Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.633
DOI: https://doi.org/10.24867/28BE08Milosavljevic

SOFTVERSKI ALAT ZA OPORAVAK DATOTEKA U SISTEMU DATOTEKA NTFS

SOFTWARE TOOL FOR NTFS FILE SYSTEM FILE RECOVERY

Nikola Milosavljević, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – *Gubitak podataka može da ima ozbiljne posledice u savremenom informatičkom okruženju, kako za organizacije tako i za pojedince. Iz tog razloga, oporavak podataka predstavlja ključni aspekt upravljanja podacima. Ova oblast zahteva stalno ažuriranje tehnika oporavka podataka i alata koji implementiraju te tehnike, s obzirom na brz razvoj hardvera i softvera. U radu su analizirane različite tehnike oporavka podataka, a naglasak je stavljen na NTFS fajl sistem. Rad takođe prikazuje implementaciju alata za oporavak podataka, koji je baziran na analizi metapodataka fajl sistema. Demonstrirano je kako se koristi implementirani alat i analizirane su njegove prednosti i mane u odnosu na druga slična rešenja.*

Ključne reči: *Oporavak podataka, digitalna forenzika, fajl sistema, NTFS*

Abstract – *Data loss can have serious consequences in modern environments, both for organizations and individuals. For this reason, data recovery is a crucial aspect of data management. This field requires continuous updates to data recovery techniques and tools that implement these techniques, considering the rapid development of hardware and software. This paper analyzes various data recovery techniques, with focus on the NTFS file system. The emphasis is placed on implementation of a data recovery tool based on the analysis of file system metadata. The paper demonstrates how to use the implemented tool and analyzes its advantages and disadvantages compared to other similar solutions.*

Keywords: *data recovery, digital forensics, file systems, NTFS*

1. UVOD

Gubitak podatak se može smatrati ozbiljnom pretnjom za individue, preduzeća i organizacije. Oblast oporavka podataka zahteva ažuriranje metodologija i tehnika u skladu sa brzim tempom razvoja hardvera i softvera. Različite metode oporavka podataka uključuju softverski i hardverski pristup i mogu se koristiti u različitim slučajevima i situacijama.

Motivacija za ovaj rad predstavlja napredovanje informacionih tehnologija i njihov uticaj na bezbednost podataka. Postojeći alati za oporavak podataka nisu u potpunosti ispratili korak napretka tehnologija, što je dodatni stimulans za rad na ovu temu.

NAPOMENA:
Ovaj rad proistekao je iz master rada čiji mentor je bio dr Stevan Gostojić, red. prof.

Izvršeno je i poređenje implementiranog rešenja sa trenutnim standardima na tržištu i izvučene su jasne crte napretka i prostor za poboljšanje implementiranog rešenja.

2. FAJL SISTEMI

Fajl sistem je metod i način strukturiranja podataka koji operativni sistemi koriste kako bi kontrolisali kako se podaci čuvaju i dobavljaju. Odgovorni su za organizaciju, skladištenje i upravljanje podacima, čineći ih esencijalnim za svakodnevnu upotrebu računara, pametnih telefona i drugih uređaja. Uprkos njihovoj vitalnoj ulozi, fajl sistemi takođe postaju izvor problema i izazova kada se suočavamo sa situacijama gubitka podataka ili potrebom za digitalnom forenzikom.

2.1. Skladišni medijumi

Skladišni medijum je fizički uređaj koji prima i čuva elektronske podake koji stižu od operativnog sistema i korisnika i omogućava dostupnost i trajnost podataka. Sami skladišni medijumi se mogu nalaziti u unutrašnjosti kompjutera, kada se zovu interni ili mogu biti priključeni računaru sa njegove spoljašnosti i tada se zovu eksterni skladišni medijumi.

