Analiza informațiilor din fișiere executabile .NET

Ionuț-Alexandru Avram

Rezumat

Malware-ul sau software-ul malițios reprezintă acele programe care au ca scop infiltrarea într-un sistem și producerea de pagube, scopul declarat al programului fiind diferit de cel real. În esență, se consideră ca fiind malware orice software rău intenționat, care nu urmărește interesele utilizatorului.

Primele programe malițioase scrise folosind .NET au apărut odată cu apariția framework-ului în 2002, dar până în 2009, malware-ul care are ca țintă această platformă nu a reprezentat o parte prea mare din total. Însă, din 2009 și până în 2015 s-a înregistrat o creștere de 1600 % în numărul de fișiere malițioase unice create folosind tehnologii .NET. Portarea pe diferite platforme, existența a numeroase packere care pot fi folsite la obfuscarea executabilelor, a unui IDE gratuit pentru dezvoltare rapidă, API-ul pus la dispoziție, dar mai ales prezența framework-ului pe toate calculatorele ce folosesc Windows ca și sistem de operare sunt principalele motive ale răspândirii vertiginoase.

Lucrarea de față își propune să prezinte un program capabil să ajute în analiza oricărui program sau set de programe scrise într-un limbaj din universul .NET. Aplicația poate fi folosită atât la analiza individuală a fișierelor, dar și pentru grupuri de fișiere, permițând cluster-izarea sau stabilirea apartenenței unui fișier la un cluster deja existent.

Analiza individuală are rolul de a evidenția caracteristicile unui executabil, cum ar fi informații de la compilare, structurile de cod definite de programator, string-urile folosite în mod direct în cod sau API-ul folosit. Toate aceste informații pot fi folosite de un programator care dorește să vadă ce ajunge în executabil în urma compilării sau de un analist malware pentru a dobândi o înțelegere preliminară a posibilului comportament al unui fișier întâlnit.

Partea de cluster-izare permite clasificarea seturilor de fișiere în familii pe baza unor criterii de similaritate implementate. Această parte are numeroase aplicații practice, cum ar fi cluster-izarea fișierelor malițioase noi apărute, stabilirea apartenenței unui fișier malițios la o familie deja cunoscută, determinarea diferitelor versiuni ale unui program legitim sau detecția unor teme copiate.

Operația de cluster-izare ușurează mult munca unui analist malware permițând o clasificare automată a diverselor sample-uri. În această manieră analistul poate analiza câte un reprezentat pentru fiecare cluster generat sau poate elimina de la analiză acele fișiere care aparțin de familii binecunoscute, putându-și concentra atenția pe familiile noi apărute. Astfel, datorită uneltelor de analiză automată, sarcinile de lucru pot fi ușurate, lucrându-se mai eficient și ajungându-se, în final, la rezultate mai bune. Lucrarea de față subliniază importanța aplicației în domeniul analizei malware, întrucât prin cluster-izare și stabilirea apartenței la o familie malware, tehnicile de analiză și detecție pot fi îmbunătățite substanțial.

Pentru a extrage informațiile necesare, aplicația prelucrează metadata prezentă în fișierele executabile rezultate folosind instrumente aparținând platformei .NET, fișierele executabile fiind numite și fișiere MSIL.

Securitatea aplicației reprezintă un aspect crucial, din prisma gradului ridicat de aplicabilitate în domeniul analizei fișierelor malițioase. Din acest motiv, operația de

despachetare, care ar putea avea consecințe negative asupra sistemului, se realizeaza într-un mediu izolat, un sandbox. Procedura responsabilă de extragerea restului de informații necesare a fost implementată în Python, fișierele date spre analiză fiind tratate ca și fișiere binare din care se citesc octeți, fără a realiza vreo emulare sau încărcare a fișierului în memorie. Aceste măsuri implementate, ajută la creșterea gradului de siguranță al aplicației și conferirea unui grad de încredere pentru utilizatori.

Pentru implementarea operației de cluster-izare, sunt definite multiple opțiuni implementate prin algoritmi proprii. Stabilirea cluster-elor se poate face după tipul fișierului, a unui hash generat folosind metadatele prezente sau prin determinarea similarității dintre API-ul, șirurile de caractere folosite sau structurile de cod definite de programator. Stabilirea similarităților dintre API-uri și șirurile de caractere folosite se realizeză folosind ca metrică distanța Jaccard, iar similaritățile dintre structurile de cod definite de programator se calculează folosind un algoritm propriu ce încearcă găsirea unor echivalențe.

Lucrarea detaliază toate operațiile descrise anterior, prezentând pe larg procedura de extragere a informațiilor și descrierea algoritmilor implementați. De asemenea, sunt prezentate problemele întâlnite în dezvoltarea aplicației, dar și soluțiile găsite. În finalul lucrării se realizează un studiu de caz pe partea de cluster-izare a fișierelor pe o campanie malware binecunoscută și anume „Operation Cleaver”, rezultatele fiind comparate cu cele obținute prin mijloace de cluster-izare deja existente și documentate. De asemenea, se prezintă posibile utilizări ale aplicației, direcții de dezvoltare și îmbunătățiri care ar putea fi aduse.