View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Digital Library of the University of Pardubice

POČET MĚŘENÍ PRO VYTVOŘENÍ VZORU IDENTIFIKAČNÍHO POLE DYNAMIKY PSANÍ KRÁTKÉHO TEXTU NA KLÁVESNICI

Miloslav Hub
Ústav systémového inženýrství a informatiky, FES, Univerzita Pardubice

Abstract
Keystoke dynamics is biometrics authentification. This biometrice usually does not involve some special hardware and it is advantage os this way how to prove you are really you. In this article it is suggested the criterium whitch determines how many measumerent is necessary for creating template of keystroke dynamics. This criterium is simultaneously used in experiments.

Keywords
Keystroke dynamics, biometrice, authentification, identity verification, authentification template

Úvod

Slovo autentizace pochází z řeckého slova authentikos, latinsky authenticus, což znamená pravý, původní, hodnověrný. Pojem autentizace v České republice definuje vyhláška Národního bezpečnostního úřadu č. 56/1999 Sb. o zajištění bezpečnosti informačních systémů nakládajících s utajovanými skutečnostmi, provádění jejich certifikace a náležitostech certifikátu, kde se autentizací subjektu rozumí proces ověření jeho identity splňující požadovanou míru záruky (§ 2 písm. f) [14].

Cílem autentizace je tedy prokázat nebo vyvrátit tvrzení subjektu o své identitě. Děje se tomu tak, že subjekt předkládá předem dohodnutý identifikační znak (případně více identifikačních znaků), který je pro daný subjekt jedinečný. Na základě jisté ověřovací informace je tento identifikační znak jako důkaz tvrzené identity přijat, případně odmítnut [13].

Podle druhu předkládaných identifikačních znaků je autentizace dělena na znalostní autentizaci, autentizaci prostřednictvím autentizačního předmětu, na biometrickou autentizaci (zkráceně biometrika) a na vícefaktorovou autentizaci, kdy je současně předkládáno několik nezávislých identifikačních znaků.

Jedna z možností biometrické autentizace je použití dynamiky psaní na klávesnici, která je pro každou osobu jedinečná, podobně jako například u vlastnoručního podpisu. Navíc tento způsob prokázání identity osoby nevyžaduje zpravidla žádné speciální hardwarové zařízení, jako je tomu u ostatních biometrik.

Analýza současného stavu

Už v 19. století operátoři telegrafu podle této dynamiky dokázali rozeznat, kdo telegrafickou zprávu vysílá [6], [10]. V roce 1975 navrhl Spillan užití charakteristik dynamiky psaní na klávesnici pro autentizaci [9].

Prvním pokusem v této oblasti byl však až experiment R. Gainese a jeho spolupracovníků v roce 1980 [4]. Experiment byl proveden se 7-mi sekretářkami, které psaly texty o délce 300-400 slov, přičemž byla měřena prodleva mezi sousedními znaky (bigramy). Bylo zjištěno, že jednotlivé sekretářky dané bigramy píší podobnou rychlostí a to i v různých

textech, což bylo i statisticky dokázáno. Tento pokus měl několik nedostatků, zejména příliš malý počet testovaných osob.

Na Gainesův experiment navázalo v 90. letech mnoho dalších výzkumů, [7,8,9]. Tyto výzkumy však stále pracovaly s dlouhým textem a měření bigramů. Zaměřeny byly zejména na zvýšení počtu osob, které se pokusu účastnily, na vyřazení nevhodných bigramů, na vytvoření šablony a ošetření extrémních hodnot.

Převratem byl až Gariciův patent založen na odlišném přístupu [5]. Předpokladem bylo, že nejvhodnější pro tento způsob autentizace je samotné uživatelské jméno. Měřena byla prodleva mezi stiskem dvou kláves. Uživatelské jméno bylo několikrát za sebou napsáno, naměřené časy zprůměrovány, čímž byl vytvořen vektor šablony (vektor, jehož jednotlivými prvky jsou průměrné doby prodlev mezi jednotlivými stisky dvou kláves). Autentizovaný uživatel poté později napsal své uživatelské jméno a tento nový vektor byl porovnán s vektorem šablony. Jako míra podobnosti byla zvolena Mahalanobisova vzdálenost.

