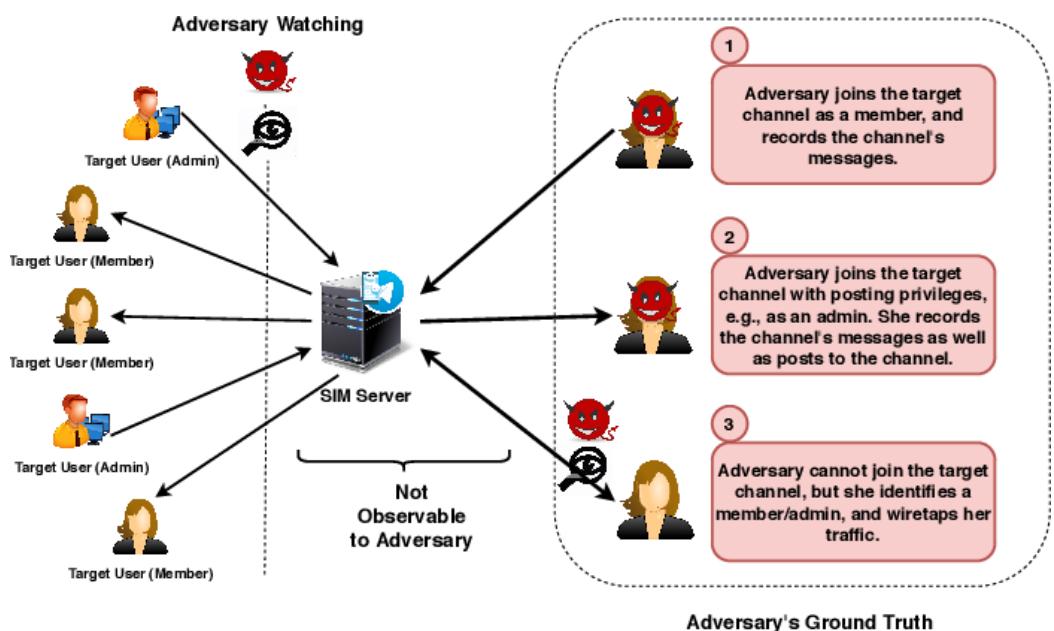


Practical Traffic Analysis

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 :

הצטיפות לערוץ פתוח: אם ערוץ MI היעד הוא ערוץ פתוח או ציבורי, התוקף יכול להציגו לערוץ חבר. 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 המידיות 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 והযוצאות, 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 לubahrot הרשות של החבר/מנהל המזוהה, ומירט את התקשרות המוצפנת שלהם. על ידי רישום דפוסי התנועה של החבר/מנהל הזוגה, התוקף יכול לשמש במידע זה כאמת בסיסית כדי להתאים לדפוסי התנועה של זרים יורטו אחרות.

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

יתר על כן התוקף מצוותת לubahrot הרשות של משתמשי MI כדי להזמין את כתובות ה-IP של החברים/מנהל המערכת של ערוץ ה-MI המשמש כיעד. השיטות הספציפיות שהזמין כוללות:

האזנה לספקן שירות אינטרנט (ISP) או ISPs או נקודות ISP (IXI) או נקודות אינטרנט (Internet Exchange). על ידי צוותת לubahrot הרשות העוברת דרך נקודות בקרה אלו, התוקף יכול לירט ולפקח על תעבורת הרשות המוצפנת של משתמשי הודעות מידיות.

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

כדי לציין צו צוותת לubahrot רשות בדרך כלל כרוך בליך וניתוח של מנוגת הנתונים המשודרות ברשות. דבר זה יכול להתבצע באמצעותים שונים, כגון פרישת התקני ניטור רשות, שימוש בכלי תוכנה או חומרה מיוחדים, או להתאפשר על תשתיית הרשות או על נקודות הקצה כדי לקבל גישה לubahrot. הפרטים המדוייקים כיצד התוקף מתקשר לubahrot רשות יכולים להיות תלויים ביכולות הטכנולוגיות והמשאבים הזמינים לתוקף.

ניתן לראות במחקר כי הטבלה II מספקת סקירה על התפלגות סוגי ההודעות השונים בתעborות ההודעות המידיות 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 תעborות ההודעות מידיות. לדוגמה, הודעות טקסט מהוות כ-29.4% מהסכום הכללי, ולאחריה תמנוגות (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 מידיים, המידיעים אותנו על בניית מודלים וניתוח של תעborות ההודעות מידיות, כמו גם פיתוח של מתפקידות ניתוח תעborות.

במערכות תקשורת, כל אירוע SIM, כגון שליחת תמונה, מייצר בועת של חבילות בגודל MT-U בקצבת MT-U בתנועת הקריפטוגרפיה. החבילות אלו עם דוחים צעירים בין החבילות וגודלה תואם את גודל ה-SIM.

הבושים מייצגות אירוע SIM, תוך שבחילות פחרות בגודל קטן מייצגות הודעות פרוטוקול ה-SIM כגון הודעות התראה, ידניות, עדכונים ועוד. כדי להזמין את הבושים 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, האובי 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'.
האובי משלב אירועים שכורבים יותר מ-te כאשר הוא מפעיל גידול בערך היעד, הגישה זו מאפשרת לאובי
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 הווואטסאפ, ובכלל כל תקשורת המתאפשרת לאפליקציה.

gilim Ci הדבר לא כל כך פשוט, CIDOU, וואטסאפ הינה אפליקציה המוצפנת קצת -

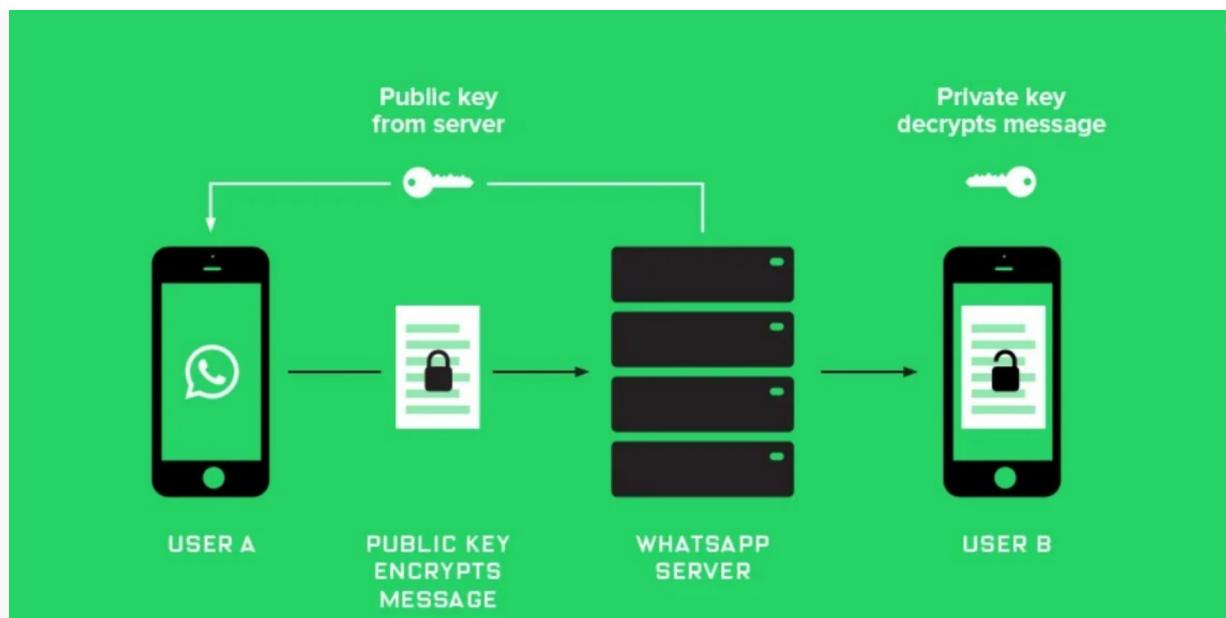
מתוך ויקיפדיה:

בנובמבר 2014 הכריזה חברת Open Whisper Systems על שיתוף פעולה עם וואטסאפ להקמת תשתיית הצפנה מקצה-לקצה עם פרוטוקול סיגנל. לאחר הקמת התשתיית באפריל 2015 הצהירה החברה חברת וואטסאפ שמתבצעת הצפנה חזקה של כל צורות התקשרות של היישוםן (שיחות וואטסאפ מצפנות באמצעות פרוטוקול SRTP) באופן צזה, שלפי טענת החברה, אף אחד כולל חברות פייסבוק עצמה שוואטסאפ בעולותה ואףלו ה-NSA או כל ארגון ממשלי אחר לא יכול לקבל גישה לתוכן של מיסרון או שיחה שלהם.

באוקטובר 2015 החלה וואטסאפ להעביר את כל הודעות המשתמשים ביישומון לשרתינו גוגל באמצעות גוגל דרייב כברירת מחדל. צעד זה העלה תהיות רבות בנוגע להצהרתת של החברה כי אף צד שלישי לא יוכל לקרוא את הודעות המשתמשים, משום שככל הודעות המשתמש מועלות לשירותים של חברות גוגל ללא הצפנה מצד וואטסאפ אלא רק אם קיימת הצפנה על ידי השרת.

