Vizuelizacija upozorenja o otkrivanju mrežnih napada u formi grafa

Stefan Aleksić

Sadržaj – Mrežni napadi mogu imati ogroman negativni uticaj na infrastrukturu jednog informacionog sistema. Standardni pristup za detekciju napada jeste upravo ugradnja sistema za detekciju neautorizovanih pristupa mreži i generisanja upozorenja (*eng. Intrusion Detection System, IDS*). Međutim, ovi sistemi mogu da generišu ogroman broj poruka upozorenja (*eng. alert*) koje centar za bezbednost nije u mogućnosti da analizira, čak i uz pomoć njihove korelacije. Ovaj rad nastoji da kroz vizuelizaciju poruka upozorenja u formi grafa, dobijenog klasterizacijom poruka na osnovu sličnosti korisnički odabranih atributa, analizatoru pomogne pri procesu detektovanja potencijalnih napada na mrežni sistem. Korisnik pre pokretanja programa bi trebalo da navede putanje (URL ili File) do xml log fajlova generisanih od strane IDS-a, kao i mogućnost konfiguracije samog udela koji svaki od atributa poruke upozorenja ima pri računanju njihove sličnosti, čime se konfiguriše sama klasterizacija. Program, uz pomoć *Networkx* i *Matplotlib* biblioteka za programski jezik *Python*, iscrtava formirane grafove, dinamički menjajući njihovu strukturu na osnovu parsiranih podataka. Konfigurabilni *FIFO buffer* je iskorišćen kako bi se ograničio broj prikazanih upozorenja, dok se u pozadini generišu preseci u vidu slika trenutnog stanja grafa sa uparenim csv fajlovima u kojima su naznačeni detalji svake od prikazanih poruka. U radu je pored teorijske osnove iskorišćene za klasterizaciju, opisana i sama implementacija uz priloženo uputstvo rukovanja programom.

I. Uvod

Mrežna komunikacija je postala glavni način za razmenu informacija u mnogim sferama današnjeg društva, kako u lične svrhe, tako i za prenos važnih informacija unutar i između kompanija pri njihovom poslovanju. Iz ovog razloga, kompanije ulažu u razvijanje svojih informacionih sistema (*eng. Informational Technology, IT systems*), koji za cilj imaju između ostalog i održavanje bezbednosti na mreži preko koje se posluje.

Nažalost, kako se razvijaju metode za očuvanje bezbednosti, tako se na drugoj strani razvijaju metode koje istu narušavaju. Veoma notoran, sajber kriminal, može da podrazumeva i naizgled vrlo bezazlene aktivnosti, poput nerelevantne elektronske pošte, neprikladnih reklama, kao i onih koje zvuče previše dobro da bi bile istinite i sl. Sve od nabrojanog se koristi kao mamac za mnogo ozbiljnije zakonske prekršaje, poput na primer *ransomware*-a (*eng. ransom* – otkup, ucena) odnosno malicioznog softvera ili *malware*-a koji onemogućuje žrtvi da pristupi svom računaru, fajlovima, sistemu, ili mreži, dok ne ispuni ucenu sajber kriminalca. Krađa identiteta, *spoofing*, *phishing* i mnoge druge aktivnosti su sve češće na mreži, što je zastrašujuće s obzirom da je njihova meta prosečan korisnik Interneta.

Tema ovog rada su napadi mnogo ozbiljnije branše, napadi koji se planiraju i bivaju *deploy*-ovani na čitave organizacije, pa čak i vojne infrastrukture. Najpoznatiji napad ovog tipa je *DoS – Denial of Service*, koji podrazumeva da se mreža koja pruža usluge u vidu servisa preoptereti nevalidnim zahtevima, tj. beskorisnim saobraćajem na mreži, kako jednostavno od prevelikog broja zahteva ne bi imala procesorsko vreme, ili *bandwidth* da pruži usluge stvarnom korisniku. Još napredniji, *DDoS – Distributed Denial of Service* podrazumeva da se na distribuiran način, od strane većeg broja „*zaraženih*“ hostova sistem obori kako bi napadač mogao da se infiltrira i prikuplja osetljive informacije iznutra.

Kako bi se ovako kobni ishodi sprečili, mnoge organizacije plaćaju izvrsne programere da upravo obore sistem na najkreativniji način, u svrhu identifikacije slabih tačaka sistema i upotrebljenih obrazaca, koje dalje koriste kao izvor informacija kako bi napade koji bi se potencijalno odigrali u budućnosti prvo mogli da identifikuju, odnosno zaustave pre nego što postanu ozbiljna pretnja.