2.1.1 Tvrdi diskovi

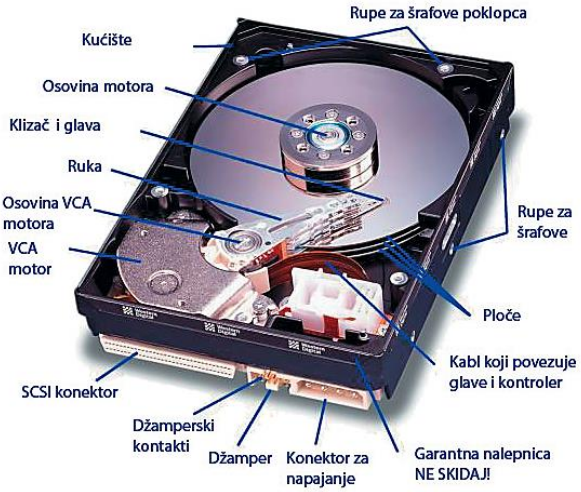
Moderni tvrdi diskovi se ne razlikuju bitno u pogledu delova koji ih sačinjavaju, slika 1. Gledano spolja, na prosečnom tvrdom disku se najpre uočava štampana ploča na kojoj su smeštene komponente koje upravljaju radom uređaja i obezbeđuju stabilno napajanje svih mehaničkih i elektronskih komponenti. Na ovoj ploči se nalaze stabilizatori napona, kontroler, read-only memory (ROM) i random-access memory (RAM). ROM diska je posebno značajan jer sadrži softver koji prilagođava rad konkretnog diska.

Po otvaranju diska vide se ploče na kojima se smeštaju podaci. Moderni tvrdi diskovi koriste ploče koje su najčešće izrađene od feromagnetnog materijala.

Pored ploča vide se i glave koje služe za čitanje i upisivanje podataka. Kako bi se izbeglo trenje i štetno habanje ploča, na tvrdim diskovima glave ne dodiruju ploču već se nalaze na jako malom rastojanju od ploča.

2.1.2. Solid-state diskovi

Solid-state diskovi koriste tip memorije pod nazivom „fleš memorija“ koja je slična RAM-u, ali za razliku od RAM memorije podaci ostaju tu nakon isključivanja računara. Koriste mrežu stranica za brzo slanje i primanje podataka gde je svaka mreža razdvojena na delove koje nazivamo stranice. SSD mogu pisati samo na prazne stranice u bloku.



Slika 1. *Prikaz tvrdog diska*

Kako je brisanje esencijalna funkcionalnost nekog diska, postavlja se pitanje kako SSD briše podatke. Kada je dovoljno stranica u bloku označeno kao neiskorišćeno, SSD će uzeti sadržaj tog bloka, staviti ga u RAM memoriju i obrisati ceo blok. Onda će uzeti prethodno kopirani sadržaj i upisati ga na novi blok bez neiskorišćenih stranica.

2.2 Definicija fajl sistema

Fajl sistemi donose mehanizam za korisnike da čuvaju podatke u hijerarhiju fajlova i direktorijuma. Fajl sistemi se sastoje od strukturalnih i korisničkih podataka koji su organizovani tako da računar zna gde da ih nađe. U velikom broju slučajeva fajl sistem je nezavistan od bilo kog specifičnog kompjutera [1].

Slično kao i razlika između klase i instance, postoje razlike između tipa fajl sistema i fajl sistema. Tipovi fajl sistema su mnogi i među njih spada New Technology File System (NTFS) koji predstavlja fokus ovog rada. Fajl sistem je konkretan sistem datoteka na nekom konkretnom skladišnom medijumu. Predstavlja instancu jednog tipa fajl sistema sa konkretnim podacima.

2.3 Aspekti fajl sistema

Postoji mnogo aspekata jednog fajl sistema kao što su dozvoljena imena fajlova, upravljanje prostorom, metapodaci i drugi. Fajl sistem detaljno raspoređuje prostor, obično pomoću više fizičkih jedinica na skladišnom medijumu. Njegova osnovna funkcija jeste organizacija datoteka i direktorijuma.

