BIOMETRIJSKI SUSTAVI – GREŠKE I RANJIVOSTI

Zlatko Sirotić
Istra informatički inženjering d.o.o., Pula
e-mail: zlatko.sirotic@iii.hr

SAŽETAK

Biometrija je fascinantno znanstveno i inženjersko područje, koje koristi znanja iz različitih drugih područja, kao što su biologija, matematička statistika, forenzika, psihologija, sigurnost i dr. U radu se daje prikaz biometrije neovisan od tehnologije, a naglasak se daje na greške i ranjivosti općeg biometrijskog sustava. Prvo se daju neki osnovni pojmovi iz područja biometrije. Dalje se prikazuje biometrijski sustav kao sustav za obradu podataka. U nastavku se prikazuju statističke osnove biometrije, pa biometrijska verifikacija i identifikacija. Na kraju se daje prikaz ranjivosti biometrijskog sustava.

Biometry is a fascinating scientific and engineering field which uses knowledge from a number of diverse scientific fields, such as biology, mathematical statistics, forensics, psychology, security etc.This paper shows biometry independent from technology and emphasizes the flaws and faults of the general biometric system. First we familiarize with general terms from this field. Later is shown biometric system as system for data processing. Then, statistical basis of biometry as well as verification and identification. System faults are shown at the end.

UVOD

Biometrija je fascinantno znanstveno i inženjersko područje, koje koristi znanja iz različitih drugih područja, kao što su biologija, matematička statistika, forenzika, psihologija, sigurnost i dr. Biometrija je vrlo mlada znanost, iako ima davne početke. Ozbiljna biometrijska istraživanja počela su šezdesetih godina prošloga stoljeća. Od sredine devedesetih godina prošloga stoljeća biometrija se značajno primjenjuje u praksi.

Biometrija koristi fiziološke ili ponašajne (eng.behavioral) karakteristike određene ljudske individue da bi ga automatski identificirala. Postoje različite biometrijske tehnologije. Najstarija je tehnologije ona otiska prsta, nastala oko 1960., komercijalizirana oko 1980. Slijedile su je mnoge druge biometrijske tehnologije, npr. mrežnice i šarenice oka, geometrije lica, geometrije ruke, otiska dlana, geometrije krvnih žila ruke ili prsta, glasa i dr.

Svaka od ovih tehnologija koristi različite biometrijske senzore i različite algoritme za uparivanje (eng.matching) biometrijskih podataka pročitanih senzorom i (prije) snimljenih biometrijskih podataka. No, bez obzira na različitost, sve ove biometrijske tehnologije imaju nešto zajedničko, a to je proces: ulaz podataka – obrada podataka – izlaz podataka. Pritom se biometrija obilato koristi oruđem matematičke statistike.

U ovom se radu daje prikaz biometrije neovisan od tehnologije, uglavnom na temelju literature [4], a naglasak se daje na greške i ranjivosti općeg biometrijskog sustava.

Prvo se daju neki osnovni pojmovi iz područja biometrije. Dalje se prikazuje biometrijski sustav kao sustav za obradu podataka. U nastavku se prikazuju statističke osnove biometrije, pa biometrijska verifikacija i identifikacija. Na kraju se daje prikaz ranjivosti biometrijskog sustava.

1. OSNOVNI POJMOVI

Kao i u drugim znanstvenim i inženjerskim područjima koja se brzo razvijaju, tako je i u području biometrije česta pojava nekonzistentnosti u korištenim pojmovima. Često je terminologija ovisna o konkretnoj biometrijskoj tehnologiji. Jedan od najboljih pokušaja za postavljanje standarda na tom području je ISO/IEC 19795-1, Information technology – Biometric performance testing and reporting. U nastavku se daju neki osnovni opći pojmovi i pojmovi o biometrijskim podacima iz tog standarda, na temelju literature [4]. Neki će se pojmovi naknadno uvesti u sljedećim točkama, kada bude posebno govora o njima.

Opći pojmovi

Biometrija (biometrics) je automatska identifikacija osobe na temelju njenih fizioloških (bioloških) ili ponašajnih karakteristika.

Biometrijska karakteristika (eng.biometric) je mjerljiva fiziološka (biološka) ili ponašajna karakteristika koja se koristi za prepoznavanje osobe. Biometrijska karakteristika ima ova četiri svojstva: svaka je osoba mora imati, treba biti dovoljno različita kod različitih osoba, treba ostati konstantna kroz vrijeme, mora biti mjerljiva kvantitativno (a ne samo opisno).

Biometrijski sustav (eng.biometric system) je cjelokupni sustav za autentikaciju (ovjeru autentičnosti osobe). Sastoji se od integriranog hardvera i softvera koji se koristi za uzimanje biometrijskog uzorka, obavljanje biometrijske verifikacije ili identifikacije i vraćanja rezultata.

Biometrijski modalitet (eng.modality) je vrsta biometrijske karakteristike. Npr., tri uobičajena biometrijska modaliteta su lice, otisak prsta, glas.

Multibiometrija (eng.multibiometrics) je automatska identifikacija osobe na temelju dvije ili više biometrijskih karakteristika. To mogu biti dvije (ili više) karakteristike koje pripadaju istom biometrijskom modalitetu, npr. otisci dva prsta, ili dvije (ili više) karakteristike koje pripadaju različitim modalitetima, npr. istovremena biometrija lica i otiska prsta. Multibiometrija ima dvije glavne prednosti pred pojedinačnom biometrijom. Kao prvo, osobe s poteškoćama, koje nemaju određenu karakteristiku, mogu se prijaviti na sustav koristeći drugu karakteristiku. Kao drugo, primjena multibiometrije može smanjiti greške kod verifikacije i identifikacije.

Biometrijski podaci

Uzorak (eng.sample) je instanca korisničke biometrijske karakteristike, prikupljena pomoću senzora. Uzorci su neprocesirani digitalni podaci. Uzorak se može konvertirati u (biometrijski) predložak ili se može upariti s predloškom kod verifikacije ili identifikacije.

Predložak (eng.template) je konvertirani i kompaktni biometrijski uzorak, pohranjen u digitalnom obliku, a kreiran je za vrijeme procesa zapisivanja (eng.enrollment) korisnika u biometrijski sustav. Postoje dvije prednosti u pohrani predloška, a ne originalnog uzorka. Prvo, predložak sadrži samo ona svojstva koja se koriste u konkretnom biometrijskom algoritmu. Budući da je "izvlačenje" tih svojstava iz originalnog uzorka procesorski dosta zahtjevno, to se radi samo jednom (kada se od uzorka stvara predložak), što je jako važno kod procesa identifikacije, gdje se jedan uzorak uspoređuje s mnogobrojnim predlošcima u bazi podataka. Drugo, predložak bi trebao sadržavati samo relevantne informacije, bez šuma i nepotrebnih informacija, što znači da bi trebao biti kompaktniji od originalnog uzorka.

Rezultat usporedbe (eng.match score) je numerička vrijednost koja predstavlja stupanj sličnosti dva biometrijska uzorka (odnosno tekućeg uzorka i predloška). Na temelju rezultata usporedbe i zadanog **praga** (eng.threshold) donosi se **odluka** (eng.match decision) o prihvaćanju ili neprihvaćanju. Npr. ako je rezultat usporedbe 50, a prag je 40, donosi se odluka o prihvaćanju uzorka kao valjanog (prihvaća se da je osoba koja je dala uzorak ona kojom se predstavlja). Kod prihvaćanja ili neprihvaćanja moguće su dvije greške: odbiti pravu osobu ili prihvatiti pogrešnu. Ispravno je prihvatiti pravu osobu i odbiti pogrešnu. Idealan biometrijski sustav bi uvijek radio bez greške. No, ne postoji idealan biometrijski sustav, tj. biometrijski sustavi uvijek rade dvije navedene greške. Na žalost, kako će se kasnije vidjeti, kada pokušamo smanjiti jednu grešku (odbijanje prave osobe), u pravilu utječemo na povećanje druge greške (prihvaćanje pogrešne osobe) i obrnuto.

Lista kandidata (eng.candidate list) je lista koja se može dobiti kod biometrijske identifikacije (ne kod verifikacije), a predstavlja skup predložaka čije je uspoređivanje s tekućim uzorkom dalo rezultate veće od praga, ili (u drugom slučaju) skup predložaka koji su nabolje rangirani kod usporedbe s tekućim uzorkom (npr. prvih pet), neovisno da li su prošli prag.

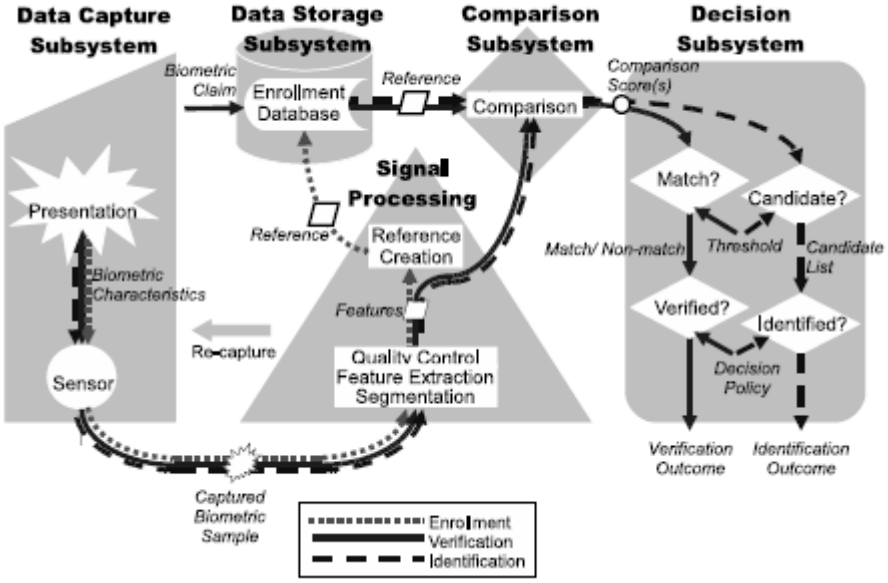
2. BIOMETRIJSKI SUSTAVI KAO SUSTAVI ZA OBRADU PODATAKA

U prethodnoj točki date su definicije nekih osnovnih pojmova, pa i pojmova (biometrijski) uzorak i (biometrijski) predložak. No, biometrijske podatke može se sa stanovišta njihovog spremanja podijeliti i na sirove (raw), token i predložak (template).

Sirovi podaci su drugi naziv za (prije spomenuti) biometrijski uzorak. To su podaci direktno preuzeti sa senzora, bez ikakvog procesiranja. **Token** se dobije iz sirovih podataka određenom minimalnom količinom procesiranja. Za razliku od tokena, **predložak** je jako različit od sirovih podataka, jer sadrži samo ona svojstva koja se koriste u konkretnom biometrijskom algoritmu. Kako je prije navedeno, to je prednost predloška (kompaktniji je i traži manje naknadno procesiranje). No, predložak ima i jednu manu – ako se u međuvremenu nađe neki bolji biometrijski algoritam, on se u pravilu ne može primijeniti na predložak, ali može na token (i na sirove podatke). Zato je u nekim slučajevima korisnije pohraniti token (umjesto predloška), a često se pohranjuju i token i predložak – predložak služi za brzi rad s tekućim algoritmom, a token se čuva za pretvorbu u novi predložak, nakon pojave boljeg algoritma.

