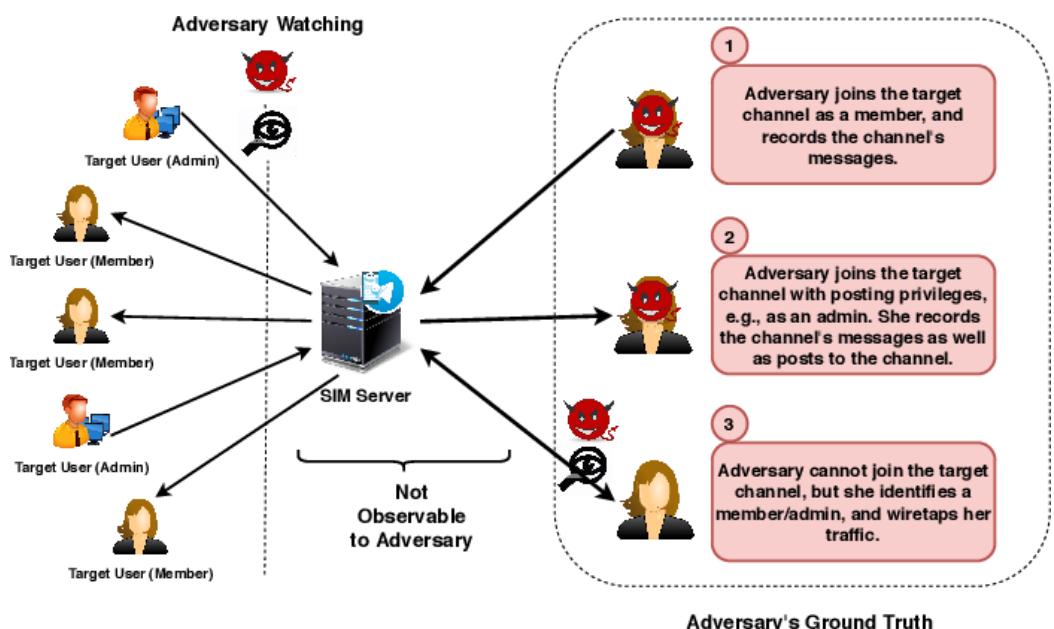


Practical Traffic Analysis

Attacks on Secure Messaging Applications



מגישי:

ויליאן אומנסקי 322880857
לירז בלס 211801220

תוכן עניינים

3-4	-----	חלק א'-שאלות
5	-----	חלק ב'-מבוא
6-8	-----	תחילת המחבר
9-22	-----	תיעוד וניתוח
23-24	-----	סיכום ומסקנות

סעיף א'

התוקף מшиיג את האמת הבסיסית על התנועה של ערוץ MI היעד באמצעות שלוש דרכיים :
הצטרפות לערוץ פתוח: אם ערוץ MI היעד הוא ערוץ פתוח או ציבורי, התוקף יכול להציגו לערוץ חבר. על ידי השתתפות פעילה בערוץ, התוקף יכול להקליט את ההודעות שנשלחו בערוץ יחד עם המטה-נתונים שלהם, כגון השעה והגודל של ההודעות. מידע זה מספק מידע אמיתי אודות דפוסי התנועה של הערוץ.

שליחת הודעות לערוץ: יתכן שבמקרים מסוימים התוקף כבר הצטרף לערוץ ההודעות המיידיות של המטרה וקיבל את היכולת לפרסם הודעות. דבר זה יכול לknור את הערוץ הוא קבוצה סגורה המאפשרת לכל חבר לפרסם הודעות או אם התוקף קיבל תפקיד מנהל לערוץ. על ידי הקבלת ההודעות הנכונות והויאות, התוקף יכול ללכוד את דפוסי התנועה של הערוץ ואפיילו להציג את ההודעות שלו עם דפוסי תעבורת שוניים.

האזנה לחבר/מנהל מערכת: אם התוקף אינו מסוגל להציגו לערוץ היעד כחבר או כמנהל מערכת, הוא עדין יכול לקבל מידע אמית בסיסית על-ID'יו כתובות ה-IP של אחד החברים או המנהלים של הערוץ. לאחר מכן, התוקף צוותת לubahורת הרשות של החבר/מנהל המזויה, ומירט את התקשרות המוצפנת שלהם. על ידי רישום דפוסי התנועה של החבר/מנהל הזוגה, התוקף יכול להשתמש במידע זה כאמת בסיסית כדי להתאים לדפוסי התנועה של זרים אחרות.

לסיכום, התוקף מшиיג את האמת הבסיסית על התנועה של ערוץ ההודעות המיידיות על ידי השתתפות פעילה בערוץ, יירות התעבורת של מנהלים/חברים, או על ידי האזנה לubahורת הרשות של אנשים ספציפיים הקשורים לערוץ.

יתר על כן התוקף מצוותת לubahורת הרשות של משתמשי MI כדי להזמין את כתובות ה-IP של החברים/מנהל המערכת של ערוץ MI המשמש כיעד. השיטות הספציפיות שהוזכרו כוללות:
האזנה לשפיקי שירות אינטרנט (ISP) או ISPs (ISP) או נקודות ZIX (ZIX). התוקף יכול לשולח בתשתיות הרשות או לקבל גישה אליה, כגון ספיקי שירות אינטרנט (ISP) או נקודות ZIX (ZIX). על ידי צוותת לubahורת הרשות העוברת דרך נקודות בקרה אלו, התוקף יכול לירט ולפקח על תעבורת הרשות המוצפנת של משתמשי הודעות מיידיות.

האזנה לאנשים מסוימים: במקרים מסוימים, התוקף עלול למתקד אנשים מסוימים, כגון פעילים חשובים. לאחר קבלת צו צוותות או קבלת גישה בלתי מורשית, התוקף מירט את התעבורת ברשות של אנשים אלה כדי לטען את דפוסי התנועה שלהם.

כדי לציין צו צוותות לubahורת רשות בדרך כלל כרוך בלמידה וניתוח של מנוט הנתונים המשודרות ברשות. דבר זה יכול להתבצע באמצעותים שונים, כגון פרישת התקני ניטור רשות, שימוש בכלי תוכנה או חומרה מיוחדים, או להתאפשר על תשתיות הרשות או על נקודות הקרה כדי לקבל גישה לubahורה. הפרטים המדוייקים כיצד התוקף מתקשר לubahורת רשות יכולים להיות תלויים ביכולות הטכנולוגיות והמשאבים הזמינים לתוקף.

ניתן לראות במחקר כי הטבלה II מספקת סקירה על התפלגות סוגי ההודעות השונים בתעborת ההודעות המיידיות הנאספת. הטבלה מציגה נתונים סטטיסטיים כגון הספירה, עצמת הקול (ב-MB), טווח הגודל והגודל הממוצע עבור כל סוג הודעה, כולל טקסט, תמונה, וידאו, קובץ והודעות שמע. מהטבלה, אנו יכולים להתבונן בפרופורציות היחסיות של סוגי הודעות שונים במאהר הנתונים של תעborת הודעות מיידיות. לדוגמה, הודעות טקסט מהוות כ-29.4% מהסכום הכללי, ולאחריהן תמונות (48%), סרטונים (15.4%), קבצים (2.1%) והודעות שמע (5.1%).

בנוסף, הטבלה מספקת תובנות על מאפייני הגודל של סוגי הודעות שונים, כגון טווח הגודלים והערכים המומוצעים שלהם.

נתונים אלה חיוניים לבנת הרכיב והמאפיינים של התקשרות מסרים מיידיים, המיידעים אותנו על בניית מודלים וניתוח של תעborות הודעות מיידיות, כמו גם פיתוח של מתפקידות ניתוח תעborה. במערכות תקשורת, כל אירוע SIM, כגון שליחת תמונה, מייצר בועת של חבילות בגודל MT-U בקצב גודל MT-U בתנועת הקריפטוגרפיה. החבילות אלו עם דוחים צעירים בין החבילות וגודלה תואם את גודל ה-SIM. הבועות מייצגות אירוע SIM, תוך שבחילות פחרות בגודל קטן מייצגות הודעות פרוטוקול ה-SIM כגון הודעות התראה, ידניות, עדכונים ועוד. כדי להזמין את הבועות הללו ולהשוף את אירוע ה-SIM, החוקרים משתמשים בסוף זמן בין חבילות (IPD), שמשמעותו "te".

שתי חבילות עם מרחק זמן קצרן מ-te יחשבו להיות חלק מאותו בועה.
ערך te הוא מונחה היפר-פרמטר במודל, והבחירה שלו נדונה במאמר.
לכל בועה המזוהה באמצעות te, האיבר יכול לחסוף אירוע SIM.
זמן של הגעת החבילה האחורה בבועה מצין את זמן ההגעה של האירוע, וכך כל הגודלים של החבילות
בבועה נותנים את גודל האירוע. בנוסף, שתי הודעות SIM שנשלחות עם מרחק בין הודעות (IMD) קצרן מ-te
'יחשבו להיות חלק מאותו אירוע'.
האיבר משלב אירועים שכורבים יותר מ-te מאשר הוא מפעיל גידול בערוץ היעד, הגישה זו מאפשרת לאיבר
לזהות אירועי SIM על ידי חיפוש אחר בועות של חבילות בגודל MTU, למרות שתוכן החבילות נשאר מוצפן
ולא נגיש.

Wet part-Whatapp Web

בחלק זה של הפרויקט, ניסינו כיצד ניתן לראות באופן מופשט ביותר את הקשר של המחשב שלנו עם שרת
הוואטסאפ, ובכלל כל תקשורת המתאפשרת לאפליקציה.

