Kivonat

Az internetes böngészés rohamos terjedése és az egyre tágabb körű online ügyintézés ösztönöz, hogy felhasználói fiókokat hozzunk létre. Ezek által adataink biztonságosan eltárolódnak és több funkcióhoz is jutunk egy adott oldalon. Ahhoz, hogy ezekhez az adatokhoz más ne férhessen hozzá, különféle azonosítási módszereket dolgoztak ki a felhasználók védelme érdekében. Ennek ellenére, ha egy illetéktelen személy belépést nyer saját online fiókunkba, akkor nem csak felhasználói adatainkat, de akár anyagi javainkat is veszélyeztetheti.

Diplomamunkám célja egy olyan másodlagos hitelesítő rendszer létrehozása webes alkalmazások számára, amely a felhasználónév-jelszó párossal való beléptetés biztonságát növeli. Ezt úgy valósítsa meg, hogy összehasonlítsa a felhasználó által beírt karakterek leütési mintáját a már korábban bevitt mintákkal. Az alkalmazást egy PHP keretrendszerben, Laravelben készítettük el és az adatok tárolására a Firebase valós idejű adatbázist használtuk.

Az eredmény egy gyűjtő felület, ahol a kísérletben résztvevők néhány szöveg begépelésével egy billentyűzési adathalmazt szolgáltatnak. Ez által saját adathalmazra teszünk szert és kísérleteket hajthatunk rajtuk végre, amelyek hasznos eredményeket szolgáltathatnak a jövőbeli kutatásokra nézve. Emellett egy azonosító felületet is implementáltunk, ahol le tudjuk tesztelni éles helyzetekben is az alkalmazást.

***Kulcsszavak:*** billentyűzés alapú felismerés, másodlagos hitelesítés, Laravel, webes biztoság

Tartalomjegyzék

[**1.** **Bevezető** 2](#_Toc517834273)

[**2.** **A projekt célja** 3](#_Toc517834274)

[**3.** **Bibliográfiai tanulmány** 4](#_Toc517834275)

[**4.** **Elméleti megalapozás** 5](#_Toc517834276)

[**4.1.** **A biometrián alapuló azonosítás áttekintése** 5](#_Toc517834277)

[**4.2.** **Billentyűzési biometria webes alkalmazásban** 6](#_Toc517834278)

[**4.2.1.** **Bemeneti mezők megválasztása** 6](#_Toc517834279)

[**4.2.2.** **Jellemzők kiemelése** 7](#_Toc517834280)

[**4.2.3.** **Az egyenletes hibaarány** 7](#_Toc517834281)

[**4.2.4.** **Felhasznált algoritmusok** 8](#_Toc517834282)

[**4.3.** **Felhasznált technológiák** 8](#_Toc517834283)

[**4.3.1.** **Adattárolás Firebaseben** 8](#_Toc517834284)

[**4.3.2.** **Más technológiák** 8](#_Toc517834285)

[**5.** **A rendszer specifikációi** 8](#_Toc517834286)

[**5.1.** **Követelmény specifikáció** 8](#_Toc517834287)

[**5.1.1.** **Kivonat** 8](#_Toc517834288)

[**5.1.2.** **Bevezető** 8](#_Toc517834289)

[**5.1.3.** **Áttekintés** 8](#_Toc517834290)

[**5.1.4.** **Rendszer követelmények** 8](#_Toc517834291)

[**5.1.5.** **Funkcionális követelmények** 8](#_Toc517834292)

[**5.1.6.** **Nem funkcionális követelmények** 9](#_Toc517834293)

[**5.1.7.** **Interfész követelmények** 9](#_Toc517834294)

[**6.** **Részletes tervezés** 9](#_Toc517834295)

[**6.1.** **Adatok tárolása Firebaseben** 9](#_Toc517834296)

[**6.2.** **Architektúra** 9](#_Toc517834297)

[**6.3.** **A rendszer tervezése és bemutatása** 9](#_Toc517834298)

[**6.3.1.** **A kontroller rész** 9](#_Toc517834299)