Ime datoteke koristi se kako bi se identifikovalo mesto za skladištenje u fajl sistemu [2]. Većina fajl sistema ima ograničenja u vezi sa dužinom imena datoteke i najčešće su osetljivi na mala i velika slova.

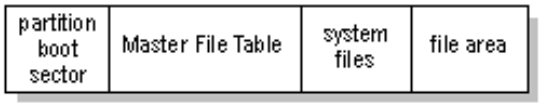
Pored fajlova, fajl sistemi čuvaju i metapodatke vezane svaku datoteku. Dužina fajla, vreme kreiranja, vreme poslednje modifikacije, permisije i tome slično predstavljaju samo neke od metapodataka koje fajl sistem upotrebljava za efikasan rad sa podacima.

3. NTFS

New Technology File System (NTFS) [3] je fajl sistem dizajniran za pouzdanost, bezbednost i efikasnost prilikom rukovanja sa velikom količinom podataka. Razvijen je od strance Microsoft-a i počevši od Windows NT 3.1, on je standardni fajl sistem Windows NT porodice zamenivši File Allocation Table (FAT) [4]. Na žalost, ne postoji oficijalna specifikacija toga kako NTFS čuva podatke na disku. Deskripcije visokog nivoa postoje, dok za niži nivo postoje samo nagađanja koja su dobijena analizom.

3.1 Struktura

Formatiranje particije NTFS fajl sistemom kao rezultat daje kreiranje nekoliko sistemskih fajlova.



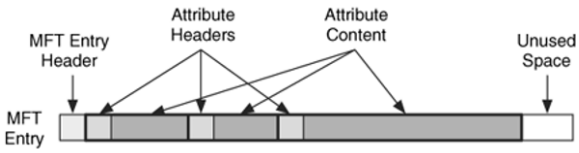
Slika 2. *Izgled NTFS particije nakon formatiranja*

Na slici 2. se vidi da se NTFS sastoji od nekoliko komponenti. Prva je PBS (partition boot sector) i ovde se nalaze informacije o tome kako fajl sistem treba da koristi particiju. Druga se zove MFT (Master File Table) [5] koja čuva zapis o svim fajlovima i direktorijuma na fajl sistemu. Sledeći je niz drugih metafajlova koji pomažu u efikasnom radu fajl sistema nakon kojeg dolazi deo za fajlove.

NTFS sadrži nekoliko fajlova koji organizuju fajl sistem. Neki od njih su $MFT i $MFTMirr koji predstavljaju master fajl tabelu i njenu rezervnu kopiju koji služe da vode računa o tome gde se nalaze fajlovi na skladišnom medijumu. $Volume fajl sadrži informacije o particiji, konkretno identifikator objekta particije, oznaku particije i flegove. $BitMap čuva niz bitova gde svaki bit pokazuje da li je odgovarajući klaster u upotrebi (alociran) ili slobodan (dostupan za alokaciju). Pored ovih ima i drugih koji u zbiru daju efikasan rad fajl sistema i sveukupno dobro iskustvo upotrebe istog.

3.2 MFT zapis i atributi

Najbitniji metafajl za ovaj rad je $MFT koji se sastoji od MFT unosa. Svaki unos predstavlja jednu fajl strukturu na sistemu. NTFS je usvojio pristup da je sve fajl, te je i zapis o $MFT fajlu sačuvan u samom $MFT fajlu.



Slika 3. *Izgled MFT zapisa*

Na slici 3. se vidi prikaz jednog zapisa. Zaglavlje zapisa je fiksne dužine 42 bajta u okviru kojih je specificirana svojstva samog zapisa. Prvo predstavlja magične bajtove NTFS fajla a to su bajtovi „FILE“ ili „BAAD“. Osim ovih, čuvaju se podaci vezani za broj linkova na fajl sistemu, ofset do prvog atributa, identifikator atributa i tome slično. Svaki zapis je 1024 bajta i sav prostor izuzev zaglavlja je rezervisan za atribute.