Osim sirovih podataka, tokena i predložaka, u bazi podataka se često čuvaju i pomoćni podaci, koji se uobičajeno nazivaju **metapodaci**. To su podaci koji opisuju biometrijske podatke (npr. podatak da je osoba imala naočale), podaci o procesu uzimanja uzorka (npr. datum i vrijeme), te podaci o osobi (npr. kojeg je spola).

Kako je prikazano u [4], svi biometrijski sustavi mogu se opisati generalnim modelom prikazanim na sljedećoj slici:



Slika 1. Komponente općeg biometrijskog sustava; Izvor: [4]

Na slici su prikazani različiti podsustavi biometrijskog sustava, neovisni od određene biometrijske tehnologije. Točkastom, punom ili isprekidanom crtom prikazani su različiti tokovi, u ovisnosti od toga da li se radi o procesu zapisivanja (eng.enrollment) korisnika u biometrijski sustav, procesu verifikacije (da li je to stvarno osoba kojom se predstavlja?) ili procesu identifikacije (tko je ova osoba?).

Prvi korak je prikupljanje podataka (eng.Data Capture), gdje korisnik prezentira biometrijskom sustavu određenu biometrijsku karakteristiku, a sustav pomoću senzora uzima biometrijski uzorak. Kod najvećeg broja biometrijskih karakteristika, ulazni analogni signal pretvara se u digitalni. Digitalni signal se čisti od nepotrebnih smetnji i provjerava se njegova kvaliteta. Ukoliko kvaliteta nije dovoljna, postupak prezentacije i uzimanja uzorka se mora ponoviti.

Ako je uzorak kvalitetan, onda se pretvara iz sirovog oblika u referencu, koja se kod procesa zapisivanja zapisuje kao predložak, a kod procesa verifikacije i identifikacije se uspoređuje s predloškom. Kod kreiranja predloška, ponekad se primjenjuje tehnika kriptiranja podataka, kako bi se sačuvala privatnost (privatnost kao pojam informacijske sigurnosti) pohranjenih predložaka. Ako je riječ o verifikaciji ili identifikaciji, referenca se uspoređuje s pohranjenim predloškom.

Kod verifikacije, gleda se da li je rezultat usporedbe (eng.match score) veći od definiranog praga (eng.threshold). Ako je veći, onda je osoba uspješno verificirana (iako je to možda i greška – prihvaćena je kriva osoba, varalica (eng.impostor)), a ako ne, onda se osoba nije uspjela verificirati (iako i tada može biti greška – odbijena je osoba koja je prava (eng.genuine)).

Kod identifikacije, proces je složeniji, jer se ovdje mora proći cijela baza predložaka, svaki predložak se uspoređuje s tekućim uzorkom (odnosno referencom tog uzorka), te se na kraju uzme jedan predložak

koji ima najveći rezultat usporedbe, ili se uzme prvih nekoliko najbolje rangiranih predložaka (onih koji su prešli prag, ili neovisno od praga – postoje različite varijante).

Postoje još dvije komponente koje ovim dijagramom nisu prikazane, a to su transportni podsustav, koji na siguran način transportira podatke između prikazanih komponenti, te administracijski podsustav, koji omogućava npr. da se odredi prag za identifikaciju, da se unose ili brišu predlošci i sl. Također, nije prikazan podsustav za detekciju "životnosti" (eng.liveness) biometrijskog uzorka, tj. podsustav koji onemogućava korištenje lažnih biometrijskih uzoraka, npr. slike očiju / slike lica umjesto živog oka / lica, otiska prsta izvedenog gumom i sl.

3. MATEMATIČKA STATISTIKA I BIOMETRIJA

Matematička statistika je temelj biometrije. U ovoj točki se prikazuju neki najosnovniji pojmovi iz matematičke statistike i njene primjene u biometriji, na temelju literature [2], [4], [7].

3.1. Diskontinuirane slučajne varijable

Diskontinuirana (ili diskretna) slučajna varijabla (prema [7]) takva je varijabla *x* koja prima niz vrijednosti:

$x_1, x_2, \ldots$

ali svaku od njih s određenom vjerojatnošću

$p(x_1), p(x_2), \ldots$

pri čemu vjerojatnosti $p(x_i)$ zadovoljavaju jednakost

$\sum p(x_i) = 1$

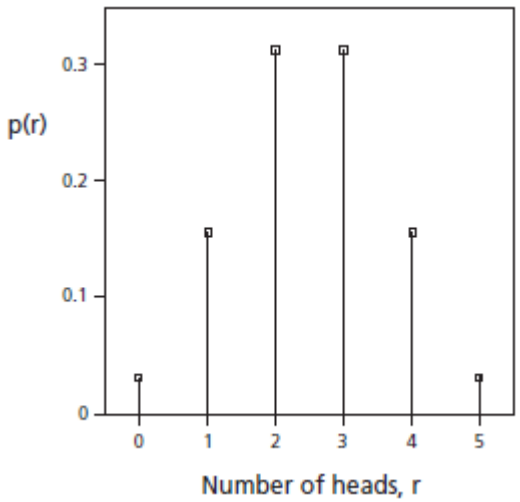
Primijetimo da niz vrijednosti koje poprima slučajna varijabla može biti i beskonačan.

Skup svih parova

$\{x_i, p(x_i)\}, \quad i = 1, 2, \ldots$

tvori **razdiobu (distribuciju) slučajne varijable** x. Zakon p(x) po kojem svakoj vrijednosti x_i pripada vjerojatnost $p(x_i)$ zovemo **funkcijom vjerojatnosti slučajne varijable** x.

Funkcija vjerojatnosti se može prikazati i grafički. Npr. na sljedećoj slici prikazana je funkcija vjerojatnosti slučajne varijable r koja predstavlja broj glava kod bacanja pet novčića (koji su pošteni, tj. kod kojih je ista vjerojatnost padanja glave i pisma):



Slika 2. Funkcija vjerojatnosti varijable r (broj glava kod bacanja pet novčića); Izvor: [2]

I kod diskontinuirane i kod kontinuirane slučajne varijable, najvažnije su njene sljedeće veličine: **matematičko očekivanje** (μ ili E(x)), **varijanca** (σ^2 ili V(x)) i **standardna devijacija** (σ). Umjesto pojma *matematičko očekivanje*, u praksi se često koristi pojam *aritmetička sredina*, ali je prvi pojam primjereniji za teoretske, a drugi za empiričke distribucije.

Kod diskontinuirane slučajne varijable, te se veličine definiraju ovako:

$\mu = \sum x_i \, p(x_i)$ (matematičko očekivanje)

$\sigma^2 = \sum (x_i - \mu)^2 \, p(x_i) = \sum x_i^2 \, p(x_i) - \mu^2$ (varijanca)

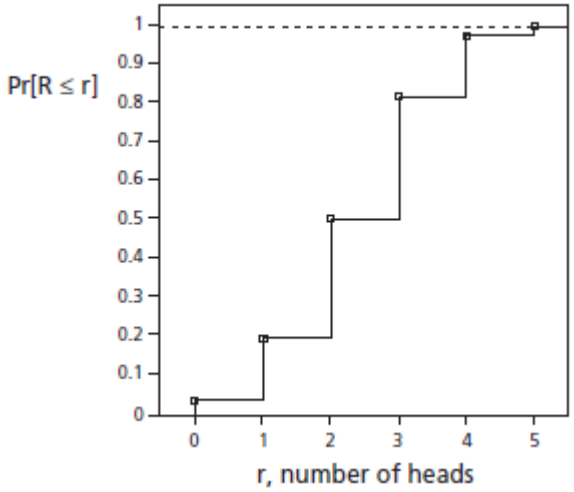
$\sigma = \sqrt{\sigma^2}$ (standardna devijacija).

Varijanca se može definirati i kao prosječno kvadratno odstupanje pojedinačnih vrijednosti od matematičkog očekivanja. Kvadrat se uzima kako se ne bi poništile pozitivne i negativne razlike, a lakše je raditi s kvadratom nego sa apsolutnim vrijednostima razlika. Kako bi se na neki način eliminirao utjecaj kvadriranja, uvela se standardna devijacija - kao (drugi) korijen varijance.

Osim funkcije vjerojatnosti, kod diskretnih slučajnih varijabli važna je **funkcija distribucije slučajne varijable** (ili funkcija kumulativne distribucije, eng. Cumulative Distribution Functions). Funkcija pokazuje kolika je vjerojatnost da slučajna varijabla x poprimi bilo koju vrijednost $\leq x_0$:

$F(x_0) = \sum_{x_i \leq x_0} p(x_i)$ tj. $F(x_0) = P\{x \leq x_0\}$

Na donjoj slici prikazana je funkcija distribucije koja odgovara prethodno prikazanoj funkciji vjerojatnosti slučajne varijable r (koja predstavlja broj glava kod bacanja pet novčića):



Slika 3. Funkcija distribucije varijable r (broj glava kod bacanja pet novčića); Izvor: [2]

3.2. Kontinuirane slučajne varijable

Diskontinuirane varijable služe za izučavanje diskontinuiranih obilježja elemenata nekog skupa ili niza pojava. Iako se može postaviti filozofsko pitanje da li (izvan matematike) uopće postoji beskonačnost, a pogotovo beskonačni kontinuum, u praksi radimo kao da kontinuirana obilježja stvarno postoje (npr. visinu ili težinu neke osobe smatramo kontinuiranim obilježjima) i za njih koristimo kontinuirane slučajne varijable. Kod kontinuirane slučajne varijable vjerojatnost ne pridružujemo pojedinoj vrijednosti varijable (jer bi ta vjerojatnost bila 0), već intervalu vrijednosti na brojevnom pravcu.

Prema [7], **funkcija vjerojatnosti kontinuirane slučajne varijable** x (ili funkcija gustoće vjerojatnosti, eng.Probability Density Function) je takva funkcija f(x) koja ima svojstva:

1. $f(x) \geq 0$ za svaki x iz domene funkcije [a, b] (a može biti $-\infty$, b može biti $+\infty$)

2. $\int_a^b f(x)\,dx = 1$ (površina ispod funkcije unutar domene [a, b] je 1)

3. $\int_{x_1}^{x_2} f(x)\,dx = P\{x_1 \leq x \leq x_2\}$ (površina ispod funkcije unutar domene $[x_1, x_2]$ jednaka je vjerojatnosti da slučajna varijabla poprimi vrijednost iz te domene).

Kod kontinuirane slučajne varijable, najvažnije veličine definiraju se ovako:

$\mu = \int_a^b x\,f(x)\,dx$ (matematičko očekivanje)

$\sigma^2 = \int_a^b (x - \mu)^2 f(x)\,dx = \int_a^b x^2 f(x)\,dx - \mu^2$ (varijanca)

$\sigma = \sqrt{\sigma^2}$ (standardna devijacija).