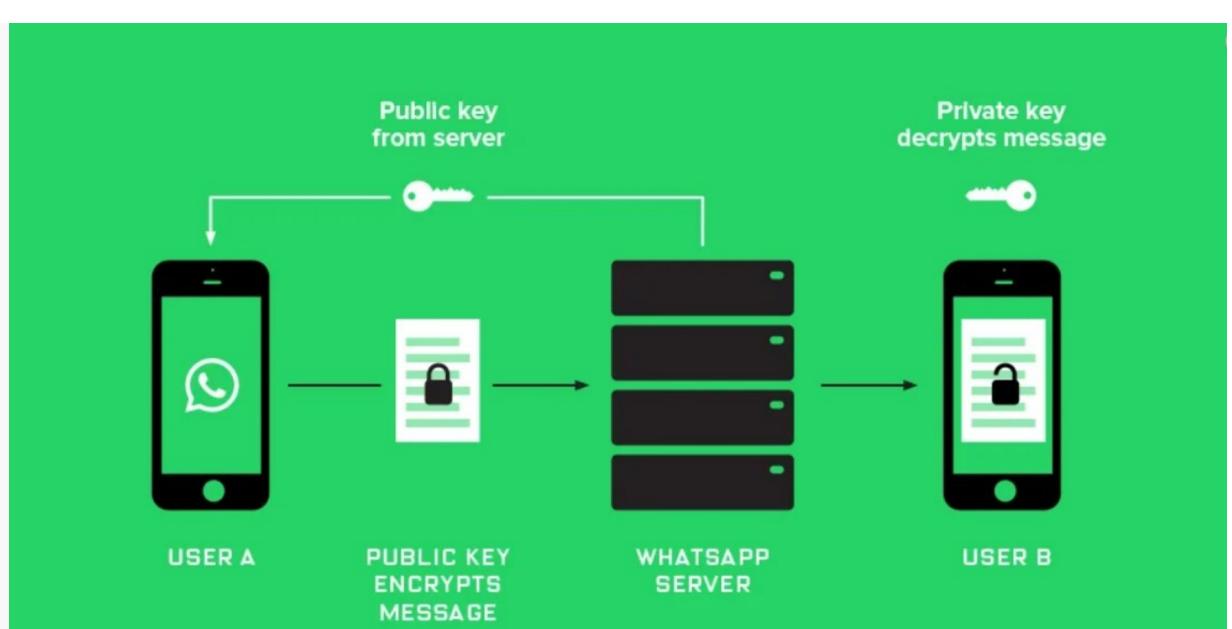
גילינו כי הדבר לא כל כך פשוט, כמובן, וואטסאף הינה אפליקציה המוצפנת מקופה לקשה -

מתוך ויקיפדיה:

בנובמבר 2014 הכריזה חברת Open Whisper Systems על שיתוף פעולה עם וואטסאפ להקמת תשתית להצפנה מkc-לקצה עם פרוטוקול סיגנל. לאחר הקמת התשתייה באפריל 2015 הצהירה החברה וואטסאפ שמתבצעת הצפנה חזקה של כל צורות התקשרות של היישוםן (שיחות וואטסאפ מוצפנות באמצעות פרוטוקול SRTP) באופן כזה, שלפי טענת החברה, אף אחד כולל חברת פיסבוק עצמה שוואטסאפ בבעלותה ואףלו ה-NSA או כל ארגון ממשלתי אחר לא יוכל לקבל גישה לתוכן של מיסרון או שיחה שלהם.

באוקטובר 2015 הינה וואטסאפ להעביר את כל הودעות המשמשים בישום לשרת גולן באמצעות גולן דרייב כברירת מחדל. עד זה הינה תהיות רבות בנוגע להצהרתת של החברה כי אף צד שלישי לא יוכל לקרוא את הודעות המשמשים, משומש שככל הודעות המשמשים מועלות לשרתים של חברות גולן ללא הצפנה מצד וואטסאפ אלא רק אם קיימת הצפנה על ידי השרת.

החל מапрיל 2016 החלו וואטסאפ להציג את כל ההודעות, התמונות, הסרטונים, ההודעות הקוליות, המسمכים והשיחות של משתמשים, באמצעות הצפנה מקיצה-לקצה. וטענה כי: "cashatam Sholahim hodaah, adam ha-yichid shiv'el le-krao oto ha-adam ao ha-kibutz alayhem atem sholahim at ha-hodaah. Af achad la-yicel le-krao at ha-hodaah shelchem: la-poushi siber, la-akrimim, la-mashterim medcaim, afilu la-anchno".



POCHILAH

חלק מהניסיונות שלנו, לא הצליחו לטעוד ע"י Wireshark חבילות אשר עברו בראשת שלנו כאשר אנו נוכנים לאפקטיביזה מהננייד, כל הניסיונות שלנו במסגרת הידע שלנו, ככלו מלהבין כיצד ניתן לזיהות פעילות כלשהי מהננייד. ולכן התמקדנו בחקירה החבילות אשר מעוברות בין המחשב שלנו (שמחובר דרך הדפסן לאפקטיביזה) לשרת הוואטסאפ. תחילית, להבנת התהיליך, תיעדנו שיחה בין שני אנשים בוואטסאפvr קר שאנו

מחוברים לאפליקציה דרך המחשב שלנו, והצד השני לא נמצא בסביבתנו. בשיחה זו הועברו הודעות טקסט, הודעות קוליות, תמונות סרטונים וקבצים. בנוסף, ניסינו להבין באיזה פרוטוקולים האפליקציה משתמשת, **ומצאנו כי ישנו כמה פרוטוקולים שוואטסאפ משתמש בהם כחלק משירותיה:** WhatsApp משתמש בשילוב של פרוטוקולים כדי לספק שירותים מסוימים מאובטחים ויעילים. פרוטוקולים אלה מבטיחים סודיות ויעילות.

1. TLS: מצפין את החיבור בין מכשיר המשתמש לשרת השירות WhatsApp. וכך מבטיח פרטיות ומונע גישה לא מורשית לנזונים במהלך השידור.

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1 0.000000000	157.240.253.60	192.168.86.38		TLSv1.2	343 ✓			Application Data
2 0.000000000	192.168.86.38	157.240.253.60		TCP	66 ✓			48446 → 443 [ACK] Seq=1 Ack=278 W

2. TCP: מספק מסירה אמינה, מבטיח קבלת הודעות בסדר הנכון ללא אובדן נתונים, במיוחד בתקשורת מבוססת טקסט.

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
7 2.424211149	157.240.253.60	192.168.86.38		TCP	66 ✓			443 → 48446 [ACK] Seq=278 Ack=159
8 2.457914510	157.240.253.60	192.168.86.38		TCP	66 ✓			443 → 48446 [ACK] Seq=278 Ack=234

3. UDP: העברת נתונים מהירה יותר, משמש לתקשורת בזמן אמת, כגון שיחות קוליות ואולי קצט שיתוף מולטימדיה. מושפר את חווית המשתמש על ידי הפחיתה עיכובים באינטראקציות רגישות בזמן.

4. QUIC: מיעיל את העברת הנתונים ברשות לא אמינים, תוך שימוש תכונות של TCP ו-UDP. במיוחד הודעות ושיחות קוליות.

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1 0.000000000	192.168.86.38	157.240.253.60		QUIC	1399 ✓		pps.whatapp...	0-RTT, DCID=6ec9a8ed868e50cf, SCI
2 0.120043722	157.240.253.60	192.168.86.38		QUIC	1274 ✓			Initial, DCID=2b7bde, SCID=b51d00

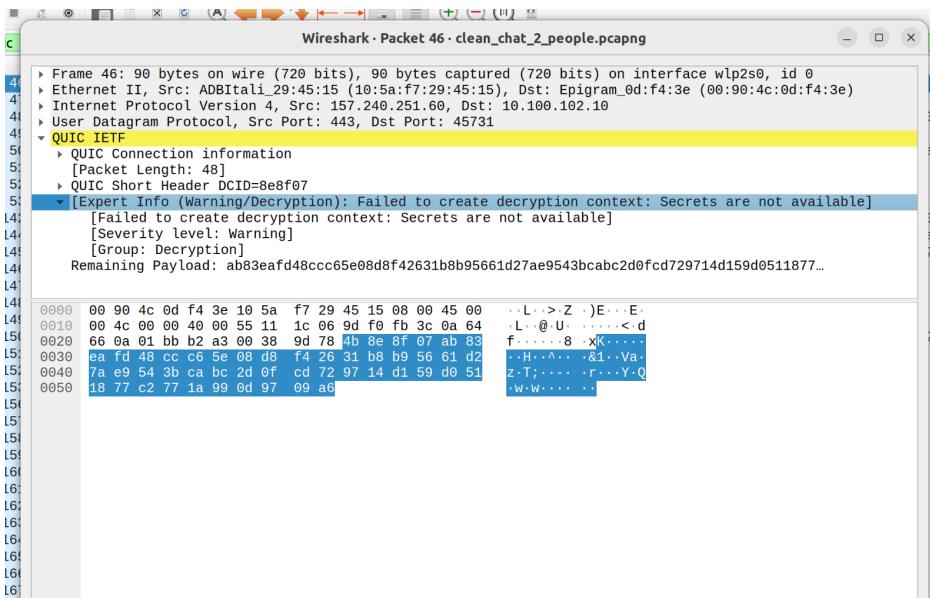
ובחזקה לשיחה שתיעדנו.

