Spašavanje oštećenih podataka

CIS-FER

Viktor Kvaternjak

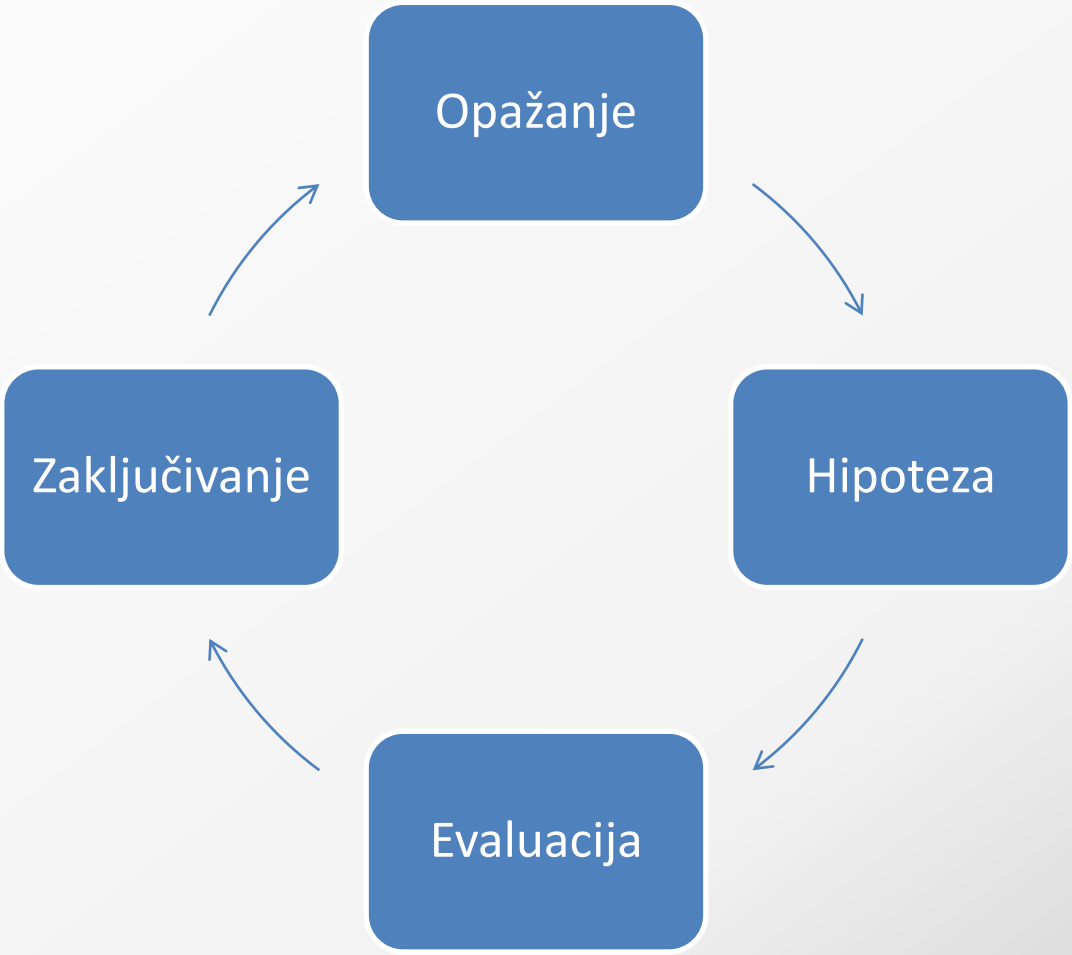
16. lipnja 2016.

Plan predavanja

- Cilj: vidjeti što sve možemo saznati o podacima i gdje se to može primijeniti

1. Kako sve podaci mogu biti oštećeni
2. Kako prikupiti podatke
3. Kako rekonstruirati podatke
4. Koje alate primijeniti

Digitalna forenzika

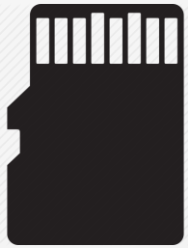
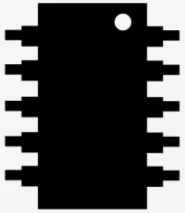
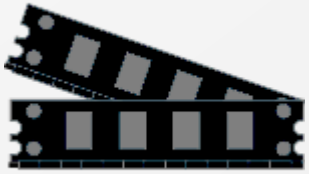


Oštećenje digitalnih medija

- Fizičko
- Logičko
  - RAID
  - Tablica particija
  - Datotečni sustav
  - Opisnik datoteke