החל מאפריל 2016 החלה וואטסאפ להציג את כל ההודעות, התמונות, הסרטונים, ההודעות הקוליות, המסמכים והשיחות של המשתמשים, באמצעות הצפנה מקצה-לקצה. וטענה כי: "כשאתם שולחים הודעה, האדם היחיד שיוכל לקרוא אותה הוא האדם או הקבוצה אליו הם אמורים שולחים את הודעה. אף אחד לא יוכל לקרוא את הודעה שלכם: לא פושעי סייבר, לא האקרים, לא מטרים מדכאים, אפילו לא אנחנו".

[תמונה מהאתר: <https://www.urtech.ca/2020/07/how-to-verify-that-whatsapp-messages-are-encrypted>](https://www.urtech.ca/2020/07/how-to-verify-that-whatsapp-messages-are-encrypted)



תחילת מחקר:

חלוקת מהניסיונות שלנו, לא הצליחנו לטעוד ע"י wireshark חבילות אשר עוברות ברשות שלנו כאשר אנו ננסים לאפליקציה מהניד, כל הניסיונות שלנו במסגרת הידע שלנו,فشل מלהבין כיצד ניתן לזהות פעילות כלשהי מהניד. ולכן התמקדמנו בחקירה החבילות אשר מעוברות בין המחשב שלנו (שמחובר דרך הדפדפן לאפליקציה) לשרת הוואטסאפ. תחילה, להבנת התהילך, תיעדנו שיחה בין שני אנשים בוואטסאפvr שאנו מחוברים לאפליקציה דרך המחשב שלנו, והצד השני נמצא בסביבתנו. בשיחה זו הועברו הודעות טקסט, הודעות קוליות, תמונות סרטונים וקבצים. בנוסף, ניסינו להבין באיזה פרוטוקולים האפליקציה משתמשת, ומיצאנו כי ישנו כמה פרוטוקולים שוואטסאפ משתמש בהם כחלק משירותיה:

בנוסף, WhatsApp משתמש בשילוב של פרוטוקולים כדי לספק שירותים הודעות מאובטחים ויעילים. פרוטוקולים אלה מבטיחים סודיות ויעילות.

1. TLS: מצפין את החיבור בין מכשיר המשתמש לשרת הוואטסאפ WhatsApp.

ורק מבטיח פרטיות ומונע גישה לא מורשית לנזונים במהלך השידור.

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

3. UDP: העברת נתונים מהירה יותר, משמש לתקשורת בזמן אמת, כגון שיחות קוליות ואולי קצט שיתוף מולטימדיה.

משפר את חווית המשתמש על ידי הפחיתה עיכובים באינטראקציות רגישות בזמן.

4. QUIC: מיעיל את העברת הנתונים ברשות לא אמינות, תוך שילוב תכונות של TCP ו-UDP.

במיוחד הודעות ושיחות קוליות.

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1 0.0000000000	157.240.253.60	192.168.86.38		TLSv1.2	343 ✓			Application Data
2 0.000028341	192.168.86.38	157.240.253.60		TCP	66 ✓			48446 → 443 [ACK] Seq=1 Ack=278 W

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1 0.0000000000	157.240.253.60	192.168.86.38		TLSv1.2	343 ✓			Application Data
2 0.000028341	192.168.86.38	157.240.253.60		TCP	66 ✓			48446 → 443 [ACK] Seq=1 Ack=278 W
3 0.000028341	157.240.253.60	192.168.86.38		TCP	66 ✓			48446 → 48446 [ACK] Seq=278 Ack=234

از כיצד בחרנו את החבילות לקביעות שתעדנו? הסתמכנו על הפרוטוקלים הללו, כל החבילות אשר היו מקושرات לשרת וואטסאפ, והגיעו מהמחשב שלנו או אל המחשב שלנו או דרך port 443.

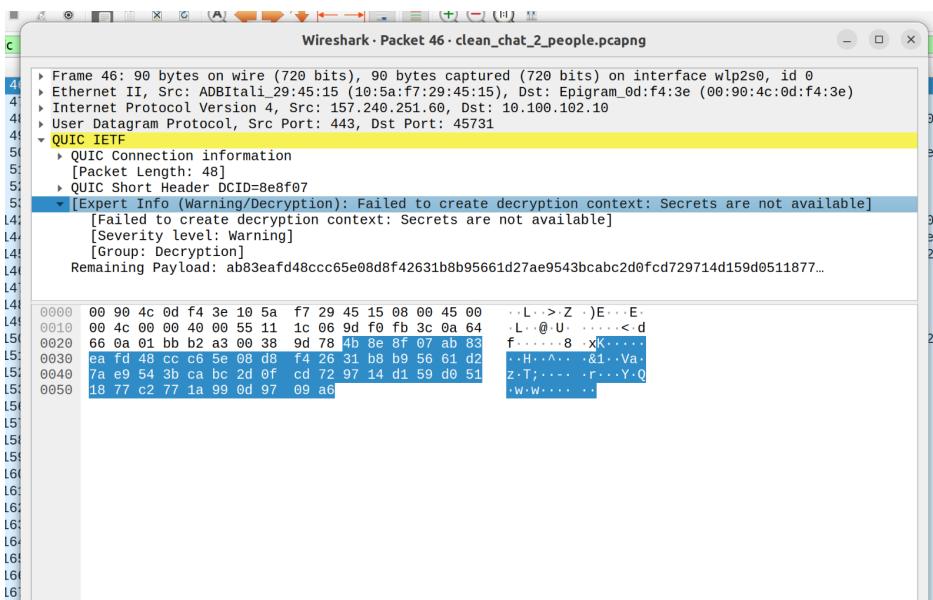
ניתנו את כל התعبורה אשר מפרקיה לנו (בהתאמה לממשיריהם אחרים שמחוברים לרשות שלנו).

ובעזרה לשיחה שתיעדנו

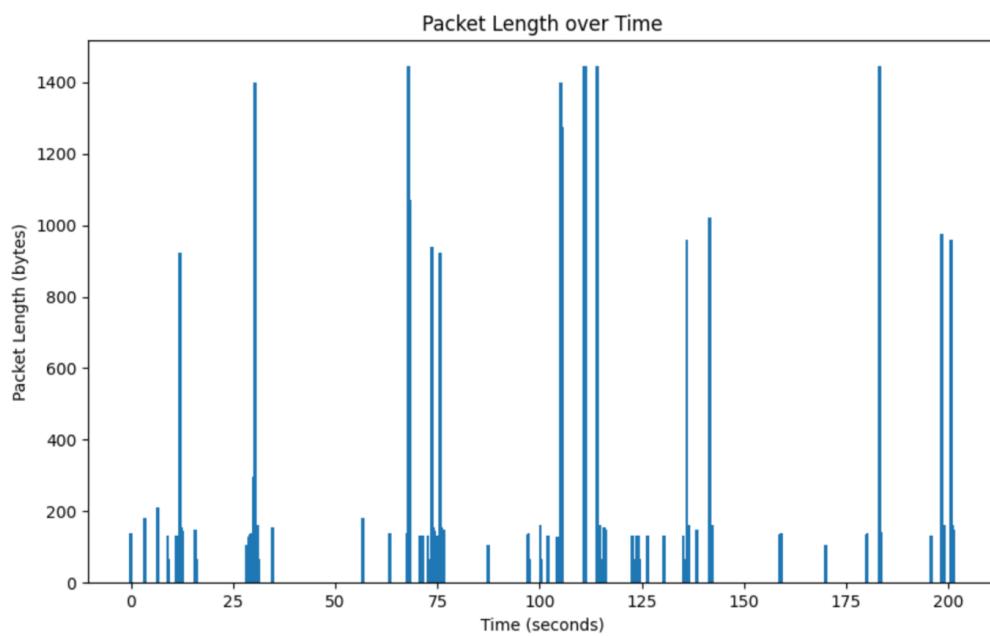
בווירשארק ניתן לראות את החבילות הרלוונטיות לשיחה באופן הבא:
בקלטה הנ"ל ניתן לראות הרבה חבילות שעוברות בפרוטוקולים שונים מהמחשב שלנו (10.100.102.10) אל הרשת 157.240.251.60 שהינו השירות של WHATAPP WEB.

