Warszawa, 22 listopad 2019 r.

Prof. dr hab. inż. Ewa Niewiadomska-Szynkiewicz
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska
ens@ia.pw.edu.pl

RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY WYDZIAŁU CYBERNETYKI WOJSKOWEJ AKADEMII TECHNICZNEJ

Tytuł rozprawy: Clustering-based method for botnet detection.

Autor rozprawy: mgr inż. Hubert Ostap

1. Ogólna charakterystyka rozprawy. Cel badań.

Systematycznie wzrastająca liczba oraz wzrastająca złożoność prób działań nieuprawnionych w sieciach teleinformatycznych powoduje konieczność opracowania nowych mechanizmów oraz technik wykrywania i przeciwdziałania potencjalnym skutkom naruszeń bezpieczeństwa informacyjnego. Jedna ze szczególnie groźnych technik ataków polega na wykorzystaniu zainfekowanych komputerów użytkowników sieci do prowadzenia zmasowanych ataków na wybrane zasoby, w tym z wykorzystaniem coraz bardziej złożonych sposobów ich maskowania.

Przedmiotem badań udokumentowanych w rozprawie są metody i techniki wykrywania botnetów, tj. grup zainfekowanych szkodliwym oprogramowaniem komputerów pozostających w ukryciu przed użytkownikiem i pozwalających jego twórcy na sprawowanie zdalnej kontroli nad wszystkimi komputerami w ramach grupy. Zainfekowane komputery w sieci stają się trudnymi do wykrycia źródłami przyszłych ataków. W ostatnich latach obserwuje się znaczny wzrost liczby i nowych typów botnetów oraz wzmożenie ich aktywności. W tym kontekście, uważam podjętą w rozprawie problematykę za bardzo aktualną i istotną, zarówno z poznawczego, jak i praktycznego punktu widzenia.

Doktorant w rozdziale wprowadzającym do pracy podaje przykłady różnych podejść do detekcji botnetów, szczególną uwagę zwraca na metody wykorzystujące pewne charakterystyki ruchu sieciowego, nie wymagające wykonania analizy zawartości przepływów. W rozdziale 2.5 „Main goal and task to resolve" formułuje główny cel rozprawy oraz zadania badawcze prowadzące do osiągnięcia tego celu. Jako główny cel prowadzonych przez siebie badań podaje opracowanie, realizację i weryfikację eksperymentalną metody detekcji botnetów polegającej na identyfikacji sytuacji, w której dochodzi do synchronicznej komunikacji generowanej przez wiele urządzeń w sieci. Tego typu zachowanie, jak wskazują źródła literaturowe, jest istotnym czynnikiem wskazującym na podejrzenie zainfekowania maszyn przez twórcę botnetu. Wymagania postawione przed wytworzonym rozwiązaniem

obejmowały jego uniwersalność, tj. zapewnienie działania niezależnie od stosowanego w sieci protokołu komunikacyjnego, wysoką skuteczność detekcji zarówno znanych, jak i jeszcze nieznanych botnetów, większą efektywność od podejść proponowanych w literaturze. Ponadto założono, że proponowane rozwiązanie będzie mogło posłużyć działaniom prewencyjnym, tj. uniemożliwiającym przyszły atak, dzięki wykryciu wszystkich zainfekowanych maszyn, nie tylko maszynę twórcy botnetu (C&C).

Podsumowując, rozprawa ma głównie charakter konstrukcyjny i doświadczalny. Rozważania teoretyczne koncentrują się na opracowaniu modeli matematycznych oraz sformułowaniu zadań klasteryzacji, filtrowania i klasyfikacji. Część konstrukcyjna pracy obejmuje realizację systemu oprogramowania BotTROP do detekcji botnetów oraz wykonanie symulatorów do generowania ruchu sieciowego tworzonego przez botnety i komponenty sieci działające w normalnych warunkach. Część doświadczalna zawiera wyniki licznych eksperymentów obejmujących zarówno badania wykonane z wykorzystaniem wspomnianych symulatorów, jak i eksperymenty w fizycznej (rzeczywistej) sieci. Wyniki badań pozwalają uznać, że Doktorant osiągnął założony cel oraz potwierdzają ogólną słuszność działań podjętych przez autora pracy.