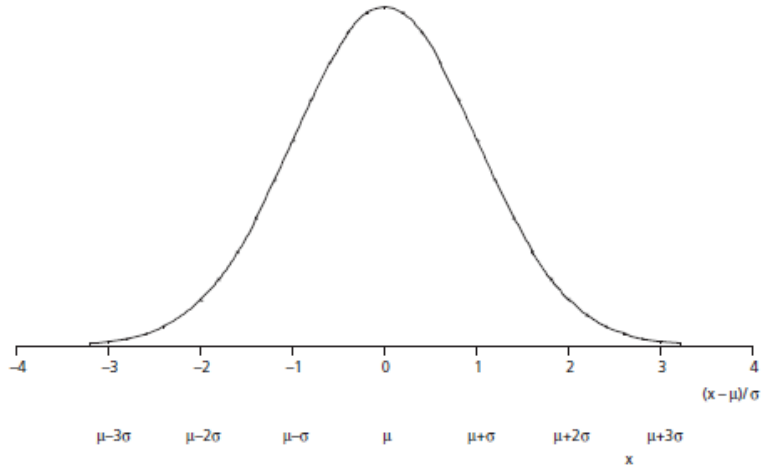
Jedna od najpoznatijih funkcija vjerojatnosti (kontinuirane varijable) je tzv. **normalna razdioba** (poznata i kao Gaussova krivulja, po matematičaru Gaussu), koja je važna po tome što mnoge druge razdiobe (diskontinuirane i kontinuirane) graniče prema njoj ako neki parametri rastu u beskonačnost. Posebno postoji tzv. **jedinična (ili standardna) normalna razdioba**, kod koje je matematičko očekivanje = 0, a standardna devijacija = 1. Funkcija vjerojatnosti jedinične (ili standardne) normalne razdiobe je

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

a funkcija vjerojatnosti općenite normalne razdiobe je

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

Na slici je prikazana **jedinična normalna razdioba**:



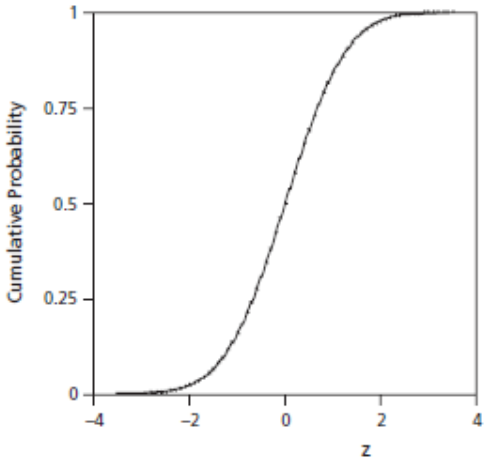
Slika 4. Normalna razdioba (najpoznatija funkcija vjerojatnosti); Izvor: [2]

Kao što kod diskontinuirane varijable svaka funkcija vjerojatnosti ima odgovarajuću funkcija distribucije, tako je i kod kontinuirane varijable.

Funkcija distribucije kod kontinuirane slučajne varijable definira se ovako:

$F(z) = \int_a^z f(x)\,dx = P\{a \le x \le z\}$

Funkcija distribucije jedinične normalne razdiobe prikazana je na sljedećoj slici:



Slika 5. Funkcija distribucije kod (jedinične) normalne razdiobe; Izvor: [2]

Ne treba raditi grešku (koja se ponekad radi u praksi) i smatrati da se sve pojave u prirodi ponašaju po zakonu normalne razdiobe. Tako i u biometriji ne smijemo automatski pretpostaviti da se neko obilježje pokorava zakonu normalne razdiobe. Za provjeru te hipoteze potrebno je izvršiti odgovarajuće testiranje, a često se za to koristi tzv. **χ^2-test**. Inače, χ^2-test se koristi i za testiranje podudaranja empirijskih rezultata u odnosu na druge teoretske razdiobe (npr. binomnu i Poissonovu), te za druge namjene, npr. usporedbu varijance uzorka i hipotetične varijance kod normalne razdiobe. Kod statističkih testiranja javljaju se, neminovno, dvije vrste grešaka, prikazane u nastavku.

3.3. Statistički testovi

U statistici, ako imamo funkciju vjerojatnosti slučajne varijable x koja (funkcija) ovisi o nekom parametru P (parametara može biti više, npr. parametri ne-jedinične normalne razdiobe su matematičko očekivanje i standardna devijacija, ali jedan parametar promatramo kao varijablu, a ostale kao konstantu) onda možemo postaviti hipotezu H_0 da je vrijednost tog parametra npr. P_0, te alternativnu hipotezu H_1, da je vrijednost parametra P_1 (moguće su i drugačije varijante postavljanja hipoteza). Odluku o tome da li prihvaćamo hipotezu H_0 ili H_1 donosimo na temelju testiranja uzorka, koji je uvijek konačan. Kod testiranja moguće su četiri situacije, dvije u kojima smo donijeli ispravnu odluku i dvije u kojima smo donijeli pogrešnu odluku:

Tablica 1. Moguće kombinacije istinitosti hipoteze H_0 i pravilnosti zaključivanja; Izvor: [7]

Hipoteza H_0	Istinita	Neistinita
Odbacuje se	**Greška 1.vrste (njena vjerojatnost je α)**	Pravilan zaključak
Prihvaća se	Pravilan zaključak	**Greška 2.vrste (njena vjerojatnost je β)**

Postoji još i pojam **jakost testa**, a računa se kao $1 - \beta$.

Budući da biometrijska verifikacija i identifikacija nisu deterministički procesi, tj. procesi koji mogu dati 100% pouzdan odgovor, već su to stohastički procesi (koji se proučavaju pomoću matematičke statistike), kod njih uvijek postoje te dvije vrste grešaka. Veličina tih grešaka može se izračunati pomoću matematičke statistike, pod pretpostavkom da je realna situacija dobro aproksimirana statističkim modelom.

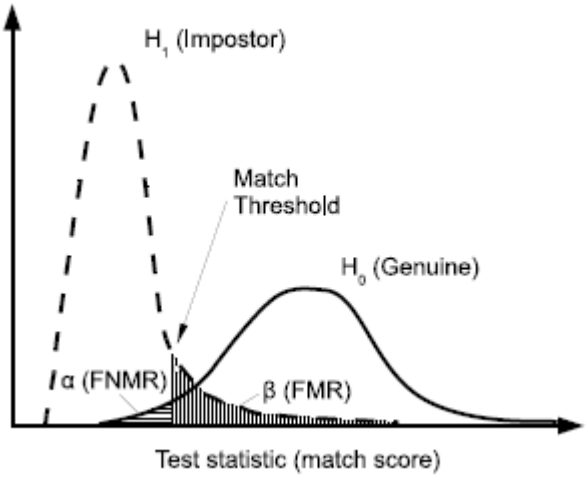
4. BIOMETRIJSKA VERIFIKACIJA

Kod biometrijske verifikacije želimo utvrditi da li je osoba koja je dala biometrijski uzorak zaista ta osoba kojom se predstavlja. Rezultat usporedbe biometrijskog uzorka s biometrijskim predloškom (match score) možemo shvatiti kao jednu konkretnu vrijednost slučajne varijable. (Napomena: ova je točka napravljena na temelju točke 7.1 iz [4].)

4.1. Biometrijska hipoteza

Postavljamo hipotezu H_0 da je riječ o pravoj osobi. Ako je rezultat usporedbe veći ili jednak od zadanog praga (eng.threshold), prihvaćamo hipotezu H_0, a ako je manji, odbacujemo hipotezu H_0 i prihvaćamo hipotezu H_1 (smatramo da je osoba lažna).

No, primijetimo da je funkcija vjerojatnosti (kontinuirane) slučajne varijable (koja predstavlja rezultat usporedbe) jedna u slučaju da je osoba prava, a druga u slučaju da je osoba lažna. Na sljedećoj slici prikazane su te dvije funkcije vjerojatnosti, te su prikazane vjerojatnosti grešaka 1. i 2.vrste (α i β):



Slika 6. Funkcije vjerojatnosti ako su istinite hipoteze H_0, odnosno H_1; Izvor: [4]

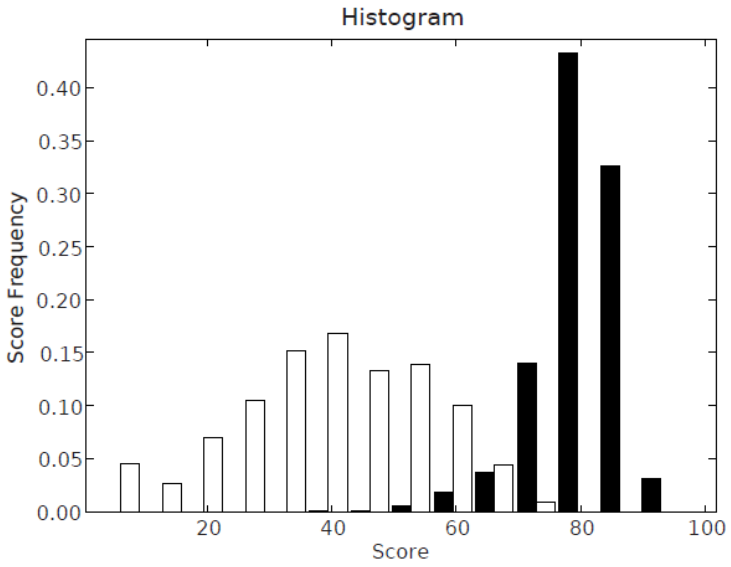
Ako je točna hipoteza H_0 (osoba je prava), njoj pripada odgovarajuća funkcija vjerojatnosti na slici desno, a površina ispod te krivulje koja se nalazi lijevo od zadanog praga (Match Threshold) predstavlja vjerojatnost greške 1.vrste (α). U biometriji se vjerojatnost te greške (umjesto α) naziva **FNMR (eng.False Non-Match Rate),** slobodno prevedeno - stopa pogrešnog odbacivanja.

Ako je točna hipoteza H_1 (osoba je lažna), njoj pripada odgovarajuća funkcija vjerojatnosti na slici lijevo, a površina ispod te krivulje koja se nalazi desno od zadanog praga predstavlja vjerojatnost greške 2.vrste (β). U biometriji se vjerojatnost te greške (umjesto β) naziva **FMR (eng.False Match Rate)**, slobodno prevedeno - stopa pogrešnog prihvaćanja.

Treba napomenuti da u praksi npr. FNMR stopa od 3% najčešće ne znači da će svaki (pravi) korisnik imati problema kod prijave u oko 3% slučajeva, već je puno vjerojatnije da će oko 3% (pravih) korisnika imati problema u skoro 100% slučajeva!

