Prof. dr hab. inż. Bohdan Macukow
Politechnika Warszawska
Wydział Matematyki I Nauk Informacyjnych
Zakład Systemów Przetwarzania Informacji
00-662 Warszawa, Koszykowa 75

Warszawa, 10 listopada 2019 r.

Recenzja Rozprawy doktorskiej mgra inż. Fryderyka Darnowskiego pt. „Metoda odtwarzania zmian danych na dysku z systemem plików NTFS wykorzystująca informacje zawarte w pliku systemowym $MFT”

Niniejsza recenzja została przygotowana na zlecenie Rady Wydziału Cybernetyki Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego przekazane przez Przewodniczącego Rady dr hab. inż. Kazimierza Worwę, prof. WAT (uchwała nr 226/WCY/2019 z dnia 24 września 2019 r.).

1. Opinia ogólna.

Najogólniej biorąc, praca jest oryginalna i niestandardowa. Należy do obszaru informatyki śledczej. Zawiera ważne wyniki naukowe, a jednocześnie ma istotne walory praktyczne włącznie z praktycznymi zastosowaniami – działań w zakresie zabezpieczania i analizy cyfrowych nośników danych. Jest to opracowanie, które należy ocenić zdecydowanie pozytywnie. Co więcej, podjęty temat jest bardzo aktualny i istotny.

2. Struktura rozprawy

Praca napisana jest poprawnym językiem z użyciem poprawnej terminologii. Jest bardzo obszerna - liczy 240 stron tekstu podstawowego oraz cztery załączniki (63 strony) i płytę CD zawierającą omówienie i wyniki działania opracowanego przez doktoranta programu. Składa się z 12 rozdziałów, zawiera wykaz najważniejszych oznaczeń, spis stosowanych akronimów, spisy rysunków i tabel, wykaz załączników oraz bibliografię.

Rozdział 1 jest wprowadzeniem do tematyki rozprawy, zawiera ustosunkowanie się do obecnego stanu wiedzy na podstawie analizy literatury, a ponadto sformułowanie problemu naukowego i przyjętą metodę jego rozwiązania. Rozdział 2 dotyczy organizacji danych na dysku z systemem plików NTFS (New Technology File System). Główną uwagę

zwrócono przy tym na plik systemowy $MFT (Master File Table) który zawiera najistotniejsze informacje o plikach zapisywanych na dysku. W rozdziale 3 przedstawiono dwa typy zapisu plików na dysku oraz rozmieszczenie danych na dysku w wyniku wykonywania ciągu operacji zapisów i kasowań plików. Rozdział 4 zawiera opracowany przez autora model matematyczny dysku, na którym wydzielono obszary zajmowane przez dane znajdujące się w różnych stanach, oraz związków pomiędzy tymi obszarami dysku, a plikiem systemowym $MFT. Rozdział 5 uzupełnia model matematyczny o sformalizowany opis akcji zapisu i kasowania pliku w zdefiniowanych przypadkach. Rozdziały 6 i 7 to formalny opisu opracowanej metody odtwarzania danych na dysku. Autor zdefiniował tutaj przebieg zmian na dysku jako deterministyczny proces zmian stanu dysku w wyniku wykonywania akcji zapisów i kasowań. Następnie zaproponował skonstruowanie pewnego odwzorowania stanu bieżącego w zbiór możliwych stanów poprzedzających stan analizowany - nazwane odwrotną funkcją przejścia. Rozdział 8 zawiera matematyczny opis opracowanej metody wyznaczania tego zbioru możliwych wcześniejszych stanów dysku przy założeniu znajomości pewnego stanu. Zwrócono tu uwagę na dużą złożoność obliczeniową opracowanego algorytmu, a rozdział 9 omawia dwie modyfikacje algorytmu z rozdziału 8. Pierwsza modyfikacja spowodowana jest faktem, że w rzeczywistości na ogół nie jest znana kolejność wykonywania wszystkich analizowanych operacji zapisów i kasowań, a tylko następstwo czasowe części akcji, o których informacje znajdują się w pliku $MFT. Celem drugiej modyfikacji jest zmniejszenie złożoności obliczeniowej algorytmu na drodze zrezygnowania z analizowania niektórych właściwości zapisywanych i kasowanych plików. Wykonano implementację metody w języku Java. Opracowany model matematyczny oraz algorytmy zweryfikowane zostały w rozdziałach 10 i 11. W pierwszym przypadku przeprowadzono obliczenia dla w pełni znanego ciągu akcji zapisów i kasowań oraz porównywano wyniki działania programu z kolejnymi znanymi stanami. Przypadek drugi dotyczy dysków rzeczywistych, dla których przeprowadzono siedem testów. Rozdział 12 zawiera wnioski końcowe będące omówieniem efektów realizacji sformułowanych zadań cząstkowych, w szczególności ustosunkowano się do możliwości i zakresu stosowania opracowanej metody oraz jej doskonalenia.