2. Syntetyczna analiza treści rozprawy. Charakter rozprawy

Zasadnicza część rozprawy składa się z sześciu rozdziałów. Rozdział pierwszy (*Introduction*) zawiera wprowadzenie w tematykę pracy. Obejmuje przedstawienie kontekstu rozważanego zagadnienia, genezę i uzasadnienie podjętego problemu badawczego oraz jego usytuowanie na tle aktualnego stanu badań. Zawarty w nim materiał stanowi punkt wyjścia dla treści prezentowanych w dalszej części pracy. W opinii recenzenta należało w tym rozdziale wyraźnie przedstawić i sformułować (wyróżnić) postawiony problem badawczy i cel badań – co ostatecznie autor chce osiągnąć. Doktorant wspomina, że efektem prac jest nowa metoda BotTROP, ale podaje niewiele informacji, które mogłyby sugerować, że będzie to rozwiązanie skuteczniejsze od zaproponowanych przez innych badaczy. Nie informuje również w jaki sposób zamierza wykazać jej efektywność. Pożądane byłoby silniejsze uzasadnienie celowości tworzenia nowego narzędzia wykrywania botnetu, w tym wskazanie słabości istniejących rozwiązań, ich krytyczna analiza, a także sformułowanie miar skuteczności ochrony sieci.

W rozdziale drugim Doktorant opisał ogólnie zagrożenia bezpieczeństwa sieci komputerowej wywoływane przez grupy zainfekowanych komputerów. Przedstawił podstawowe definicje, komponenty botnetu, opisał proces tworzenia grupy oraz jej późniejszego szkodliwego działania, czyli pełny cykl życia systemu. Omówione są różne warianty architektury systemu, tj. scentralizowana, rozproszona i hybrydowa. Szczególna uwaga jest zwrócona na botnety zdecentralizowane wykorzystujące sieci społeczne. Druga część rozdziału zawiera przegląd metod detekcji botnetu. Autor przytacza liczne publikacje naukowe przedstawiające różne podejścia i techniki wykrywania tego typu ataku sieciowego. Przedstawiona jest przyjęta w literaturze naukowej klasyfikacja metod oraz omówione są wybrane rozwiązania. W tej części rozdziału zaprezentowana jest bardzo ważna cecha chrakteryzująca aktywność zainfekowanych maszyn – synchroniczna komunikacja. Jest to cecha, która stanowi podstawę metody opracowanej przez autora oraz rozwiązań zaproponowanych przez innych badaczy.

Istotną częścią rozprawy jest rozdział trzeci zawierający opis autorskiej metody wykrywania botnetu - BotTROP. Jej istotą jest wykrycie synchronicznej komunikacji zachodzącej w sieci, wskazującej na potencjalne zainfekowanie grupy maszyn przez twórcę botnetu. Zaproponowane podejście zakłada wykorzystanie kilku technik, tj. klasteryzacji

wysyłanych pakietów, a konkretnie znaczników czasowych przypisanych inicjacji sesji, filtrację pakietów i adresów IP oraz klasyfikację adresów IP pod względem typu komunikacji (synchroniczna, asynchroniczna).

W rozdziale czwartym Doktorant opisał autorską wersję systemu do detekcji botnetu oferującego dwie metody: BotTROP (metoda zaproponowana przez Doktoranta) oraz BotGAD (metoda z literatury). Przedstawił również opis dwóch wykonanych przez siebie symulatorów generujących sztuczny ruch sieciowy w warunkach normalnej pracy oraz w przypadku zainfekowania maszyn przez twórcę botnetu. Wytworzenie tych narzędzi było niezbędne do przeprowadzenia wnikliwej analizy skuteczności zaproponowanych rozwiązań.

Rozdział piaty zawiera szczegółową prezentację oraz omówieniem syntetycznych wyników licznych eksperymentów przeprowadzonych dla symulowanego i rzeczywistego ruchu sieciowego. Szczególna uwaga jest zwrócona na porównanie autorskiego rozwiązania (BotTROP) i metody BotGAD. Należy podkreślić, że Doktorant przeprowadził bardzo wnikliwą analizę eksperymentalną, rozważył różne warianty działania metod i różne dane wejściowe. Szkoda jedynie, że do badań porównwczych wybrał tylko jedną metodę z literatury, w tym bazującą na bardzo podobnym podejściu. W rozdziale szóstym autor pisze „The BotTROP method was created as a modification of the BotGAD algorithm".

Podsumowanie rozprawy zawiera rozdział szósty. Omówione są podstawowe własności metody BotTROP oraz zakres prac wykonanych w celu zbadania jej skuteczności i efektywności. Szkoda, że dopiero w tym rozdziale Doktorant wyraźnie podaje, że metoda BotTROP jest modyfikacją metody BotGAD (zdanie w paragrafie powyżej) i wypunktowuje podstawowe elementy różniące obie metody. Ten opis zdecydowanie powinien się pojawić w rozdziale trzecim. Rozdział kończy syntetyczne omówienie wyników oraz kilka wniosków dotyczących możliwego zastosowania wykonanego narzędzia w sieci produkcyjnej. Autor wskazuje możliwe udoskonalenia systemu. Wydaje się, że opisy te powinny znaleźć się raczej w części poświęconej implementacji – są nieco zbyt szczegółowe i techniczne na podsumowanie pracy. Oczekuje się, że wnioski formułowane w podsumowaniu powinny mieć raczej bardziej ogólną naturę.