4.2. Procjena grešaka kod biometrijske verifikacije

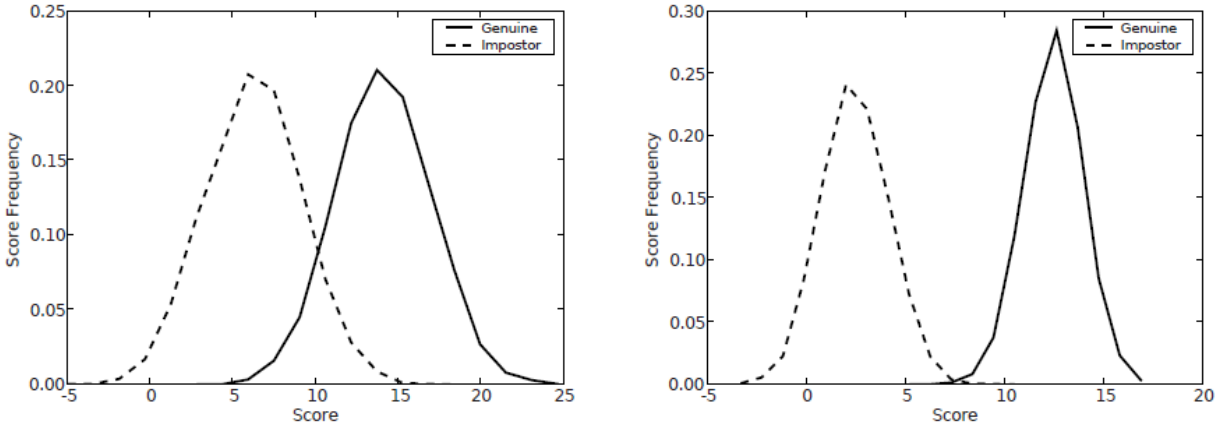
Na prethodnoj slici (6.) imali smo razdiobe (funkcije vjerojatnosti) koje su bile "glatke", a jedna od njih, ona za pravu osobu, liči na normalnu razdiobu. Međutim, u biometrijskoj praksi rijetko postoje unaprijed zadane teoretske razdiobe (pogrešno je pretpostaviti da se sve zbiva npr. po normalnoj razdiobi), već se stvarne razdiobe moraju utvrditi empirijski, pokusom. Pokus se radi na temelju postojeće baze predložaka i/ili sirovih biometrijskih uzoraka. Što je više elemenata u bazi podataka i što se više njih uzme u pokus, to će empirički rezultati pokusa biti bliži stvarnom stanju. Na temelju pokusa dobije se **histogram**, zapravo dva – za verifikaciju prave i lažne osobe, kao na sljedećoj slici:



Slika 7. Histogram rezultata usporedbe (match score).
Crni pravokutnici su rezultati usporedbe prave osobe, a bijeli pravokutnici lažne; Izvor: [4]

Histogrami nisu kontinuirane razdiobe. No, na temelju njih mogu se napraviti kontinuirane razdiobe, koje mogu biti više ili manje "glatke" (bez oštrih vrhova).

Kad tako dobijemo kontinuirane razdiobe, vrlo je važno vidjeti na koji se način one preklapaju. Dva različita slučaja preklapanja možemo vidjeti na sljedećoj slici:



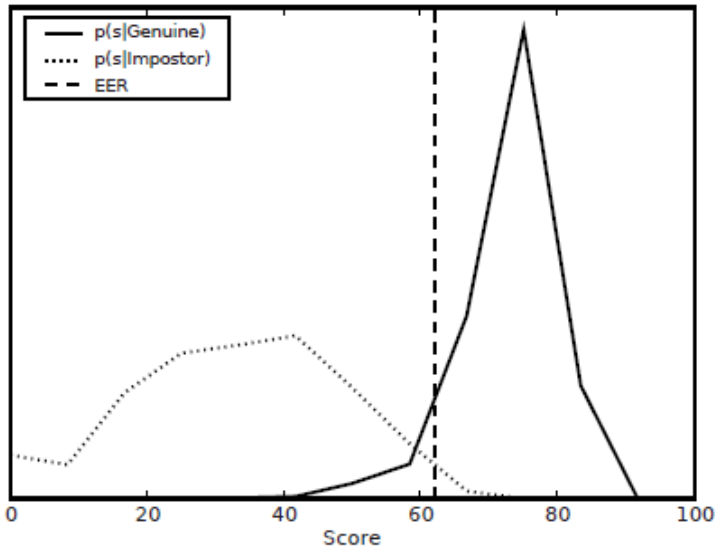
a) b)

Slika 8. Veza između preklapanja razdioba rezultata usporedbe i grešaka.
(a) Preklapanje je veliko, FNMR i FMR su veliki.
(b) Preklapanja skoro nema, FNMR i FMR su vrlo mali, praktički zanemarivi; Izvor: [4]

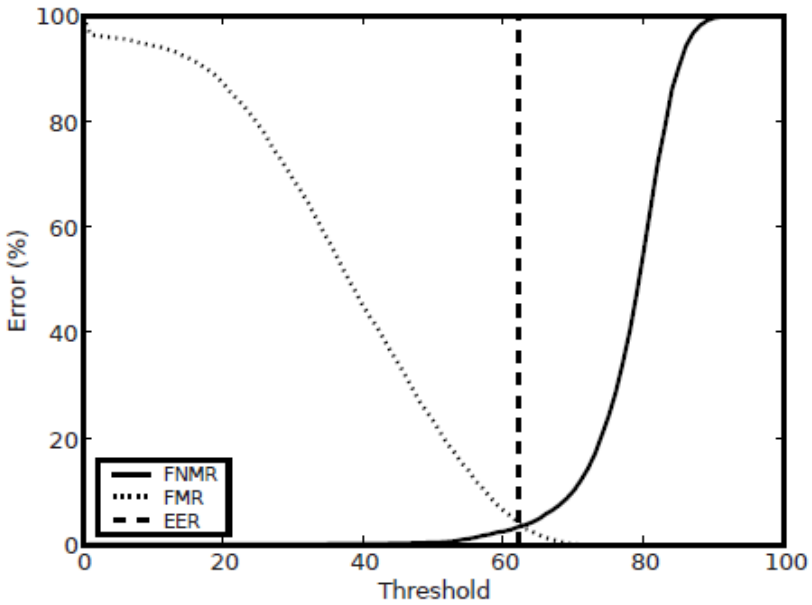
U slučaju (a) preklapanje između obje razdiobe je veliko, tako da smo u dilemi koju veliku grešku napraviti. Ako odaberemo niski prag (više ulijevo) onda ćemo dobiti malu FNMR grešku (pogrešno odbacivanje), jer će biti obuhvaćena skoro cijela površina desne krivulje (za pravu osobu), ali će zato biti vrlo velika FMR greška (pogrešno prihvaćanje), jer će biti obuhvaćen i dobar dio lijeve krivulje (njena desna strana), što znači da će biti malo onih pravih osoba koje se neće prijaviti na sustav, ali će biti puno lažnih osoba koje će se moći prijaviti na sustav. Ako odaberemo suprotno, tj. da prag postavimo visoko (više udesno), onda ćemo imati situaciju da će biti malo lažnih osoba koje će se moći prijaviti na sustav (mali FMR), ali će biti puno slučajeva da se prava osoba neće moći prijaviti (veliki FNMR). Nažalost, ovakav izbor između dvije loše mogućnosti nije rijedak u praksi. U slučaju (b) imamo idealnu situaciju. Razdiobe se skoro ne preklapaju, što znači da možemo istovremeno imati i mali FNMR i mali FMR.

Često nas zanima da vidimo kod kojeg praga (threshold) je FNMR = FMR. Takav prag naziva se **EER (eng.Equal Error Rate)**, slobodno prevedeno – stopa jednakih grešaka, jer su kod njega greške međusobno jednake. Treba naglasiti da to *nije* onaj prag kod kojeg se sijeku dvije razdiobe (ona za pravu i lažnu identifikaciju). No, iz razdioba se mogu dobiti dvije krivulje čije sjecište je zaista na EER pragu. Uobičajeno je da se razdiobe prvo "preskaliraju", tako da se raspon pragova (koji je npr. na slici 8. između –5 i 20) svede na raspon od 0 do 100 (to ne znači 0% do 100%, moglo je biti npr. i od 0 do 10).

Takve razdiobe prikazuje sljedeća slika, pod (a). Na temelju tako pripremljenih razdioba mogu se napraviti **FNMR i FMR krivulje**, na sljedećoj slici su prikazane pod (b):



(a) Genuine and impostor probability distribution functions



(b) The false match rate and false non-match rate for the distributions in (a)

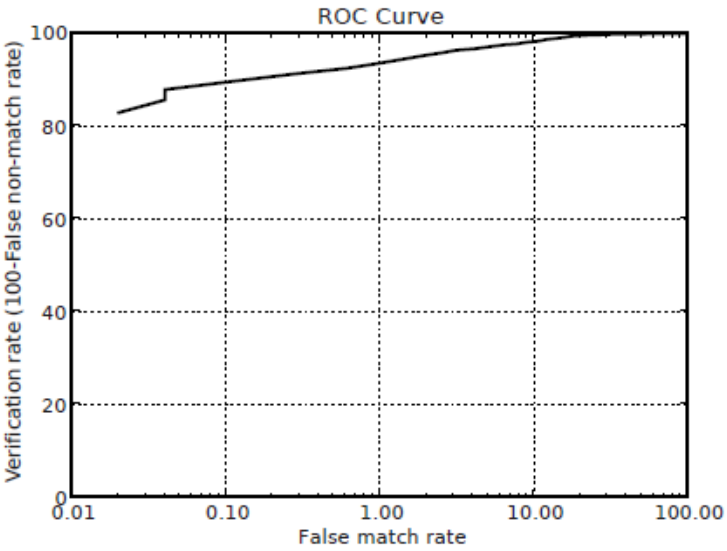
Slika 9. (a) Razdioba kod prave osobe (Genuine) i lažne (Impostor)
(b) Vrijednost FNMR i FMR u ovisnosti od praga (threshold); Izvor: [4]

FNMR / FMR krivulje pokazuju (na osi y) kolika je vrijednost (postotak) FNMR / FMR greške, za zadanu vrijednost praga. Kako je već rečeno, te se krivulje uvijek sijeku kod EER praga, za razliku od razdioba, koje se općenito ne sijeku u EER, kao što prikazuje i slika (a).

4.3. Evaluacija performansi sustava za biometrijsku verifikaciju

Kako bi se u praksi mogla lakše odabrati odgovarajuća kombinacija FNMR i FMR greške, ili kako bi se lakše uspoređivala dva biometrijska sustava s obzirom na njihovu preciznost, rade se tzv. **ROC krivulje (eng. Receiver Operating Characteristic)**, koje prikazuju odnos između FMR stope i tzv. verifikacijske stope (eng.verification rate) koja se dobije kao 100 – FNMR.