No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info
1	0.000000000	10.100.102.10	157.240.251.60	QUIC	1399	✓	web.whatapp...	0-RTT, DCID=f793efb3a38fd9b5, SCID=
2	0.066538928	157.240.251.60	10.100.102.10	QUIC	1274	✓		Initial, DCID=8e8f07, SCID=401d43
3	0.067126543	10.100.102.10	157.240.251.60	QUIC	82	✓		Handshake, DCID=401d43339543eb66,
4	0.068743158	157.240.251.60	10.100.102.10	QUIC	254	✓		Handshake, DCID=8e8f07, SCID=401d
5	0.068770229	157.240.251.60	10.100.102.10	QUIC	99	✓		Protected Payload (KP0), DCID=8e8
6	0.068796877	157.240.251.60	10.100.102.10	QUIC	122	✓		Protected Payload (KP0), DCID=8e8
7	0.069496744	10.100.102.10	157.240.251.60	QUIC	127	✓		Protected Payload (KP0), DCID=8e8
8	0.098465541	10.100.102.10	157.240.251.60	QUIC	77	✓		Protected Payload (KP0), DCID=401
9	0.126852705	157.240.251.60	10.100.102.10	QUIC	84	✓		Handshake, DCID=8e8f07, SCID=401d
10	0.127559714	157.240.251.60	10.100.102.10	QUIC	154	✓		Protected Payload (KP0), DCID=8e8
11	0.127640502	157.240.251.60	10.100.102.10	QUIC	314	✓		Protected Payload (KP0), DCID=8e8
12	0.128539741	10.100.102.10	157.240.251.60	QUIC	77	✓		Protected Payload (KP0), DCID=401
13	3.999840426	10.100.102.2	10.100.102.255	UDP	77	✓		60420 - 15600 Len=35
14	10.000697953	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	28.005024097	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 (KP0), DCID=c15
25	38.052146425	62.0.32.32	10.100.102.10	QUIC	122	✓		Protected Payload (KP0), DCID=c15
26	38.052153572	62.0.32.32	10.100.102.10	QUIC	99	✓		Protected Payload (KP0), DCID=c15
27	38.053587686	10.100.102.10	62.0.32.32	QUIC	161	✓		Protected Payload (KP0), 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 (KP0), DCID=c15
30	38.081402193	10.100.102.10	62.0.32.32	QUIC	77	✓		Protected Payload (KP0), DCID=400
31	38.262324140	62.0.32.32	10.100.102.10	QUIC	1274	✓		Protected Payload (KP0), 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 ברשת לאורך כל השיחה, ניתן לראות את הקפיצות החודשת כאשר מדובר בשיחת של תמונות/קבצים או סרטונים והודעות קוליות. החביבות הקטנות יותר משמעותית יכולות להיות חלק מכל הودעה הגדולה או חביבות שנשלחות כחלק מהפרוטוקול (לחיצות ידיים וכדומה).

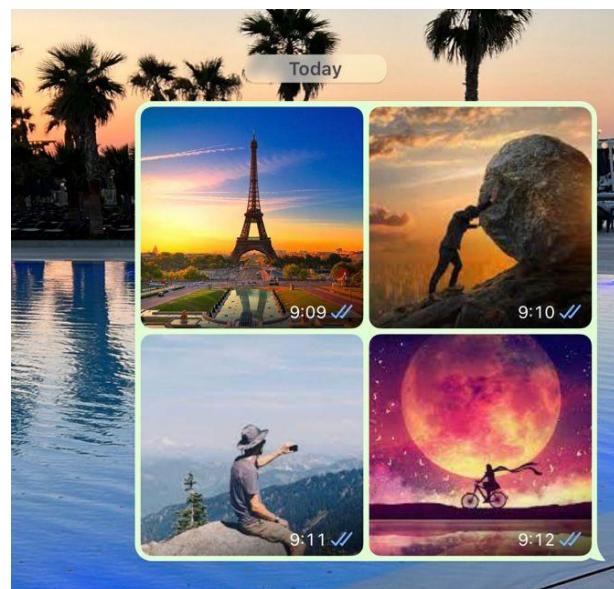


תיעודים-Event-Based Detector

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

קבוצת מס' 1:

קבוצת זו מתאפיינת בשליחת תמונות, קבוצה שבה המנהל שולח תמונות, ככלומר העברת התוכן הינה באמצעות תמונות בלבד בלבד באמצעות המנהל אל משתמשים. ביצענו מאורע שבו בכל דקה נשלחה תמונה בקבוצה מהמנהל(שמחבר ל web whatsapp).



באופן דומה כמו שראינו מקודם אם נסנן את כל החבילות שעוברות בראשת שלנו ונשאר את אלו שימושת ב프וטוקול QUIC, הקלטת wireshark תראה כך:

photo_messaging_clean.pcapng									
No.	Time	Source	Destination	Protocol	Length	Frame	Server Name	Info	
10.0.031720834	192.168.86.38	157.240.253.60	192.168.86.38	QUIC	1399	✓	media.fhfa1-...	0-RTT, DCID=8db1ed47c1406cf310a7e	
11.0.034848246	213.57.24.97	192.168.86.38	192.168.86.38	QUIC	84	✓		Handshake, DCID=d44ae09, SCID=400105	
12.0.035991238	213.57.24.97	192.168.86.38	192.168.86.38	QUIC	378	✓		Protected Payload (KPO), DCID=d44	
13.0.056828967	192.168.86.38	213.57.24.97	192.168.86.38	QUIC	77	✓		Protected Payload (KPO), DCID=d400	
14.0.099303939	213.57.24.97	192.168.86.38	192.168.86.38	QUIC	282	✓		Protected Payload (KPO), DCID=d44	
15.0.099642186	192.168.86.38	213.57.24.97	192.168.86.38	QUIC	73	✓		Protected Payload (KPO), DCID=d400	
16.0.100116917	192.168.86.38	213.57.24.97	192.168.86.38	QUIC	390	✓		Protected Payload (KPO), DCID=d400	
17.0.104752081	157.240.253.60	192.168.86.38	192.168.86.38	QUIC	1274	✓		Initial, DCID=25b693, SCID=bb1d00	
18.0.104871817	157.240.253.60	192.168.86.38	192.168.86.38	QUIC	1274	✓		Handshake, DCID=25b693, SCID=bb1d	
19.0.105467601	192.168.86.38	157.240.253.60	192.168.86.38	QUIC	87	✓		Handshake, DCID=bb1d008f41700b34	
20.0.105837234	157.240.253.60	192.168.86.38	192.168.86.38	QUIC	1274	✓		Handshake, DCID=25b693, SCID=bb1d	
21.0.105969898	192.168.86.38	157.240.253.60	192.168.86.38	QUIC	87	✓		Handshake, DCID=bb1d008f41700b34	
22.0.120450175	192.168.86.38	213.57.24.97	192.168.86.38	QUIC	77	✓		Protected Payload (KPO), DCID=400	
23.0.148213008	157.240.253.60	192.168.86.38	192.168.86.38	QUIC	809	✓		Handshake, DCID=25b693, SCID=bb1d	
24.0.148240308	157.240.253.60	192.168.86.38	192.168.86.38	QUIC	122	✓		Protected Payload (KPO), DCID=25b	
25.0.148550387	213.57.24.97	192.168.86.38	192.168.86.38	QUIC	90	✓		Protected Payload (KPO), DCID=d44	

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 (24:05:88:29:2d: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

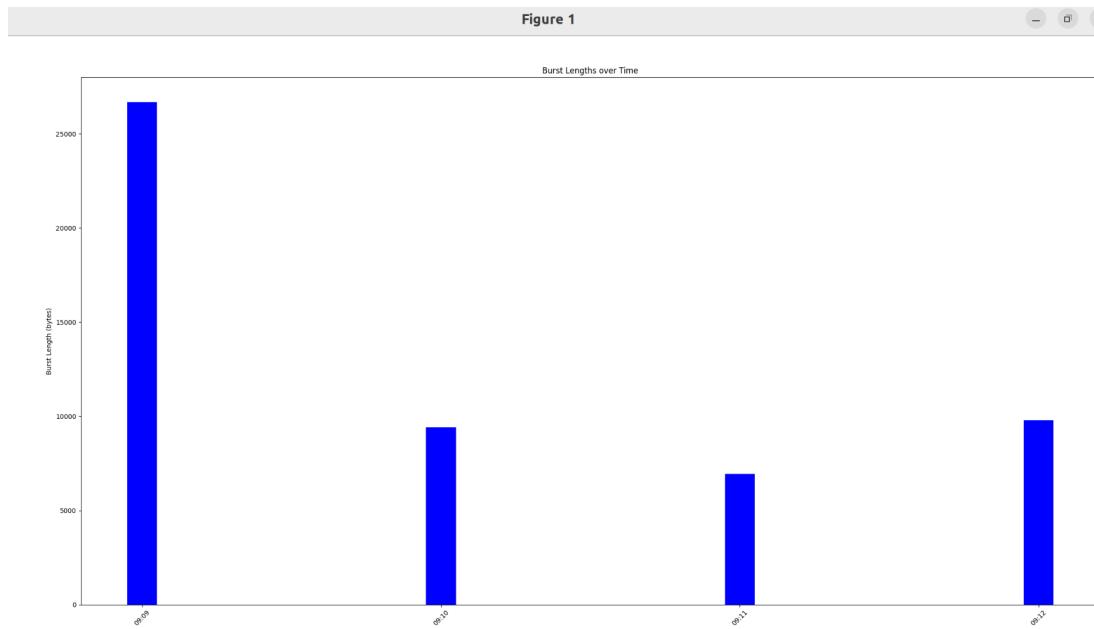
בתקלה אנו מתעדים דבר מסקרן:

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

אנו מניחים כי שליחת התמונות מהמחשב שלנו לשרת הוואטסאפ נעשית באופן הבא: המחשב שלנו (הלקוח), מתחילה את התהליך על ידי גישה ל-Web WhatsApp. זה כולל פתיחת דף אינטרנט ובקור במכשיר האינטראקט של WhatsApp. WhatsApp עשויה להשתמש ברשומות להעברת תוכן (CDNs) כדי ליעל את אספקת המדיה, כגון תמונות. CDNs אלה מורכבים משרתים המופצים ברחבי העולם. ניתן שהשרת המדובר הינו השירות שמצאו בתקלה. ואשר אנו שולחים תמונה היא מנוטבת דרך שרת זה.