Cilj ovog rada je upravo da na osnovu prepoznatih napada, za koje je dobro poznato kako izgledaju, odnosno koje obrasce primenjuju i kakva upozorenja IDS sistem za njih generiše, prvo identifikuje odnosno odvoji od ostalog saobraćaja na mreži uz pomoć klasterizacije, a onda tako klasterizovanu grupu upozorenja vizuelno prikaže analizatoru koji održava bezbednost mreže. Na ovaj način analizator ne treba da prolazi kroz detaljne informacije svakog upozorenja koje biva generisano, već samo da definiše atribute, tj. njihov udeo pri računanju sličnosti, na osnovu kojih program u pozadini formira klastere. Tek kada uoči klaster srodnih upozorenja koji potencijalno mogu da predstavljaju jednu od faza napada, analizator može da dejstvuje na odgovarajući način kako bi se napad tu i zaustavio.

U narednom poglavlju će biti opisana teorijska osnova rada, dalje sama implementacija programa, kako bih u četvrtom poglavlju prikazao dobijene rezultate i na kraju izneo sopstvene zaključke vezane za temu.

II. Teoretska osnova

A. IDS mrežna upozorenja

Poruka mrežnog upozorenja (*eng. alert*) $a \in A$ može biti generisana od strane IDS sistema pokrenutog za mrežu ili konkretan mrežni uređaj i služi kao indikacija potencijalno malicioznih aktivnosti na mreži. Svako upozorenje se može posmatrati kao vektor atributa $a = (a_1, a_2, \dots, a_n)$, ovi atributi mogu biti IP adresa i broj porta uređaja koji je poslao zahtev, ili uređaja kome je zahtev upućen, vremenska markica kada je upozorenje generisano, protokol koji se koristi, tip upozorenja itd. Svaka poruka upozorenja takođe sadrži i jedinstveni identifikator (*eng. Unique Identifier – UID*) na osnovu koga se skladišti u bazi, odnosno log fajlu.

$$a = (uid, ts, src\_ip, src\_port, dst\_ip, dst\_port, proto, type)$$

B. IDS meta upozorenje

Napad, preciznije sva upozorenja koja su generisana za istu malicioznu akciju, odnosno jedna korak u napadu i, će za ishod imati skup svih istinski pozitivnih (*eng. true positive*) upozorenja S_i, pa će onda $\hat{S} = \{S_0, S_1, \dots, S_{n-1}\}$ predstavljati skup svih koraka u vidu skupova upozorenja za jedan napad.

Korelacijom je moguće redukovati ove skupove na apstraktna upozorenja, takozvana meta upozorenja, čime bi se smanjila količina redundantnih informacija, jer će jedno meta upozorenje referencirati ceo skup IDS upozorenja za jednu akciju. U ove svrhe se koristi funkcija korelacije, koja klasterizuje skup upozorenja A u skup klastera $\hat{C}$, gde se jedan klaster $Ci \in \hat{C}$ sastoji od svih upozorenja koja predstavljaju jedan korak u napadu.

Meta upozorenje se sastoji od sopstvenog identifikatora *UID-a*, vremenske markice u vidu vremena i datuma kada je kreiran, skupa svih identifikatora upozorenja koje apstrahuje, kao i skupa svih IP adresa žrtve i napadača koje se javljaju kroz upozorenja. Na kraju, meta upozorenju je pridružena i poruka koja opisuje napad. Ovim možemo da zaključimo da je IDS meta upozorenje zapravo skup svih upozorenja koje je generisao IDS kao odgovor na istu akciju.

$$m = (uid, ts, alert\_ids, attackers, victims, message)$$

Napad sačinjen od većeg broja koraka (*eng. multi-step attack*) možemo da predstavimo kao $M_j \subseteq \hat{S}$, pa skup svih takvih napada dalje obeležavamo sa $\widehat{M} = \{M_0, M_1, \dots, M_{m-1}\}$.

C. Klasterizacija upozorenja

Proces klasterizacije upozorenja predstavlja razdvajanje poruka upozorenja u odgovarajuće grupe upozorenja, odnosno u klastere $\hat{C} = \{C_0, C_1, \dots, C_{n-1}\}$, gde svaki klaster $Ci \in \hat{C}$ predstavlja napad i njemu pridružena upozorenja. Klasteri bi trebalo da modeluju stvarne napade $Si \in \hat{S}$, pa bismo u najboljem slučaju imali $\hat{S} = \hat{C}$. U ove svrhe se primenjuju dva zadatka: filtriranje upozorenja i izolacija napada.

Filtriranje upozorenja uklanja prividno tačna (*eng. false positive*) upozorenja, tako da u najboljem slučaju važi:

$$\forall a \in A: \quad \exists Si \in \widehat{S} \ \wedge \ a \in Si \iff \exists Ci \in \hat{C} \ \wedge \ a \in Ci$$

Definicija prividno tačnih upozorenja zavisi od konteksta, odnosno napada za koji se posmatra problem. Ovaj korak podrazumeva da se upozorenja dodele klasteru i samim tim dalje formiraju ulazne podatke za sledeći stepen analize ako i samo ako su prepoznata u skupu $\hat{S}$.