Dalším přelomem byla práce J. Younga a R. W. Hammona, jejíž výsledkem byl další patent [11]. Na rozdíl od předešlých přístupů nebyly jako identifikační znaky uvažovány pouze doby mezi stiskem jednotlivých kláves, ale také doby stisku kláves. Porovnání vektoru šablony a vektoru přístupu je provedeno prostřednictvím Euklidovské vzdálenosti. Tito autoři se však zaměřili pouze na průběžnou autentizaci.

Později následovalo mnoho dalších výzkumů. Mezi nejcitovanější patří výzkumy R. Joice [7] a S. Blehy [1]. Joice navrhuje používat pro zjišťování dynamiky psaní na klávesnici kombinaci uživatelského jména, jména, příjmení a hesla. Jako kritérium podobnosti volí součet absolutních vzdáleností jednotlivých složek vektorů (vektoru šablony a vektoru přístupu). Bleha navrhuje použití normalizované vzdálenosti a normalizovaného Bayesovského klasifikátoru. Joice i Bleha používají jako identifikační znaky dobu mezi stiskem příslušných kláves.

Nyní se v důsledku rozvoje metod výpočtové inteligence objevují i návrhy použití těchto metod i pro účely biometrické autentizace, např. teorie fuzzy množin [2], [3].

V současné době je jedinou komerčně používanou aplikací pro statickou autentizaci prostřednictvím biometriky psaní na klávesnici program BioPassword patentovaný [12] americkou firmou BioNet Systemes. Algoritmus programu je však utajený.

Definice problému

Samotnému procesu autentizace předchází fáze vytvoření vzoru biometrických charakteristik pro každého uživatele, který bude v budoucnu autentizován. Dynamika psaní na klávesnici patří do skupiny stochastických identifikačních znaků, protože na hodnotu tohoto znaku bude mít vliv mnoho drobných nekontrolovatelných vlivů, např. únava uživatele, jeho případné zranění a podobně. Je proto třeba zodpovědět otázku, kolikrát je třeba zopakovat měření těchto charakteristik, aby vytvořený vzor co nejlépe představoval typickou dynamiku psaní určitého hesla daným uživatelem.

Volba kritéria

K určení vhodného počtu měření pro vytvoření vzoru identifikačního pole se nabízí přístup matematické statistiky, jenž by maximalizoval počet těchto měření, neboť prostřednictvím většího výběrového souboru lze spolehlivěji odhadnout vlastnosti základního souboru. Tedy např. aritmetický průměr se stává s rostoucí velikostí výběrového souboru spolehlivějším odhadem střední hodnoty. Základní soubor je v tomto případě představován „typickou dynamikou psaní určitého hesla daným uživatelem“, tedy vzorem identifikačního

pole, výběrový soubor představují jednotlivá měření dynamiky psaní určitého hesla daným uživatelem (identifikační pole předkládané při autentizaci).

Základním nedostatkem tohoto přístupu je předpoklad náhodného výběru pro vytvoření výběrového souboru. Aplikujeme-li tento předpoklad na řešený problém, pak bychom museli předpokládat, že uživatel daný text píše způsobem, jenž je nezávislý na způsobu psaní textu při předešlých měření. Jelikož však jednotlivá měření probíhají postupně za krátké časové úseky, není tento předpoklad správný. Jistě při tomto psaní dochází u uživatele k procesu učení a navíc hrozí únava, která může mít za následky překlepy a podobně. Nehledě k tomu, že opakované psaní stejného textu je pro uživatele nepříjemné a klade na něho nároky.

Z tohoto důvodu je třeba zvolit takový počet měření pro vytvoření vzoru, který bude kompromisem mezi dostatečným množstvím relevantních dat a přijatelností pro uživatele.

