# Review 15: Bringing a GAN to a Knife-Fight: Adapting Malware Communication to Avoid Detection

**Paper: https://arxiv.org/abs/2109.04344v3**

פינת הסוקר:

המלצת קריאה מעדן ומייק: מומלץ לאנשים העוסקים בתחום ה-Cybersecurity או לאנשים שאוהבים GAN

בהירות כתיבה:

ידע מוקדם: GAN

יישומים פרקטיים אפשריים: זיהוי Malware

פרטי מאמר:

פורסם בתאריך: 6.8.2018, בארקיב.

הוצג בכנס: (IEEE Security and Privacy Workshops (SPW

תחומי מאמר:

זיהוי אנומליות

GAN

נוזקות

כלים מתמטיים, מושגים וסימונים:

GAN

מבוא:

כלי אבטחה שונים הנמצאים היום בשוק מצליחים להתמודד בהצלחה רבה עם נוזקות (Malwares) ידועים. כלים אלו כוללים חוקים סטטים שנכתבו בצורה ידנית על ידי חוקרי אבטחה, או לחילופין מערכות גילוי אנומליות שמבוססות על אלגוריתמים של למידת מכונה. אך, הצלחה זו בהווה אינה מבטיחה הצלחה בעתיד שכן כפי שכלי הגנה שונים יודעים להשתמש בלמידת מכונה בשביל גילוי נוזקות, כך גם התוקפים יכולים להשתמש בלמידת מכונה כדי ליצור נוזקות מתוחכמות יותר שיוכלו להתחמק מכלל כלי הגנה הקיימים בשוק. הנוזקה תוכל לשנות את התנהגותה בהתאם לסביבה ולתנאים אותן היא רואה ובכך לחמוק מכלי הגנה שונים. נוזקה שכזו יכולה להוות בעיה גדולה עבור תעשיית הסייבר.

כיום, נוזקות שמשנות את התנהגותן בצורה דינמית מסתמכות בעיקר על שיטות סטטיות (למשל אלגוריתם DGA). אך שיטות אלו אינן מספיקות כדי להתחמק מכלי ההגנה הטובים ביותר בשוק. כדי להתחמק מכלי הגנה אלו הנוזקה צריכה להיות מסוגלת לפתח לוגיקה מורכבת וללמוד את הסביבה שלה. כאן, נכנסים כותבי המאמר שמציעים שימוש ב- (Generative Adversarial Networks (GAN שישמש את הנוזקה כיצד לפעול תחת תנאים שונים. הרעיון מאחורי השימוש ב-GAN הוא ש-GAN יוכל לעזור ללמוד את תעבורת הרשת בה היא נמצאת ובכך לחכות אותה בצורה טובה יותר. בכך, היא תצליח להתחמק מכלל כלי ההגנה הפרוסים בארגון ולא להתגלות.

הסבר על השיטה:

כותבי המאמר מציינים שתעבורת הרשת אותה ה-GAN ינסה לחקות הינה של פייסבוק ובפרט צ'ט של אפליקציית פייסבוק. זאת מכיוון שתעבורת רשת מסוג זה היא נפוצה ולרוב מאופשרת על ידי כלים כמו חומת אש. לאחר שה-GAN ילמד כיצד תעבורת פייסבוק נראת, הוא יוכל להנחות את הנוזקה כיצד לפעול כדי להתחמק מחסימה. המטרה של הנוזקה היא לתקשר עם שרת התוקף. הסביבה (המגן) תוכל לחסום את ערוץ התקשורת של הנוזקה ועליה לדעת להתמודד עם מצב זה. GAN נועד בדיוק בשביל לענות על אותו מצב.

ה-GAN יבצע אימון מקדים(pretraining) לפי הכנסתו לארגון, כאשר הפלט אותו ינסה ליצור הינה הקונפיגורציה ,בדמות שלושה פרמטרים, של הנוזקה שמכתיבה לו כיצד להתנהג. הפלט כולל שלושה פיצ'רים: כמות הבייטים המקסימלית שתעבור ב-flow, הזמן של flow וכמה זמן לחכות בין ה-flow הנוכחי לבא אחריו (כאשר הכוונה ל-network flow).

ה-GAN עובד בשני חלקים:

Generator: מקבל רעש כלשהוא ומנסה לבנות ממנו את הקונפיגורציה.

Discriminator: מקבל קונפיגורציה זו וקונפיגורציה אמיתית של פייסבוק וצריך להחליט האם הקונפיגורציה שנוצרה על ידי G הינה אמיתית או מזויפת.

לאחר ההחלטה של ה-D מחושב ה-Loss ותהליך של Backprop עדכן את המשקלים של D ו-G. כותבי המאמר מציינים שב-GAN קשה לקבוע מתי ה-GAN מתכנס שכן פונקציית ה-Loss לא בהכרח מצביעה על איכות הפלט שה-GAN מייצר. כדי להתמודד עם בעיה זאת הכותבים השתמשו בכלי הגנה חיצוני מסוג Intrusion Protection System ,או בקיצור IPS, כדי שיבחן את הפלט של ה-GAN.