גרף זיהוי האירועים:

עבו כל גרפ מסונן בכחול את הזמן `th` על מנת לקבל את גרפ האירועים המתאים על מנת לבצע את אלגוריתם התקיפה נשתמש בתוכנית `event_extraction.py` אשר יוצרת לנו את הפרדת האירועים (הסביר מפורט ניתן למצוא מפורט בטיבר github.org). נקבל את הגרף הבא (0.01):



אשר מראה לנו באופן מתאים את שליחת 4 התמונות בכל דקה בקצבה כאירוע נפרד. ניתן להסיק כי אופי הקבוצה הינו צזה שנשלחות בו קבצים גדולים (יותר מ-25,000 בתים) בזמןים רציפים ואחדים ולכן ניתן להסיק כי מדובר בין אדם אחד ששולח את המידע. בKİוסק הגרף מתרגם גם באופן דומה את גודל התמונות, חשוב לציין כי 3 התמונות האחרונות נשלחו באיכות מאוד גבוהה וכן גודל בהתקדים.

אר נשים לב שגודל התמונות כאן לא מתרגם מדויק את גודל המקורי, ניסינו להבין מדויק הדבר כך והגענו למסקנות הבאות:

כאשר 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 התמונה, הפשטה מטה נתוניים וڌחיסה נוספת המימושת על ידי WhatsApp.

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

2. הוסרת 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. לסתיכום, ההבדל בגודל

הקבצים נובע מטכניות דחיסת התמונות והאופטיימיזציה של WhatsApp, שמקטיננות את גדי הקבצים להעברת רשת יעילה וחווית משתמש.

קבוצה מס' 2

קבוצה שבה נשלחו 5 סרטונים מאיתנו למשתתפים בכל דקה.



הקלטה:

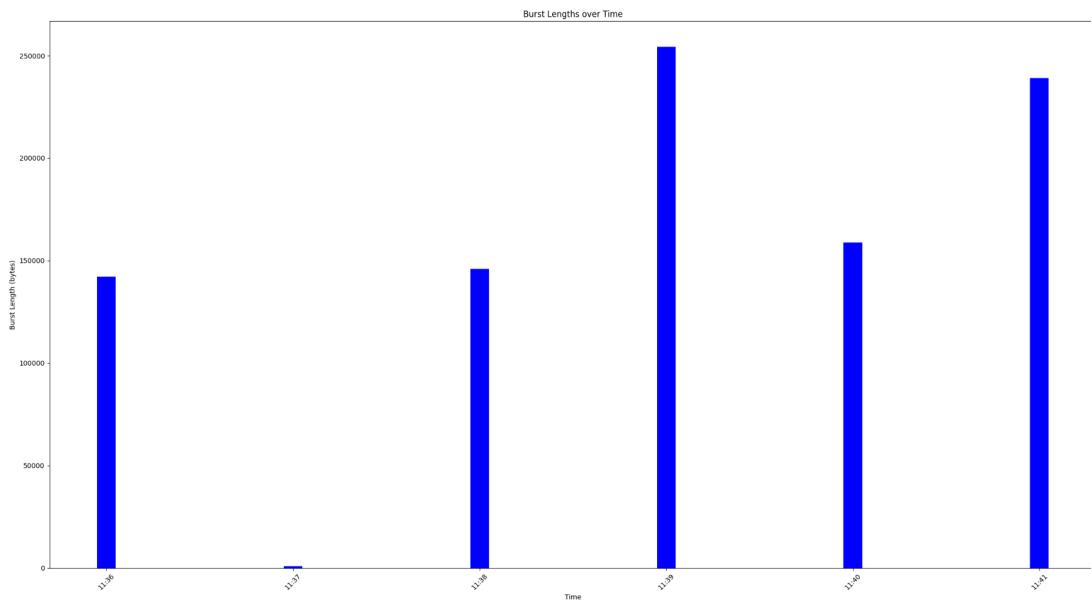
Wireshark Network Traffic Analysis							
No.	Time	Source	Destination	Protocol	Length	Frame	Server Name
1	0.000000000	172.20.10.6	34.107.221.82	TCP	66 ✓		
2	0.000020767	172.20.10.6	34.107.221.82	TCP	66 ✓		
3	0.216525935	34.107.221.82	172.20.10.6	TCP	66 ✓		
4	0.220537165	34.107.221.82	172.20.10.6	TCP	66 ✓		
5	2.739466935	172.20.10.6	157.240.195.56	TLSv1.2	195 ✓		
6	2.880021419	157.240.195.56	172.20.10.6	TCP	66 ✓		
7	3.009454072	157.240.195.56	172.20.10.6	TLSv1.2	213 ✓		
8	3.009491655	172.20.10.6	157.240.195.56	TCP	66 ✓		
9	8.114367551	172.20.10.6	34.117.65.55	TLSv1.2	105 ✓		
10	8.336031295	34.117.65.55	172.20.10.6	TLSv1.2	105 ✓		
11	8.336108458	172.20.10.6	34.117.65.55	TCP	66 ✓		
12	10.240066676	172.20.10.6	34.107.221.82	TCP	66 ✓		
13	10.244075117	172.20.10.6	34.107.221.82	TCP	66 ✓		
14	10.278830747	34.107.221.82	172.20.10.6	TCP	66 ✓		
15	10.278877106	34.107.221.82	172.20.10.6	TCP	66 ✓		
16	10.945864765	172.20.10.6	157.240.195.56	TLSv1.2	186 ✓		
17	11.058334835	157.240.195.56	172.20.10.6	TLSv1.2	101 ✓		
18	11.058374638	172.20.10.6	157.240.195.56	TCP	66 ✓		
19	11.070705891	157.240.195.56	172.20.10.6	TCP	1446 ✓		
20	11.070748907	172.20.10.6	157.240.195.56	TCP	66 ✓		
21	11.070758783	157.240.195.56	172.20.10.6	TLSv1.2	206 ✓		
22	11.070767048	172.20.10.6	157.240.195.56	TCP	66 ✓		
23	11.070876667	157.240.195.56	172.20.10.6	TCP	1446 ✓		
24	11.070888369	172.20.10.6	157.240.195.56	TCP	66 ✓		
25	11.072097006	157.240.195.56	172.20.10.6	TLSv1.2	1446 ✓		
26	11.072145476	172.20.10.6	157.240.195.56	TCP	66 ✓		
27	11.072164507	157.240.195.56	172.20.10.6	TCP	1446 ✓		
28	11.072179499	172.20.10.6	157.240.195.56	TCP	66 ✓		

Frame 1: 66 bytes on wire (528 bits), 66 bytes captured (528 bits) on interface wlp2s0, id 0

```
0000  c2 2c 5c d4 33 64 00 90 4c 0d 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
```

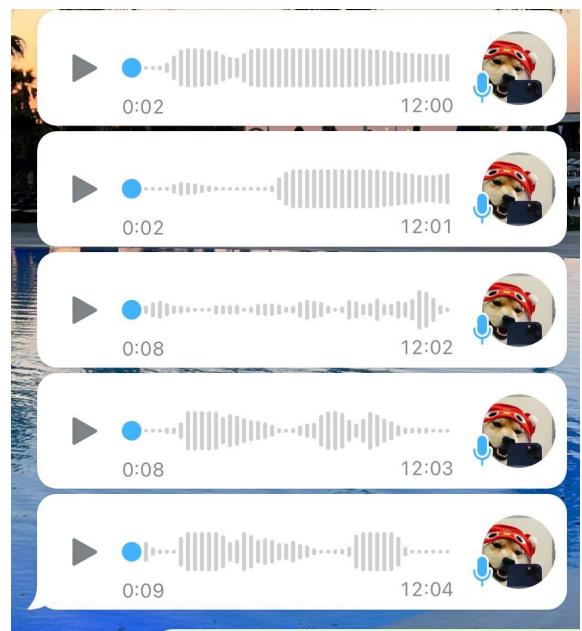
רואים כי ישן חבילות מפרוטוקלים שונים כמו שרהינו מקודם, ב��וף כאן אנו גם חווים בעיה בראש DUP ACK שעליה למדנו בקורס(באינטרנט עם השרת של גול. 34.107.221.82).