3. Ocena merytoryczna

3.1 Znaczenie problematyki podjętej w recenzowanej rozprawie

Informatyka śledcza/kryminalistyczna to ogólna nazwa czynności zabezpieczenia i analizy cyfrowych nośników danych. Gwałtowny rozwój technologii internetowych, przechowywanie informacji w „chmurze” oraz różnorodność nośników cyfrowych sprawiają, że wciąż nie ma ustalonych reguł jak je zabezpieczyć. Podobnie istnieje ogromna ilość metod analizy cyfrowych nośników danych oraz programów do ich analizy.

W dalszej części rozprawy zaprezentowana została ta część analizy nośników danych, która odpowiada za odzyskiwanie wykasowanych plików (*carving*). Współczesne metody

analizy pomijają w rozważaniach system plików i zmiany w nim zachodzące, natomiast koncentrują się na obszarach danych. Podobnie, dla odzyskanego pliku z przestrzeni niezaalokowanej, niewiadome pozostają informacje o tym kiedy plik był nagrany, kiedy był wykasowany czy też, w przypadku plików pofragmentowanych, gdzie znajdują się jego pozostałe części.

Autor rozprawy sformułował następujący problem naukowy:

Należy wykorzystać dostępne informacje systemowe ($MFT) i zmiany w nim zachodzące dla odzyskiwania informacji o plikach, które były zapisywane i wykasowywane w pewnym analizowanym przedziale czasu.

Przy rozwiązywaniu powyższego problemu naukowego autor opracował oryginalną metodę odzyskiwania informacji o zapisywanych i kasowanych plikach na dysku z systemem NTFS, w tym charakterystykach czasowych ich nagrywania i wykasowywania.

Rozwiązanie powyższego problemu naukowego wymagało jego podzielenia na zadania cząstkowe:

1) analiza organizacji danych na dysku z systemem plików NTFS,
2) opracowanie modeli matematycznych dysku oraz akcji podejmowanych na dyskach,
3) przedstawienie zmian w procesie zapisów i kasowań plików na dysku jako procesu deterministycznego,
4) opracowanie metody odczytywania stanu dysku z przeszłości,
5) oszacowanie złożoności obliczeniowej zaproponowanych algorytmów,
6) analiza zakresu zastosowań opracowanej metody.

Zaproponowana metoda wykorzystuje modelowanie matematyczne analizowanego procesu, a na podstawie skonstruowanego, obszernego i szczegółowego modelu matematycznego zaproponowany został algorytm odzyskiwania informacji. Zastosowano tu podejście stosowane w teorii automatów skończonych.

Sformułowany problem i metodyka jego rozwiązania świadczą o swobodzie poruszania się w obszarze informatyki i matematyki w zakresie jej wykorzystania i zastosowania praktycznego w informatyce.

Podjętą problematykę oraz postawione tezy, rozprawy można sklasyfikować, jako teoretyczną o charakterze silnie aplikacyjnym.

3.2 Metodyka badawcza i samodzielny dorobek autora

Autor wprowadził pojęcie stanu, który mówi o zawartości pliku systemowego $MFT oraz zawartości obszarów danych na dysku przed akcją zapisu lub kasowania kolejnego pliku.