Svaki MFT zapis može imati više atributa vezanih za sebe. Sami atributi imaju tip, opciono ime atributa i vrednost predstavljenu kao niz bajtova. Atributi mogu biti rezidentni ili nerezidentni u zavisnosti od broja bajtova koje njihova vrednost uzima. Ako vrednost atributa može stati u jedan MFT zapis, tada je atribut rezidentan.

Ako ne može, što je slučaj sa DATA atributom, tada se kao vrednost atributa čuva mapa alokacija koja pokazuje gde se na disku zapravo nalaze podaci vezani za vrednost ovog atributa i koje su oni dužine.

Pored već spomenutog DATA atributa postoje razni, od kojih je bitan STANDARD_INFORMATION atribut koji čuva informacije o tome kad je fajl napravljen, poslednji put otvaran, permisije i tome slično. ATTRIBUTE_LIST je atribut koji se umeće kada bi broj atributa prevazišao veličinu od 1024 bajta kako bi se zapis jedne datoteke raširio na više MFT zapisa.

3.3 Interoperabilnost

NTFS fajl sistem ima mogućnost i drajvere za upotrebu na različitim operativnim sistemima. Iako je napisan za Windows, i najčešće je u potpunosti kompatibilan unazad i unapred, postoje tehnički faktori koje treba uzeti u obzir prilikom mautovanja novih NTFS fajl sistema na starijim verzijama Windows operativnog sistema.

Verzije Linux kernela od 2.1.74 uključuju drajver napisan od strane Martina fon Luisa koji ima sposobnost čitanja NTFS particija. Tek od verzije 5.15 [6] Paragonov NTFS drajver je integrisan i podržava čitanje i pisanje običnih i kompresovanih datoteka.

4. OPORAVAK DATOTEKA

U računarstvu, oporavak datoteka je proces vraćanja obrisanih, nedostupnih, izgubljenih, oštećenih ili formatiranih podataka sa nekog sladišta ili medija kada se podacima u njima ne može pristupiti na običan način. Oporavak može biti potreban zbog fizičkih ili logičkih oštećenja na uređajima za skladištenje ili logičkih oštećenja na fajl sistemu koje sprečavaju da se fajl sistem mauntuje od strane operativnog sistema.

Logičke greške se dešavaju kada su tvrdi diskovi funkcionalni ali korisnik ili operativni sistem ne može povratiti ili pristupiti podacima sačuvanim na njima. Mogu nastati zbog izgubljenih partcija, oštećenja čipa, greške u radu firmware-a ili greške pri formatiranju.

Širok spektar neispravnosti može izazvati fizičku štetu na skladišnim medijumima. Ona se mogu desiti zbog ljudskih grešaka i prirodnih katastrofa. CD-ROM može imati oštećen metalni substrat ili sloj boje, tvrdi diskovi mogu patiti od mnogih mehaničkih neispravnosti rada glava za čitanje, motora ili se trake jednostavno mogu slomiti.

4.1 Metode oporavka podataka

Postoje dve metode oporavka podataka, oporavak podataka analizom metapodataka i oporavak podataka pretragom za poznate tipove podataka. Obe imaju svoje prednosti i mane i koriste se u različitim slučajevima.

Oporavak podataka analizom metapodataka je najčešće prvi način kojim program za oporavak podataka pokušava da oporavi podatke. Čitanjem metapodataka iz fajl sistema mogu se oporaviti fajlovi sa svojim originalnim imenom, putanjama, datumom kreiranja i poslednjim modifikacijama i permisijama. U ovom slučaju, program pokušava da pronađe informacije o fajlovima koji su dealocirani (prostor koji oni zauzimaju je označen kao slobodan za alociranje), a kada ih nađe, on prekopira sadržaj fajla na neko novo mesto, prateći instrukcije iz metapodataka za to gde se ti podaci nalaze. Ovaj metod ne radi ako je po disku puno pisano otkako su podaci izgubljeni ili ako je instanciran novi fajl sistem na disku jer se svi metapodaci tada brišu. Često daje zadovoljavajuće rezultate, ali u forenzičkim istragama se nekad pribegava i sofisticiranijoj drugoj metodi.