[**6.3.2.** **A modell rész** 9](#_Toc517834300)

[**6.3.3.** **A nézet (view) rész** 9](#_Toc517834301)

[**6.3.4.** **Segítő osztályok** 9](#_Toc517834302)

[**6.4.** **AJAX hívások bemutatása** 9](#_Toc517834303)

[**6.5.** **Adatok mozgása** 9](#_Toc517834304)

[**7.** **A rendszer felhasználása** 9](#_Toc517834305)

[**7.1.** **Adatgyűjtő felület** 9](#_Toc517834306)

[**7.2.** **Bejelentkezési felület** 10](#_Toc517834307)

[**8.** **Gyűjtő üzembe helyezése és kísérleti eredmények** 10](#_Toc517834308)

[**8.1.** **Gyűjtő üzembe helyezése** 10](#_Toc517834309)

[**8.2.** **Felmerült problémák és megoldásaik** 10](#_Toc517834310)

[**8.3.** **Kísérleti eredmények** 10](#_Toc517834311)

[**9.** **Következtetések** 10](#_Toc517834312)

[**9.1.** **Megvalósítások** 10](#_Toc517834313)

[**9.2.** **Összehasonlítás hasonló rendszerekkel** 10](#_Toc517834314)

[**9.3.** **További fejlesztési irányok** 10](#_Toc517834315)

[**10.** **Irodalomjegyzék** 10](#_Toc517834316)

[**11.** **Függelékek** 10](#_Toc517834317)

1. **Bevezető**

Az internet rohamos fejlődése által egyre több olyan tevékenységet vagyunk képesek lebonyolítani otthon, ami személyes jelenlétet igényelt volna évekkel ezelőtt. Manapság a legújabb okos telefonok megvásárlásától a banki átutalásokig bármit képesek vagyunk pár mozdulattal lebonyolítani és ezzel arányosan hatalmasra nőtt a felhasználói fiókjainknak az értéke is. Bárki, aki az adatainknak, vagy éppen azonosító tárgyainknak (pl. bankkártya, beléptető tóken) a tulajdonában van, képes arra, hogy hozzáférjen a személyes adatainkhoz, vagy akár az anyagi értékeinkhez is.

Tegyük fel, hogy nem biztonságosan kezeltük egy online banki alkalmazásnál használt fiókunkat és a belépési adatainkat megszerezte egy harmadik fél, akinek szándéka ezt a lehetőséget kihasználni. Ilyen esetben segít egy másodlagos hitelesítő rendszer, ami képes a detektálására az esetleges betolakodóknak. A rendszer észleli azt, hogy a belépett személynek egyes tulajdonságai nem egyeznek meg a regisztrált felhasználóéval, ezért jelzést küld és további megerősítést kér. Amennyiben a megerősítés nem megy végbe, a rendszer elkönyvelheti, hogy feltörés ment végbe és segítséget tud nyújtani a felhasználónak a további biztonsági lépések megtételéhez.

Egy ilyen rendszernél például ajánlatos viselkedési biometrián alapuló eszközökkel ellenőrizni, hogy valódi-e a felhasználó. Ezek az eszközök képesek arra, hogy a felhasználót a vele született, egyedi viselkedés mintái és mikro-mozzanatai alapján összehasonlítsák a korábbi belépéseihez és megállapítsák, hogy mekkora hasonlóság mérhető nála.

Általában egy ilyen rendszernél nem csak egyfajta viselkedést ellenőriznek, ugyanis a viselkedési biometrián alapuló hitelesítők nem képesek száz százalékos eredményt garantálni és megeshet, hogy a hamis felhasználót beengedik, vagy éppen a valós felhasználót kizárják. Ennek elkerülésére a rendszer több viselkedést is kell, hogy figyeljen és az egyes tulajdonságok (billentyűzés, egérmozgás, szóhasználat, stb.) mind visszaadnak egy szavazatot, amit összegezve megkapjuk, hogy mi a teendő az adott felhasználóval.