Digitalni mediji




Optički mediji

- ISO9660, UDF
- Ograničen vijek trajanja – oko 10 godina
- Ogrebotine, atmosferski uvjeti
- Podaci za detekciju i oporavak od pogreške zauzimaju gotovo 1/3 površine

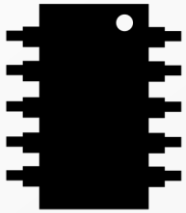
HDD

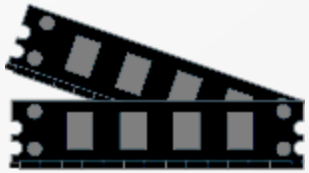
- NTFS/HFS+/EXT4
- 1 TB disk sadrži oko 90 GB podataka za oporavak od pogrešaka
- Fizički udarac, kvar na elektronici
- Otvaranje u čistoj sobi, izravno čitanje s ploča

Stalna memorija

- SSD, USB stickovi, memorijske kartice, NAND, NOR
- Degradacija čipa, električko oštećenje, nadogradnja firmwarea
- Oporavak: odlemljivanje čipa
- Enkripcija na sklopovskoj razini

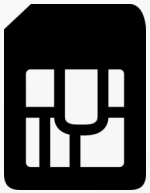
Privremena memorija





- Potrebno održavati stalno napajanje
- Akvizicija kroz OS
- “Cold boot” napad, napad preko sabirnica

Manje istražene tehnologije



- SIM, SCADA, IoT
- DFRWS: Software Defined Network Forensics Challenge, 2016

Logička oštećenja

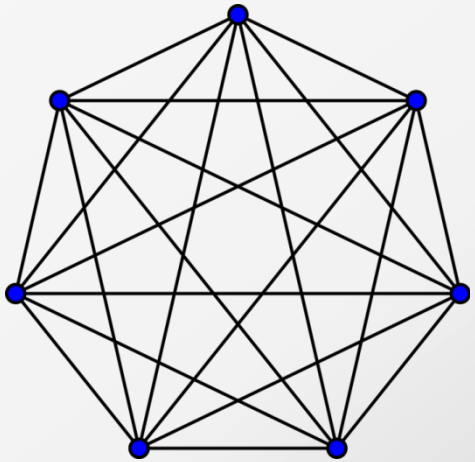
- Pod pretpostavkom da hardver ispravno radi, preostaju logička oštećenja
- RAID – ovisno o načinu rada, dizajniran za brzinu ili sigurnost (ili oboje)
- Tablica particija – potražiti karakteristični indikator početka
- Datotečni sustav – izazov
- Opisnik datoteke – samo oznaka brisanja

Postupak rekonstrukcije

- Enkripcija
- Izgubljeni metapodaci
- Fragmentacija

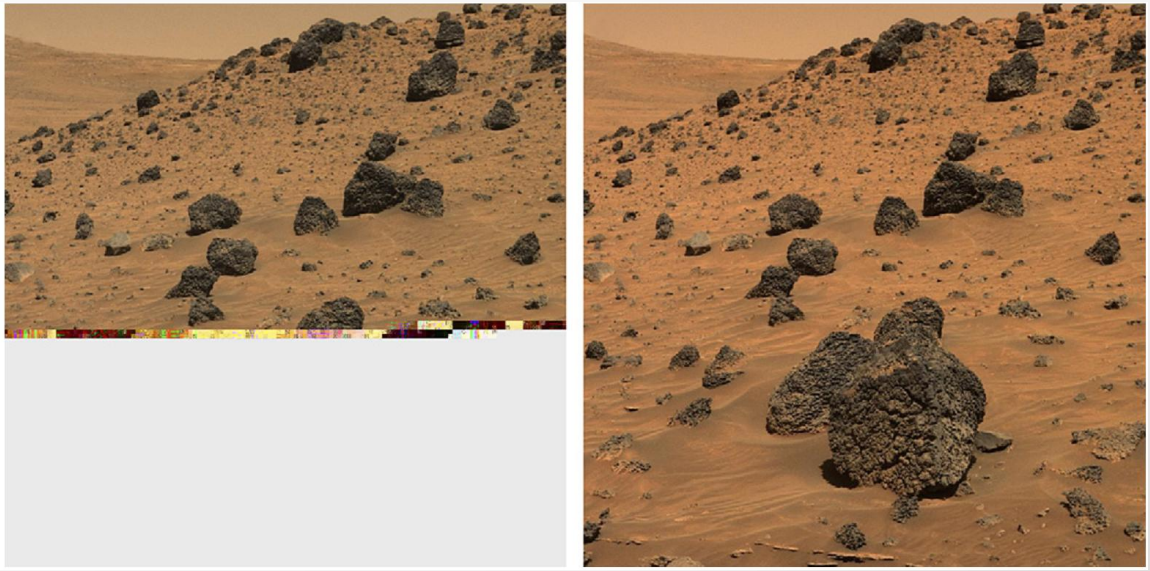
Rekonstrukcija uz strukturu datoteke

- Karakteristični početak (i završetak) datoteke
  - JPEG: `0xFFD8` do `0xFFD9`
  - PE: `MZ\0\0`
- Algoritam hamiltonovskog puta
- Algoritam paralelnog jedinstvenog puta