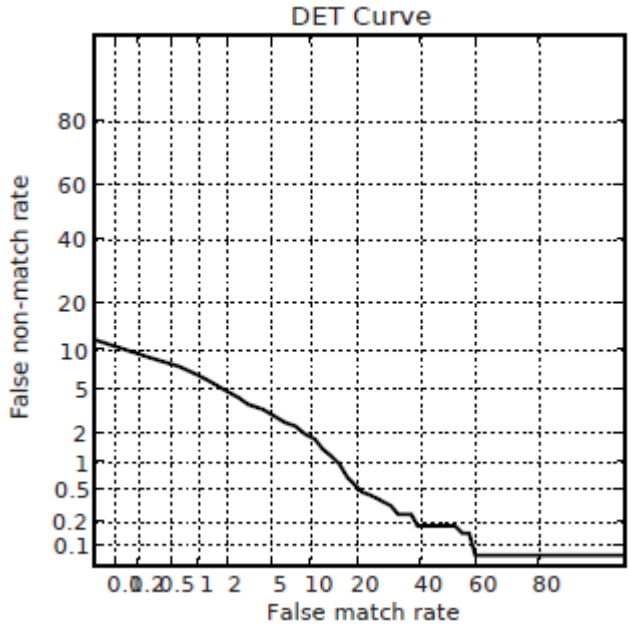
FMR stopa se prikazuje na osi x, uobičajeno u logaritamskom mjerilu, a verifikacijska stopa se prikazuje na osi y, pri čemu može biti isto u logaritamskom mjerilu, ili u linearnom mjerilu, kao na sljedećoj slici:



Slika 10. ROC krivulja; Izvor: [4]

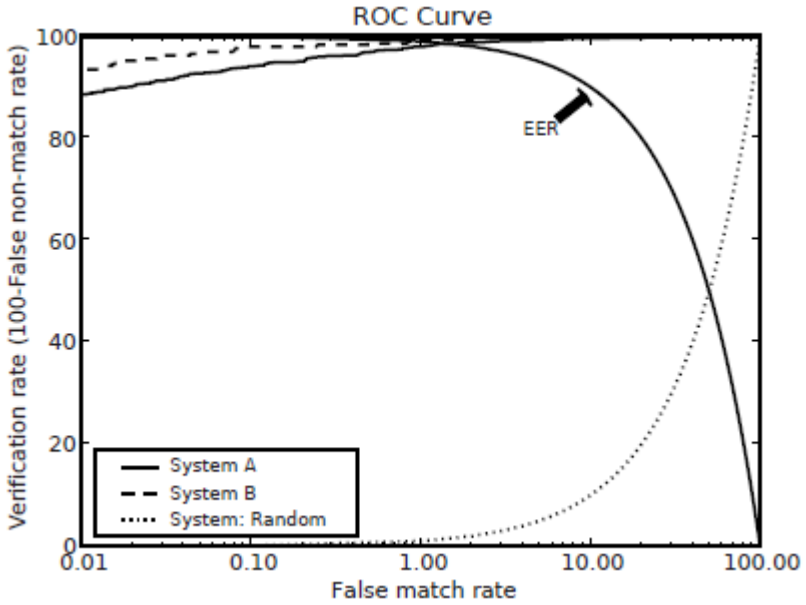
Iz slike se može vidjeti da je npr. kod FMR od 0,1% verifikacijska rata oko 90% (odnosno FNMR je oko 10%). Primjećuje se da je ROC krivulja uvijek rastuća, jer za veću verifikacijsku ratu (os y) treba prihvatiti veću FMR grešku (os x). Idealna bi bila ROC krivulja kod koje bi za bilo koju vrijednost FMR, verifikacijska rata bila 100% (dužina paralelna sa osi x).

Ponekad se na os y umjesto verifikacijske stope stavlja FNMR stopa, pa se dobiva tzv. **DET krivulja (eng.Detection Error Trade-off)**. Ona nije neka nova vrsta krivulje, već jedna varijanta ROC krivulje. Kao i kod prethodne ROC krivulje, često je os x prikazana u logaritamskom mjerilu. Druga je varijanta da se obje osi prikazuju u mjerilu na temelju normalne razdiobe (kao na slici 11.), čime se, u slučaju da je razdioba zaista normalna, dobiva DET krivulja kao pravac. DET krivulja je padajuća, jer većoj FMR odgovara manja FNMR greška:



Slika 11. DET krivulja; Izvor: [4]

Sljedeća slika pokazuje na koji se način ROC krivulja može koristiti za usporedbu performansi dvaju biometrijskih sustava. Vidi se da je sustav B (isprekidana crta) bolji od sustava A (puna crta), jer za isti FMR ima bolju verifikacijsku stopu (tj. manji FNMR):



Slika 12. ROC krivulje kod tri hipotetska biometrijska sustava; Izvor: [4]

Na slici je ROC krivulja (prikazana točkasto) i za treći hipotetski biometrijski sustav (Random), koji radi slučajno, tj. bilo koja osoba (prava ili lažna) se uspijeva prijaviti u 50% slučajeva. Kod takvog sustava je FMR uvijek jednak verifikacijskoj stopi. Njoj zrcalna krivulja (označena sa EER) sastoji se od točaka kod kojih je FMR = FNMR.

5. BIOMETRIJSKA IDENTIFIKACIJA

U prethodnoj točki promatrao se problem biometrijske verifikacije, tj. utvrđivanja da li je osoba koja se prijavila zaista prava osoba. Drugim riječima, verifikacija odgovara na pitanje "Da li je to zaista ta osoba kojom se predstavlja?". Za razliku od toga, biometrijska identifikacija odgovara na pitanje "Tko je ova osoba?". (Napomena: ova je točka napravljena na temelju točke 7.2 iz [4].)

5.1. Biometrijska identifikacija je složenija od biometrijske verifikacije

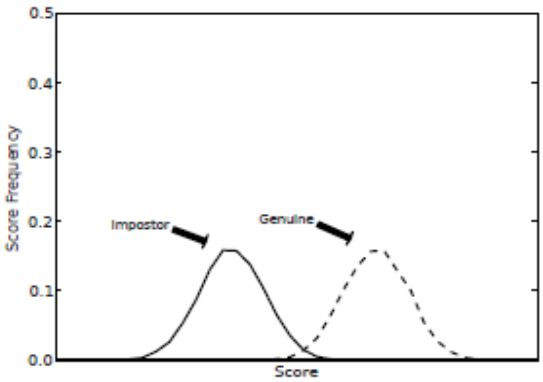
Moguće je da osoba koja se prijavljuje u biometrijski sustav (sa ili bez znanja da to radi, npr. kod nadzora ulica kamerama osoba ne zna da ju identificiraju) uopće nije upisana u biometrijsku bazu. To još više otežava postupak identifikacije, koji je i inače puno složeniji nego postupak verifikacije.

Verifikacija je 1:1 usporedba između biometrijskog uzorka i predloška iz biometrijske baze. Identifikacija je 1:N usporedba, jer se biometrijski uzorak u općem slučaju mora usporediti sa svakim predloškom iz baze. U biti, verifikacija se može shvatiti kao poseban slučaj identifikacije, kod kojeg sistem ima samo jedan predložak u bazi.

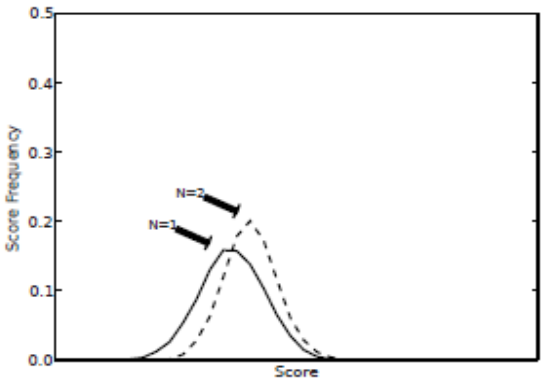
Identifikacijski sustavi najčešće rade tako da operator pogleda rezultate koje mu je predočio biometrijski sustav. Sustav daje rezultate na dva uobičajena načina. Jedan je način da se operateru prikažu sve osobe čiji je prag usporedbe veći od zadanog praga, a drugi je način da se operateru prikaže prvih nekoliko osoba koje imaju najveći rezultat usporedbe.

Identifikacijski sustav je bitno različit od verifikacijskog sa stanovišta preciznosti. I kod verifikacije, preciznost je ovisila o preklapanju razdioba za pravu i lažnu osobu. No, kod identifikacije će preciznost padati sa porastom veličine biometrijske baze. Pretpostavimo da se na sustav prijavljuje lažna osoba i da sustav kod identifikacije vraća samo jednu osobu (tzv. Rang 1 identifikacija, što ne smanjuje općenitost razmatranja). U slučaju da u bazi postoji samo jedan predložak, razdioba za lažnu osobu će biti ista kao kod verifikacije, tj. javljat će se iste greške zbog preklapanja razdiobe za lažnu osobu s onom za pravu osobu.

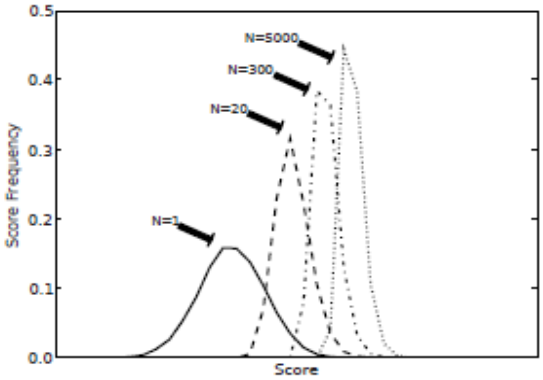
Međutim, kako raste broj predložaka, tako se razdioba za lažnu osobu pomiče udesno, što znači da se povećava preklapanje sa razdiobom za pravu osobu, pa se povećavaju i greške. To se lijepo vidi na sljedećoj slici (13.):



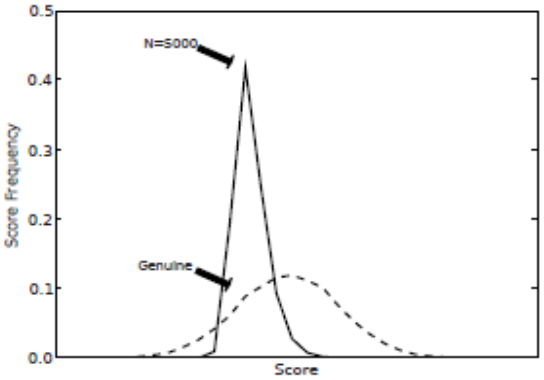
(a) The original genuine and impostor score distributions.



(b) The Rank 1 impostor score distributions for databases of size 1 and 2.



(c) Rank 1 impostor score distributions for a range of database sizes.



(d) The Rank 1 impostor score distribution for a database of size 5000, along with the genuine score distribution.

Slika 13. Efekt veličine baze podataka na Rang 1 razdiobu lažne osobe; Izvor: [4]

Pod (a) vide se originalne razdiobe za pravu i lažnu osobu.

Pod (b) vidi se kako se razdioba za lažnu osobu pomaknula udesno i suzila kada je broj predložaka u bazi povećan sa 1 na 2 (napomena: pod (b) i (c) nisu prikazane razdiobe za pravu osobu, već samo za lažnu).

Pod (c) prikazano je daljnje pomicanje udesno i sužavanje razdiobe za lažnu osobu, kako se povećava broj predložaka u bazi.