Iako uspešnija od prve metode po količini pronađenih podataka ova metoda ne može povratiti same metapodatke kao što su ime, putanja, permisije i tome slično. Zasniva se na ideji da se ceo disk skenira i traže se potpisi fajlova. Potpisi fajlova su magični bajtovi koji opisuju određen tip fajlova. Kako ovi potpisi nisu standardizovani ovo predstavlja veliki problem za ove alate. Postoje fajlovi koji imaju početni potpis, neki imaju samo krajnji potpis, dok treći imaju i početni i krajnji potpis. Veliki problem je oporavak fajlova koji su isprekidani na disku. Međutim i ovi problemi su donekle prihvatljivi jer se najčešće ovom metodom oporavlja dobar deo neophodnih informacija iako je ona dosta sporija od prve metode.

4.2 Stepen uspešnosti oporavka podataka

Faktori koji igraju ključnu ulogu za ishod oporavka podataka su kvalitet softvera za oporavak podataka, ispravnost akcija prilikom vršenja oporavka podataka i stepen oštećenosti skladišnog medijuma. Prva i druga stavka su u velikoj meri kontrolisane i svi uključeni imaju za cilj da oporavak prođe dobro. Poslednji faktor je izvan svačije kontrole i činjenica je da je on od presudnog značaja.

Dobra stvar je što se velika oštećenja na disku dešavaju jako retko.

Samo pisanje velike količine podataka na disk može da ga ošteti do te mere da se ne može oporaviti. U velikom broju slučaja se izgube samo metapodaci i moguće je primenom druge metode oporaviti veliku količinu nefragmentiranih podataka.

4.3 Faze oporavka podataka

Proces se sastoji od četiri faze. Prva faza je popravka tvrdog diska gde se on osposobljava za rad. Mogu se zameniti neispravne glave, motor i tome slično.

Druga faza je pravljenje kopije diska na drugi disk ili u fajl sa slikom diska, kako bismo osigurali da su podaci očuvani na originalu i da se oni neće menjati a dalji rad se vrši sa kopijom.

Treća faza je logički oporavak fajlova, particija, MBR i drugih struktura fajl sistema. Ako je disk logički neispravan, postoji nekoliko razloga za to i koristeći klon moguće je popraviti tabelu particija ili MBR kako bi se pročitala struktura podataka fajl sistema.

Poslednja faza je vezana za popravku oštećenih fajlova koji su povraćeni. Oštećenje može da se desi kada se podaci upišu na logički neispravan sektor na disku. U nekim slučajevima je moguće u potpunosti oporaviti oštećene fajlove.

5. ALAT ZA OPORAVAK PODATAKA

Cilj alata jeste da podmladi stara rešenja za oporavak podataka i da usput postigne idealno iste ili bolje rezultate upotrebom metode oporavka podataka na osnovu metapodataka.

Okvir samog projekta je postavljen tako da alat podržava samo NTFS fajl sistem sa kojeg će pokušati oporavak podataka. Takođe, alat može biti pokrenut isključivo sa Linux operativnih sistema koji ujedno predstavlja i očekivano razvojno okruženje.

Funkcionalnosti alata su detekcija fajl sistema, pri čemu ako fajl sistem nije NTFS alat trenutno prekida sa radom, parsiranje PBS (partition boot sector) kako bi utvrdio gde se nalazi MFT, čitanje i parsiranje MTF tabele i oporavak obrisanih podataka iz rezidentnih i nerezidentnih atributa. Pored toga, alat je namenjen za upotrebu putem komandne linije, gde nudi mali grafički interfejs za selekciju pronađenih fajlova koje treba povratiti.