3. Ocena analizy źródeł i sposobu sformułowania wniosków wynikających z analizy źródeł.

Ogółem Autor w całej rozprawie odwołuje się do 101 pozycji literatury związanych z tematyką pracy. Przegląd jest dość obszerny i obejmuje liczne pozycje z ostatnich lat. Doktorant analizuje literaturę głównie w rozdziałach pierwszym i drugim. Na podstawie literatury omawia charakterystykę botnetu oraz różne techniki do detekcji tego typu ataków. Szkoda, że Doktorant nie przedstawił obszerniejszej krytycznej oceny przydatności i/lub ograniczeń istniejących narzędzi do rozwiązania sformułowanego przez niego problemu badawczego.

Podsumowując, uważam, że mimo pewnych braków przedstawiona w rozprawie analiza źródeł oraz postawione na podstawie przeglądu wnioski świadczą o wiedzy autora w przedmiocie rozprawy. Doktorant analizuje przedstawione w literaturze podejścia i rozwiązania do detekcji botnetów, formułuje zagadnienia badawcze i proponuje autorską metodę rozszerzającą propozycję innych badaczy. Należy podkreślić, że problematyka cyberbezpieczeństwa stanowi tak obszerny i dynamicznie rozwijający się nurt badań, że jakikolwiek przegląd literatury jest z góry skazany na bycie niekompletnym i stosunkowo szybko nieaktualnym.

4. **Analiza poprawności rozwiązania przedstawionego zadania, poprawność przyjętych założeń i wybranych metod.**

Doktorant rozwiązał postawione w pracy zagadnienia. Wykorzystując obserwacje innych badaczy opracował autorską metodę i algorytm detekcji botnetów. Wykonał odpowiednie środowisko oprogramowania do detekcji botnetu oraz badania skuteczności i efektywności różnych metod do identyfikacji tego typu ataków. Dołączył odpowiednie symulatory do generowania danych wejściowych dla scenariuszy badawczych, inspirowanych danymi rzeczywistymi. Wykonał, wnikliwie przeanalizował i dobrze udokumentował liczne eksperymenty badawcze. Należy podkreślić, że część testów została przeprowadzona w rzeczywistej sieci, w typowych warunkach jej funkcjonowania.

Zdaniem recenzenta przyjęta metoda badawcza obejmująca przedstawienie zapisu formalnego proponowanej metody, sformułowanie odpowiednich algorytmów oraz ich weryfikację przez eksperyment symulacyjny, a następnie eksperyment w środowisku docelowym jest właściwa dla badanego zagadnienia. Uzyskane wyniki potwierdzają poprawność rozwiązania. Autor w rozprawie doktorskiej uzasadnia, że zaproponowana technika może wspierać detekcję zagrożeń w systemach teleinformatycznych, a uzyskane rezultaty pokazują wyższą skuteczność w stosunku do rozwiązań proponowanych w literaturze.

5. **Oryginalność rozprawy, samodzielny dorobek autora, pozycja rozprawy w stosunku do stanu wiedzy prezentowanego w literaturze światowej.**

Rozprawa zawiera nowe oryginalne rezultaty. Do szczególnie ważnych należy zaliczyć:

- Opracowanie uniwersalnej metody detekcji botnetu polegającej na identyfikacji urządzeń nadających synchronicznie, zakładającej możliwość współpracy z różnymi protokołami komunikacyjnymi w sieci.
- Opracowanie i implementacja algorytmu detekcji botnetu realizującego autorską metodę.
- Opracowanie i realizacja środowiska do badania i oceny skuteczności i efektywności algorytmów detekcji botnetu.
- Wykazanie poprawności i skuteczności zaproponowanej techniki za pomocą eksperymentów wykonanych w symulatorze i w rzeczywistej sieci teleinformatycznej.

Przedstawione wyniki mają istotne znaczenie dla rozwoju metod detekcji ataków sieciowych.