Pod (d) vidi se razdioba za lažnu osobu ako je broj predložaka 5000, te razdioba za pravu osobu.

Dakle, jasno se vidi kako je relativno malo preklapanje između razdiobi za pravu i lažnu osobu, ali sa samo jednim predloškom u bazi, preraslo u veliko preklapanje ako je broj predložaka u bazi 5000.

5.2. Procjena grešaka kod biometrijske identifikacije

Zbog prethodno navedenih razlika između identifikacije i verifikacije, razumljivo je da je različita i metrika za mjerenje performansi (prije svega preciznosti) sustava za identifikaciju u odnosu na sustav za verifikaciju.

Kao prvo, FNMR stopa (pogrešno odbijanje) kod identifikacijskog sustava može se dobiti na temelju sljedećeg razmatranja. Pretpostavimo da u bazi veličine N za svakog korisnika postoji *m* predložaka (naravno, u praksi je najčešće m = 1). Tada se $\mathbf{FNMR_N}$ (tj. FNMR kod identifikacije nad bazom veličine N) može logičkim razmatranjem procijeniti kao:

$FNMR_N = FNMR^m$

Dakle, $FNMR_N$ uopće ne ovisi od N, već ovisi od m. Što više predložaka korisnik ima u bazi, to je manja mogućnost pogrešnog odbijanja (jer je $x^m < x$ ako je $x < 1$ i $m > 1$).

FMR_N (FMR, stopa pogrešnog prihvaćanja, ali kod identifikacije nad bazom veličine N) može se procijeniti na temelju sljedećeg razmatranja. Vjerojatnost da nema pogrešnog prihvaćanja u bazi s jednim predloškom je (1- FMR), a vjerojatnost da nema pogrešnog prihvaćanja u bazi sa N predložaka je $(1- FMR)^N$. Prema tome, vjerojatnost pogrešnog prihvaćanja je:

$FMR_N = 1- (1- FMR)^N$

Dakle, uz zadani FMR, što je veći N, to će biti manji $(1- FMR)^N$ pa će biti veći FMR_N. Također, uz zadani N, što je veći FMR, to će biti manji $(1- FMR)^N$ pa će biti veći FMR_N. Prema tome, FMR_N možemo smanjiti tako da smanjimo FMR ili/i N (ako možemo).

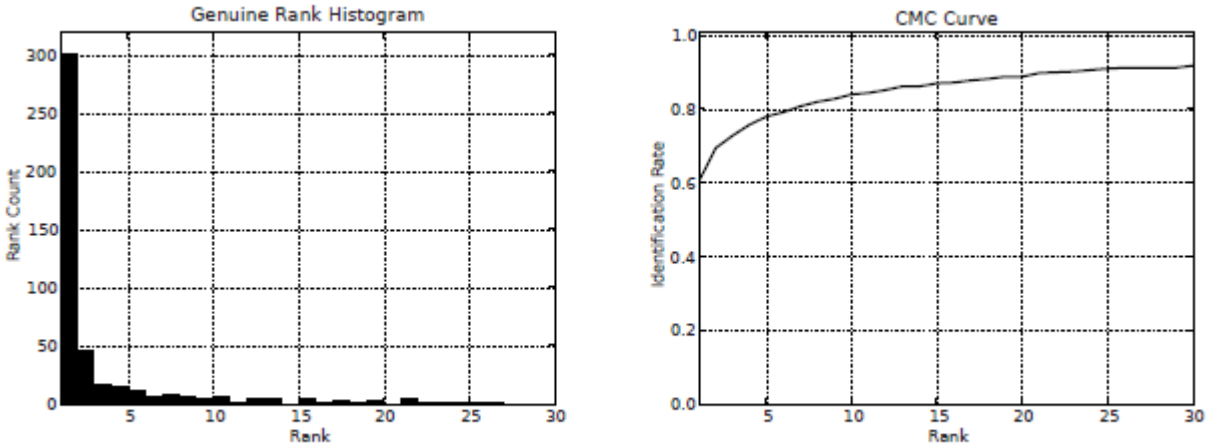
Kako kažu autori u [4], prethodne formule su samo logičke procjene grešaka, a stvarne stope grešaka treba testirati u svakom konkretnom biometrijskom sustavu.

Kod identifikacije, vrlo je velika razlika da li osoba koju treba identificirati sigurno postoji u sustavu, pa je to **identifikacija na zatvorenom skupu** (eng.closed-set identification) ili možda ne postoji, pa je to **identifikacija na otvorenom skupu** (eng.open-set identification). U ovisnosti od toga, bira se **identifikacija temeljena na rangu** (eng.rank-based), ili **identifikacija temeljena na pragu** (eng.threshold-based). Ako se radi o identifikaciji na zatvorenom skupu, pogodnija je identifikacija temeljena na rangu, gdje se operateru prikazuje prvih nekoliko kandidata, rangiranih na temelju rezultata usporedbe. S druge strane, ako se radi o identifikaciji na otvorenom skupu, pogodnija je identifikacija temeljena na pragu, gdje se operateru prikazuju svi kandidati koji prelaze izabrani prag.

5.3. Sustavi za identifikaciju temeljeni na rangu (rank-based)

Kako je rečeno u 5.2, identifikacija temeljena na rangu najčešće se bira kada se radi identifikacija na zatvorenom skupu. Najvažnija krivulja koja se tada koristi je **CMC krivulja (eng.Cumulative Match Characteristic)**, koja prikazuje ovisnost identifikacijske stope od veličine ranga (tj. od veličine kandidacijske liste, broja kandidata za identifikaciju).

CMC krivulja se najčešće dobiva testiranjem na testnoj bazi predložaka. Prvo se napravi odgovarajući broj probnih transakcija nad tom bazom. Za svaku probnu transakciju, gleda se da li je prava osoba dobila rang 1, 2, … i na temelju toga se napravi histogram. Budući da je skup zatvoren, osoba mora dobiti neki rang, makar i zadnji (tj. onaj koji odgovara broju predložaka u testnoj bazi). Na sljedećoj slici, pod (a) prikazan je histogram dobiven testiranjem na bazi sa 5000 predložaka, nad kojom je napravljeno 500 transakcija. Histogram pokazuje da je kod 300 transakcija prava osoba rangirana na 1.mjestu, kod (oko) 50 transakcija na 2.mjestu itd.:

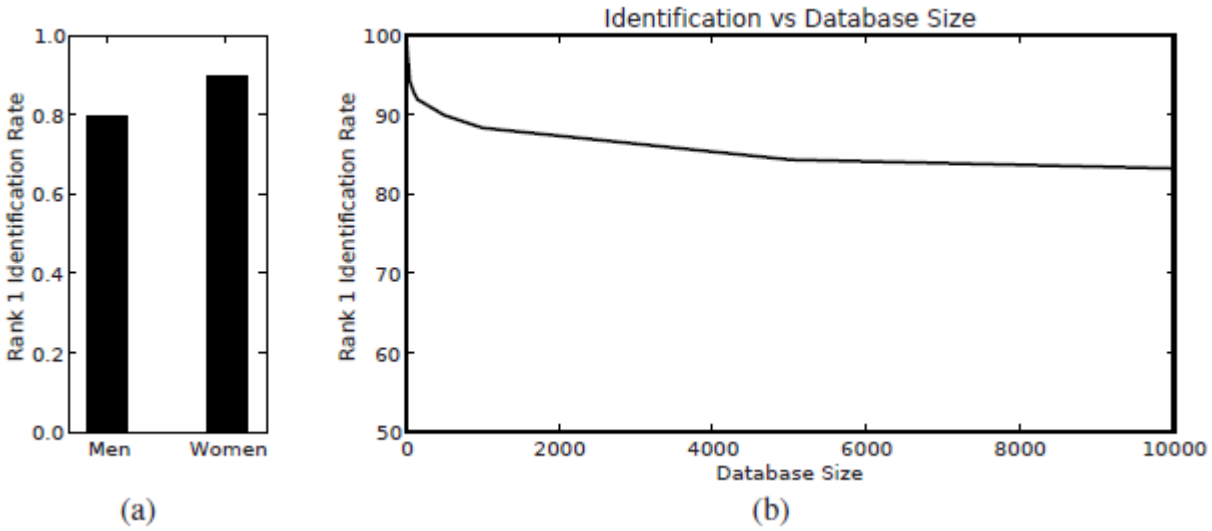


Slika 14. Rezultati rangiranja u sustavu sa 500 transakcija nad bazom od 5000 predložaka;
Izvor: [4]

Na temelju dobivenog histograma, lako se sastavi CMC krivulja, na slici 14. pod (b). Npr. ako se 300 (broj transakcija sa rangom 1) podijeli sa 500 (ukupni broj transakcija), dobiva se 0,6 (što znači da će u navedenom sustavu u oko 60% slučajeva prava osoba biti na 1.mjestu). Zatim se zbroji 300 + 50 (broj

transakcija sa rangom 1 i 2) podijeli sa 500 i dobiva se 0,7 (što znači da će u navedenom sustavu u oko 70% slučajeva prava osoba biti na 1. ili 2.mjestu) itd. Tako dobivene točke (rang na osi x, identifikacijska rata na osi y) daju CMC krivulju.

Druga krivulja koja je korisna za evaluaciju identifikacijskih sustava temeljenih na rangu je Rang 1 identifikacijska krivulja (eng.Rank 1 identification graph). Krivulja uobičajeno pokazuje kako Rang 1 identifikacija ovisi o veličini baze. Na sljedećoj slici, pod (b), vidi se kako se smanjuje identifikacijska stopa s porastom veličine baze (u nekom konkretnom sustavu):



Slika 15. Rang 1 identifikacijska krivulja; Izvor: [4]

Npr. ako je veličina baze blizu 1, u više od 95% slučajeva prava osoba dobit će rang 1. Ako je veličina baze oko 2000, identifikacijska stopa se smanjuje ispod 90%, a kod baze veličine 10000 manja je od 85%.

Slika (a) pokazuje nešto drugo – kako se Rang 1 identifikacija ponaša u ovisnosti od određene demografske karakteristike, u ovom slučaju spola. U konkretnom primjeru biometrijski sustav bolje identificira žene nego muškarce, barem sa stanovišta ranga 1.

Treba napomenuti da su u praksi zatvoreni sustavi vrlo rijetki, tako da je razmatranje iz ove podtočke više akademske nego praktične naravi. U praksi su daleko češći otvoreni biometrijski sustavi, prikazani u sljedećoj podtočki.

5.4. Sustavi za identifikaciju temeljeni na pragu (threshold-based)

Kako je rečeno u 5.2, kod identifikacije na otvorenom skupu ne znamo da li je osoba koju želimo identificirati uopće zapisana u bazu podataka. Kod identifikacije na zatvorenom skupu, imali smo vrlo veliku vjerojatnost da prava osoba bude uključena u listu kandidata, samo smo trebali odabrati dovoljno veliki rang. Kod otvorenog skupa to ne vrijedi (jer osoba možda nije zapisana u bazu podataka), pa nam je pogodnija identifikacija na temelju praga.