5.1 Funkcionalni zahtevi

Korisnik treba da može da izabere opciju između oporavka podataka sa mauntovanog diska ili oporavka na osnovu slike diska. Alat mora da ima podršku za skeniranje diska/-slike. Ovo predstavlja mogućnost detekcije fajlova na disku bez njihovog oporavka. Alat mora da ima mogućnost da zaista oporavi podatke koji su pronađeni na osnovu skeniranja diska/slike upotrebom metode oporavka na osnovu metapodataka. Takođe, treba da može da oporavi njihovu tačnu strukturu u što je većem stepenu moguće. Kao sigurnosnu funkcionalnost alat mora da omogući log zapis svih akcija i koraka koje preduzima da ispuni svoj cilj.

Da bi ispunio ove funkcionalnosti, alat nakon parsiranja validira unesene parametre i ako su validni inicijalno pokušava da skenira disk i pronađe MFT i zapise u njemu. Nakon toga prolazi kroz iste i prezentuje korisniku one koji su označeni kao nealocirani, nakon čega korisnik treba da izabere koje fajlove želi da oporavi. Ako alat radi u režimu koji nije „dry-run“ režim, alat ih oporavlja.

6. ZAKLJUČAK

Iako je alat za oporavak podataka implementiran uspešno u okvirima koji su predviđeni specifikacijom, postoje određena ograničenja. Prvo ograničenje je starost skladišnog medijuma. Što je skladišni medijum stariji i što se više koristi, manje su šanse za savršen oporavak. Stepen degradacije se može umanjiti samo ako se zaplenjeni skladišni medijum odmah klonira i samim tim onemogući da se ošteti kao digitalni dokaz.

Implementirani alat je testiran na tvrdom disku (koji je magnetni) i na USB fleš disku (koji predstavlja fleš memoriju). Nisu primećene značajne razlike u brzini između ova dva skladišta podataka, ali to može biti zbog približno jednakih veličina MFT.

Stepen fragmentacije nije nešto što utiče na rad alata direktno. Alat je dizajniran tako da prolazi kroz sve mape alokacije vezane za fajl i spaja ih u jedan kontinualni fajl na odredišnom fajl sistem. Kod fragmentiranih fajlova, ponašanje alata slično je kao i ponašanje njemu sličnih.

7. LITERATURA

[1] B. Carrier, File System Forensic Analysis: What is a File System?, Massachusetts: Addison-Wesley Professional, 2005.

[2] B. Carrier, File System Forensic Analysis: File name category”, Massachusetts: Addison-Wesley Professional, 2005.

[3] B. Carrier, File System Forensic Analysis: NTFS Concepts”, Massachusetts: Addison-Wesley Professional, 2005.

[4] “File Allocation Table” – FAT16 F, FATX F. [Online]. Available: http://cs.williams.edu/~jannen/teaching/s19/cs333/readings/FAT/File_Allocation_Table-Wikipedia.pdf. [Accessed: Sep. 16, 2023]

[5] “NTFS Master File Table (MFT)” [Online]. Available: https://www.ntfs.com/ntfs-mft.htm. [Accessed: Sep. 27, 2023]

[6] “Linux kernel version history” Jan. 11, 2020. Available: https://en.wikipedia.org/wiki/Linux_kernel_version_history#Releases_5.x.y. [Accessed: Sep. 23, 2023]

Kratka biografija:

Nikola Milosavljević rođen je 4.7.1999. u Novom Sadu, Vojvodini. Pohađao je gimnaziju „Isidora Sekulić“. Fakultet tehničkih nauka upisao je 2018. na smeru Računarstvo i automatika. Master studije na studijskom programu Računarstvo i automatika upisuje 2022. godine.