גרף זיהוי האירועים:



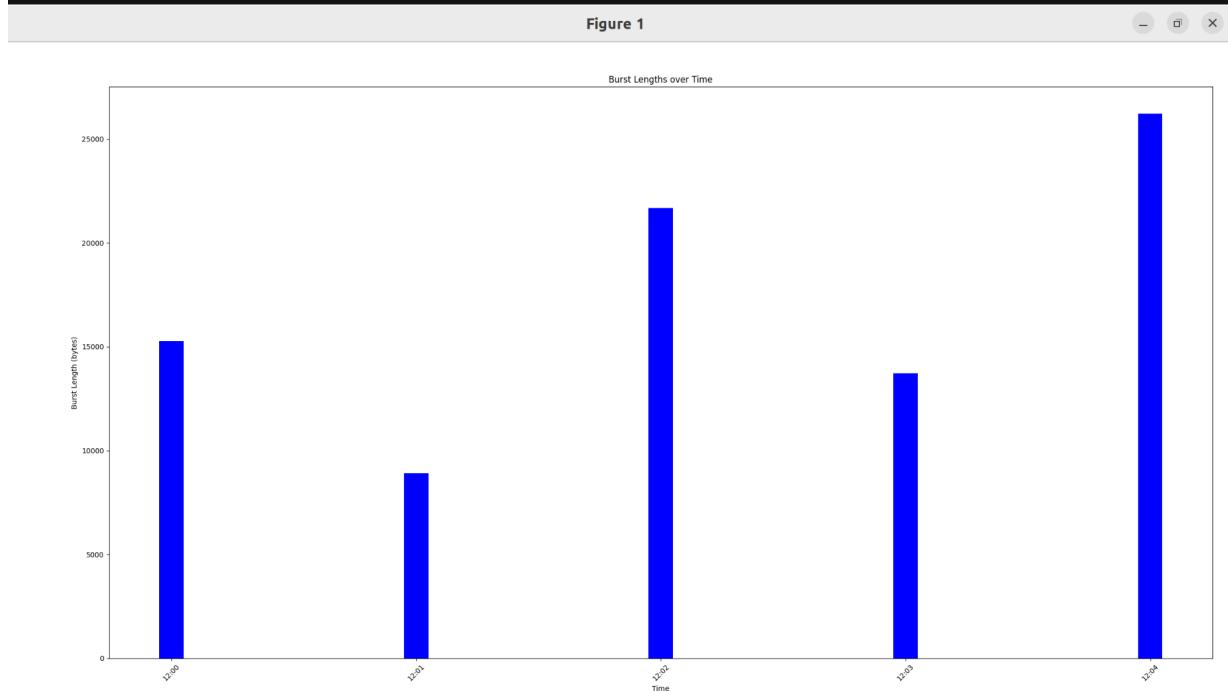
ע"י הצלחנו לzechות את הגודלים וליצא את האירועים המתאים של שליחת סרטון בכל דקה, בנוסף כל אירוע מתאים לאודל סרטונים וב9:11 שלחנו סרטון ארוך יותר וכן גם רואים זאת מהגרף. בנוסך הייתה לנו הפסקה קטנה בין שליחת ההודעות בהתחלה ורואים זאת בדקה 37 שיש שליחה מינימלית של חבילות.

קבוצה מס' 3: בקבוצה זו אנו קיבלו הודעות קוליות בכל דקה ממשתף בקבוצה



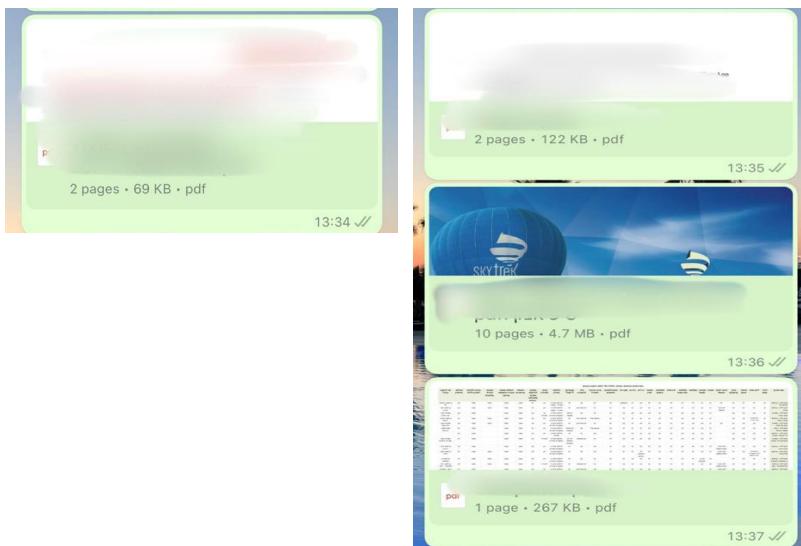
גרף זיהוי אירועים:

לפי $th=0.1$ אנו רואים כי אפילו שיש הודעות שזמן ההקלטה זהה אך יש ערך שונה בגודלי האירועים ולכן אנו יכולים להניח כי הגודל לא רק תלוי בזמן אלא גם "ברצף הקולי" המשפיע על גודל החבילה.



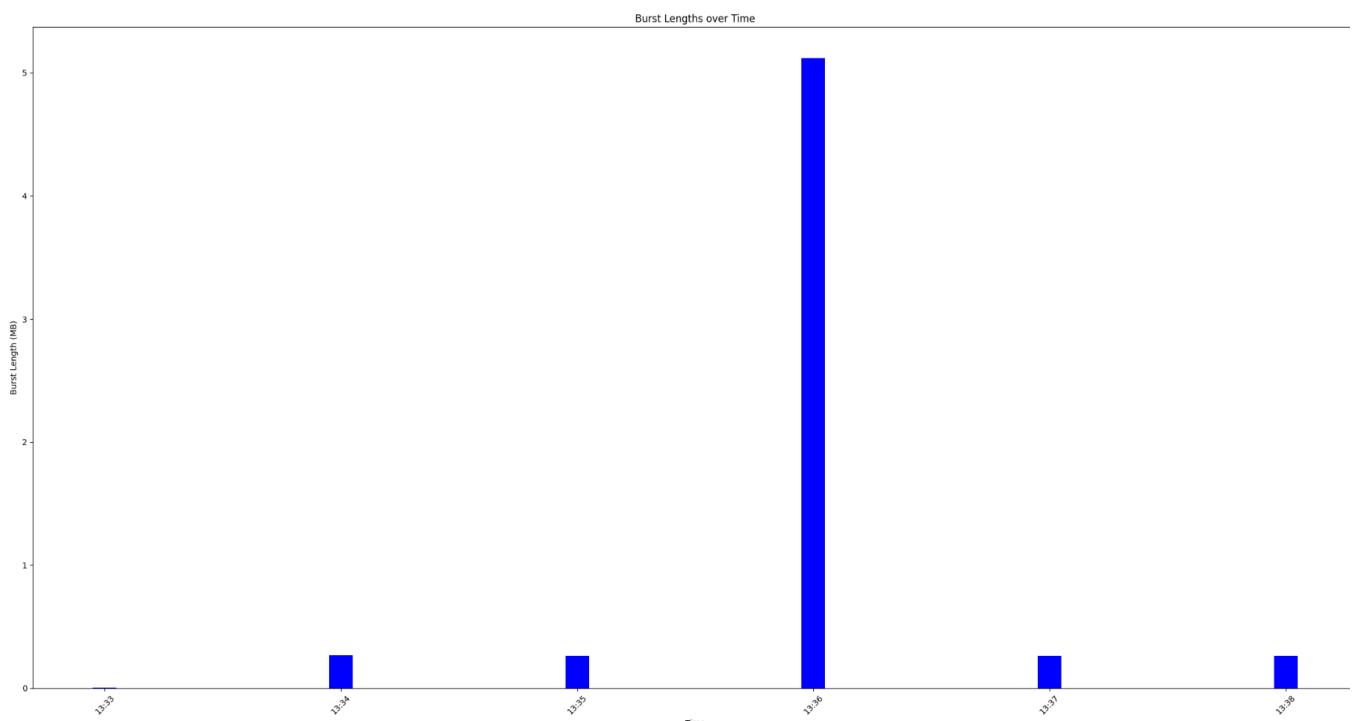
קובוצה מס' 4:

בקבוצה זו שלחנו קבצי PDF שונים בגודלים שונים בכל דקה.



גרף זיהוי אירוחים:

מכיון שישנו קובץ של 5MB ייצרנו תוכנית נוספת שתטראה לנו את המידע ב5MB. רואים بصورة מאוד מובהקת



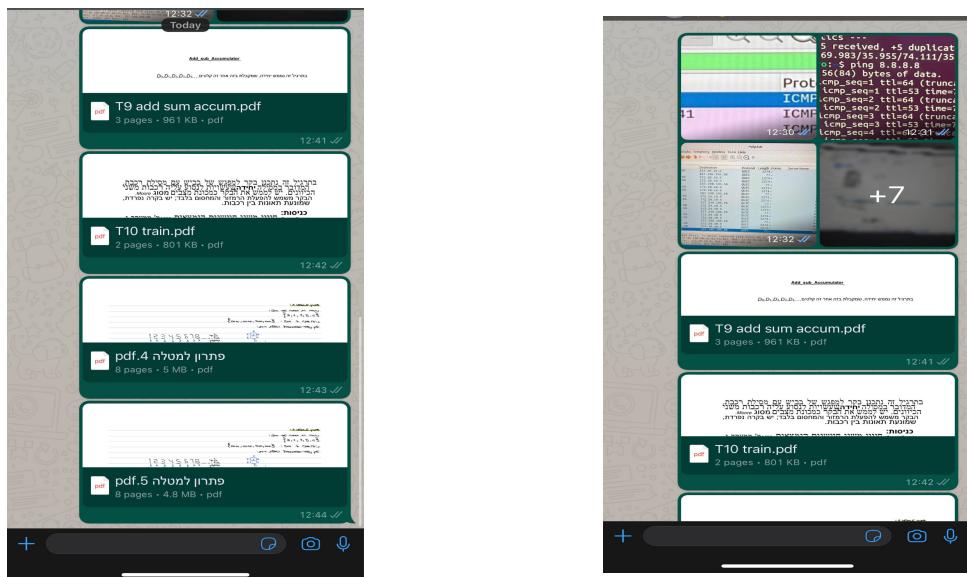
את השליחה של הקובץ הגדול ביותר ביחס לשאר ההודעות. בנוסף ב13:38 אנו רואים כי מקבלת עוד הודעה שאינה קשורה לקובוצה (קובוצה אחרת) ואנו רואים כיצד הדבר מפיע לנו לנתח את המידע מהקובוצה שאנו חוקרים.

