PRIRODOSLOVNO-MATEMATIČKI FAKULTET, SVEUČILIŠTE U ZAGREBU
MATEMATIČKI ODSJEK

Detekcija pakiranih datoteka

Jurica Miletić, Roko Kokan, Ante Sosa

April 29, 2018

1 UVOD

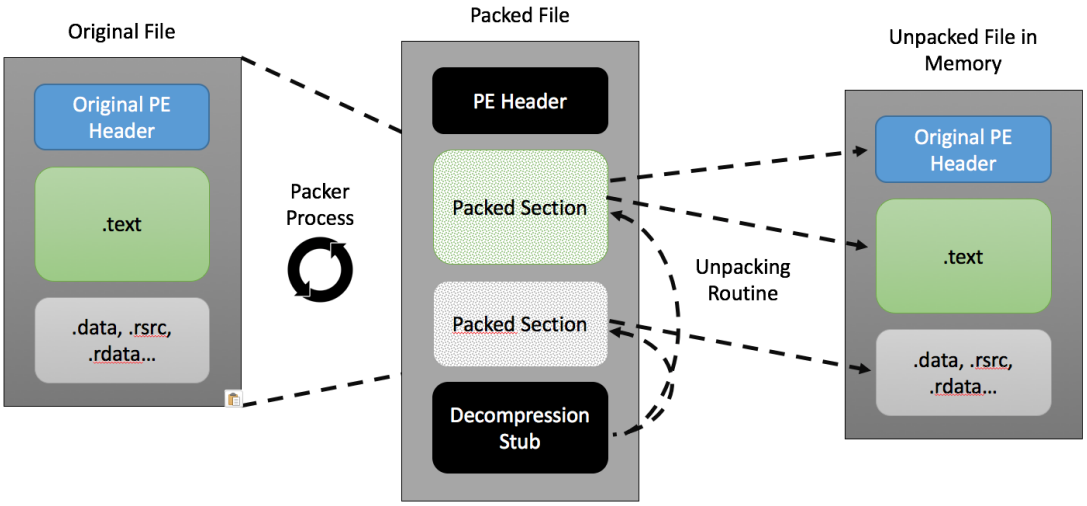
Pakiranje je metoda izmjene izvršnih datoteka bez mijenjanja njihove izvorne funkcionalnosti, ali na način da se datoteka zaštiti od reverznog inženjeringa, da se smanji veličina originalne izvršne datoteke, ili da se prikrije zlonamjeran izvršni kod. Pakiranje podrazumijeva izmjenu sadržaja datoteke te dodavanje instrukcija koje će prilikom izvršavanja taj sadržaj obnoviti.

Packeri modificiraju originalnu izvršnu datoteku na razne načine:

- Kompresijom podataka
- Enkripcijom podataka
- Prikrivanjem (obfuscate)
- Dodavanjem detekcije izvršavanja unutar debuggera ili virtualnog računala
- Modificiranjem raznih dijelova formata izvršne datoteke

Na Slici 1.1 je prikazan proces pakiranja.

U području računalne sigurnosti posebno su učestali packeri za Windows Portable Executable, tj. PE datoteke.



Slika 1.1:

1.1 Opis problema

PE datoteke su povijesno najčešći nositelji malicioznog koda u obliku virusa, ransomwarea, trojanskih konja, itd., te se packeri koriste da bi se taj maliciozni kod prikrio. Klasična statička analiza (bez pokretanja datoteka) koju provode antivirusi bazira se na potpisima. Oni nastaju tako da se prikupe primjeri nekog malwarea te se pronađe niz byteova specifičan za taj malware, koji se zatim traži prilikom skeniranja datoteka antivirusom.

Primjenom packera mijenja se sadržaj datoteke, zbog čega potpisi mogu prestati biti prisutni. Na taj se način iz jedne maliciozne datoteke može napraviti više različitih inačica. Packeri koji imaju nemalicioznu primjenu vrlo su rijetki u odnosu na packere koji se koriste za malware.

1.2 Skup podataka

Skup podataka bazirat će se na skupu raznovrsnih packera i dodatnim nepakiranim datotekama. Svaki primjer se sastoji od dva dijela: originalne datoteke i TitaniumCore izvještaja za tu datoteku.

Cilj zadatka je odvojiti samo pakirane datoteke od nepakiranih i raspakiranih, ali će sudionici za razvoj modela dobiti detaljnije informacije, poput generalnih vrsta packera. U podacima mogu biti varijante višestruko pakiranih datoteka, poput dvostruko pakirane datoteke koja se tijekom TitaniumCore procesiranja zatim jednom raspakira (i dalje se označava kao pakirana).

2 Cilj i hipoteze istraživanja problema

3 Pregled dosadašnjih istraživanja

Analiza entropije se koristi za uvid u sadržaj PE datoteka, primarno za detekciju kompresije i kriptografije tipično vezanih uz packere. Dio pristupa se bazira na analizi svojstava PE zaglavlja i jednostavnim klasifikatorima. Također, znatno poboljšanje točnosti detekcije malwarea dobiveno je koristeći stablo odluke za detekciju pakiranja.

Radovi koji koriste ovaj pristup kvalitetan su izvor analize značajnosti komponenata PE formata, ali su isto tako skloni pretreniranju zbog ograničenosti podataka s kojima rade.