Iako postoje različite terminologije, kod identifikacije temeljene na pragu često se definira pojam **alarm** kao situacija u kojoj dobivamo nepraznu listu kandidata (dobivenu na temelju praga, a ne rangiranja), što je znak operateru da je tražena osoba možda pronađena (jer se nalazi u listi kandidata).

Odabrani prag često se naziva **prag alarma** (eng.alarm threshold).

Ako se osoba zaista nalazi u bazi podataka, to se naziva **detekcija**.

Ako se osoba ne nalazi u bazi podataka, to je **lažni alarm**.

Uz pojam detekcija, definiraju se još pojmovi **identifikacija,** te **korektna detekcija i identifikacija**, a njihov odnos bi se mogao prikazati ovako:

1. Detekcija = osoba iz liste kandidata se zaista nalazi u bazi podataka
2. Identifikacija = prava osoba je prva po rezultatu, ali rezultat može biti ispod praga (lista kandidata može biti prazna)
3. Korektna detekcija i identifikacija = detekcija + identifikacija (prava osoba nalazi se u listi

kandidata i to na prvom mjestu).

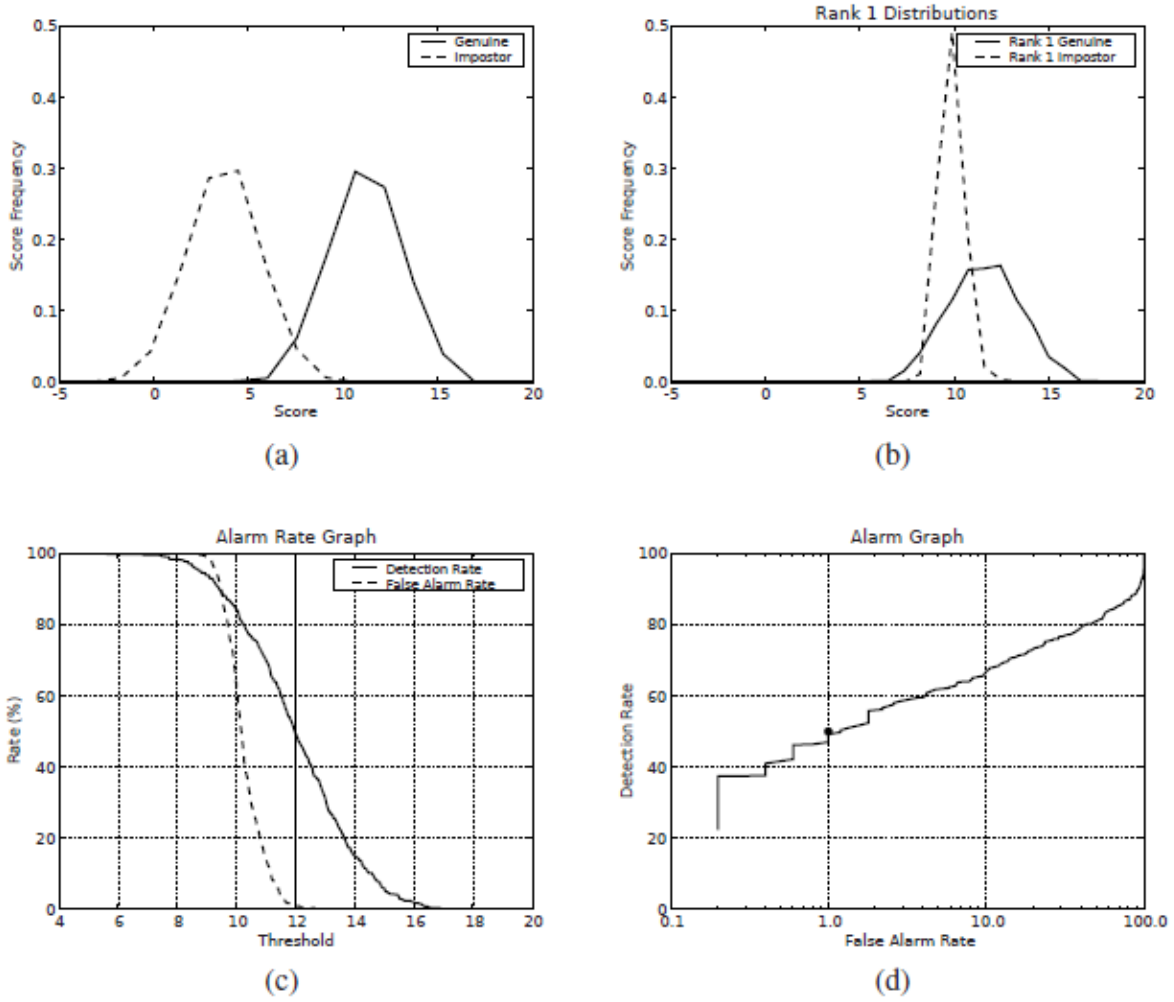
Kada projektiramo ili evaluiramo neki identifikacijski sustav temeljen na pragu, moramo odabrati koji od gornjih zahtjeva želimo zadovoljiti. Najveći zahtjev je korektna detekcija i identifikacija.

Za evaluaciju performansi, najzanimljivije su nam dvije stope:

Stopa detekcije pokazuje vjerojatnost da će prava osoba biti uključena u listu kandidata i ta stopa ovisi samo o pragu, a ne ovisi o veličini baze.

Stopa lažnog alarma pokazuje vjerojatnost pojave lažnog alarma (osoba iz liste kandidata ne postoji u bazi). Ova stopa ovisi o veličini baze (veća baza, veća stopa).

Kod evaluacije sustava za identifikacije temeljenu na pragu koristi se krivulja **alarmiranja (eng.Alarm Curve)**. Ta je krivulja posebna vrsta ROC krivulje (opisane u 4.3.). Krivulja alarmiranja prikazuje ovisnost stope detekcije od stope lažnog alarma (ali samo za određenu veličinu baze). Postupak dobivanja te krivulje prikazuje sljedeća slika:



Slika 16. Temelji krivulje za alarmiranje; Izvor: [4]

Na prethodnoj slici su pod (a) prikazane razdiobe za pravu i lažnu osobu.

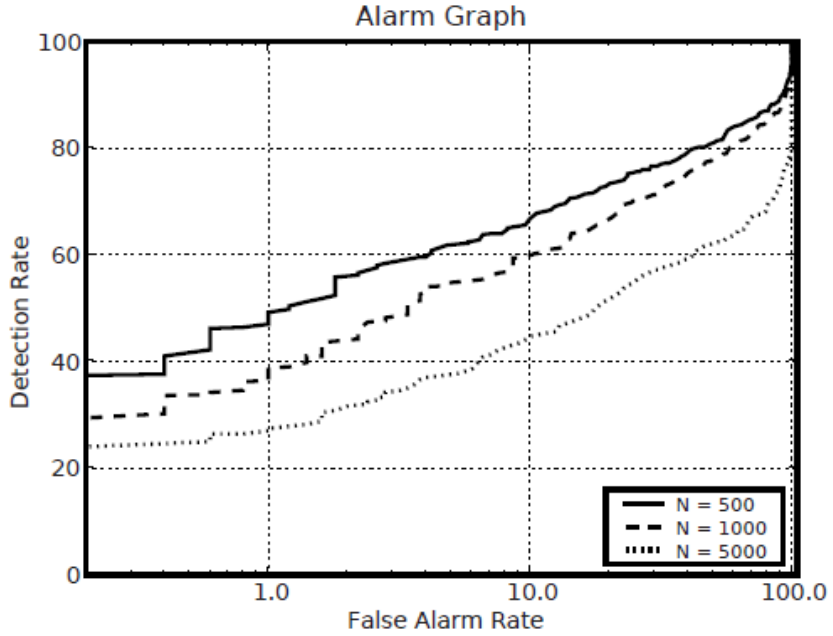
No, izračun se dalje ne temelji izravno na tim razdiobama, već na Rang 1 razdiobama, prikazanim pod (b). Razlog je taj što nas zanima samo najviši rang, jer dok god postoji barem jedan rezultat (pravi ili lažan) iznad praga alarma, to je dovoljno za pojavu detekcije ili lažnog alarma. Primjena Rang 1 razdiobe je razlog zašto stopa lažnog alarma ovisi o veličini baze – zato jer se Rang 1 razdioba za lažnu osobu (crtkana linija) pomiče udesno s povećanjem baze (kako smo vidjeli na slici 13.c).

Pod (c) su izračunate krivulje ovisnosti stope detekcije, odnosno stope lažnog alarma, od odabranog praga alarma. Krivulja stope detekcije dobije se tako da se na os x nanese odabrani prag, a na os y površina ispod Rank 1 razdiobe za pravu osobu (na slici (b) označena punom linijomo) koja se (površina) nalazi desno od tog praga. Slično tome, krivulja stope lažnog alarma dobije se tako da se na os x nanese

odabrani prag, a na os y površina ispod Rank 1 razdiobe za lažnu osobu (na (b) označena crtkanom linijomo) koja se (površina) nalazi desno od tog praga. Na slici (c) je posebno prikazan (kao vertikalna crta) prag na 12, a odgovarajuće stope lažnog alarma, odnosno detekcije, su oko 1%, odnosno oko 50%.

Na kraju, pod (d), napravljena je krivulja alarmiranja, koja prikazuje ovisnost stope detekcije (os y) od stope lažnog alarma (os x), ali samo za unaprijed zadanu veličinu baze (jer je tom veličinom baze određena Rang 1 razdioba kod lažne osobe). Krivulja se dobije tako da se uzmu točke iz krivulja pod (c), npr. jedna točka (ona za prag alarmiranja 12) je točka (1, 50). Os x dana je u logaritamskom mjerilu, kao što je uobičajeno kod svih ROC krivulja.