קובץ מס' 5

קובץ שבה המנהל (אנחנו) מעבירים תמונות וקבצים. תוך כדי שבוחן הפתוח הגימיל שלו מקבל מיילים.
בקבוצה השיחה נראה כך :

(2)

(1)

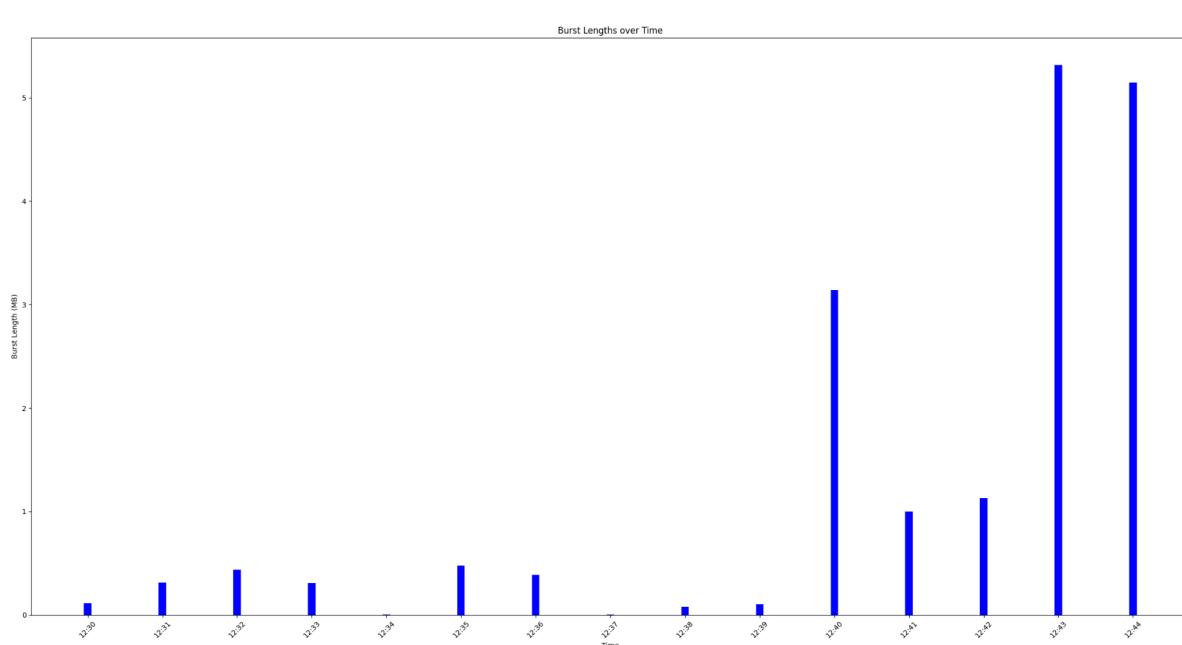


בקבוצה זאת התבצע רשימת תמונות וקבצים באופן הבא :

נשלחו תמונות בהפרשים של דקה אחד אחרי השני, ולאחר 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.vr שכאן אנו מודמים סיטואציה שבה התוקף מאמין לקובוצה מסוימת אך למתקף יש רושע לא רק מהאפקטiva אלא גם מהרשota.

גרף אירועים(MB):

הגרף מראה לנו את שילוח התמונות כל דקה ולמעשה ניתן לראות כי ממוצע התוקף יכול להסיק כי ההודעות שנשלחות כל דקה ממוצען הן פחות או יותר מבינת משקל אותו דבר , ולאחר מכן יש קפיצה גדולה בגרף ב 12:40 נס' שהtokf יכול להבין שמדובר בקפיצה חשודה , לעומת זאת עליות משקל יותר גדול ולהבין שמדובר בסוג קובץ אחר ולא תמונה נס' למשה tokf מקבל המשחה לאיזה סוג הודעות ישן בקובוצה ובכך להבין מה ה"אופי שלה", האם זאת נס' למשה tokf מקבל המשחה לאיזה סוג הודעות ישן בקובוצה ובכך להבין מה ה"אופי שלה", האם זאת הקבוצה הרציה יוכל להסיק גם שבגלל הגודל החיריג מדובר בהפרעה.($th=10$)

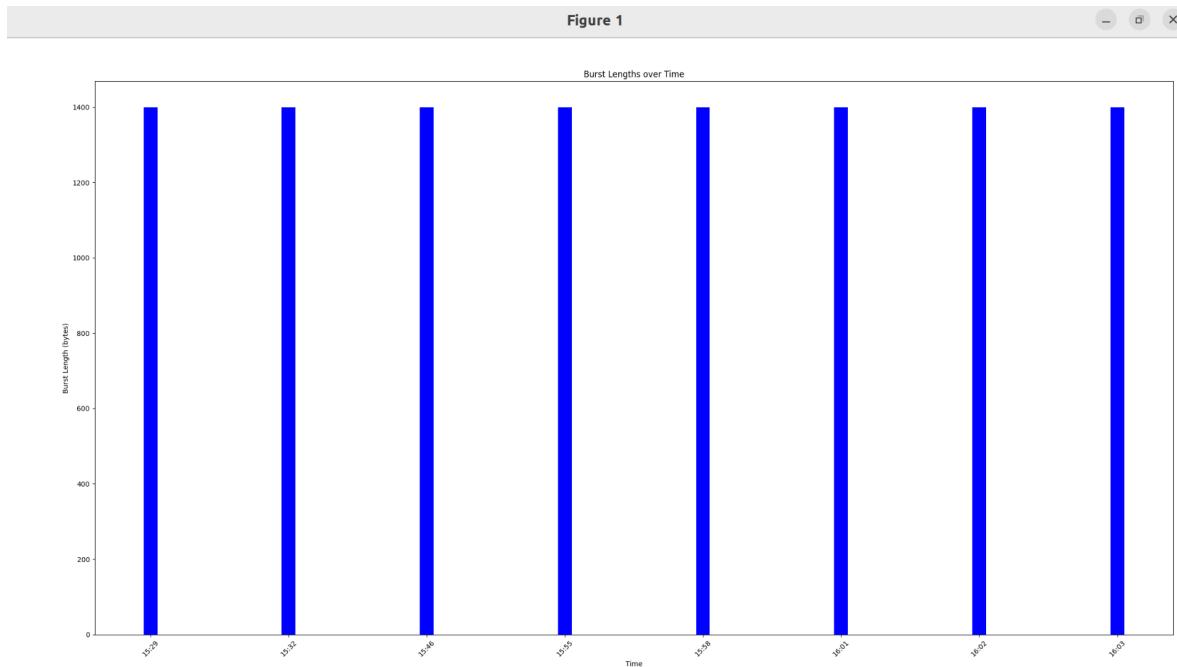


קבוצה מס' 6:

תיעדנו קבוצה של קורס מהתוואר לאחר מבחן בקורס זה הוא זמן מאוד "תועט".



גרף האירועים:



כאן רואים שוב שיש סוג של "קצב" בהודעות וכל דקה שולחים כמה אנשים הודעה, עדין האירועים קטנים מאוד ולכן ניתן להבין כי מדובר בהודעות טקסט, בקבוצת הודעות עם הרבה משתתפים בשיחה צפופה. [th 0.000001](#)

גרף PDF:

מה אנו רואים בעצם?

עמודות כחולות: מראות באיזו תדירות מתרחשים עיכובים שונים בין הודעות. העמודות הקצרות מייעדות להודעות מהירות, והעמודות הגבוהות מייעדות לעיכובים מעט ארוכים יותר. כל עמודה מייצגת טווח של זמן עיכוב.

עיקונה אדום: הקן האדום המעווק הוא קן מיוחד שעוזר לנו להבין את דפוס העיכובים הללו. זה מראה לנו כמה סבירים זמני עיכוב מסוימים.

גובה העמודות והעיקונה: ככל שהפסים הכהולים או העיקול האדום גבוהים יותר בנקודה מסוימת, כך מתרחשים לעיתים קרובות יותר עיכובים מסווג זה. לכן, אם סרגל גובה מאד או שהעיקונה גבוהה בהתחלה, זה אומר שההודעות מהירות הן באמת נפוצות.