Izolovanje napada zahteva da se svako upozorenje iz skupa A dodeli jednom klasteru $Ci \in \hat{C}$. Skup korelisanih klastera $\hat{C}$ bi trebalo da isprati originalne korake u napadu $Si \subseteq A$, $Si \in \hat{S}$. Ovaj zadatak zavisi od toga da li za cilj imamo visoku homogenost ili heterogenost unutar klastera, odnosno izazov je pronaći korelaciju koja najbolje opisuje realnu situaciju, a optimalna je iz perspektive data mining-a.

D. Dodavanje konteksta

U ovom koraku, svakom klasteru $Ci \in \hat{C}$ se dodeljuje labela $li \in L$ koja sadrži dodatne informacije vezane za klaster, na primer opis klastera, opis tačaka koje su podložne napadu itd. Na ovaj način se olakšava analiza i kasnije uklapanje koraka u kompleksne napade.

E. Povezivanje napada

Kao poslednji korak u procesu korelacije upozorenja, povezivanje napada nastoji da pronađe relacije između klastera u skupu $\hat{C}$. Rezultat su $Ii \in \hat{I}$ koji reflektuju uklapanja iz skupa $\hat{S}$ u skup napada sa većim brojem koraka $\widehat{M}$. Primer ovoga je skeniranje nezaštićenih web server-a u okviru jedne podmreže, koje je dalje praćeno napadom na sam server, odnosno ugradnjom malicioznog softvera i eventualnom preuzimanju kontrole. Kako bi ova dva koraka povezali, koristi se kombinacija sekvencijalnih i kauzalno-zasnovanih mehanizma korelacije.

Ova faza je jako zavisna on samog opisa dodeljenog klasterima, jer se na osnovu toga najbolje mogu odrediti potencijalne veze između grupa upozorenja.

F. Transformacija upozorenja u strukturu grafa

Skup upozorenja A sa svojim atributima se može transformisati u težinski graf $G = (A, E)$ gde je za skup čvorova grafa iskorišćen skup A, a skupom E označeni svi potezi između čvorova, odnosno poruka upozorenja, u obliku (a_i, a_j) čija se težina potega dobija na osnovu vrednosti funkcije sličnosti između para upozorenja a_i i a_j:

$$s = F_{sim}(a_i, a_j), \in [0, 1]$$

Funkcija F_{sim} poredi svih n atributa $(a^0, a^1, \dots, a^{n-1})$ između dva upozorenja respektivno:

$$F_{sim}(a_i, a_j) \ = \sum_{k=0}^{n-1} c^k \cdot h^k \left(a_i^k, a_j^k\right)$$

Za svaki od atributa se primenjuje odgovarajuća funkcija poređenja $h^k \in [0, 1]$ čija vrednost biva skalirana odgovarajućim faktorom vektora $c = (c^0, c^1, \dots, c^{n-1})$ za koji važi $\sum c^j = 1$, odnosno funkcija sličnosti dva upozorenja zapravo predstavlja skalarni proizvod vektora c i h, gde vrednosti elemenata vektora h^k dobijamo na osnovu pridruženih funkcija za poređenje specifičnih atributa.

Na osnovu izračunate težine potega s se dalje odlučuje da li će odgovarajući poteg biti prisutan u grafu ili ne. Sličnost između dva čvora (poruka upozorenja) je potrebno da bude veća ili jednaka od minimalnog praga sličnosti τ, za poteg (a_i, a_j) da bi bio deo skupa potega E^τ odnosno grafa G^τ. Iz ovoga možemo zaključiti da τ kontroliše broj prikazanih potega $|E|$, time što uklanja potege između čvorova koji su najverovatnije nepovezani.

Težina potega zavisi isključivo od funkcije F_{sim} odnosno od vektora c i funkcija poređenja vektora h, dakle ovo će biti parametri konfiguracije kako bi se dobijali grafovi za različite vrste napada.

G. Klasterizacija upozorenja unutar strukture grafa

III. Implementacija

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections

A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. *Abbreviations and Acronyms*

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. *Unitsransome*

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

C. *Equations*

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. *Some Common Mistakes*

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

IV. PRIKAZ DOBIJENIH PODATAKA

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. *Authors and Affiliations*

The template is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

Identify applicable funding agency here. If none, delete this text box.

1) For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) For papers with less than six authors: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) Deletion: Delete the author and affiliation lines for the extra authors.

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1", "Heading 2", "Heading 3", and "Heading 4" are prescribed.

C. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a.] Sample of a Table footnote. (*Table footnote*)

Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

V. ZAKLJUČAK

ZAHVALNICA

Želeo bih da se zahvalim profesoru dr Vladimiru Ćiriću, kao i [insert title here] Nađi Gavrilović za mentorstvo pri pisanju rada.

LITERATURA

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.