Kad ocjenjujemo performanse određenog sustava za identifikaciju temeljenu na pragu, ili uspoređujemo dva takva sustava, najčešće nam nije dovoljna samo jedna krivulja alarmiranja, već nam treba više krivulja, za različite veličine baze. Npr. sljedeća slika prikazuje tri krivulje alarmiranja, za baze veličine 500, 1000 i 5000 predložaka, na istom biometrijskom sustavu:



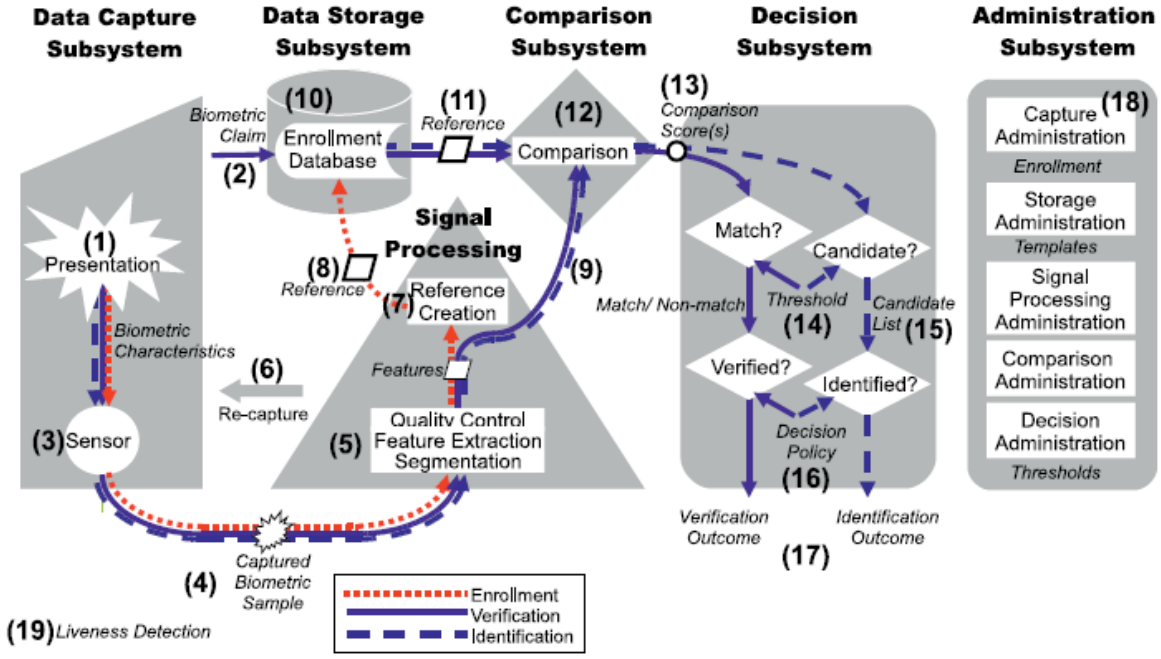
Slika 17. Različite krivulje alarmiranja (iz istog biometrijskog sustava), za različite veličine baze podataka; Izvor: [4]

Vidi se kako performanse opadaju s veličinom baze, jer kod veće baze ista stopa pogrešnog alarma ima manju stopu detekcije (krivulja alarmiranja za veću bazu nalazi se ispod krivulje za manju bazu).

6. RANJIVOST BIOMETRIJSKIH SUSTAVA

Procjena ranjivosti biometrijskog sustava je vrlo važna za biometrijsku sigurnost i taj je koncept različit od koncepta biometrijske preciznosti, koji je prikazan u prethodne dvije točke. Naime, savršeno precizan sustav može biti vrlo ranjiv na napad neautoriziranog korisnika, koji može naći alternativne načine da se lažno prijavi na sustav. Npr. jedan primjer je prijava na sustav pomoću lažnog otiska prsta, napravljenog od gume. Ranjivosti biometrijskog sustava se do sada posvećivala manja pažnja nego preciznosti, ali je sve veće korištenje biometrije u praksi natjeralo da se ovom pitanju posveti dužna pažnja. (Napomena: ova je točka napravljena na temelju točke 12.3. iz [4].)

U 2.točki, na slici 1. bio je prikazan model općeg biometrijskog sustava. Sada se, na slici 18., ponovno prikazuje taj model, ali ovaj put sa upisanim točkama (potencijalnog) napada (eng.Points of Attack), kojih ima ukupno 19:



Slika 18. Točke ranjivosti u općem biometrijskom sustavu; Izvor: [4]

U nastavku se prikazuje ranjivost sustava u točkama (potencijalnog) napada.

1. **Prezentacija** (Presentation): Korištenje lažne biometrijske karakteristike je najčešći napad u toj točki. Može se desiti i kod kreiranja predložaka i kod verifikacije / identifikacije. Primjer: otisak prsta od gume.

2. **Zahtjev za identifikacijski dokument** (Identity Claim): Lažan zapis/dokument koristi se da bi se dobio drugi, valjan biometrijski zapis/dokument. Primjer: Lažna putovnica koristi se da bi se dobio novi identitet u aplikaciji temeljenoj na biometriji šarenice oka.

3. **Senzor** (Sensor): Primjer: Napadač makne kameru i zamijeni ju s uređajem koji dalje šalje uvijek istu sliku.

4. **Prijenos uzorka** (Transmission – Sample): Primjer: Napadač preko nesigurne veze uzima podatke o otisku prsta osobe koja se prijavila, ili suprotno, šalje dalje lažan otisak prsta umjesto pravog.

5. **Kontrola kvalitete i ekstrakcija biometrijskih značajki** (Quality control and feature extraction): Potrebno je kontrolirati kvalitetu kod kreiranja predložaka, kako se u bazi ne bi našli predlošci koji omogućavaju laganu lažnu prijavu. Primjer: Korisnik se prijavljuje na sustav s prljavim prstima.

6. **Ponovno preuzimanje uzorka** (Re-capture): Primjer: Neki sustav za geometriju ruke dozvoljava korisniku da se prijavi neograničeni broj puta, pa je velika šansa da to uspije lažnoj osobi.

7. **Kreiranje reference** (Reference creation): Primjer: Haker unese novi ili mijenja postojeći programski kod za kreiranje reference, tako da omogućava prijavu lažnim osobama.

8. **Prijenos od reference do predloška** (Transmission - Reference to enrollment): Primjer: Haker preko nesigurne linije zamijeni predložak koji se treba unijeti u bazu.

9. **Prijenos biometrijskih značajki u bazu** (Transmission - Features to database): Kao kod kreiranja predložaka (8.), samo je ovdje riječ o verifikaciji/identifikaciji.
10. **Baza predložaka** (Enrollment database): Ranjivosti su kao kod svake baze podataka. Primjer: Zlonamjerni administrator baze unese predložak za napadača.

11. **Prijenos reference iz baze** (Transmission - Reference from database): Primjer: Haker zamjenjuje pravi predložak dobiven iz baze s lažnim, prije usporedbe s tekućim uzorkom.

12. **Proces usporedbe** (Comparison process): Primjer: Haker mijenja programski kod za usporedbu, tako da se u određenom periodu (dovoljnom za napad) dobivaju samo visoki rezultati usporedbe.

13. **Prijenos rezultata** (Transmission – Score): Primjer: Haker preko nesigurne linije mijenja pravi rezultat usporedbe s lažnim, višim.

14. **Proces definiranja praga** (Threshold process): Primjer: Haker postavlja prag za verifikaciju na 0, omogućavajući da se svi uspješno prijave.

15. **Lista kandidata** (Candidate list): Primjer: Haker osigurava da se određeni identiteti nikada ne rangiraju visoko na listi kandidata (pa ih se ne može identificirati).

16. **Politika odlučivanja** (Decision policy): Primjer: Zlonamjerni administrator baze mijenja poslovna pravila, lažno označavajući nekoga kao osobu koja se ne mora identificirati na sustav.

17. **Prijenos ishoda** (Transmission – Outcome): Primjer: Senzor otiska prsta za otključavanje vrata šalje nekriptirani kod koristeći jednostavan relej. Napadač miče senzore i kratko spaja žice, pa se vrata otvaraju.

18. **Administracija** (Administration): Primjer: Zlonamjerni administrator zamjenjuje prag i omogućava napadaču da uđe u sustav.

19. **Detekcija životnosti** (Liveness detection): Primjer: Sustav koji detektira životnost prsta na temelju temperature prevari se na taj način da napadač zagrije lažni prst, napravljen od gume.

ZAKLJUČAK

Biometrija koristi fiziološke ili ponašajne (eng.behavioral) karakteristike određene ljudske individue (osobe) da bi ga automatski verificirala ili identificirala.

Postoje različite biometrijske tehnologije i svaka koristi različite biometrijske senzore i različite algoritme za uparivanje (eng.matching) biometrijskih podataka pročitanih senzorom i biometrijskih podataka iz biometrijske baze podataka. Prikazano je da sve ove biometrijske tehnologije imaju zajednički proces ulaz podataka – obrada podataka – izlaz podataka i da se biometrijski sustav može prikazati kao skup istih komponenti, neovisno od tehnologije.

Biometrijska verifikacija i identifikacija su najvažniji biometrijski procesi. To nisu deterministički procesi, tj. procesi koji mogu dati 100% pouzdan odgovor, već su to stohastički procesi, koji se proučavaju pomoću matematičke statistike.

Prikazano je kako se kod verifikacije javljaju dvije vrste grešaka: pogrešno odbacivanje prave osobe, čija je stopa FNMR (eng.False Non-Match Rate), te pogrešno prihvaćanje lažne osobe, čija je stopa FMR (eng.False Match Rate). Prikazane su i krivulje za evaluaciju performansi sustava za verifikaciju: ROC krivulja (eng. Receiver Operating Characteristic) i DET krivulja (eng.Detection Error Trade-off). Općenito je u biometriji jedno od glavnih pitanja kako naći mjeru između dvije greške (jer manji FNMR obično znači veći FMR i obrnuto).

Prikazano je kako je identifikacija puno složenija od verifikacije. Za razliku od verifikacije, kod identifikacije je stopa pogrešnog prihvaćanja FMR_N ovisna o veličini baze - veća baza, veća greška. Također, kod identifikacije osoba (koju se identificira) ne mora postojati u bazi. Ako postoji u bazi, to je identifikacija na zatvorenom skupu (eng.closed-set identification), u suprotnom je to identifikacija na otvorenom skupu (eng.open-set identification). Kod zatvorenog skupa (koji je u praksi vrlo rijedak) koristi se identifikacija temeljena na rangu (eng.rank-based), kod otvorenog skupa koristi se identifikacija temeljena na pragu (eng.threshold-based). Kod identifikacije temeljene na rangu prikazana je CMC krivulja (eng.Cumulative Match Characteristic), a kod identifikacije temeljene na pragu prikazana je krivulja alarmiranja (eng.Alarm Curve), jedna vrsta ROC krivulje.

U više je navrata naglašavano da biometrijske razdiobe ne moraju odgovarati nekoj poznatoj teoretskoj statističkoj razdiobi (npr. normalnoj razdiobi), te da uvijek treba empirički provjeriti konkretne razdiobe i parametre konkretnog biometrijskog sustava.

Ranjivosti biometrijskog sustava se do sada posvećivala manja pažnja nego preciznosti, ali je sve veće korištenje biometrije u praksi natjeralo da se ovom pitanju posveti dužna pažnja. Prikazane su točke (potencijalnog) napada (njih 19) na općeniti biometrijski sustav.

LITERATURA

[1] Bača, M. (2004), Uvod u računalnu sigurnost, Narodne novine, Zagreb

[2] DeCoursey, W.J. (2003): Statistics and Probability for Engineering Applications, Newnes, Woburn Massachusetts

[3] De Marsico, M. i ostali (2009): Fractal Indexing in Multimodal Biometric Contexts (sažetak članka, iz knjige Intelligent Computing Based on Chaos), Springer, New York

[4] Dunstone , T. i Yager N. (2009): Biometric System and Data Analysis, Springer, New York

[5] Garnett, W.P. (1997): Chaos Theory Tamed, Joseph Henry Press, Washington

[6] Jain, K.A., editor (2008): Handbook of Biometrics, Springer, New York

[7] Pavlić, I. (2000): Statistička teorija i primjena, Tehnička knjiga, Zagreb