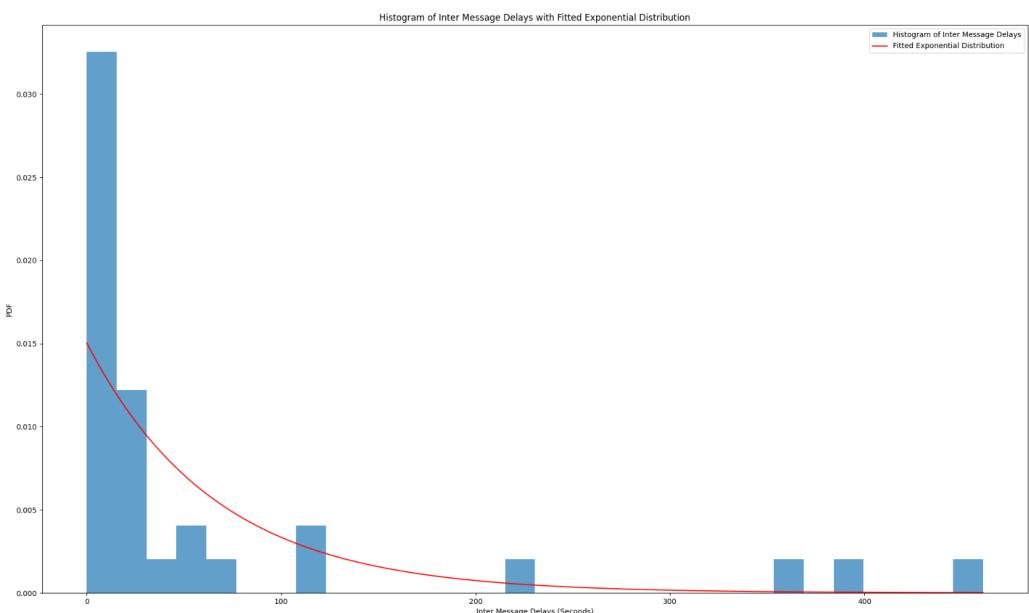
עיקול העיקונה כלפי מטה: משמעות הדבר היא שעיכובים מהירים של הודעות נפוצים מאד, אך ככל הזמן ההשניה מתארך, הוא הופך פחות נפוץ. זה כמו לומר, "ברוב המקרים, הודעות נשלחות במהירות, אבל עיכובים ארוכים יותר קוררים בתדריות נמוכה יותר".

הגרף מראה לנו את צפיפות הה הודעות, בהתחלה רואים את העמודות הכהולות גבוהות ובעות ולאחר מכן מכך הן נפחחות, הדבר מראה על כך שבכל דקה הדיל"י הארוך בין ההודעות פחות נפוץ, מה שאומר לנו שהקבוצה פעליה.

כאן אנו רואים איך זמן השהייה יורד כאשר זמן השיחה נמשך (יותר אנשים מצטרפים לשיחה וניהם חלק ממנה לאחר הייציאה מהבינה).

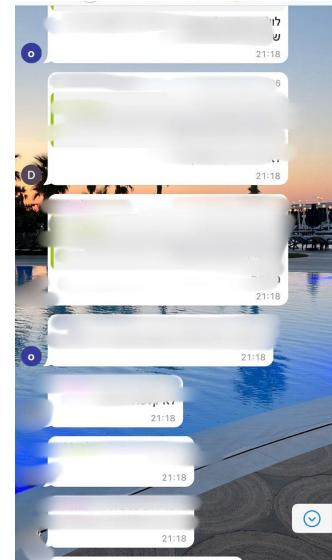
חשוב לציין כי גרף ה-[ה-הפק נוצר ע"י ליקחת היציאת מאפליקציה](#), ולכן לא העלינו את הקבצים הללו על מנת לשמור על פרטיותם של המשתתפים.

Figure 1

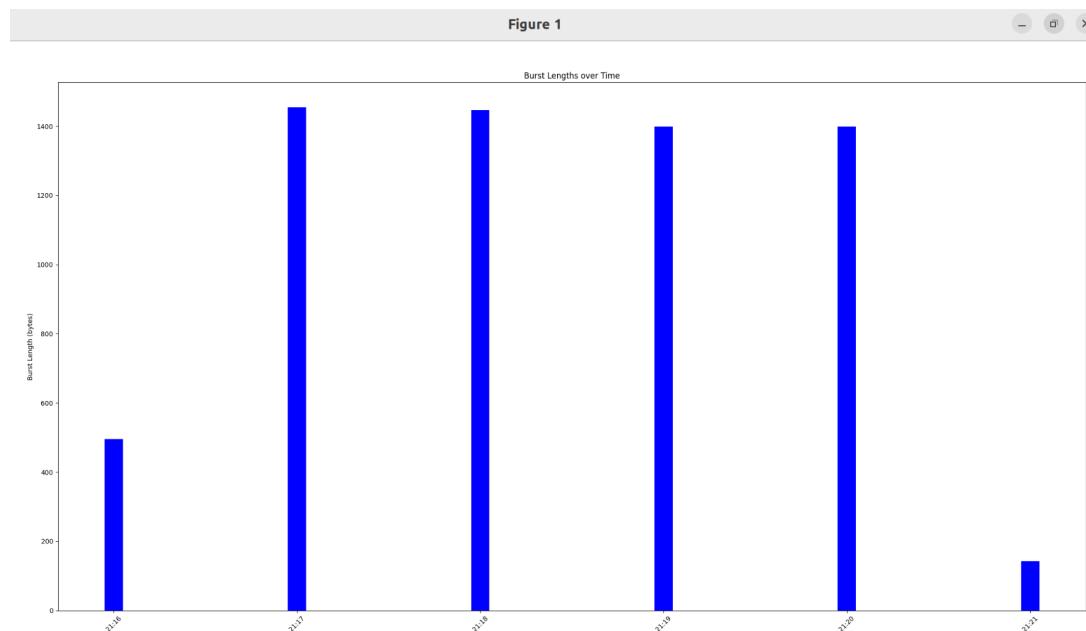


קבוצה מס' 7:

קבוצה נוספת (אחרת) אשר הקלטנו לאחר מבחן וניתן להבחן באופן רועש יחסית של הודעות שנשלחו יחד בדקה!



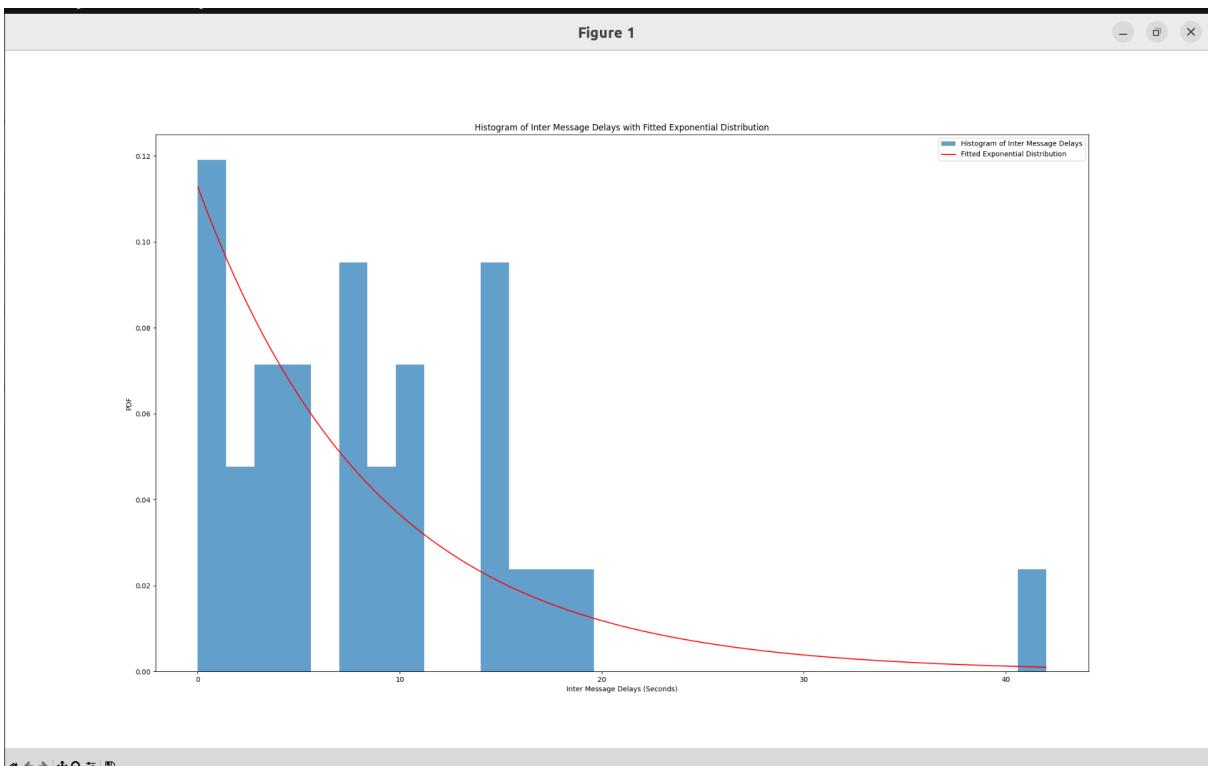
גרף אירועים:



ראאים כי בכל דקה יש אירוע אק האקט יחסית, בהתייחס לכך שהודעת טקסט גודלה בין 200 ל-100 בתים ניתן לראות כי נשלח כאן רצף הודעות גדול יחסית בכל דקה (הודעות טקסט).