בווירשאך ניתן לראות את הabilities הרלוונטיות לשיחה באופן הבא:
בקלטה הנ"ל ניתן לראות הרבה abilities שונות בפרוטוקולים שונים מהמחשב שלנו (10.100.102.10) אל השרת 60 WHATAPP WEB של 157.240.251.60.

group_2.pcapng

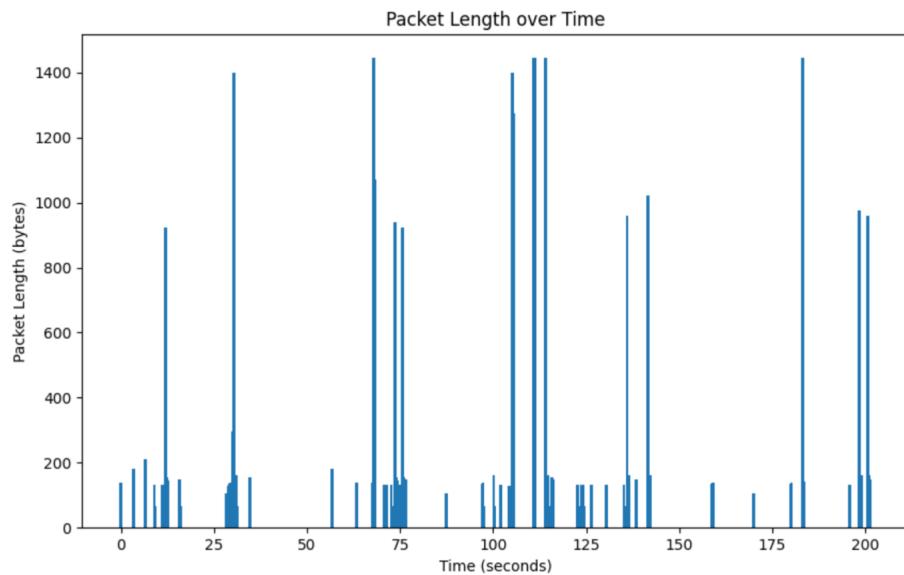
No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1	0.0000000000	10.100.102.10	157.240.251.60	QUIC	1399 ✓		web.whatssapp...	0-RTT, DCID=f793efb3a38fd9b5, SCID=Initial, DCID=8e8f07, SCID=401d43:
2	0.066538928	157.240.251.60	10.100.102.10	QUIC	1274 ✓			Handshake, DCID=8e8f07, SCID=401d43339543eb66,
3	0.067126543	10.100.102.10	157.240.251.60	QUIC	82 ✓			Handshake, DCID=8e8f07, SCID=401d43339543eb66,
4	0.068743158	157.240.251.60	10.100.102.10	QUIC	254 ✓			Handshake, DCID=8e8f07, SCID=401d43339543eb66,
5	0.068770229	157.240.251.60	10.100.102.10	QUIC	99 ✓			Protected Payload (KPO), DCID=8e8f07, SCID=401d43339543eb66,
6	0.068796877	157.240.251.60	10.100.102.10	QUIC	122 ✓			Protected Payload (KPO), DCID=8e8f07, SCID=401d43339543eb66,
7	0.069496744	10.100.102.10	157.240.251.60	QUIC	127 ✓			Handshake, DCID=8e8f07, SCID=401d43339543eb66,
8	0.090465541	10.100.102.10	157.240.251.60	QUIC	77 ✓			Protected Payload (KPO), DCID=401d43339543eb66,
9	0.126852705	157.240.251.60	10.100.102.10	QUIC	84 ✓			Handshake, DCID=8e8f07, SCID=401d43339543eb66,
10	0.127559714	157.240.251.60	10.100.102.10	QUIC	154 ✓			Protected Payload (KPO), DCID=8e8f07, SCID=401d43339543eb66,
11	0.127640502	157.240.251.60	10.100.102.10	QUIC	314 ✓			Protected Payload (KPO), DCID=8e8f07, SCID=401d43339543eb66,
12	0.128539741	10.100.102.10	157.240.251.60	QUIC	77 ✓			Protected Payload (KPO), DCID=401d43339543eb66,
13	3.999840426	10.100.102.2	10.100.102.255	UDP	77 ✓			60420 - 15600 Len=35
14	10.060697953	10.100.102.2	10.100.102.255	UDP	77 ✓			51535 - 15600 Len=35
15	16.096889705	10.100.102.2	10.100.102.255	UDP	77 ✓			35742 - 15600 Len=35
16	22.137491520	10.100.102.2	10.100.102.255	UDP	77 ✓			57452 - 15600 Len=35
17	28.050524097	10.100.102.2	10.100.102.255	UDP	77 ✓			41310 - 15600 Len=35
18	34.005752453	10.100.102.2	10.100.102.255	UDP	77 ✓			43096 - 15600 Len=35
19	38.037849477	10.100.102.10	62.0.32.32	QUIC	1399 ✓		media.fhfa2...	0-RTT, DCID=d29fb84c5815270a206e,
20	38.039450743	10.100.102.10	62.0.32.32	QUIC	473 ✓			0-RTT, DCID=d29fb84c5815270a206e,
21	38.044303729	62.0.32.32	10.100.102.10	QUIC	1274 ✓			Initial, DCID=c154c7, SCID=400708,
22	38.045215490	10.100.102.10	62.0.32.32	QUIC	82 ✓			Handshake, DCID=4007088e7233960d,
23	38.052091078	62.0.32.32	10.100.102.10	QUIC	255 ✓			Handshake, DCID=c154c7, SCID=4007,
24	38.052133599	62.0.32.32	10.100.102.10	QUIC	99 ✓			Protected Payload (KPO), DCID=c15,
25	38.052146425	62.0.32.32	10.100.102.10	QUIC	122 ✓			Protected Payload (KPO), DCID=c15,
26	38.052153572	62.0.32.32	10.100.102.10	QUIC	99 ✓			Protected Payload (KPO), DCID=c15,
27	38.053587686	10.100.102.10	62.0.32.32	QUIC	161 ✓			Protected Payload (KPO), DCID=400,
28	38.057809789	62.0.32.32	10.100.102.10	QUIC	84 ✓			Handshake, DCID=c154c7, SCID=4007,
29	38.060182444	62.0.32.32	10.100.102.10	QUIC	378 ✓			Protected Payload (KPO), DCID=c15,
30	38.081402193	10.100.102.10	62.0.32.32	QUIC	77 ✓			Protected Payload (KPO), DCID=400,
31	38.262324140	62.0.32.32	10.100.102.10	QUIC	1274 ✓			Protected Payload (KPO), DCID=c15,

כל החבילות אשר מועברות ע"י פרוטוקול QUIC, כנראה בתמונה, הינן החבילות שמעבירות את המידע של שיחת הוואטסאפ בין שני אנשים. ניתן לראות כי כל המידע הינו מוצפן ולא ניתן לגשת אליו.



בנוסף, יצרנו את התוכנית `messages_over_time.py` אשר מראה לנו באופן פשוט את גודל החבילות שנשלחות בראש לתוך כל השיחה, ניתן לראות את הקפיצות החודשת כאשר מדובר בשיחת תמונות/קבצים או סרטונים והודעות קוליות.

החבילות הקטנות יותר משמעותית יכולות להיות חלק מכל הودעה הגדולה או חבילות שנשלחות חלק מהפרוטוקול (לחיצות ידיים וכדומה).



Event-Based Detector

בפרויקט שלנו התמקדנו באlgorigthm התקיפה של Event-Based Detector עפ"י המאמר. لكن בשבייל להבין את הנושא שעליינו לחקור המשכננו לטעוד קבוצות וואטסאפ פשוטות יחסית(קבוצה של מעת אנשים עם אופי מובהק כמו שליחת תמונות בלבד), על מנת להבין כיצד להסיק את המסוקנות ולהגיע לתוכאות טבות, משם המשכננו לקבוצות יותר מורכב בפרויקט שלנו, אנו מתחילהים בחקר הדברים הפחותים והסקת המסוקנות מתוכם על מנת להמשיך לחלקים המורכבים יותר.

קובוצה מס' 1:

קובוצה זו מתאפיינת בשילוח תמונות, קבוצה שבה המנהל שולח תמונות, ככלומר העברת התוכן הינה באמצעות תמונות בלבד במאזענות המנהל אל המשתמשים.
ביצענו מאורע שבו בכל דקה נשלחה תמונה בקובוצה מהמנהל(שמוחובר ל web whatsapp).



באופן דומה כמו שראינו קודם אם נסנן את כל החבילות שעוברות בראשת שלנו ונשאר את אלו שימושísticas בפרוטוקול QUIC, הקלטת wireshark תראה כך:

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
10.0.000000000	192.168.86.38	213.57.24.97	QUIC	1399 ✓			media.fhfa1...	0-RTT, DCID=8db1ed47c1406cf310a7e, 0-RTT, DCID=8db1ed47c1406cf310a7e,
20.0.002112761	192.168.86.38	213.57.24.97	QUIC	476 ✓				Initial, DCID=d44ae0, SCID=400105
30.0.018320302	213.57.24.97	192.168.86.38	QUIC	1274 ✓				Handshake, DCID=d44ae0, SCID=400105
40.0.018373967	213.57.24.97	192.168.86.38	QUIC	256 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
50.0.019213872	213.57.24.97	192.168.86.38	QUIC	90 ✓				Handshake, DCID=400105127193df0c, DCID=d44ae0, SCID=400105
60.0.019601540	192.168.86.38	213.57.24.97	QUIC	127 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105127193df0c, DCID=d44ae0, SCID=400105
70.0.020348564	213.57.24.97	192.168.86.38	QUIC	122 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105127193df0c, DCID=d44ae0, SCID=400105
80.0.021241132	213.57.24.97	192.168.86.38	QUIC	90 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105127193df0c, DCID=d44ae0, SCID=400105
90.0.021380312	192.168.86.38	213.57.24.97	QUIC	76 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105127193df0c, DCID=d44ae0, SCID=400105
10.0.031720834	192.168.86.38	157.240.253.60	QUIC	1399 ✓			pps.whatsapp...	Initial, DCID=323192de96e91a0a66e, DCID=d44ae0, SCID=400105
11.0.034848246	213.57.24.97	192.168.86.38	QUIC	84 ✓				Handshake, DCID=d44ae0, SCID=400105
12.0.035991238	213.57.24.97	192.168.86.38	QUIC	378 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
13.0.056820967	192.168.86.38	213.57.24.97	QUIC	77 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
14.0.099303939	213.57.24.97	192.168.86.38	QUIC	282 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
15.0.099642186	192.168.86.38	213.57.24.97	QUIC	73 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
16.0.100116917	192.168.86.38	213.57.24.97	QUIC	390 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
17.0.104752081	157.240.253.60	192.168.86.38	QUIC	1274 ✓				Initial, DCID=25b693, SCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
18.0.104871017	157.240.253.60	192.168.86.38	QUIC	1274 ✓				Handshake, DCID=25b693, SCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
19.0.105467601	192.168.86.38	157.240.253.60	QUIC	87 ✓				Handshake, DCID=bb1d008f41700b34, SCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
20.0.105837234	157.240.253.60	192.168.86.38	QUIC	1274 ✓				Handshake, DCID=25b693, SCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
21.0.105969898	192.168.86.38	157.240.253.60	QUIC	87 ✓				Handshake, DCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
22.0.120450175	192.168.86.38	213.57.24.97	QUIC	77 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
23.0.148213008	157.240.253.60	192.168.86.38	QUIC	809 ✓				Handshake, DCID=25b693, SCID=bb1d008f41700b34, DCID=d44ae0, SCID=400105
24.0.148240308	157.240.253.60	192.168.86.38	QUIC	122 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105
25.0.148550387	213.57.24.97	192.168.86.38	QUIC	90 ✓				Protected Payload (KP0), DCID=d44ae0, SCID=400105