A fiókunk biztonsága érdekében tehát érdemes minél jobban megerősíteni a védelmet és erre egy módszer a billentyűzési ritmus figyelése. A felhasználónak tudnia nem kell róla, mert a háttérben működik. Emellett már egyre jobb és jobb eredményeket érnek el ilyen téren, tehát valószínű, hogy hamarosan egyre több oldalon lesz lehetőség ez a fajta másodlagos azonosítás is.

1. **A projekt célja**

Az első fejezetben említett billentyűzés alapú azonosítási modul megvalósításában nyilvánul meg a célom. Ez magába foglalja egy gyűjtő rendszer létrehozását, mely segítségével a további kísérletekhez tudunk tanító és tesztelő adathalmazokat biztosítani, valamint egy azonosító rendszer létrehozását, ami segítségével a felhasználót egy második biztonsági réteggel is meg tudjuk védeni.

Jelenleg számos kísérlet folyik jobb és jobb eredményekért a billentyűzési ritmus alapú felismeréssel kapcsolatosan. Már a telegráf korában felfigyeltek az emberek a tényre, hogy be tudják egymást azonosítani a beütési ritmusuk alapján. Ez többnyire annak is köszönhet, hogy hogy az emberi agynak ugyanazok a neuro-fiziológiai részei kerülnek használatba gépeléskor, mint amik az aláírások esetében is. Bár egyre több kísérlet folyik ilyen irányban, teljes pontosságú megoldást még nem találtak és viszonylag kevés rendszernél alkalmazzák a módszert.

Az alkalmazás célközönsége az érzékeny adatokat kezelő alkalmazások köre, mivel ilyen esetekben kifejezetten fontos az, hogy az belépett felhasználó az legyen, aki a tulajdonosa a fióknak.

­Megvalósítás céljából a PHP egyik keretrendszerét, a Laravelt használtam. A választás azért esett a Laravel-re, mert számos ponton megkönnyíti a rendszer implementációját és amennyiben más PHP alapú rendszerbe szeretnénk a modult integrálni, kevés átalakítással képesek vagyunk rá. Az alkalmazás egyaránt képes gyűjteni adatokat a valós felhasználótól, valamint külön oldal van létre hozva arra is, hogy pár, a kísérletbe beavatott személy próbáljon belépni a tesztelt felhasználó adataival. Ezeket a belépési kísérleteket szintén elmentjük. A teljes adatbázis JSON formátumban van eltárolva a Firebase valós idejű adatbázisban, ezért a kísérletekhez, amiket Matlabban folytattunk, létrehoztunk egy átalakítót is, ami a JSON formátumban megadott adatainkat csv formátumba konvertálja.

Egy másik célja az államvizsga munkámnak, hogy minél hatékonyabb algoritmust, illetve jellemző kiemelő módszert találjak kísérletek által. Emellett az alkalmazást működővé tenni mobil eszközökön illetve táblagépeken is.

1. **Bibliográfiai tanulmány**
2. **Elméleti megalapozás**
   1. **A biometrián alapuló azonosítás áttekintése**

Az éppen bejelentkező felhasználóról eldönteni, hogy tényleg ő birtokolja-e az adott fiókot, több módon is lehetséges. Az azonosítási módszereket az alábbi három kategóriába soroljuk:

1. Tudás alapú azonosítás - Jelszó, minta illetve kérdés alapú ellenőrzések tartoznak ide. Legnagyobb előnyük hogy könnyű őket implementálni, míg hátrányuk, hogy el lehet őket felejteni, vagy éppen fel is törhetik pár algoritmus segítségével.
2. Token alapú azonosítás - Általában egy tárgy segítségével azonosítsák a felhasználót, ami csak az ő birtokában lehet. Ezt feltörni nehezebb és általában az elkészítése olcsó, de könnyű elveszíteni. Ide tartoznak a bankkártyák és a beléptető csipogók (például a Sapientia marosvásárhelyi székhelyének a bentlakásánál).
3. Biometrián alapuló azonosítás – A felhasználónak egy adott viselkedését vagy vele született jellemzőjét vizsgáljuk az azonosítás során. Két fajtája lehet:
   1. fiziológiai, amelyet nehéz hamisítani és magas pontossággal rendelkezik, mind például az ujjlenyomat vizsgálása vagy a retina ellenőrzése. Ezek a rendszerek viszont költségesek és mivel a felhasználó vele született fizikai tulajdonságát nézi, amennyiben egyszer sikerül megszereznie egy harmadik félnek a mintát, a felhasználó többé nem használhatja az adott módszert.
   2. viselkedési, melynél azt nézzük, hogy a felhasználó „hogyan” csinál valamit, nem pedig azt hogy mit csinál. Ilyen azonosításokat végezhetünk az egérmozgás, járás, hang, billentyű leütési ritmus és még sok egyéb egyedi mintát alkotó viselkedés elemzésével. Ezek a rendszerek általában nem biztosítanak teljes pontosságot, mivel megeshet, hogy kizárják a valódi felhasználót, vagy éppen beengedik a hamisat.