Autor przedstawił model akcji zapisu danych oraz akcji ich kasowania. Przyjęto przy tym, że kolejność wykonywania tych akcji jest w pełni znana. Przeanalizowano różne przypadki akcji zapisu. W szczególności uwzględniono fragmentację zapisywanego pliku lub jej brak oraz dwa typy zapisu: zapis wyłącznie w obszarze dysku niezaalokowanym oraz zapis możliwy zarówno w obszarze niezaalokowanym jak i w obszarze zaalokowanych przez dane, o których podano informację, że dane te dotyczą plików wcześniej wykasowanych.
Nowatorskim elementem modelu matematycznego jest formalne opisanie działania algorytmów FFA (First Fit Algorithm) w którym przydział następuje zgodnie z zasadą „pierwszy wolny" , oraz algorytmu BFA (Best Fit Algorithm) ustalającego rozmieszczenie zapisywanych plików na dysku.

W wyniku działania tych algorytmów powstaje deterministyczny proces zapisu danych pliku o znanej wielkości. Zapisy plików o wielkości nieznanej w chwili rozpoczynania ich zapisywania nie są w pracy analizowane. Na bazie modelu stanu, modelu akcji zapisu oraz modelu akcji kasowania pliku zdefiniowany został pewien deterministyczny proces zmian stanu dysku i pliku $MFT w wyniku wykonania ciągu akcji zapisów i kasowań. Zmianę stanu przedstawiono w postaci funkcji przejścia dla ciągu analizowanych przypadków wystąpienia akcji zapisu plików oraz akcji kasowania plików.

Zgodnie z powyższym założeniem, że w ciągu akcji występują wyłącznie pliki o znanej wielkości i jest możliwość ich zapisu ze względu na wielkość miejsca dostępnego na dysku, znajomość stanu początkowego procesu oraz ciągu wykonywanych akcji pozwala na jednoznaczne wyznaczenie stanu końcowego. Istotą opracowanej metody odzyskiwania informacji o danych jest odwrócenie tego procesu. Autor wykazał, że nie jest to jednoznaczne, czyli że dla ustalonego stanu po tzw. cofnięciu znanej akcji - uznaniu jej za niewykonaną - poprzedników tego stanu może być wiele. Stąd wynika duża złożoność pamięciowa algorytmu opartego na takiej analizie procesu. Autor pokazał jednak, że w wyniku kolejnych cofań akcji można stwierdzić, że wynikające z tego ciągi kolejnych stanów są niedopuszczalne np. z powodu sprzeczności z działaniem algorytmów FFA oraz BFA. Pozwala to na istotne zmniejszenie liczby rozpatrywanych stanów w kolejnych krokach cofania akcji.

Samodzielny, szczególnie istotny dorobek autora rozprawy polega m.in. na:

1. Opracowaniu modelu matematycznego obejmujący opis dysku, pliku systemowego $MFT, działania algorytmów FFA i BFA, wykonywania akcji zapisu oraz kasowania. Model skonstruowany został przy przyjęcie pewnych dodatkowych założeń, ale słuszność tego podejścia potwierdziły wyniki przeprowadzonych testów. Wszystkie istotne związki między modelowanymi obiektami zostały poprawnie zidentyfikowane. Model posługuje się językiem logiki matematycznej, teorii mnogości oraz teorii grafów i sieci. Szczególnym walorem modelu jest matematyczny opis pliku systemowego oraz działania podstawowych algorytmów rozmieszczających dane na dysku.

2. Opracowaniu na bazie modelu matematycznego autorskiego algorytmu wyznaczania stanu dysku w przeszłości, przy założeniu znajomości kolejności wykonywanych akcji zapisów i kasowań. Algorytm ten ma raczej charakter badawczy, ale jego idea może być wykorzystywana dla wielu przypadków wynikających z modyfikacji założeń przyjętych w modelu matematycznym.
3. Opracowanie zmodyfikowanego algorytmu wyznaczania stanu dysku w przeszłości dostosowanego do dysków rzeczywistych i uwzględniającego ograniczenia związane z występowaniem niektórych zjawisk pominiętych w modelu matematycznym. Algorytm ten jest prezentacją oryginalnego rozwiązania postawionego problemu naukowego.
4. Opracowanie niezbędnego dla celów badawczych oraz praktycznego wykorzystania algorytmu programu w języku *Java*.
5. Przeprowadzenie badań symulacyjnych przypadków testowych oraz badań na dyskach rzeczywistych. Pozwoliły one na wskazanie głównych zalet opracowanej metody, ale także jej najważniejszych ograniczeń.