Předpokládejme následující model. Biometrické charakteristiky uživatelů naměřené při samotné autentizaci představují požadované hodnoty, které odhadujeme prostřednictvím jejich vzoru získaného při registraci z měření. Potom můžeme jako kritérium zvolit součet druhých mocnin reziduálních hodnot mezi odhadem a skutečnou hodnotou standardizovaných dat definovaný vztahem (1).

$$s^2(n) = \sum_{i=1}^{u} \sum_{j=1}^{m} \frac{\left(x_j^i - \frac{1}{n} \cdot \sum_{k=1}^{n} t_{j,k}^i \right)^2}{\frac{1}{n-1} \cdot \sum_{k=1}^{n} \left(t_{j,k}^i - \frac{1}{n} \cdot \sum_{k=1}^{n} t_{j,k}^i \right)^2} \quad \text{pro } n \geq 2 \qquad (1)$$

$s^2(n)$ .....reziduální rozptyl mezi odhadem a skutečnou hodnotou standardizovaných dat

x_j^i .....hodnota j-té biometrické charakteristiky i-tého autentizovaného uživatele

$t_{j,k}^i$ .....hodnota j-té hodnoty biometrické charakteristiky i-tého uživatele při k-tém měření pro vytvoření vzoru typických hodnot identifikačních znaků

u .....počet registrovaných uživatelů

m .....počet měřených charakteristik

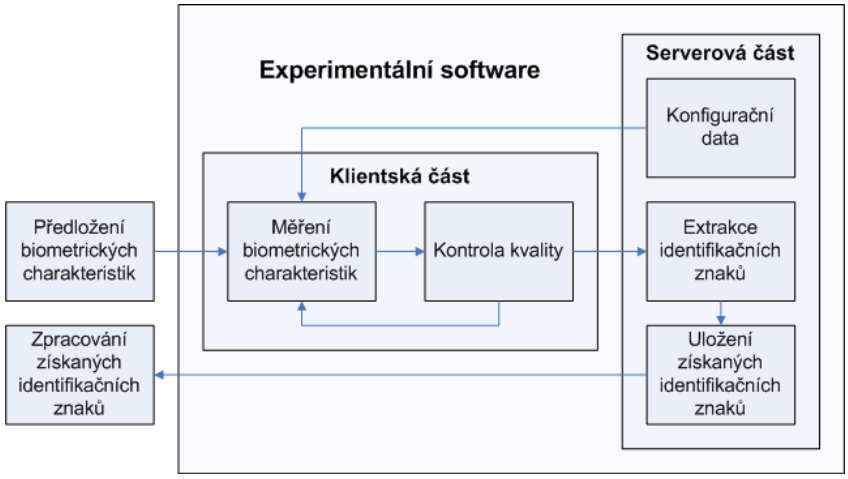
n .....počet měření pro vytvoření vzoru identifikačního pole

Snahou je pak nalézt takový počet měření n pro vytvoření vzoru typického identifikačního pole, při kterém bude $s^2(n)$ nejnižší.

Základní informace o provedeném experimentu

Pro účely zjištění použitelnosti autentizace prostřednictvím dynamiky psaní na klávesnici byl vytvořen speciální software, který měří biometrické identifikační charakteristiky dynamiky psaní zadaných hesel na klávesnici a extrahuje a ukládá získané identifikační znaky pro další zpracování.

Tento software se skládal z klientské části vytvořené v programovacím jazyku Java, která načítala konfigurační data uložené na serveru (např. testované heslo) a odesílala naměřená data serverové části pro další zpracování. Pro komunikaci klientské části se serverovou byl použit protokol HTTP. Serverová část byla kromě konfiguračního souboru tvořena PHP programem uloženým na webovém serveru Apache a databázovým serverem MySQL. Princip tohoto experimentálního software znázorňuje obr. 1.