Bár a viselkedési biometrián alapuló rendszerek képesek arra, hogy egymagukban, viszonylag magas pontossággal azonosítsák a felhasználót, egy ilyen rendszer önmagában nem hatékony. Ezért is az alkalmazásnál ez a fajta hitelesítés nem veszi át a felhasználónév-jelszó páros szerepét, hanem az azonosított felhasználókat ellenőrzi le, hogy tényleg ők-e azok, akik eddig is folyamatosan bejelentkeztek.

Egy másik fontos része a billentyűzés alapú hitelesítésnek, mint bármely más viselkedésen alapuló biometriai hitelesítőnek, hogy egymagában mivel téves adatot ad, több ilyen ellenőrző modullal együtt képesek valós képet alkotni a felhasználó kilétéről. Például a billentyűzés mellett még lehet nézni az egér mozgását is. Ezért ajánlatos az alkalmazásban megírt azonosító modult egy szavazó komponensként alkalmazni az ellenőrzés során.

* 1. **Billentyűzési biometria webes alkalmazásban**

Az dolgozatmunkám egy webes alkalmazásra épül, mivel sokkal könnyebben meg lehet valósítani az adatgyűjtést, ha egy szerveren rajta van az oldal és egyidejűleg korlátlan számú felhasználó írhatja be az adatokat. Ennek ellenére a felhasználónak lehetősége van arra, hogy ha nincs felügyelve, akkor egy rövid időre megszakítsa a gépelést és ez elrontsa az adott mintát. Emiatt párhuzamosan gyűjtöttem felügyelet nélküli beviteleket és általam felügyelteket is.

Az alkalmazás gyűjtésénél a JavaScriptben megírt JQuery könyvtár függvényeit használtam, melyek által le lehet kérni az adott billentyűnek a lenyomási és felengedési idejében kapott időbélyeget. Ezeket egy rejtett mezőben tárolva tovább lehet küldeni az űrlap beküldése során és a szerver feldolgozza, majd tárolja őket.

Egy másik előnye a webes alkalmazásoknak ilyen téren, hogy kevés módosítással képesek vagyunk arra, hogy a számítógép/laptop billentyűzetei mellett a rendszert működésre bírjuk okos telefonokon és táblagépeken is.

* + 1. **Bemeneti mezők megválasztása**

Tekintsünk most el pár pillanatra a programozás résztől és nézzünk rá a problémára úgy, hogy közben próbáljuk az emberi viselkedésnek a jellemzőit felhasználni.

Már számos adathalmazt hoztak létre, ezek közül én két ismertebbet próbáltam ötvözi, hogy a lehetséges jobb eredmény elérése mellett vizsgáljam meg azt is, mekkora különbség jelentkezik a felhasználó által megadott és a felhasználót a születésétől kísérő adatok beírásának hasznosságában billentyűzési ritmus alapú ellenőrzésnél. A két billentyűzés dinamikai adatbázis a következő:

* ATVS-Keystroke database [1] – 63 felhasználónak 12 valós és 12 imposztor általi belépése van tárolva. Minden belépés során a felhasználók a személyes adataikat adták meg, amik a vezetéknév, keresztnév, e-mail cím, nemzetiség és személyi igazolvány szám. Két szesszióban folyt le a valós felhasználótól az adatgyűjtés, szessziónként 6 belépéssel. Az adathalmaz előnye, hogy a felhasználó egy olyan szöveget kell, hogy beírjon, amit már jól ismer és ezért összefüggőbbek lesznek a mintái.
* CMU database [2] – 51 felhasználó 8 szesszióban adta le az adatait és két szesszió között legalább egy nap telt el. Ennél az adathalmaznál a felhasználók csak egy jelszavat gépeltek be, ami mindenkinek előre meghatározott volt és tartalmaz kis- és nagybetűt, speciális karaktert valamint számot. Az adathalmaz előnye, hogy nem volt szűkség külön imposztorokra, mert mindenki ugyanazt a jelszavat ütötte be.

Ismerve a fent említett két adathalmazt, olyan bemeneti mezőket választottam, amik ötvözik ezeknek a tulajdonságait és általuk olyan kísérleteket hajthatunk végre, amiket külön a két adathalmazon nem tehettünk volna. Ehhez az első mezőnek a *felhasználónevet*, a másodiknak az oldalon ügyködő személynek a *valódi nevét*, míg a harmadik mezőnek egy *training password* nevű adatot kellett tartalmaznia. Ez utóbbi kötelező módon kell tartalmazzon kis- és nagybetűt, számot és speciális karaktert.

A felmerült kérdések a következők voltak: Ha a legszemélyesebb a felhasználónév akkor az is lesz a leghatékonyabb azonosításnál? Amennyiben igen, a jelszó, amit nem látunk milyen mértékben lesz kevésbé hasznos az azonosításnál, mint a másik két mező. Ezeknek a megválaszolását a 8.3-as szekcióban lehet megtalálni a többi kísérleti eredmény mellett.

* + 1. **Jellemzők kiemelése**

Az adatok bevitele után következik a rendszer tanításának első része, ami a jellemzők kiemeléséből áll.

Amikor az adatok megérkeznek a szerverre, még nyers formában vannak és úgy lehet őket kezelni, mint egy lista időbélyeg párosokból. Ezeket az adatokat nem lehetne felhasználni azonosításra, mert az időbélyegek folyamatosan változnak. Ezért a beérkező billentyűk leütési és felengedési időpontjait az őket követő billentyű leütési időiből kivonjuk. Ez által egy úgynevezett di-gráfot hozunk létre, mert kettőnként csoportosítva nyerjük ki az adatokat, amiket felhasználunk a minta elkészítésére, ami által később a felhasználót tudjuk azonosítani.

A tulajdonságok, amiket jelenleg kinyerünk a következők:

* Két egymás utáni leütés közti idő (DD = Down-Down)
* Két egymás utáni felengedés közti idő (UU = Up-Up)
* Az egyik billentyű leütése és a következő felengedése közti idő (DU = Down-Up)
* Az egyik billentyű felengedése és a következő leütése közti idő (UD = Up-Down)
* Az egyik billentyű lenyomása és felengedése között eltelt idő (H = Hold)

[rajz]

Ezeket a tulajdonságokat még kibővíthetjük, hogy hármanként is nézze az időzítéseket, viszont ez az információ ugyanaz, mintha összegeznénk két di-gráf elemet, ezért nem is emeltem ki több jellemzőt.

* + 1. **Az egyenletes hibaarány**

A fent említett finomított adatokat átadhatjuk egy algoritmusnak, ami a szívét fogja képezni a rendszerünknek. Egy jó algoritmus nagyságrendekkel növelheti az elért eredményt, ami alatt minél több elfogadott valódi felhasználót és elutasított hamisat értünk. Az algoritmusok jóságát fontos hogy tudjuk összemérni, amire az úgynevezett Egyenletes Hibaarányt (EER = Equal Error Rate) használjuk. [ref]

Minden kiértékelt bejelentkezési próbálkozás kap egy pontszámot, amiről a vágáspont alapján eldönti, hogy valós felhasználónak könyveli-e el az aktuálisat, vagy hamisnak. Az EER-t két tényező határozza meg: a hibásan elfogadott hamis felhasználók aránya és a hibásan elutasított valós felhasználóké. Ezt legegyszerűbben úgy tudjuk vizualizálni, ha az adott belépési kísérleteket ábrázoljuk az általuk szerzett pontszám függvényében. Ezt elvégezve a valós és hamis felhasználókra ugyanazon az ábrán két görbét kapunk, melyeknek a metszéspontja adja az Egyenletes Hibaarányt.