6. **Analiza poprawności prezentacji wyników pracy**

Recenzowana rozprawa doktorska jest napisana w języku angielskim i liczy 178 stron, zawiera 53 rysunki, 43 tabele, listę skrótów oraz bibliografię obejmującą 101 publikacji. Zasadnicza treść obejmuje 170 stron, pozostałe zawierają spis treści, wykaz rysunków, tabel, skrótów oraz bibliografię. Pewnym utrudnieniem dla systematycznego czytelnika jest brak w opracowaniu indeksu oraz spisu oznaczeń używanych w formułach matematycznych.

Generalnie praca jest zredagowana dość starannie, zarówno pod względem językowym, jak i graficznym. Występują pewne błędy językowe, gramatyczne i redakcyjne, ale nie wpływają one w sposób istotny na jakość tekstu. Główne zastrzeżenia dotyczą prezentacji formuł matematycznych. Wzory są pisane za pomocą czcionek o różnej wielkości. Brakuje odstępów między formułami pisanymi w jednej linii. Występują błędy w symbolach (np. str. 30) i wzorach (np. str. 33, 41, 43 itd.) – literówki, zgubione nawiasy itd. Brakuje dokładnych objaśnień zmiennych lub objaśnienia są niejasne.

Niektóre rysunki powinny być lepszej jakości (np. 2.8, 3.2, 3.3, 3.4). Szczególnie ważne są rysunki 3.2 – 3.4, które ilustrują działanie metody – opisy osi *Time* są zupełnie nieczytelne. Niepełne podpisy pod rysunkami 3.5 i 3.6. Zastrzeżenia można zgłosić również do opracowania bibliografii – wskazana byłaby większa staranność. W przypadku większości publikacji stanowiących materiały konferencyjne brakuje nazwy konferencji i numerów stron, w przypadku książek Doktorant nie podaje nazwy wydawnictwa.

Układ oraz zawartość większości rozdziałów są generalnie prawidłowe, aczkolwiek niektóre z nich mogłyby być lepiej napisane. Autor w klarowny sposób omówił cykl życia i możliwe architektury botnetów. Nieco zbyt skrótowo przedstawił techniki detekcji tego typu ataków. Zdaniem recenzenta, czytelność pracy poprawiłoby podzielenie podrozdziału 2.4 na dwa podrozdziały – pierwszy, poświęcony omówieniu charakterystycznych zachowań maszyn zainfekowanych przez botmastera (część podrozdziału 2.4 łącznie z rysunkiem 2.7) oraz drugi stanowiący przegląd rozwiązań proponowanych w literaturze. Brakuje również w rozdziale 2 jasnego sformułowania kryteriów według jakich jest oceniana skuteczność metod detekcji botnetu – opis jest wprowadzony dopiero w rozdziale 5.

Nie wnoszę krytycznych uwag do rozdziałów 4 i 5. Opis realizacji algorytmu oraz symulatorów jest jasny. Przebieg wykonanych eksperymentów, scenariusze testów zostały opisane zgodnie ze sztuką. Wyniki badań są dobrze przedstawione i wnikliwie przeanalizowane.

Najsłabszą część pracy stanowi moim zdaniem rozdział 3 prezentujący metodę opracowaną przez Doktoranta i realizujący ją algorytm. Opis jest chaotyczny i niejasny. Może wynikać to z faktu, że praca jest napisana w języku angielskim, przez co trudność pisania jest większa. Zaproponowana notacja jest w pewnych miejscach nieoczywista, często przypadkowa i niezgodna z przyjętymi zasadami. Często trudno jest się zorientować co autor ma na myśli. Np. str. 37, raz mamy „*k-means*", w następnym zdaniu „*K-means*", później K jest wektorem ($K[min]$), a w algorytmie K jest 2-wymiarową tablicą. Tak więc nie jest wiadome czym jest K – stałą, wektorem, czy tablicą? Pojawiają się takie niedokładne sformułowania jak „*moment of communication*", „*set of ordered numbers of clusters*" zamiast po prostu „*set of clusters identifiers*" itd. Czasami to co jest napisane słownie nie odzwierciedla prezentowanych dalej zapisów formalnych.

Pewne zastrzeżenia są również do rozdziału podsumowującego pracy (rozdział 6). Rozdział ten zawiera szereg szczegółowych informacji dotyczących zaproponowanego rozwiązania, które powinny znaleźć się w rozdziale 3. Dopiero w tym miejscu autor jasno i dokładnie podaje co jest głównym osiągnięciem w zakresie opracowanej metody detekcji – jakie zmiany zaproponował w stosunku do metody BotGAD.