Obr. 1: Experimentální software

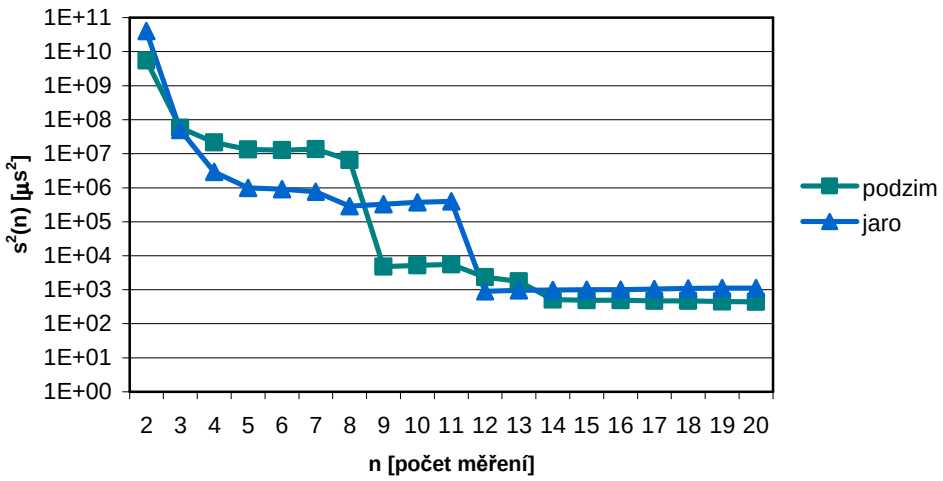
Získávanými identifikačními znaky byly doby stisku příslušných kláves a doby mezi stiskem příslušných kláves. Přesnost měření byla 1µs. Současně byly měřeny i další parametry související s experimentem (identifikátor uživatele, IP adresa počítače, datum a čas experimentu, použité klávesy,…).

Testovanými osobami byli studenti Fakulty ekonomicko-správní Univerzity Pardubice, kteří se experimentu účastnili dobrovolně v rámci výuky odborných předmětů na počítačových učebnách.

Za účelem zjištění vhodného počtu měření pro vytvoření vzoru byly provedeny experimenty, při kterých testované osoby nejprve 20 krát napsaly zadaný text a o týden později tento text napsaly ještě jednou. Prvních 20 měření složilo pro vytváření vzorů, poslední 21. měření simulovalo samotnou autentizaci. První experiment, kterého se účastnilo 36 osob, byl proveden pro text „jaro“. Druhý, jehož se zúčastnilo 44 osob, byl proveden pro text „podzim“.

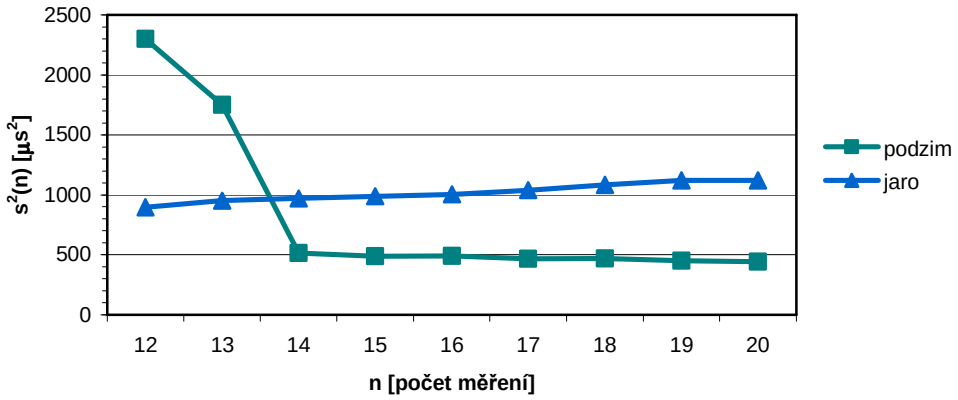
Výsledky experimentu

Výsledky provedených experimentů znázorňují Graf 1 a Graf 2.



Graf 1: Závislost reziduálního rozptylu na počtu měření pro vytvoření vzoru

Graf 2 znázorňuje stejné výsledky, narozdíl od logaritmického měřítka osy y je na rozdíl od předcházejícího grafu (viz Graf 1) použito měřítko lineární a výsledky jsou uvedeny pouze pro 12. až 20. měření.