[ábra]

Az EER értéke minél kisebb, az algoritmus annál jobban működik. Napjainkban már vannak algoritmusok, amik közel 0% EER-t érnek el, viszont nagyon sok tanító mintát igényelnek és ezért valós rendszereknél nem működnének.

Mikor küszöbértéket, vagy más néven vágáspontot választunk, egy olyan függőleges egyenest helyezünk el a grafikonon, ami vagy az EER ponton helyezkedik el, vagy abban irányba mozog, amerre több hamis felhasználót ki tud zárni, ahogy az az ábrán (ábra) is látszik. Minden következő belépési kísérlet pontszáma akkor lesz elfogadva, ha meghaladja ezt az értéket. Ez lehet minden felhasználóra egy globális érték, vagy felhasználónkét váltakozó.

Valós rendszereknél nem áll rendelkezésre az imposztor adathalmaz, ezért a vágáspontot más módon kell, hogy beállítjuk [ref]. Erre az egyik megoldás az, hogy kizárjuk a felhasználó mintáinak leggyengébb 10 százalékát. Ezzel általában elég magasra tudjuk helyezni a vágáspontot ahhoz, hogy kizárjuk a lehetséges imposztorok nagy részét. (ábra)

* + 1. **Felhasznált algoritmusok**

A következőben tárgyalt algoritmusokat először Matlab alatt teszteltük le a CMU [2] adathalmazon. Tesztelés után integráltuk az algoritmusokat a webes alkalmazásunkba is.

* + - 1. *Manhattan scaled*

Az első algoritmus, amit kipróbáltunk, mivel egyszerűsége ellenére az egy nagyon jó eredményt ad az eddigi felmérések szerint [3][ref].

* + 1. **Kimenetek fúziója**

* 1. **Felhasznált technológiák**
     1. **Adattárolás Firebaseben**
     2. **Más technológiák**

1. **A rendszer specifikációi**
   1. **Követelmény specifikáció**
      1. **Kivonat**
      2. **Bevezető**
      3. **Áttekintés**
      4. **Rendszer követelmények**
      5. **Funkcionális követelmények**
      6. **Nem funkcionális követelmények**
      7. **Interfész követelmények**
2. **Részletes tervezés**
   1. **Adatok tárolása Firebaseben**
   2. **Architektúra**
   3. **A rendszer tervezése és bemutatása**
      1. **A kontroller rész**
      2. **A modell rész**
      3. **A nézet (view) rész**
      4. **Segítő osztályok**
   4. **AJAX hívások bemutatása**
   5. **Adatok mozgása**
3. **A rendszer felhasználása**
   1. **Adatgyűjtő felület**
   2. **Bejelentkezési felület**
4. **Gyűjtő üzembe helyezése és kísérleti eredmények**
   1. **Gyűjtő üzembe helyezése**
   2. **Felmerült problémák és megoldásaik**
   3. **Kísérleti eredmények**
5. **Következtetések**
   1. **Megvalósítások**
   2. **Összehasonlítás hasonló rendszerekkel**
   3. **További fejlesztési irányok**
6. **Irodalomjegyzék**

[1] Morales, A., Falanga, M., Fierrez, J., Sansone, C., & Ortega-Garcia, J. (2015, September). Keystroke dynamics recognition based on personal data: A comparative experimental evaluation implementing reproducible research. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on* (pp. 1-6). IEEE.

[2] Killourhy, K. S., & Maxion, R. A. (2009, June). Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP international conference on* (pp. 125-134). IEEE.

[3] Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. The Scientific World Journal, 2013.

1. **Függelékek**