Frame 10: 1399 bytes on wire (11192 bits), 1399 bytes captured (11192 bits) on interface wlp2s0, id 0
 Ethernet II, Src: Epigram_0d:f4:3e (00:90:4c:0d:f4:3e), Dst: Google_29:2d:87 (24:05:88:29:d2:87)
 Internet Protocol Version 4, Src: 192.168.86.38, Dst: 157.240.253.60
 User Datagram Protocol, Src Port: 48776, Dst Port: 443
 QUIC IETF
 QUIC IETF

בהקלטה אנו מתעדים דבר מסקרן:

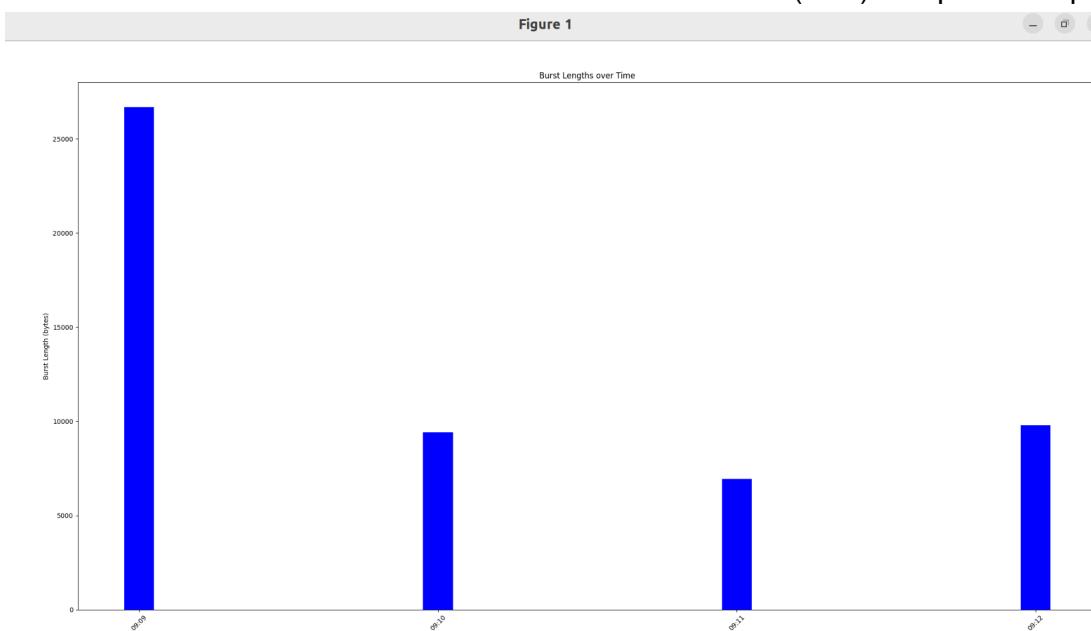
נראה כי בנוסף לחבר בין המחשב שלנו (192.168.86.38) לבין שרת הוואטסאפ (157.240.253.60) קיימם קשר נוסף בין המחשב שלנו לכתובת (213.57.24.97), לאחר בדיקה גילינו כי מדובר בשרת של HOT בישראל.

אנו מ닝חים כי שילוח התמונות מהמחשב שלנו לשרת הוואטסאפ נעשית באופן הבא: המחשב שלנו (הлокאלי), מתחילה את התהילר על ידי גישה ל-Web WhatsApp. זה כולל פתיחת דף דפן אינטרנט וביקור במכשיר האינטרנט של WhatsApp. WhatsApp עשויה להשתמש בressources תוקן (CDNs) כדי ליעל את אספקת המדייה, כגון תמונות. CDN אלה מורכבים משרתים המופצים במקומות שונים ברחבי העולם. ניתן שהשרת המדבר הינו השירות שמצאו בהקלטה. וכך אשר אנו שולחים תמונה היא מנוטבת דרך השירות זה.

גרף זיהוי האירועים:

על מנת לבצע את אלגוריתם התקיפה נשתמש בתוכנית event_extraction.py אשר יוצרת לנו את הפרדת האירועים (הסביר מפורט ניתן למצוא מפורט בithub(git)).

נקבל את הגרף הבא (0.01) :



אשר מראה לנו באופן מתאים את שליחת 4 התמונות בכל דקה בקבוצה כאירוע נפרד.

ניתן להסביר כי אופי הקבוצה הינו צזה שנשלחות בו קבצים גדולים (יחסית יותר 25,000 בתים)

בזמנים רציפים ואחדים ולכן ניתן להסביר כי מדובר בין אדם אחד ששולח את המידע.

בנוסף הגרף מתאר גם באופן דומה את גודל התמונות, חשוב לציין כי 3 התמונות האחרונות נשלו באיכות מאודו ולכך גודלן בהתאם.

אך נשים לב שהגודל התמונות כאן לא מתאר במדויק את גודל המקורי, ניסינו להבין מדוע הדבר כך והגענו למסקנות הבאות:

casar אנו שולחים תמונות דרך WhatsApp, האפליקציה דוחסת אותן לפני העברתן דרך הרשת. הדחיסה נעשית כדי להקטין את גודל התמונה, מכיוון שתמונות קטנות יותר דורשות פחות רווח פס ואחסון.

ניתן ליחס את הפער בין גודלי הקבצים למספר גורמים, כולל המרת פורמט תמונה, הפשטה מטה נתונים וڌيسה נוספת המימושת על ידי WhatsApp.

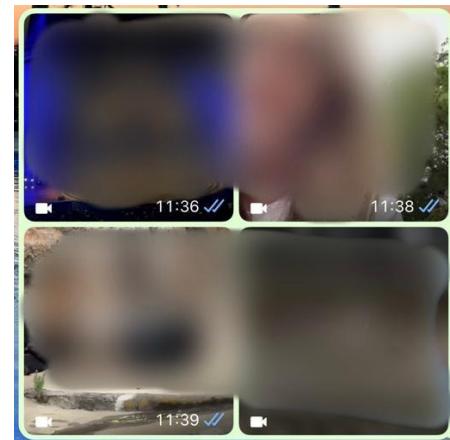
1. המרת פורמט תמונה: WhatsApp עשויה להמיר את התמונות לפורמט עיל יוטר (למשל, WEBP) שיגרם לגודלים קטנים יותר של קבצים.

2. הסרת מטה נתונים: WhatsApp עשויה להסיר כמה מטה נתונים או מידע לא חיוני מהתמונות, ולהקטין עוד יותר את גודל הקובץ. וכן אנו רואים שגודל החבילות שנשלחות הוא קטן יחסית.

לúcטום, ההבדל בגודלי הקבצים נובע מטכנייקות דחיסת התמונות והאופטיימיזציה של WhatsApp, שמקטינית את גודלי הקבצים להעברת רשת עיליה וחווית משתמש.

קבוצה מס' 2:

קבוצה שבה נשלחו 5 סרטונים מאיתנו למשתתפים בכל דקה.



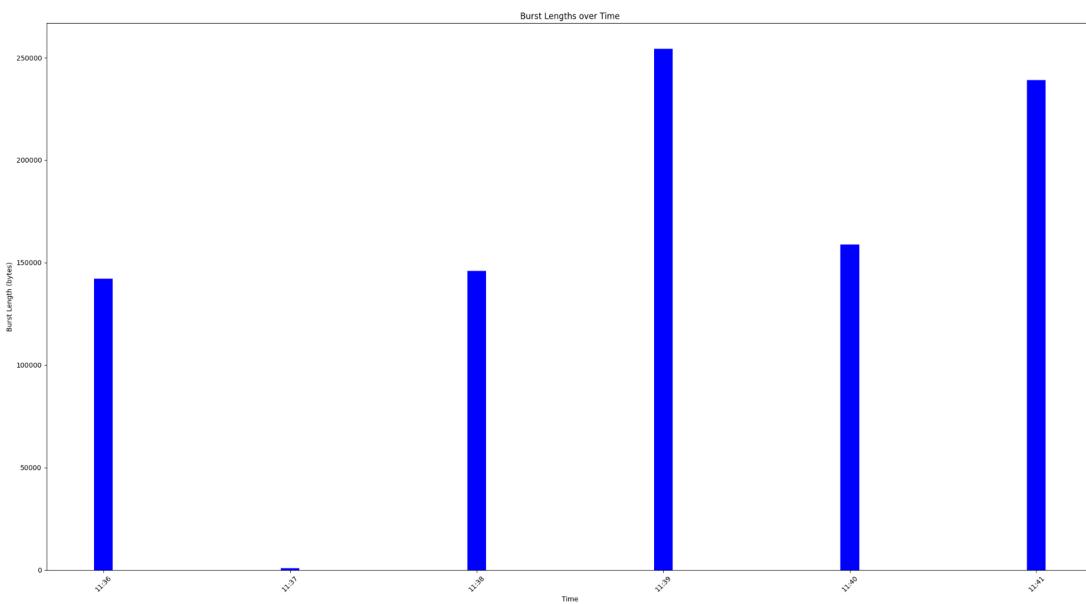
הקלטה:

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
10. 0.000000000	172.20.10.6	34.107.221.82	TCP	66 ✓				35736 → 80 [ACK] Seq=1 Ack=1 Win=
20. 0.000020767	172.20.10.6	34.107.221.82	TCP	66 ✓				43114 → 80 [ACK] Seq=1 Ack=1 Win=
30. 0.216525935	34.107.221.82	172.20.10.6	TCP	66 ✓				[TCP ACKed unseen segment] 80 → 4
40. 0.220537165	34.107.221.82	172.20.10.6	TCP	66 ✓				[TCP ACKed unseen segment] 80 → 3
52. 5.739466935	172.20.10.6	157.240.195.56	TLSv1.2	195 ✓				Application Data
62. 6.880021419	157.240.195.56	172.20.10.6	TCP	66 ✓				443 → 59972 [ACK] Seq=1 Ack=130 W.
73. 0.009454072	157.240.195.56	172.20.10.6	TLSv1.2	213 ✓				Application Data
83. 0.009491655	172.20.10.6	157.240.195.56	TCP	66 ✓				59972 → 443 [ACK] Seq=130 Ack=148
98. 0.114367551	172.20.10.6	34.117.65.55	TLSv1.2	105 ✓				Application Data
108. 0.336031295	34.117.65.55	172.20.10.6	TLSv1.2	105 ✓				Application Data
118. 0.336108458	172.20.10.6	34.117.65.55	TCP	66 ✓				34948 → 443 [ACK] Seq=40 Ack=40 W.
1210. 0.240066676	172.20.10.6	34.107.221.82	TCP	66 ✓				[TCP Dup ACK 1#1] 35736 → 80 [ACK]
1310. 0.244075117	172.20.10.6	34.107.221.82	TCP	66 ✓				[TCP Dup ACK 2#1] 43114 → 80 [ACK]
1410. 0.278830747	34.107.221.82	172.20.10.6	TCP	66 ✓				[TCP Dup ACK 4#1] [TCP ACKed unseen segment]
1510. 0.278877106	34.107.221.82	172.20.10.6	TCP	66 ✓				[TCP Dup ACK 3#1] [TCP ACKed unseen segment]
1610. 0.945864705	172.20.10.6	157.240.195.56	TLSv1.2	186 ✓				Application Data
1711. 0.058334835	157.240.195.56	172.20.10.6	TLSv1.2	101 ✓				Application Data
1811. 0.058374638	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=36
1911. 0.070705891	157.240.195.56	172.20.10.6	TCP	1446 ✓				443 → 59956 [ACK] Seq=36 Ack=121
2011. 0.070748907	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=141
2111. 0.070758783	157.240.195.56	172.20.10.6	TLSv1.2	200 ✓				Application Data
2211. 0.070767048	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=155
2311. 0.070876667	157.240.195.56	172.20.10.6	TCP	1446 ✓				443 → 59956 [ACK] Seq=1550 Ack=12
2411. 0.070888369	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=293
2511. 0.072097006	157.240.195.56	172.20.10.6	TLSv1.2	1446 ✓				Application Data [TCP segment of type SYN-ACK]
2611. 0.072145476	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=431
2711. 0.072164507	157.240.195.56	172.20.10.6	TCP	1446 ✓				443 → 59956 [ACK] Seq=4310 Ack=12
2811. 0.072179499	172.20.10.6	157.240.195.56	TCP	66 ✓				59956 → 443 [ACK] Seq=121 Ack=569

Frame 1: 66 bytes on wire (528 bits), 66 bytes captured (528 bits) on interface wlp2s0, id 0
 Ethernet II, Src: Unknown (00:0c:29:dd:52:b8), Dst: Unknown (00:0c:29:d4:20:0a)
 0000 c2 2c 5c d4 33 64 00 90 4c 00 f4 3e 08 00 45 00 ,\ 3d . L ..> .E.
 0010 00 34 32 65 40 00 40 06 52 87 ac 14 0a 06 22 6b 42e@ @. R .."K
 0020 dd 52 8b 98 00 50 fa c8 82 f0 e6 03 c9 62 80 10 R ..P .. .b..

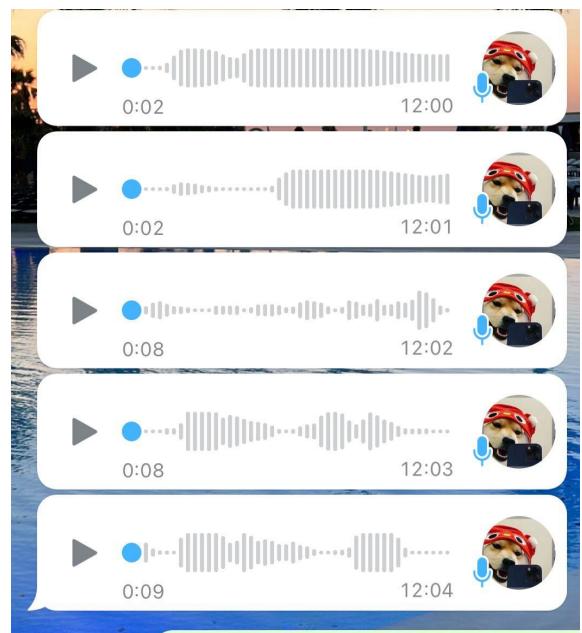
רואים כי ישן חבילות מפרוטוקלים שונים כמו שרהאים מקודם, ב באמצע CAN אנו גם חווים בעיה ברשת (DUP ACK שעליה למדנו בקורס) באינטרנט עם השרת של גוגל. 34.107.221.82.

גרף זיהוי האירועים:

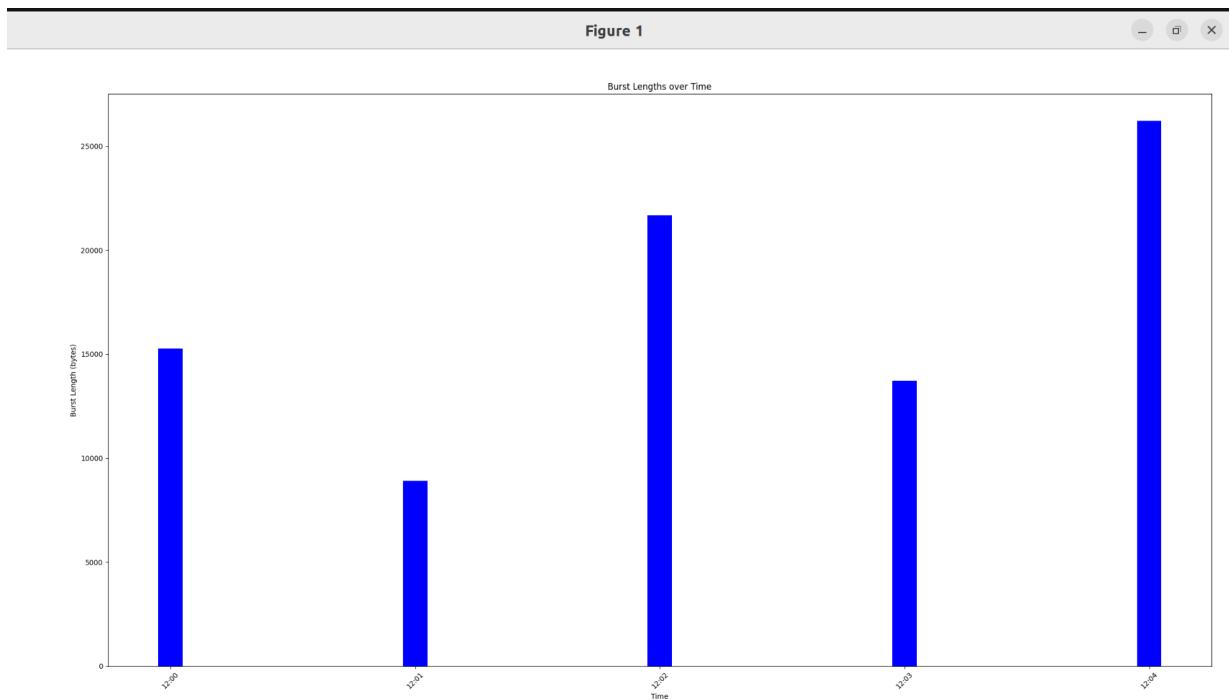


ע"י $th=0.1$ הצלחנו לזהות את הגדים וליצא את האירועים המתאים של שליחת סרטון בכל דקה, בנוסף כל אירוע מתאים לאודל הסרטוניים וב9:39:11 שלחנו סרטון ארוך יותר וכן גם רואים זאת מהגרף. בסיום הייתה לנו הפסקה קצרה בין שליחת ההודעות בהתחלה ורואים זאת בדקה 37 שיש שליחת מינימלית של חבילות.

קבוצה מס' 3: בקבוצה זו אנו קיבלנו הודעות קוליות בכל דקה ממשתתף בקבוצה



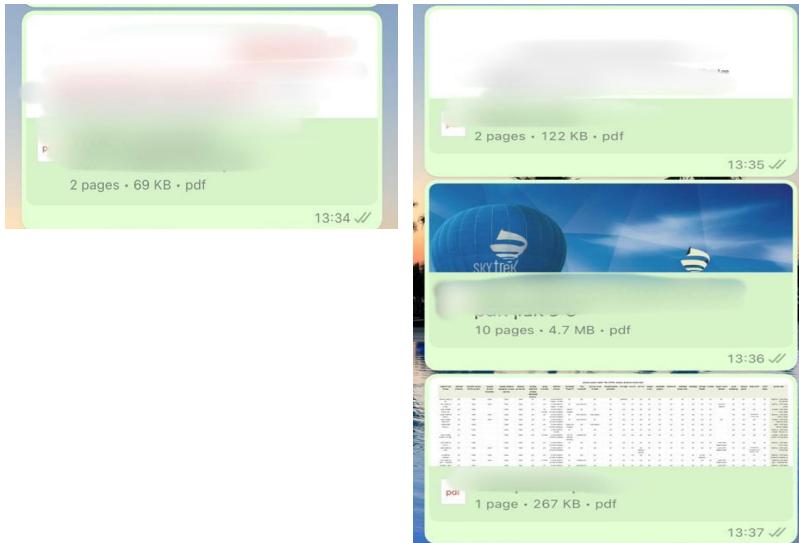
גרף זיהוי אירופין:



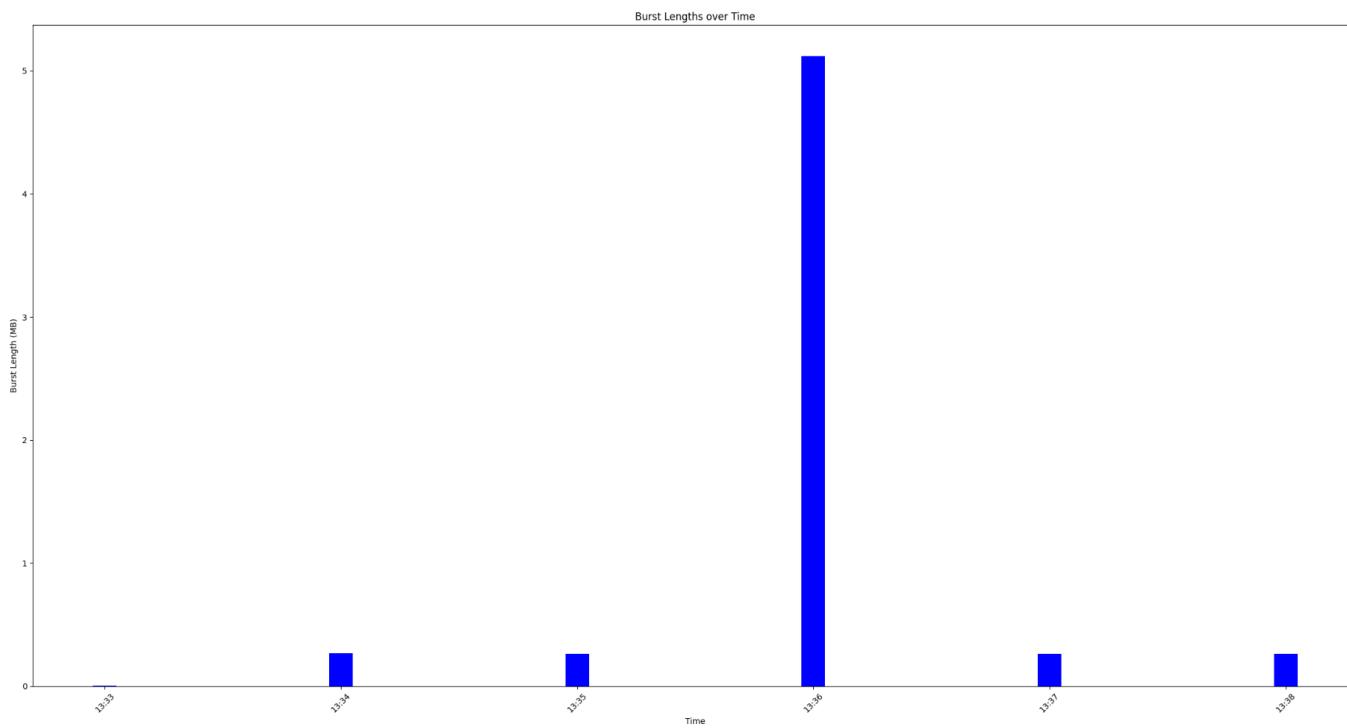
לפי 0.1=**that** אנו רואים כי אפילו שיש הודעות שזמן ההקלטה זהה אך יש עדין שינוי בגודלי האירועים ולכן אנו יכולים להניח כי הגודל לא רק תלוי בזמן אלא גם "ברצף הקולי" שמשפיע על גודל החבילה.

קבוצה מס' 4

בקבוצה זו שלחנו קבצי PDF שונים בגודלים שונים בכל דקה.



גרף זיהוי אירוחים:



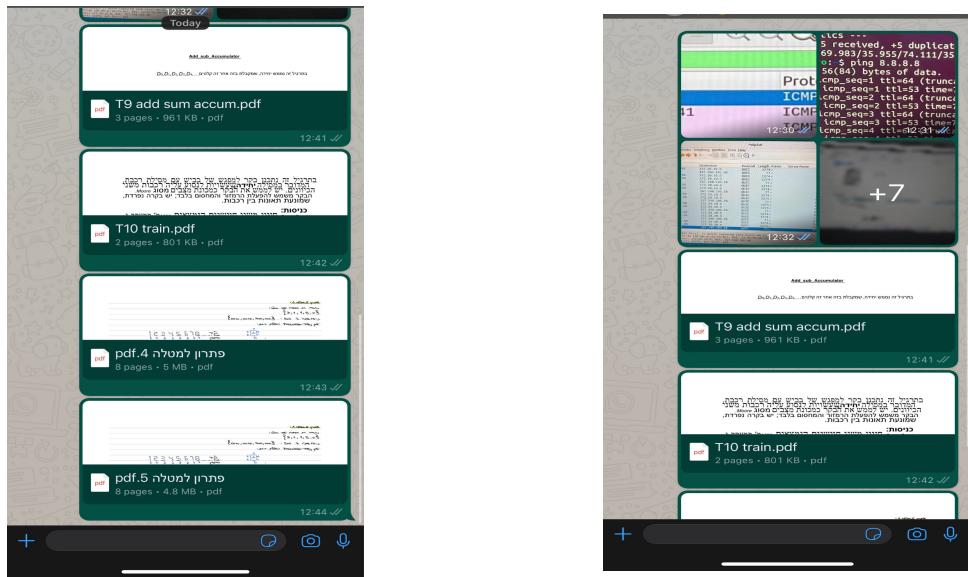
מכיוון שישנו קובץ של 5MB ייצרנו תוכנית נוספת לטעינה ב-5MB. רואים בצד שמאל מובהקת את השליחה של הקובץ הגדול ביותר ביחס לכל שאר ההודעות. בנוסף ב-13:38:13 רואים כי מקבלת עוד הודעה שאינה קשורה לקבוצה (牒בוצה אחרת) וגם רואים כיצד הדבר מופיע לנו לנוכח את המידע מה牒בוצה שאנו חווים.

牒בוצה מס' 5

קבוצה שבה המנהל (אנחנו) מעבירים תמונות וקבצים. תוך כדי שבחילון הפתוח הג'מייל שלו מקבל מיילים.
בקבוצה השיחה נראה כר :

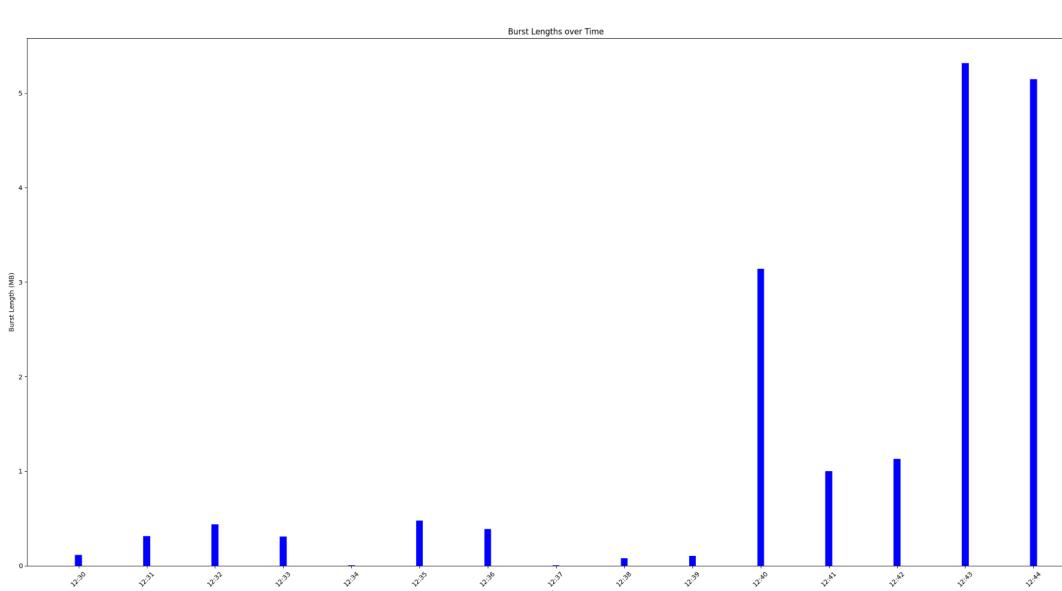
(2)

(1)



בקבוצה זאת התרבעו שליחת תמונות וקבצים באופן הבא :
נשלחו תמונות בהפרשים של דקה אחד אחרי השני, לאחר מכן נשלחו קבצים אחד אחרי השני בהפרשים של דקה . תוך כדי שליחת הודעות הללו הוואטסאפ היה פתוח ל"רushi רקע" , דוא"ל (gmail) היה פתוח תוך כדי קבלת הקבצים, בនוסף התקבלו גם הודעות טקסט מקובוצה אחרת. כך שכן אנו מודמים סיטואציה שבה התוקף מאמין לקבוצה מסוימת אך לモתקף יש רושע לא רק מהאפליקציה אלא גם מרשת.

