברוב המקרים ההתפלגות של ההשהיות תהיה קרובה להתפלגות אקספוננציאלית.

Message delays - מדד המתאר את גודל החבילות שנשלחות. לרוב לכל סוג הודעה ששולחים (כגון: תמונה, סרטון...) תהיה התפלגות ייחודית של גודל החבילות. טבלה II מתארת את חמשת סוגי ההודעות הנפוצים ביותר וההתפלגות להופעתם, גודל ממוצע של של חבילה וטווח גדלים אפשרי. גרף ה-CCDF המופיע במאמר מתאר את ההתפלגות לגודל ההודעה באותו טווח.

Communication latency - זמני ההגעה של החבילות עלול להשתנות בין שני משתמשים שונים. זה עלול לקרות בעקבות עיכובים ברשת או בגלל עיבוד מידע בשרת.

המאמר מציע שלושה אלגוריתמי תקיפה:

Event based - בשיטה נחלץ אירועים בהם מועברים הודעות בעלות משקל גדול יחסית. בשיטה זו בודקים את מספר האירועים החופפים (מבחינת זמן בתוספת ה-latency וגודל ההודעה) אצל התוקף ואצל המשתמש המותקף. במידה ויש כמות מספקת של אירועים חופפים, מסווגים את הנתקף כחבר בקבוצה, אחרת לא. בגרף מספר 8 ניתן לראות את ההודעות שנשלחו בקבוצה על ציר זמן ואת תעבורת החבילות לפי זמן וגודל החבילה. ניתן לראות שאצל המשתמש שחבר בקבוצה, יש Spikes בגרף החופפים את זמני שליחת ההודעות בקבוצה. בעוד שאצל המשתמש שאינו חבר בקבוצה, לא רואים התאמה. כמו כן, ניתן לראות שיש תעבורה של חבילות קטנות שמהן האלגוריתם מתעלם. גרף זה מוכיח שתעבורת הרשת חושפת מאפיינים אשר מאפשרים לבצע קישור בין משתמשים לבין קבוצות או ערוצים מסוימים ובכך זה מהווה סיכון לריגול אחר המשתמשים.

Shape based- אלגוריתם המוצא קורלציה בין וקטורים המתארים תעבורת רשת.

Deep corr - אלגוריתם המבוסס על למידה עמוקה לצורך מציאת קורלציה. האלגוריתם משיג תוצאות יחסית נמוכות.

בסוף המאמר מוצגות כמה שיטות המאפשרות טשטוש עקבות בתעבורת הרשת של הצד המותקף. השיטה הראשונה שבה עוסק המאמר זה שימוש בכלי עקיפה להעברת נתונים כמו VPN, עם אפשרות להוספת רעש רקע לצורך טשטוש ומחיקה של עקבות בתעבורת הרשת.

השיטה השנייה שהמאמר מציע הוא להשתמש בשרת פרוקסי מקומי ומרוחק, כאשר השרת המקומי יבצע Padding באופן רנדומלי לחבילות שנשלחות והשרת המרוחק ידע לתקן את החבילות בהתאם לפני שהן מגיעות אל השרת של ה-IM service. אפשרות אחרת שנבדקת היא להוסיף באופן אקראי delay לחבילות.