Rekonstrukcija uz strukturu datoteke

- Validacija strukture
  - Validator - PNG
  - Dekoder – ZIP, JPEG

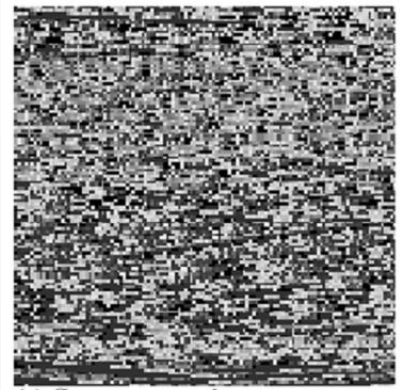


Opći rekonstruktor

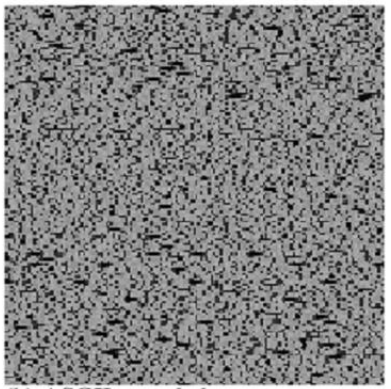
- Predobrada
  - Isključivanje otprije poznatih datoteka
- Razvrstavanje
  - Po vrsti datoteke
- Sastavljanje
  - Iskorištavanje karakteristika datotečnog sustava
  - Omjer fragmentacije manjih i većih datoteka

Identifikacija vrste datoteke

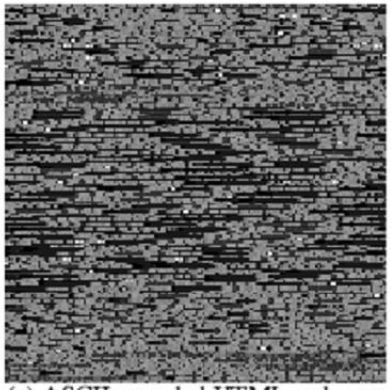
- Tehnike zasnovane na potpisima
- Statističke tehnike
- Tehnike umjetne inteligencije
- Ostali pristupi



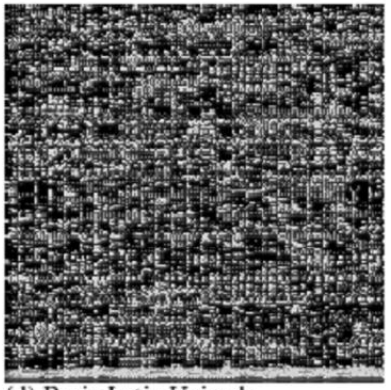
(a) C++ source code



(b) ASCII-encoded text



(c) ASCII-encoded HTML web page



(d) Basic Latin Unicode

Potpisi

- Baza definicija koja opisuje uzorke koji se često ponavljaju u pojedinoj vrsti datoteke
- Kriptografski sažetci poznatih datoteka ubrzavaju postupak
- Primjer: Yara potpisi, TrID

Statističke tehnike

- Karakteristike bloka otkrivaju vrstu podatka:
  - Koliko često se pojavljuje neki bajt
  - Kakva je distribucija bajtova
  - Entropija, standardna devijacija
  - N-grami

Neke entropijske mjere - NIST

- Frekvencijski test
  - 0101101101001000 (9,7)
- Test frekvencija po blokovima
  - 0101 1011 0100 1000 (2,2 1,3 3,1 3,1)
- Test najduljeg podniza jedinica u bloku
  - 0101101101001000 (2)

Tehnike umjetne inteligencije

- Nadogradnja statističkih tehnika
- Uvježbavanje modela na poznatim podacima
- Testiranje na nepoznatim podacima

Alati

- Foremost, Scalpel
- Undelete, Piriform Recuva
- Encase, FTK
- Entropy-based file classifier, Sceadan

Zaključak

- Područje aktivnog istraživanja
- Korisno za:
  - Istražitelje i forenzičare
  - Oporavak od pogrešaka
  - Nove primjene
- Kombinacija niza tehnika

Hvala!

Pitanja?

Grafički materijali na slajdovima 7 i 9 preuzeti su iz referencirane literature.