Hash funkcie

Vrsta: Seminarski | Broj strana: 20 | Nivo: Fakultet organizacije i informatike, Varaždin

Uvod

HASH funkcija ili hash algoritam je funkcija za sažimanje i identificiranje podataka. Takav sažetak naziva se hash vrijednost ili jednostavno hash, a proces izračunavanja te vrijednosti naziva se hashiranje (eng. hashing). Hash funkcije koje su injekcija i surjekcija a time i bijekcija nazivaju se randomizirajuće funkcije. Domena hash funkcija u većini slučajeva veća je od kodomene pa nisu bijekcija (citat Wikipedia).

Osnovno svojstvo svih hash funkcija je da ako su dva izlaza dobivena istom funkcijom različita onda su i ulazi bili različiti. To znači da su hash funkcije determinističke tj. za identičan ulaz dobivamo identičan izlaz. Ukoliko hash funkcija nije surjekcija onda dva identična izlaza ne podrazumijevaju identične ulaze.

Hash funkcija od ulaza varijabilne veličine vraća znakovni niz fiksne dužine.

Slika: Za tri različita ulaza dobivamo različite izlaze koji su uvijek jednako dugi bez obzira na dužinu ulaza (izvor http://en.wikipedia.org/wiki/Hash_function)

Riječ «hash» u engleskom jeziku znači «sjeckati i miješati» (chop and mix), a hash funkcije rade upravo to – ulaz podijele na više dijelova koje zatim miješaju koristeći različite, pažljivo odabrane matematičke operacije. Izraz je prvi put upotrijebio Peter Luhn iz IBM-a u jednom dopisu s početka 1953. godine. U upotrebu je ušao desetak godina kasnije.

Kolizija hash sažetaka

Kolizija predstavlja situaciju u kojoj za dva različita ulaza hash funkcija izračuna identične izlaze. Ta situacija posljedica je fiksne dužine hash sažetaka, a varijabilne dužine ulaza. Ukoliko je hash sažetak uvijek iste dužine ima ih ograničen broj dok ulaza ima beskonačno mnogo. Jasno je da će svakom hash sažetku biti pridruženo beskonačno mnogo ulaza.

Unatoč koliziji hash funkcije i dalje zadovoljavaju definiciju funkcije koja kaže da funkcija svakom elementu domene pridružuje jedan i samo jedan član kodomene. Međutim takva funkcija nije bijekcija (jedan na jedan) jer više različitih članova domene ima pridružen isti član domene.

Potpuno izbjegavanje kolizije moguće je samo u ograničenom broju slučajeva kada su nam unaprijed poznati ulazi u funkciju i njihov broj. Hash funkcija kod koje ne može doći do kolizije naziva se savršena hash funkcija.

Upitno je koliko je kolizija uistinu velik sigurnosni problem hash funkcija iz nekoliko razloga. Kod korištenja novijih hash funkcija ni uz pomoć superračunala napad nije moguće izvesti u razumnom vremenu. Ukoliko i pronađemo koliziju primatelja kome je upućena poruka „pošaljite tajne podatke o slučaju na ja_sam@yahoo.com“ nećemo prevariti porukom „iohasfsdfopscvpb123132#$2“. Te dvije poruke imaju isti sažetak.

Hash funkcije bile bi ozbiljno ugrožene pronalaskom algoritma kojim bi bili u mogućnosti pronaći smislenu poruku koja ima isti sažetak kao izvorna poruka. Primjerice primatelja koji je trebao primiti originalnu poruku iz prethodnog primjera možemo prevariti porukom „pošaljite tajne podatke o slučaju na netko_drugi@gmail.com lkasd934m,kwet“435345fh%&7u“.