לאחר שה-GAN עבר אימון מקדים, מחברים אותו לנוזקה והוא מתחיל לייצר את הקונפיגורציה שנשלחת לנוזקה. על ידי הקונפיגרציה החדשה, הנוזקה תוכל לשנות את מאפייני ערוץ התקשורת עם שרת התוקף. אם הנוזקה מגלה שערוץ התקשורת שלה נחסם היא מאותת ל-GAN שבתגובה מבצע אימון נוסף על הדאטה שהנוזקה שלחה לו שנחסם ומייצר לה קונפיגורציה חדשה.

הסביבה:

כדי להראות שהגישה אכן עובדת בסביבה אמיתית, כותבי המאמר ציינו שתי מגבלות:

הנוזקה בה משתמשים הינה נוזקה אמיתית

החסימה של תעבורת הרשת צריכה להיות אמיתית גם כן

הנוזקה בה השתמשו הינה FLU. הקוד שלה שונה בכך שאת הפרמטרים שקובעים כיצד לייצר את ערוץ התקשורת עם השרת היא תקבל מה-GAN. השרת של התוקף הוקם על גבי ענן AWS. נזכיר שוב שמטרת הניסוי היא לראות שהנוזקה מצליחה לתקשר עם השרת עם כמה שפחות חסימות של המגן.

המגן במקרה זה הינו IPS מסוג Stratosphere. ה-IPS הוקם בחיבור של הארגון עם האינטרנט ובכך ה-IPS יוכל לבדוק את כל התעבורה שיוצאת מהארגון ולחסום תעבורה חשודה. במקרה זה עליו לנסות לחסום את תעבורת הנוזקה אל עבר השרת שלה.

ה-GAN בו השתמשו הינו רשת מסוג LSTM גם ל-Generator וגם ל-Discriminator. בנוסף מומש רכיב WEB שיהיה אחראי על התקשורת בין ה-GAN לבין הנוזקה ובהעברת המידע של ה-GAN לנוזקה.

איור 1: סביבת הניסוי. ה-GAN ממומש ומתאמן באופן נפרד ומתקשר עם הנוזקה דרך שירות ה-WEB. הנוזקה מקבלת את הקונפיגורציה מה-GAN ומשנה את ערוץ התקשורת שלה בהתאם בזמן אמת. על ערוץ התקשורת עם השרת להיות פעיל בכל זמן נתון. ה-IPS יחסום כל תעבורת רשת שלא נראת כמו צ'ט של פייסבוק.

הניסוי:

הדאטה שנאסף בשביל אימון ה-GAN הינו דאטה אמיתי של שני משתמשים המדברים ביניהם דרך אפליקציית הצ'ט של פייסבוק. השיחה כללה תמונות, טקסט ומסמכים כדי לדמות שיחה מגוונות ואמיתית. תעבורת הרשת של השיחה הוקלטה ונשמרה בתוך הדאטה שהוזן ל-GAN.

ה-IPS קונפג בחלונות זמן שונים, כאשר גודל החלון מצביע על כמות המידע אותו בוחן ה-IPS. על כל חלון שהוא רואה הוא מבצע פרדיקציה - כן או לא צ'ט של פייסבוק. שימוש בחלונות שונים נועד לראות כיצד ה-GAN מתמודד בזמן אמת עם שינויים בסביבה.

תוצאות:

בטבלה הבאה ניתן לראות את טיב האימון של ה-GAN:

עמודת ה-Mean ועמודת ה-Std מייצגות את הממוצע וסטיית התקן של מספר הפלטים (קונפיגורציות) של ה-GAN שלא נחסמו על ידי ה-IPS. ניתן לראות שכל שה-GAN מתאמן יותר, כך מספר הקונפיגורציות שלא נחסמות בממוצע עולה. זה מצביע על כך שה-GAN מצליח ללמוד כיצד להתמחק מה-IPS.

בגרף הבא, כותבי המאמר מראים את כמות ה-Flows שהנוזקה יצרה בכל חלון זמן של חמש דקות בהתאם לההוראות שקיבלה מה-GAN:

ניתן לראות שככל שה-GAN מתאמן יותר אחוז המידע שנחסמת על ידי ה-IPS יורדת לאפס.

לבסוף כותבי המאמר מציינים שתוך כדי הניסוי, בזמן הריצה, הדאטה שה-IPS תייג, נשלח לאימון נוסף של ה-GAN כדי לנסות ולשפר אותו. הם מוסיפים ואומרים ששיטה זאתי הראתה שיפור בביצועים אך לא במידה שהיא מובהקת מספיק. סה"כ השיטה הראתה תוצאות טובות. 63% מהתקשורת לא נחסמה, דבר שהופך את השיטה לפרקטית לדברי כותבי המאמר.

סיכום:

המאמר מציג את היכולת לחבר GAN לפיתוח של נוזקה אמיתית במטרה להפוך את הנוזקה למתוחכמת ומקלמת אף יותר. בעזרת ה-GAN הנוזקה הצליחה להוריד ב-63% את סיכוי גילויה על ידי כלי הגנה אמיתי מסוג IPS. ה-GAN אומן כדי לנסות ולחקות תעבורת רשת של צ'ט פייסבוק והמשיך להתאמן תוך כדי זמן הריצה לאחר שקיבל מידע מהנוזקה אילו מפלטיו התגלו על ידי ה-IPS. חשוב לציין שה-GAN התאמן על כמות דאטה קטנה בהתחלה ועדיין הצליח להביא תוצאות טובות מה שהופך את השיטה הזו לפרקטית במציאות.

שיתוף פעולה: הסקירה נכתבה יחד עם עדן יבין.