7. Słabe strony rozprawy i jej główne wady

Słabe strony rozprawy, to nawiązując do poprzedniego punktu, mało klarowna prezentacja opracowanej metody BotTROP. Uwagi dotyczą zarówno przyjętej notacji, prezentacji formuł, jak i pewnych niezgodności opisów nieformalnych i formalnych. Utrudnia to znacznie czytelność pracy i zrozumienie koncepcji proponowanego rozwiązania. Po pierwsze z przedstawionego opisu nie wynika jasno co stanowi bazę do detekcji synchronicznej komunikacji – co jest wynikiem klasteryzacji? Na str. 30 (rozdz. 3.2) czytamy „*During the first stage the algorithm clusters all* ***the time stamps of the packets*** *sent from the source the destination IP being analyzed*". Na stronie 32 „*During the next step,* ***all connection moments*** *of every source IP are compared with the connection time of the IP at which the clusters were created*". Str. 73 „*an appropriate method for clustering* ***network traffic***....". Jednocześnie na str. 32 zapis algorytmu (Algorithm 2) wyraźnie wskazuje, że

elementami klastra są po prostu **adresy IP**, co nie do końca zgadza się z prezentacją graficzną (rys. 3.3), gdzie elementy klastra to **pary**: **znacznik czasowy i adres IP**. Autor niepotrzebnie wprowadza bardzo skomplikowaną notację. Definiując zmienne zupełnie pomija możliwość skorzystania z struktur złożonych – tablic. Właściwie je wykorzystuje, ale niejawnie. IP^S_{max}, ST itd. to tablice 2 i 3 wymiarowe. Pisząc jawnie, że posługuje się wielowymiarowymi tablicami nie raziłby zapis $ST(i,j,k)$, czyli elementu tablicy ST. Inaczej nie ma jasności, czy np. ST to zmienna, tablica, czy funkcja. Wprowadzone pewne dość dziwne i niepotrzebne konstrukcje (np. definicja $T(.,.,.)$ na str. 31) tylko zaciemniają opis. Brakuje konsekwencji w definiowaniu zmiennych, raz używane są indeksy grup i protokołów, a innym razem jawnie pełne nazwy Pr_j i G^{IP}_i (np. str. 33). Znacznie prościej można było również przedstawić definicje k^* i $e(i,j)$ jako, k^*=arg max $I_T(\ldots)$. Wszystkie zamieszczone w rozdziale 3 algorytmy prezentują działanie metody dla ustalonego indeksu grupy (i) oraz protokołu (j). Można było prowadzić dalsze rozważania z pominięciem tych parametrów, co poprawiłoby czytelność pracy. Poza uwagami dotyczącymi notacji i jasności przekazu pozytywnie oceniam sposób prezentacji metody i algorytmów z wykorzystaniem pseudokodu i ilustracją graficzną.

Ponadto autor nie zawsze dostatecznie precyzyjnie wskazuje na to, co stanowi jego oryginalne osiągnięcie. Szkoda, że we wstępie rozdziału 3 nie zostało jasno pokazane (np. w punktach) co różni metody BotGAD i autorską BotTROP. Nie odnosi się również do swoich publikacji.

W rozdziale 6 Doktorant opisuje działania jakie należy podjąć w przypadku planów zastosowania narzędzia BotTROP do wsparcia detekcji botnetów w rzeczywistej sieci. Niemniej, wykrycie synchronicznej komunikacji nie jest jednoznaczne z wykryciem botnetu (wspomina o tym na str. 20). Szkoda, że Doktorant nie zaproponował dalszych działań, które mogłyby wzmocnić przekonanie, że wykryta komunikacja synchroniczna jest związana z istnieniem sieci botnet. Wspomniane działania mogłyby być prowadzone przy wykorzystaniu technik opracowanych przez innych badaczy.

8. Przydatność rozprawy dla nauk technicznych

Mimo wymienionych powyżej słabych stron pracy uważam, że przedstawione w rozprawie wyniki wnoszą istotny wkład w dyscyplinę informatyka, a konkretnie w rozwój badań w zakresie cyberochrony sieci teleinformatycznych. Proponowana metoda detekcji botnetów jest uniwersalna, może być stosowana dla różnych protokołów komunikacyjnych. Wykonane eksperymenty badawcze pokazują jej znaczną skuteczność i efektywność. Należy podkreślić, że tematyka rozprawy cieszy się od kilku lat niesłabnącym zainteresowaniem.

9. Podsumowanie i wniosek końcowy

Uważam, że rozprawa mgr inż. Huberta Ostapa spełnia wymagania stawiane rozprawom doktorskim przez przepisy ustawy o tytułach i stopniach naukowych. Stawiam wniosek o przyjęcie rozprawy i dopuszczenie do obrony doktorskiej.

podpis