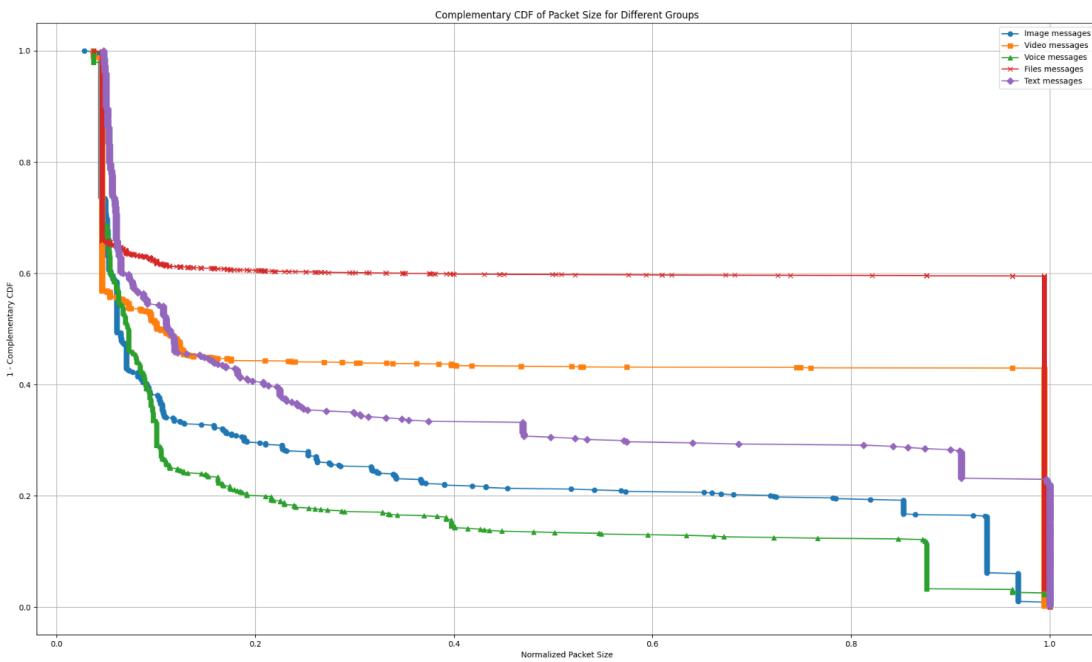
0.0000001th

גרף PDF



ניתן לראות כי זמן השהייה הקצר יותר מגיע בתדרות גבוהה יותר וכן הקבוצה מאוד "רועשת" כמו שnitן לשים לב והרוח בזמן השקט .

CDFF



ע"י קבוצות מספר 1,2,3,4,6 ייצרנו את גרפּ ה-CCDF. הגרפּ מציג את פונקציית ההפּוצה הצפּיפה הפּוכה (Complementary Cumulative Distribution Function - CCDF) של גודלי החבילות מנורמלות לארבעה קבוצות שונות. כל קבוצה מייצגת סוג אחר של העברת הودעה (טקסט, וידאו, תמונות וכו'). ציר ה-X מייצג את גודל החבילה המונרמל, הנע בין 0 ל-1, כאשר 1 מתאים לאודל המרבי שנרשם בקבוצה. ציר ה-Y מייצג את הפונקציה הפּוכה של פונקציית ההפּוצה הצפּיפה הפּוכה (1 - CCDF), שבעצם מראה את ההסתברות שגודל החבילה יהיה גדול יותר מסף מסוים או במילויים פשוטות, עבור כל קבוצה כל חבילה מנורמלת לפי גודלה ביחס לכל החבילות בקבוצה שלה ואנו יכולים לראות מהגרפּ מההסתברות שנראית אותה בקבוצה.

מסקנות וסיכום

בניתוח שלנו, חקרו את תכונות הרשות מהודעות WhatsApp בקבוצות שונות. ערכנו תוצאות מעניינות והתמודדנו עם כמה אתגרים בפענוח הנתונים.

קובוצה מס' 1: תמונות בלבד

שmeno לב שליחת תמונות בוואטסאפ הובילה לאי-התאמות משמעותיות בגודל בין נתוני לכידת החבילות המקוריות לבין גודלי הקבצים המקוריים באפליקציה. דרישת התמונה והמרת הפורמט של WhatsApp הפיכתו במידה ניכרת את גודלי התמונות במהלך השימוש, מה שמקשה להסיק במדויק את גודל התמונה המקורי מהמידע המוצג. הדבר יכול להקשות על התקוף להבוי מה היא הקבוצה המדוברת ולזהות את השicityות אליה.

קובוצה מס' 2: סרטונים

בקבוצה זו, היה לנו דפוס ברור ואמין של התנהגות הודעות. והתוקף יכול לזהות באופן ברור שאכן מדובר בקבוצת סרטונים בגודלים שנשלחו הסרטונים אם הוא הינו יכול מהקובוצה.

קובוצה מס' 3: הקלטות

בקבוצה זו גם ניתן להבין שמדובר בהקלטות, אם האויב ידע את אורך וזמןיהם.

קובוצה מס' 4: קבצים

בקבוצה זו ראיינו כי גודל הקבצים נראה באופן די ברור על הגרפ' וכן בהתחשב בגודלי הקבצים ובזמנים ניתן לנתח את הקבוצה.

קובוצה מס' 5 קבצים ותמונה: קבוצה זו הייתה עם רשות אחר וראיינו כי ישנה הפרעה אך עדין ניתן לראות כי הדבר לא פוגע באופן משמעותי בגודלי האירועים.

קובוצה מס' 6, 7 הودעות טקסט: בקבוצות טקסט לא הצליחנו ליזאץ מכוון שהודעות כוללות קורים הרבה מקרים במצבה הודיעות נשלחות אפילו בדקה אחת. אך הגרפ' לא מדויק ועדיין ניתן להבין כי מדובר באופןו של קבוצה עם גדים ייחודיים, בהשוואה לשאר סוגיה הודיעות הודיעות טקסט הן מאוד קטנות וכן בזמן של שולחים הודיעות האויב יכול לזהות את אופי הקבוצה ע"י אירועים קטנים אך רציפים.

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

באשר לשאלת האם ניתן להבין את הקבוצות שמשתתף בהם המותקף כאשר הוא מתקשר עם כמה קבוצות במקביל ניתן לראות כי הדבר יחסית מאוד מורכב. גם אם התוקף משתמש בשתי קבוצות שאנו משתתפים בהם. ניהיה "ערובוב" מאוד גדול של הודיעות, ולא ניתן לנתח לאיזה קבוצה מסוימת כל חיבור. כאשר התוקף מאמין לקבוצה יחידה קל ייחסת ליזאץ את האירועים ולנתח אותם לפי האופי של הודיעות וזמן השליחה.

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 "ירוט וניתוח תעבורת b-Web WhatsApp" נובעת מօפי פרוטוקולי התקשרות המשתמשים באפליקציה ומבנה הנתונים המועברים. Web WhatsApp משתמש בפרוטוקולים בפורמטים שונים של נתונים כדי להקל על התקשרות בין הלקות לבין השירות.

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

יותר מזה שמן לב כי גם על התקוף לחשב וליצור גרפ כמה שייתר אמיתי עם threshold מתאים על מנת לקבל את אופי האירועים בקבוצה. בנוסף לעליון לנוקות את החבילות שמספריעות לו. לכן אנו חושבים שבסופו של דבר נדרש זמן והשקעה והתקוף צריך להיות בקבוצה על מנת לא "לכלת לאיבוד" בכל המידע שעובר ברשות המותקף.

כאשר התקוף מנסה לתאם נוכחות של אדם במספר קבוצות או ערוצים, הם מתמודדים עם מספר אטגרים בשל האופי של אפליקציית הוואטסאפ שחקרנו. מדוע?

1. הצפנה מקצת לkeys: כאשר ההודעות מוצפנות, ניתן לפענה אותן רק במקשייר של מקבל ההודעה. המשמעות היא שגם אם היריב יכול לירות את ההודעות המוצפנות, הוא לא יוכל לקרוא את התוכן ללא מפתחות ההצפנה.

הצפנה זו מקשה על התקוף לגשת יישורות לתוך ההודעות, ולהבין איזו הודעות הקשורות לאיזה קבוצה. **2. ערבות הודעות:** הודעות מוצפנות או שישות שונות מתרבבות לעיתים קרובות בתעבורה הרשת. ערובה זה מאייגר את התקוף להבדיל בין מסרים השייכים לקבוצות שונות. הם יכולים לצפות בזרימת התעבורה ובגודלה ההודעות, אך לא גישה למפתחות ההצפנה, לא ניתן לקבוע את התוכן המדויק או אילו הודעות שייכות לאיזו קבוצה.

3. גודל החבילות וזמן: בעוד התוכן של הודעות מוצפן, מטא נתונים מסוימים כגון תזמון הודעות, גודל ותדירות התקשרות עדין גלויים לתקוף. עם זאת, כפי שראינו בעובדה, מטא נתונים בלבד עשויים שלא להיות מספיקים כדי לאמת באופן אמין בין נוכחות של אדם במספר קבוצות. לדוגמה, לקבוצות רבות עשויות להיות דפואים תנואה דומים, מה שהופך לאתגר.

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

המשמעות היא שגם אם התקוף צופה באינטראקציות מרובות ממключиים שונים, הוא לא יוכל לחבר אותן באופן סופי לאותו אדם. יתר על כן, משתמשים עשויים לעבור בין מכשיירים או לשנות את זהותם, מה שמסבך עוד יותר את התהילה.

ביבליוגרפיה: מפורטת בגרסה האנגלית.