3.3 Dorobek publikacyjny Doktoranta i wykorzystana literatura

Autor rozprawy jest autorem bądź współautorem 5 prac. W spisie literatury autor rozprawy wymienia 75 prac innych autorów – książek, publikacji i publikacji online, obejmujących szeroki przegląd zagadnień związanych z zagadnieniami zabezpieczania, analizy i odzyskiwania informacji zawarte na nośnikach cyfrowych. W większości są to prace anglojęzyczne świadczące o bardzo dobrej orientacji autora w trendach i osiągnięciach światowej nauki w rozpoznanej problematyce.

4. Uwagi krytyczne

Trudno oprzeć się wrażeniu, że oceniana praca jest zbyt długa. W początkowych rozdziałach (2 - 3) można było zapewne odwołać się do odpowiednich pozycji literatury, ale z drugiej strony, jest to bardzo dobre wskazanie złożoności analizowanego problemu ze względu na różnorodność i niejednoznaczność zapisu danych na nośnikach cyfrowych. Trzeba przyznać, że Autor wykonał tutaj dużą pracę analityczną.

Niektóre stwierdzenia podawane w pracy mogą być dyskusyjne. Przykładowo na str. 48 zamieszczone jest zdanie „*Wedle rozeznania w istniejącej literaturze, zapis typu U jest całkowicie pominięty. Z tego względu w pracy analizowane będą równolegle oba typy zapisu*”. Z przesłanki w tym sformułowaniu należałoby jednak wnioskować, że zapisu typu U nie powinno się modelować, wbrew konkluzji zawartej w tym zdaniu.

Autor wielokrotnie stwierdza, że w plik systemowy $MFT zawiera wyłączne dane o czasie astronomicznym akcji zapisu. Dlaczego w takim razie nie uwzględniono tego faktu

już w opisie założeń do modelu matematycznego i konsekwentnie nie opracowywano metody odtwarzania zmian danych na dysku dla takiego przypadku?

Często również rozważania w rozprawie są bardzo szczegółowe, co także wpływa na długość pracy. Powyższe drobne uwagi krytyczne nie obniżają w sposób znaczący ani redakcji rozprawy, ogólnie starannej, ani jej wyników naukowych i nie maja wpływu na poziom i ocenę pracy.

5. Wnioski i konkluzja końcowa

Podsumowując stwierdzam, że recenzowana rozprawa doktorska mgra inż. Fryderyka Darnowskiego pt. „*Metoda odtwarzania zmian danych na dysku z systemem plików NTFS wykorzystująca informacje zawarte w pliku systemowym $MFT*" spełnia ustawowe wymogi stawiane rozprawom doktorskim:

- stanowi oryginalne rozwiązanie problemu naukowego,
- wykazuje ogólną wiedze Doktoranta w dyscyplinie informatyka techniczna i telekomunikacja w zakresie informatyki,
- potwierdza umiejętność Doktoranta w zakresie samodzielnego prowadzenia badań naukowych.

Recenzja została opracowana zgodnie z:

- § 6.4 rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 19 stycznia 2018 roku *w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodach doktorskich, w postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora,*
- warunkami określonymi w art. 13 ust. 1 ustawy z dnia 14.03.2003 r. z późn. zm. (Dz.U. 2015 poz. 249) *o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki*....,
- w związku z art. 179 ust. 1 ustawy z dnia 3 lipca 2018 r. *Prawo wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1669).

Uważam, że całość wywodu naukowego, a także jakość przeprowadzonych badań są na wysokim poziomie. Biorąc pod uwagę istotność podjętego tematu, stosowane metody badawcze, jej oryginalny i nowatorski charakter i wreszcie wkład w rozwój nauk inżynieryjno-technicznych **wnoszę o dopuszczenie do dalszych etapów postępowania o nadanie stopnia naukowego doktora nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja w zakresie informatyki, w tym dopuszczenie do publicznej obrony.**