גרף אירועים(MB):



הgraf מראה לנו את שליחת התמונות כל דקה ולמעשה ניתן לראות כי בממוצע התוקף יכול להסיק כי ההודעות שנשלחות כל דקה בממוצע הן פחות או יותר מבחינת משקל אותו דבר , ולאחר מכן יש קפיצה

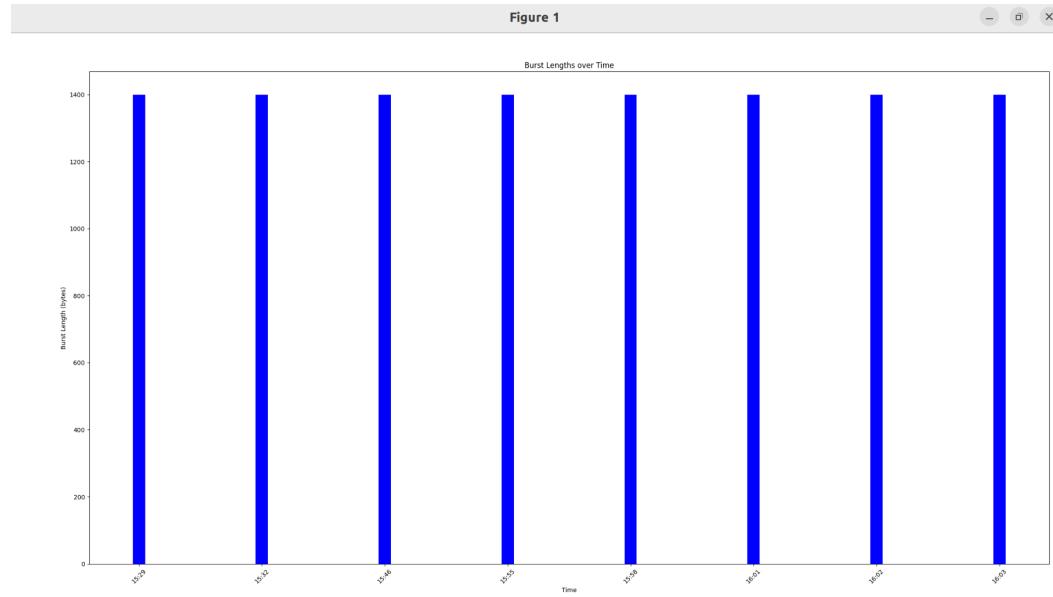
גדולה בגרף ב 12:40 כך שהתקף יכול להבין שמדובר בקפיצה חשודה , לעומת הودעה בעלת משקל יותר גדול ולהבין שמדובר בסוג קובץ אחר ולא תמונה כך למשה התקף מקבל המכחשה לאיזה סוג הודעות ישן בקובוצה ובכך להבין מה ה"אופי שלה", האם זאת הקבוצה הרצiosa יוכל להסיק גם שבגלל הגודל החיריג מדובר בהפרעה.(th=10)

קובוצה מס' 6:

תיעדנו קבוצה של קורס מהתוואר לאחר מבחן בקורס זה הוא זמן מאד "טויו".



גרף האירועים:

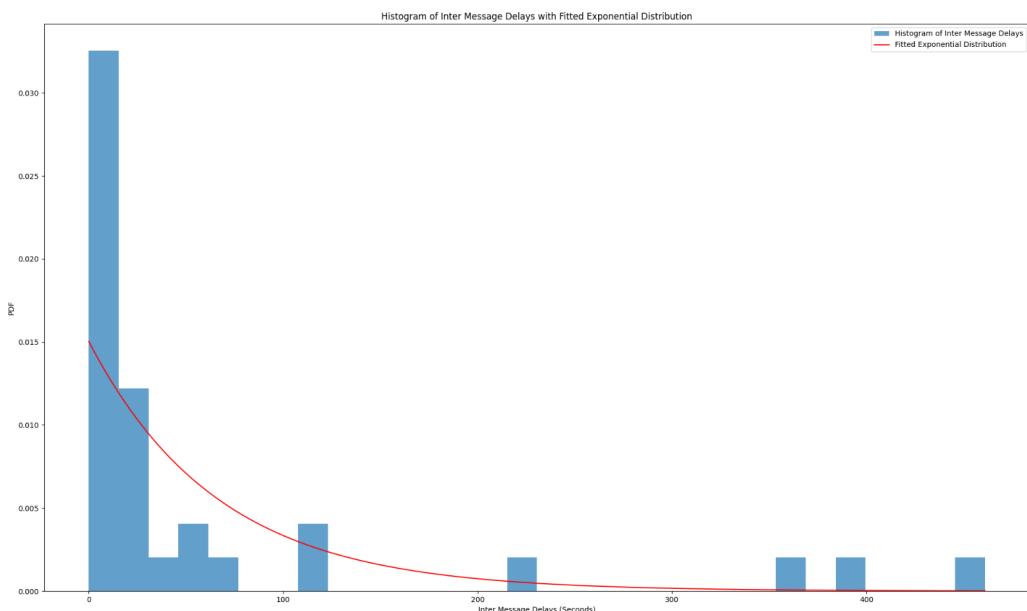


כאן רואים שוב שיש סוג של "קצב" בהודעות וכל דקה שלוחחים כמה אנשים הודעה, עדין האירועים קטנים מאוד ולכן ניתן להבין כי מדובר בהודעות טקסט, בקבוצת הודעות עם הרבה משתתפים בשיחה צפופה.

גרף PDF:

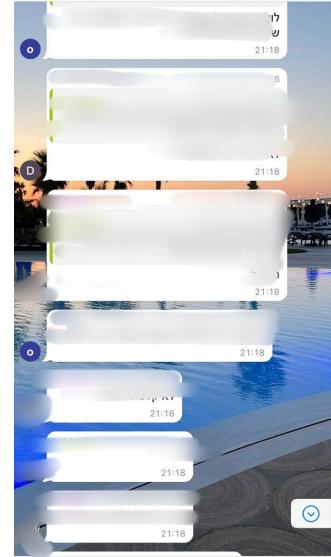
כאן אנו רואים איך זמן השהייה יורד כאשר זמן השיחה נמשך (יותר אנשים מצטרפים לשיחתונה ונהיי חלק ממנה לאחר היציאה מהבחן).

Figure 1

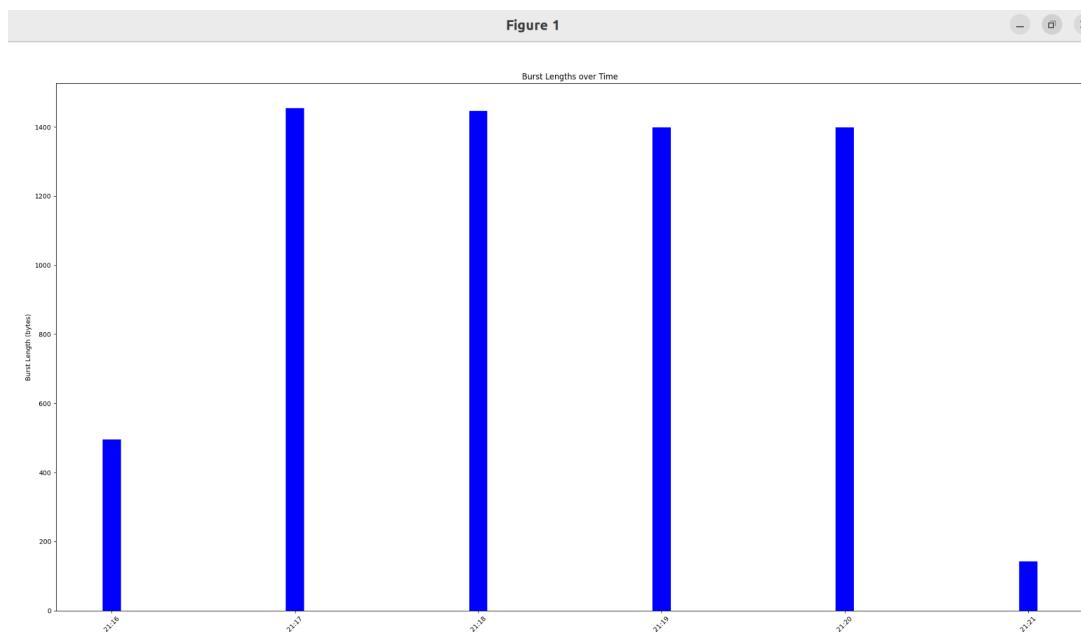


קבוצה מס' 7:

קבוצה נוספת (נוסף לאחר מכן) אשר הקלטנו לאחר מבחנים וניתן להבחן באופן רועש יחסית של הודעות שנשלחו יחד בדקה!

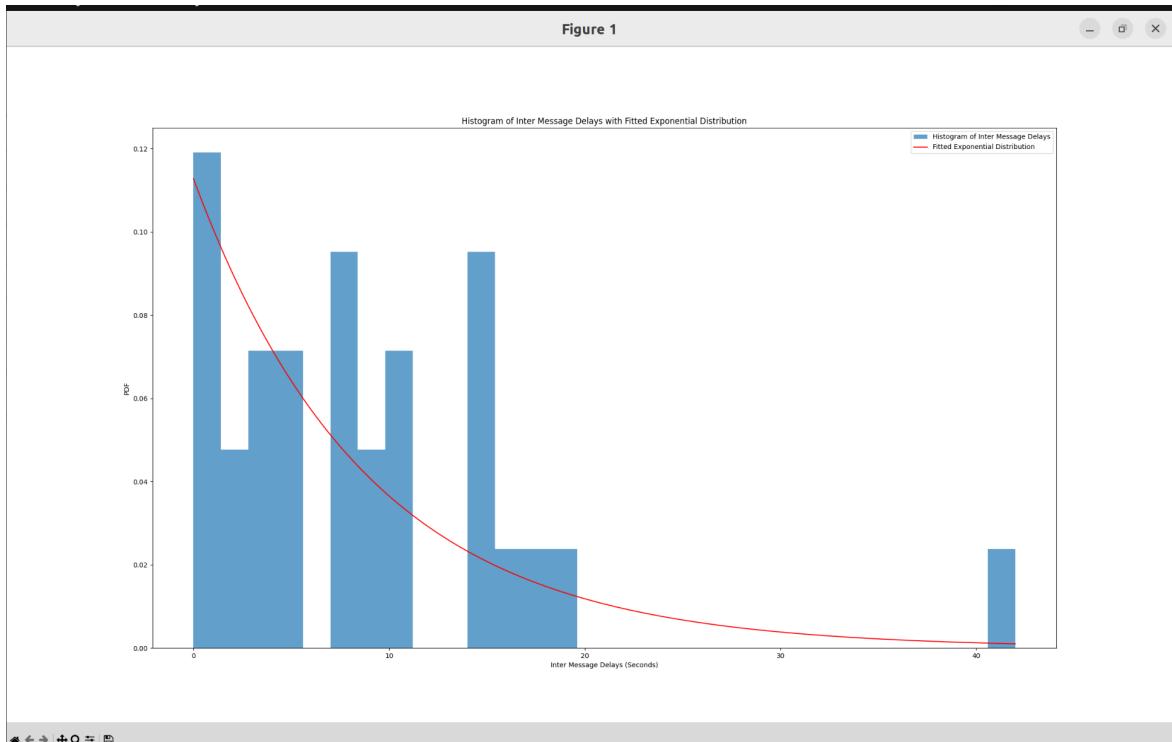


גרף אירועים:



