

Matematički fakultet
Univerzitet u Beogradu

Bezbedno traženje biomarkera korišćenjem hibridne homomorfne enkripcione šeme

Una Stanković
una_stankovic@yahoo.com

June 10, 2017



Uvod

Postavka problema

Postupak

Pretraga i izvlačenje podataka iz baze

Metod za bezbednu pretragu i izvlačenje podataka

Bezbedna pretraga biomarkera

Zaključak

Literatura



Motivacija

- ▶ brz razvoj tehnologija sekvenciranja genoma
- ▶ pristup velikim skupovima genoma
- ▶ veliki potencijal u razvoju biomedicinskih istraživanja
- ▶ primene u:
 - ▶ medicini,
 - ▶ biomedicinskim istraživanjima,
 - ▶ uslugama koje se direktno pružaju korisnicima,...



Figure : Idejni prikaz.



Osnovna ideja

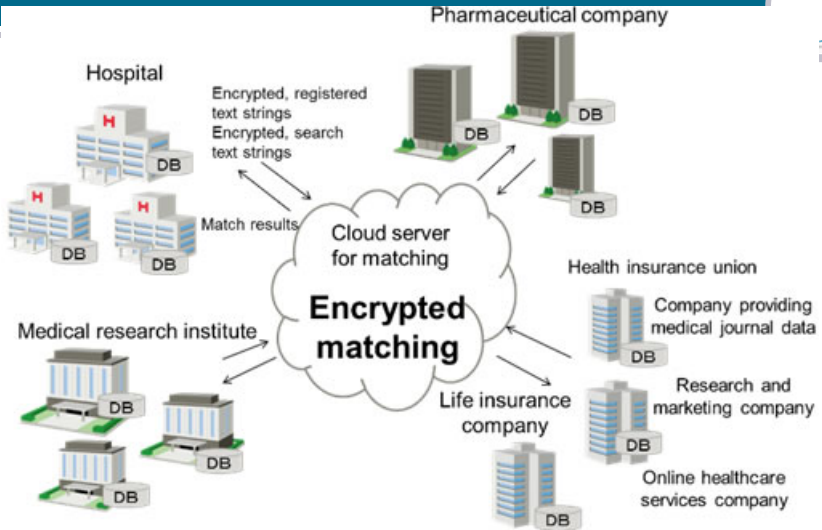
- ▶ izvodimo operacije nad enkriptovanim tekstom
- ▶ dobijamo enkriptovani rezultat
- ▶ dekriptujemo dobijeni rezultat
- ▶ konačni rezultat je isti kao da smo primenili operacije nad neenkriptovanim tekstom

Postavka problema

Zadatak



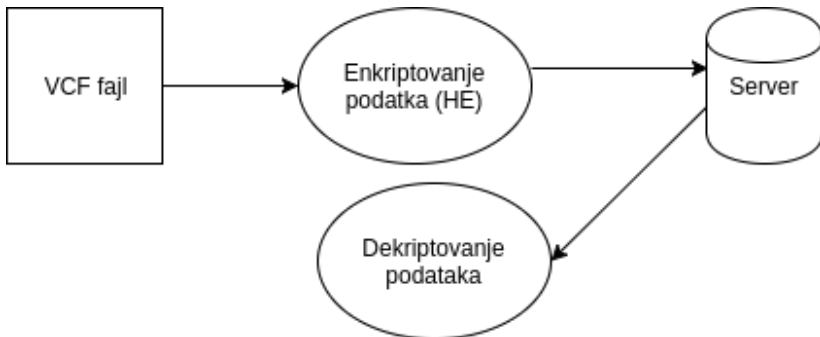
Zadatak je da se na bezbedan način izračuna verovatnoća genetskih bolesti kroz uparivanje skupa biomarkera sa enkriptovanim genomima koji se čuvaju u javnom "oblaku"(engl. cloud).





Uslovi

1. proces uparivanja mora biti izvršen korišćenjem homomorfne enkripcije
2. serveru se ne smeju otkriti informacije o bazi ili upitu





- ▶ Baza je skup n torki
- ▶ Svaka torak se sastoji iz (d_i, α_i)
- ▶ d_i pripadaju nekom domenu, α_i su odgovarajuće vrednosti atributa u prostoru običnog teksta
- ▶ Pojednostavljeni upit za pretragu : *odaberi α_i ako postoji indeks i takav da je $d_i = d$, inače nula.*



Metod za enkodiranje baze

- ▶ pogodan za efikasno izračunavanje jednakosti i izvlačenje podataka



$$DB(X) = \sum_i \alpha_i X^{d_i} \in \mathbb{Z}$$

- ▶ Postupak
 1. korisnik enkriptuje polinom sa javnim ključem
 2. šifrovani tekst se čuva na serveru
 3. u fazi ispitivanja upita sa nazivom d , korisnik enkriptuje X^{-d} sa simetričnom enkripcijom
 4. šifrovani tekst šalje ka serveru.



- ▶ VCF fajl sadrži informacije o genotipu: (ch_i, pos_i, SNP_s_i) tj. broj hromozoma, pozicije i sekvence SNP alela (koji moraju biti A, T, G ili C)
- ▶ Upit korisnika je, isto, triplet ovakvog oblika
- ▶ Cilj je da odredimo postoji li ili ne prisustvo odgovarajućeg biomarkera u fajlu iz baze
- ▶ n_{SNP} je maksimalni broj alela, koje kodiramo kao:

$$A \rightarrow 00, T \rightarrow 01, G \rightarrow 10, C \rightarrow 11$$


- ▶ Stavljamo bit 1 na početak stringa sa leve strane da označimo početnu poziciju
- ▶ Popunjavamo string dužine $l_{SNP} = 2n_{SNP} + 1$ i konvertujemo dobijeno u ceo broj, označen sa α_i



Sa razvojem bioinformatike i sve većim potrebama za čuvanjem podataka u oblaku, možemo očekivati porast u broju, kvalitetu i pouzdanosti algoritama za enkripciju podataka, kao i razvoj sve boljih i bržih mehanizama za pretragu genoma.



- ▶ Jung Hee Cheon, Miran Kim, Yongsoo Song, "Secure Searching of Biomarkers Using Hybrid Homomorphic Encryption Scheme"
<http://eprint.iacr.org/2017/294.pdf>

An abstract graphic consisting of multiple flowing, curved lines in shades of light blue and white. The lines originate from the left and curve towards the right, creating a sense of movement and fluidity. Some lines have small, glowing white dots or sparkles along their length. The overall shape is reminiscent of a stylized wave or a plume of smoke.

Hvala na pažnji!