**סיכום המאמר -**

יישומי הודעות מיידיות (Instant Messaging IM -), כגון טלגרם, Signal, ו-WhatsApp הפכו פופולריים ביותר בשנים האחרונות. שירותי IM מאפשרים למשתמשים ליצור קבוצות חברתיות, פרטיות וציבוריות ולהעביר הודעות מסוגים שונים, כולל הודעות טקסט, תמונות, סרטונים וקבצי אודיו. בפרט, יישומי IM נמצאים בשימוש נרחב להחלפת תוכן רגיש מבחינה פוליטית וחברתית. על מנת להגן על הלקוחות שלהם, שירותי IM פורסים מנגנוני הצפנה חדישים (הצפנה מקצה לקצה או מקצה לאמצע) כדי לאבטח תקשורת משתמשים, שירותים מאובטחים אלו נקראים SIM - secure IM. במאמר זה, מראים שלמרות השימוש בהצפנה מתקדמת, יישומי IM מדליפים מידע רגיש של לקוחותיהם לכל מי שרק עוקב אחר תעבורת ה-IM המוצפנת שלהם. הסיבה העיקרית שבגללה ניתן לבצע תקיפות היא העובדה שספקי ה- IM אינם פורסים מנגנונים כדי לטשטש את מאפייני התעבורה (למשל, תזמון וגדלים של מנות).

מודל סטטיסטי עבור מציאת מאפייני תעבורת IM - החוקרים הצטרפו ליותר מ-1,000 ערוצי טלגרם ציבוריים ותיעדו את התקשורת שלהם, על בסיס המידע הזה הם הפיקו מודל סטטיסטי אשר נותן אפיון לתעבורה של התקשורת המיידית. נתונים אלו מהווים בסיס ל-2 אלגוריתמים לניתוח התעבורה : **אלגוריתם event-based** - מבצע התאמה אופטימלית בין משתמשים לערוצים,  **אלגוריתם** **shaped-base** - מתאים בין צורות הזרימה של תעבורת ב-SIM. על מנת להתאים את המשתמשים לערוצי היעד (איטי יותר אך בעל ביצועי זיהוי מדויקים יותר).

אלגוריתמים אלו הינם בעלי אחוזי דיוק גבוהים מאוד (93%), האמנם, שימוש בVPN ו- TOR מורידים את אחוזי הדיוק (62%), אך רוב המשתמשים אינם פורסים אמצעי הגנה אלו. בנוסף, החוקרים בדקו ומצאו כי האלגוריתמים הסטטיסטיים טובים יותר מאלגוריתמים מבוססי למידה עמוקה.ההתקפות חלות על כל שירותי ה-IM הגדולים ומבוצעות על ידי ניתוח התנועה בלבד, כלומר התוקף לא צריך את עזרת ספק ה-SIM. היריב מבצע ניתוח תעבורה כדי לזהות את המשתתפים, המנהלים והחברים בערוצי ה-IM.

Inter-Message Delays (IMDs) - עיכובים בין הודעות - הIMD מתייחס למרווח הזמנים בין חבילות שנשלחו, באמצעות פרמטר זה ניתן לזהות סוגי הודעות ואירועים שנשלחו בקבוצה.

**Event Extraction – חילוץ אירועים -** כל אירוע (למשל תמונה שנשלחה) מייצר פרץ של מנות בתעבורה המוצפנת במרווחי זמנים קרובים מאוד. דברים אחרים כגון התראות, לחיצות ידיים, עדכונים וכו' הם חבילות מפוזרות (לא קרובות בזמנים) בגדלים קטנים.ולכן נוכל לזהות אירועים על ידי הסתכלות על פרצי המנות כפי שתואר לעיל, על אף שהמידע עצמו מוצפן.

**-"Ground Truth"** על מנת לבצע תקיפה על יעד כלשהו אנו צריכים מידע בסיסי אמיתי ונכון.

ניתן להשיג מידע זה ב-3 דרכים -

1. אם הערוץ הוא פתוח (ציבורי) אך למצטרפים אין אפשרות לפרסם בה הודעות - היריב מצטרף לערוץ ומתעד את ההודעות שנשלחו ואת ה- meta data (גודל, זמן, מי שלח למי ועוד).
2. היריב מצטרף לקבוצת היעד ומסוגל לפרסם בה הודעות. במקרה זה, היריב יכול לתעד את ההודעות ובנוסף לפרסם הודעות משלו עם דפוסי תנועה הרצויים.
3. היריב אינו יכול להצטרף כחבר/מנהל, אך הוא זיהה (את כתובת ה-IP של) אחד מחברי/מנהלי הערוץ, במקרה זה היריב מאזין לתעבורת הרשת (המוצפנת) של החבר המזוהה ומתעד את דפוסי התעבורה שלו.

**האזנת סתר של היריב-** היריב עוקב אחר תעבורת הרשת המוצפנת של משתמשי IM כדי לזהות את כתובות ה-IP של החברים/מנהלי ערוץ ה-IM יעד.

זה יכול להתבצע על ידי האזנת סתר לתעבורת הרשת של ספקי האינטרנט או ה-IXP שבהם הוא שולט.

לחלופין, היריב יכול להאזין לתנועת הרשת של אנשים ספציפיים (למשל, פעילים חשודים), אולי לאחר קבלת צו האזנת סתר. היריב משתמש באלגוריתם זיהוי על מנת להתאים את דפוסי התעבורה שהתקבלה מהאזנת הסתר למידע שמועבר בפועל בערוץ היעד.

**תהליך המחקר -** במאמר החוקרים אספו נתונים מאפליקציית **טלגרם** על מנת לבצע את המחקר.  
הבחירה בטלגרם היא מכיוון וזו אפליקציית IM פופולארית ביותר וכי הקבוצות\ערוצים הם ציבוריים וניתן להצטרף אליהם בקלות ללא אישור, בניגוד לאפליקציות IM אחרות. לכל הודעה החוקרים חילצו את: הid של הערוץ, זמן שליחת ההודעה, סוג ההודעה (תמונה, טקסט, סרט ועוד) גודל ההודעה.

החוקרים פתחו 2 חשבונות טלגרם – אחד בתור התוקף והשני הוא המטרה אותו נרצה לתקוף (היעד).

שני החשבונות הצטרפו לאותו הערוץ, התוקף אוסף את כל הנתונים מהערוץ, לאחר מכן משתמש באלגוריתם איתור (event-based, shaped-base) על מנת להחליט האם המשתמש היעד מקושר לערוץ הנוכחי.

לאור האיום שהוצג לעיל החוקרים תכננו ופרסמו מערכת אמצעי נגד, אשר זמין לציבור בקוד פתוח, הנקראת IMProxy, שיכולה לשמש לקוחות IM ללא צורך בתמיכה כלשהי מספקי IM.