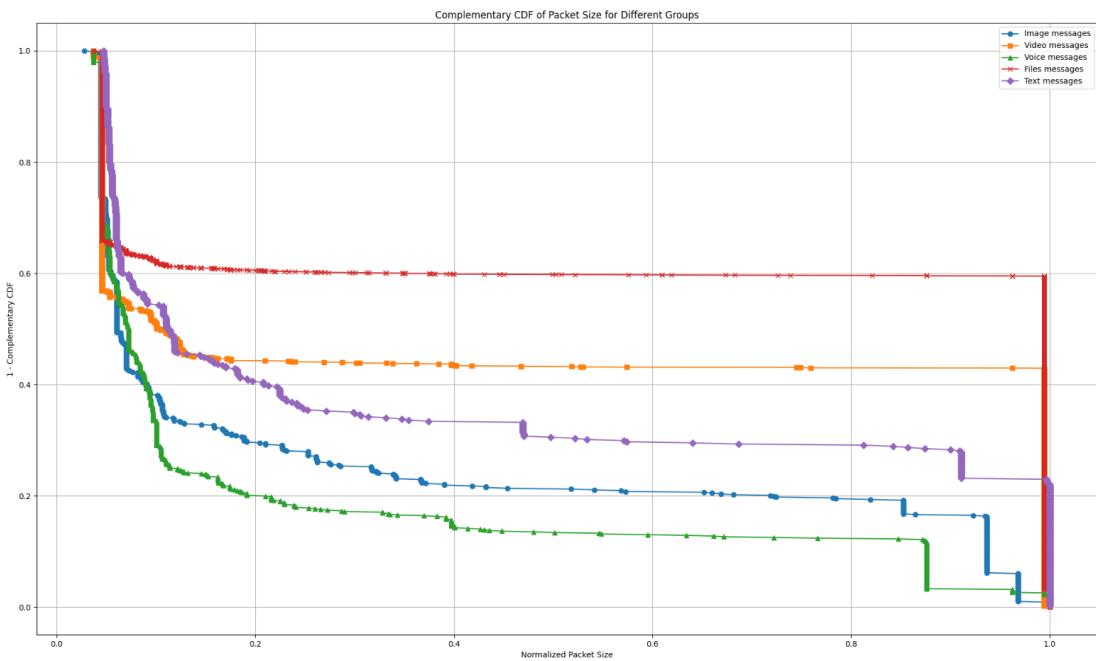
ראויים כי בכל דקה יש אירוע אחד הוא קטן יחסית, בהתייחס לכך שהודעת טקסט גודלה בין 200 ל-100 ביטים ניתן לראות כי נשלח כאן רצף הודעות גדול יחסית בכל דקה (הודעת טקסט).
0.0000001.

גרף PDF



ניתן לראות כי זמן השהייה הקצר יותר מגיע בתדריות גבוהה יותר וכן הקבוצה מאוד "רועשת" כמו שnitin לשים לב.

CDFF



ע"י קבוצות מספר 1,2,3,4,6 ייצרנו את גраф CCDF. הגרף מציג את פונקציית ההפוצה הצפיפה ההפוכה (Complementary Cumulative Distribution Function - CCDF) של גודלי החבילות מנורמלות לארבע קבוצות שונות. כל קבוצה מייצגת סוג אחר של העברת הودעה (טקסט, וידאו, תמונות וכו'). ציר ה-X מייצג את גודל החבילה המנורמל, הנע בין 0 ל-1, כאשר 1 מתאים לגודל המרבי שנרשם בקבוצה. ציר ה-Y מייצג את הפונקציה ההפוכה של פונקציית ההפוצה הצפיפה ההפוכה (CCDF - 1), שבעצם מראה את ההסתברות שגודל החבילה יהיה גדול יותר מס' מסויים. או במילים פשוטות, עבור כל קבוצה כל חבילה מנורמלת לפי גודלה ביחס לכל החבילות בקבוצה שלה ואנו יכולים לראות מהגרף מה ההסתברות שנראה אותה בקבוצה.

מסקנות וסיכום

בניטוח שלנו, חקרו את תעבורת הרשת מהודעות WhatsApp בקבוצות שונות. ערכנו תכיפות מעניינות והתמודדנו עם כמה אתגרים בענוקה הנתונים.

קבוצה מס' 1: תמונות בלבד

שmeno לב שליחת תמונות בוואטסאפ הובילה לאי-התאמות משמעותיות בגודל בין נתוני לכידת החבילות המקוריות לבין גDALי הקבצים המוצגים באפליקציה. דחיסת התמונה והמרת הפורמט של WhatsApp הפחתו במידה ניכרת את גDALי התמונות במהלך השידור, מה שמקשה להסיק במידוק את גודל התמונה המקורי מהמידע המוצג. הדבר יכול להקשות על התוקף להבין מה היא הקבוצה המדוברת ולזהות את השיעיות אליה.

קבוצה מס' 2: סוגי הודעות שונות

בקבוצה זו, היה לנו דפוס ברור ואמין של התנהלות הודעות. והתוקף יכול לזהות באופן ברור שאכן מדובר בקבוצה אם הוא הינו יכול מהקבוצה.

קבוצה מס' 3: הודעות מגוונות

בקבוצות עם סוגים מסוימים (תמונות, קבצים), הייתה ברורה יחסית ומאפשרת גילוי.

קבוצה מס' 4,5: הודעות טקסט

בקבוצות אלו אנו מסיקים כי אם התוקף נמצא בקבוצות עפ"י שכלל של שני הגרפים התוקף יכול להבין כי אנו נמצא בקבוצה.

ניתן להסיק כי האלגוריתמים במאמר אכן מציגים את הפוטנציאל להבחן בדפוסי תקשורת קבוצתיים. למרות שקיים אתגרים, ניתן זה מספק תובנות חשובות לגבי אופי קבוצות שונות ויכול להיות כל' רב ערך להבנת הדינמיקה של קבוצות WhatsApp.

באשר לשאלה האם ניתן להבין האם ניתן לזהות את הקבוצות משתתף בהם המותקף כאשר הוא מתקשר עם כמה קבוצות במקביל ניתן לראות כי הדבר יחסית מאוד מורכב. גם אם התוקף משתתף בקבוצות שאנו משתתפים בהם.

דבר נוסף שחקרנו הוא מדוע כאשר האזנו להקלטות של שיחות בוואטסאפ ניתן לראות שלפעמים רואים את החבילות כסוג HTTP QUIC TCP וכו' השאלה היא למה בכל פעם זה יכול להופיע באופן שונה ומה ניתן ללמוד מזה?

השונות בסוגי החבילות שנצפו בעת יירוט וניתוח תעבורת b-Web WhatsApp נובעת מօפי פרוטוקולו' התקשרות המשתמשים באפליקציה וממבנה הנתונים המועברים.

WhatsApp Web משתמש בפרוטוקולים בפורמטים שונים כדי להקל על התקשרות בין הלוקו' לבין הרשת.

כשאנו בדקנו את הקבוצות ניסינו לעשות זאת בזמןי "שקט" על מנת לקבל הקלטה כמה שייתר נקייה ולהיות בטוחים שלאלו אכן החבילות הקשורות לקבוצות הוואטסאפ היחידה שאנו פעילים בה בזמן קצוב.(בין אם אנו שלוחים הודעות או מקבלים הודעות). כאשר נשלחות הודעות בו זמנית מכמה שיחות, הדבר יותר מורכב ולא יהיה ניתן לזהות באופן מדויק אילו חבילות הקשורות לאיזו שיחה.

יותר מזה שmeno לב כי גם על התוקף לחשב וליציר גרף כמה שייתר אמייתי עם threshold מתאים על מנת לקבל את אופי האירועים בקבוצה. בנוסף עליו לנוקות את החבילות שמספריעות לו.

לכן אנו חושבים שבסופה של דבר נדרש זמן והשquaה והtokף צריך להיות בקבוצה על מנת לא "לlict לאיבוד" בכל המידע שעובר בראשת המותקף.

כאשר התוקף מנסה לאמת נכחות של אדם במספר קבוצות או עריצים, הם מתמודדים עם מספר אתגרים בשל האופי של אפליקציית הוואטסאפ שחקרנו. מדוע?

1. **הצפנה מקצת לנצח:** כאשר ההודעות מוצפנות, ניתן לעננה אותן רק במכשיר של מקבל ההודעה. המשמעות היא שגם אם היריב יוכל לירט את ההודעות המוצפנות, הוא לא יוכל לקרוא את התוכן ללא מפתחות הצפנה. הצפנה זו מקשה על התוקף לגשת שירותים לתוך ההודעות, ולהבין איזו הודעה קשורת לאיזה קבוצה.

2. ערבות הודעות: הודעות מקבוצות או שיחות שונות מתערבבות לעיתים קרובות בתעבורת הרשת. ערבות זה מאטגר את התקוף להבדיל בין מסרים השייכים לקבוצות שונות. הם יכולים לצפות בזרימת התעבורה ובגודל ההודעות, אך לא יכולה לפתחות ההצעה, לא ניתן לקבוע את התוכן המדויק או אילו הודעות שייכות לאיזה קבוצה.

3. גודל החבילות וזמן: בעוד התוכן של הודעות מוצפן, מטא נתונים מסוימים כגון תזמון הודעות, גודל ותדרות התקשרות עדין גולאים לתוקף. עם זאת, כפי שראינו בעובדה, מטא נתונים בלבד אינם שלא להיות מספיקים כדי לאמת באופן אמין בין נוכחות של אדם במספר קבוצות. לדוגמה, לקבוצות רבות עשויות להיות דפואים תנואה דומות, מה שהופך לאטגר.

4. זהויות מרובות: משתמשים לעיתים קרובות, משתמשים בכמה פלטפורמות כאשר הם נכנסים לאפליקציה, כמו מהמחשב/טלפון.

המשמעות היא שגם אם התקוף צופה באינטראקציות מרובות מממשיכים שונים, הוא לא יכול לקשר אותן באופן סופי לאותו אדם.

יתר על כן, משתמשים עשויים לעבור בין ממשיכים או לשנות את זהותם, מה שמסביר עוד יותר את התהילה.

ביבליוגרפיה: מפורטת בגרסה האנגלית.