Graf 2: Závislost reziduálního rozptylu na počtu měření pro vytvoření vzoru pro 12 až 20 měření

Závěr

Z předcházejících grafů (viz Graf 1 a Graf 2) je patrný nejprve prudký pokles reziduálního rozptylu, jehož strmost postupně klesá s rostoucím počtem měření pro vytvoření vzoru. V případě textu „jaro“ dokonce později dochází k opačnému efektu, kdy s rostoucím množstvím měření kvalita vzoru klesá (reziduální rozptyl roste).

Tento jev je způsoben skutečností, že v určitém okamžiku vytváření vzoru identifikačního pole přestává uživatel psát daný text takovým způsobem, jakým ho bude psát při samotné autentizaci. To může být způsobeno jeho únavou při opakovaným psaním stejného textu.

Na základě těchto experimentů lze tedy konstatovat, že pro vytvoření dostatečně kvalitního vzoru je vhodných 15 měření. Tento počet měření se jeví jako vhodný i z hlediska požadavku, aby na uživatele nebyly kladeny nepřiměřeně vysoké nároky.

Poděkování

Tento článek vznikl v rámci interního grantu FG452008 Fakulty ekonomicko-správní, Univerzity Pardubice, které tímto děkuji za poskytnuté prostředky.

Použitá literatura

[1] BLEHA S., SLIVINSKY, CH., HUSSIEN, B. Computer-Access Security Systems Using Keystroke Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 1990, vol. 12, No. 12.

[2] ČAPEK J. Identifikace uživatele informačním systémem. *Scientific papers of the University of Pardubice*, 2004, vol. 9, Ser. D, s. 21-25. ISSN 1211-555X. ISBN 80-7194-716-4.

[3] ČAPEK J., HUB M. Fuzzy Approach in Biometric Authentication by Keystroke Dynamics. *WSEAS TRANSACTIONS on SYSTEMS*, 2005, Issue 4, Volume 4. ISSN 1109-2777

[4] GAINES R., LISOWSKI W., PRESS S., et.al *Authentication by keystroke timing: Some preliminary results*. Rand Report R-256-NFS. Santa Monica, CA: Rand Corporation, 1980.

[5] GARICA J. *Personal identification apparatus*. Patent Number 4.621.334. Washington, D.C.: U.S. Patent and Trademark Office, 1986.

[6] ILONEN J. Keystroke dynamics [on-line]. Lappeenranta University of Technology, Finland [cit. 10-8-2004]. Dostupné z <http://www.it.lut.fi/kurssit/03-04/010970000/seminars/Ilonen.pdf>

[7] JOICE, R., GUPTA, G. *Identity Authentication Based on Keystroke Latencies*. Technical report #5, Australia: Department of Computer Science, James Cook University, 1989.

[8] LEGGET, J., WILLIAMS G., UMPHRESS D. *Verification of user identity via keyboard characteristics*. J.M. Carey, Ed. Ablex Publishing. New York: Norwood, 1986.

[9] LEGGETT. J, WILLIAMS G., USNIK M. Dynamic identity verification via keystroke characteristics. In *International Journal of Man-Machine Studies*, v36, Sept. 1990, s. 859-870.

[10] UMPHRESS D, WILLIAMS G. Identity verification through keyboard characteristics. In *Int. J. Man-Machine Studies*, Sept. 1985, 23, 3, s. 263-273.

[11] YOUNG J., HAMMON R. W. *Method and apparatus for verifyng an individual's identity*. Patent Number 4.805.222. Washington, D.C: U.S. Patent and Trademark Office, 1989.

[12] ZILBERMAN A. G. *Security method and apparatus empoying authentication by keystroke dynamics* (1998) United States Patent 6.442.692

[13] Struktura autentizace. ČSN ISO/IEC 10181-2 369694 ČNI, 1998.

[14] Vyhláška NBÚ č. 56/1999 Sb. ze dne 19. března 1999 o zajištění bezpečnosti informačních systémů nakládajících s utajovanými skutečnostmi, provádění jejich certifikace a náležitostech certifikátu.

Kontakt:

Ing. Miloslav Hub, Ph.D.
Ústav systémového inženýrství a informatiky
Fakulta ekonomicko-správní
Univerzita Pardubice
Studentská 84
532 10 Pardubice
miloslav.hub@upce.cz