НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

“ЗАТВЕРДЖУЮ”
Директор ФТІ
_____ Новіков О.М.
(підпис) (ініціали, прізвище)
“____”_____ 2016р.

Захист програмного забезпечення та даних

Методичні вказівки
до виконання розрахункової роботи
для студентів навчального напряму

6.170101 «Безпека інформаційних і комунікаційних систем»

Рекомендовано кафедрою
інформаційної безпеки
(Протокол № 6\2009 від „26”червня 2016 р.)

Завідувач кафедри
інформаційної безпеки
Грайворонський М.В.

Київ – 2016

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Захист програмного забезпечення та даних

Методичні вказівки
до виконання розрахункової роботи
для студентів навчального напряму

6.170101 «Безпека інформаційних і комунікаційних систем»

Київ – 2016

Розрахункова робота (РР) є видом самостійної роботи студентів. Вона виконується як семестрове завдання і містять індивідуальний набір завдань розрахункового характеру, який охоплює матеріал семестру. Варіанти завдань для розрахункової роботи співпадають з номером залікової книжки (2 останні цифри).

Ваговий бал за розрахункову роботу в РСО – 20 балів.

На захисті студент повинен:

- показати акуратно оформлений звіт -5,
- відповісти на тестове теоретичне питання з матеріалу розрахункової роботи – 5,

ЗАВДАННЯ
на розрахунково-графічну роботу
з дисципліни "Захист програмного забезпечення та даних"

1. ЦІЛЬ розрахунково-графічних робіт

Метою розрахунково-графічної роботи є освоєння методів автентифікції на підставі біометричних характеристик користувача.

2. ЗАВДАННЯ НА розрахунково-графічних робіт

Розробити програму автентифікації користувача по клавіатурному почерку.
Вимоги до програми:

1. Програма повинна працювати в двох режимах:
   - навчання (створення біометричного еталону)
   - ідентифікації (порівняння з біометричним еталоном).
2. На етапі навчання необхідно визначати еталонні статистичні параметри клавіатурного почерку – оцінки математичного чекання і дисперсії тривалості утримання клавіш. Параметри повинні записуватися у файл. Вчення виробляти по багатократному набору фіксованій контрольній фразі, символи якої рівномірно розподілені по клавіатурі.
3. На етапі ідентифікації необхідно визначити параметри введеної контрольної фрази і перевірити гіпотезу про те, що отримані оцінки математичного очікування і дисперсії належать тому ж розподілу, що і параметри біометричного еталону. На цьому етапі необхідно відображувати отримані оцінки і еталонні параметри. Рівень значущості критерію задавати в діалоговому вікні.
4. На етапах навчання і ідентифікації передбачити можливість відбракування грубих помилок (окремих вимірів)

3. Вимоги до оформлення роботи:
Оформлена робота повинна містити:
- постановку завдання
- опис і блок-схему алгоритму її рішення
- лістинг функціональної частини програми
- результати експериментальних досліджень (помилки першого і другого роду для різних осіб)
Необхідно також продемонструвати робочу версію додатку.

4. МЕТОДИЧНІ ВКАЗІВКИ

Алгоритм аутентифікація на основі вільного / фіксованого тексту для введення

Автентифікація користувача по клавіатурному почерку можлива наступними способами:

- по набору ключової фрази (пароля), на характеристики набору якої попередньо здійснюється настройка програми, завдяки багаторазовим повторам введення з клавіатури ключової фрази;
- по набору "вільного" тексту;

Обидва способи використовують два режими роботи: режим настройки (навчання) і режим автентифікації. У режимі настройки розраховуються і запам'ятовуються еталонні характеристики набору користувачем ключових фраз. У режимі автентифікації здійснюється розрахунок тимчасових параметрів користувача і порівняння отриманих результатів з еталонними на основі чого приймається рішення про легальність користувача.

Тимчасові інтервали між натисканням клавіш на клавіатурі і час утримання (натискання) клавіш дозволяють досить однозначно охарактеризувати почерк роботи користувача на клавіатурі, що підтверджується рядом експериментів. При цьому часові інтервали між натисканням клавіш характеризують темп роботи, а час утримання клавіш характеризує стиль роботи з клавіатурою (високо підняті руки -різкий удар, низько лежачі руки - плавне натискання). Сучасні операційні системи дозволяють проводити виміри часу з дуже високою точністю, що дозволяє точно визначити тимчасові характеристики користувачів при роботі з клавіатурою.

Однак існує коло обмежень на застосування даного способу на практиці. Застосування способу автентифікації по клавіатурного почерку доцільно тільки по відношенню до користувачів з досить тривалим досвідом роботи з комп'ютером і сформованим почерком роботи на клавіатурі, тобто до програмістів, секретарям і т.д. Інакше вірогідність неправильного розпізнавання «легального» користувача істотно зростає і робить непридатним даний спосіб автентифікації на практиці. Виходячи з теорії машинопису, можна визначити час становлення почерку роботи з клавіатурою $\mathbf{t_{min}}$, при якому досягається мінімально необхідна ймовірність автентифікації користувача $\mathbf{P_{min}}$ приблизно в 6 місяців.

Тимчасові інтервали між натисканням сусідніх букв при наборі ключової фрази, як правило, підпорядковуються нормальному закону розподілення. Якщо часові інтервали підкоряються нормальному закону розподілу, то існують алгоритми для:

- побудови довірчих інтервалів математичного очікування і дисперсії;

- перевірки гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей при допущенні про рівність дисперсій в генеральних сукупностях;
- перевірки гіпотези про рівність дисперсій двох нормальних генеральних сукупностей;
- виключення грубих помилок у спостереженні;

Розглянутий нижче спосіб автентифікації користувача за клавіатурного почерку припускає автентифікацію користувача по набору ключовою фрази. Даний вибір обумовлений тим, що застосування систем автентифікації користувача «на льоту» (в процесі роботи з комп'ютером) в даний час пов'язаний з певним колом проблем, і в першу чергу з тим, що в наслідок розвитку технології "мультимедія» час роботи користувача безпосередньо з клавіатурою надзвичайно мало. Крім того, застосування існуючих статистичних методів дозволяє отримати прийнятну ймовірність автентифікації навіть при невеликій довжині ключової фрази (близько 6-8 символів). Це пов'язано з тим, що у користувача з'являються ознаки автоматизму в наборі ключової фрази невеликої довжини, а при використанні такої фрази в плині тривалого періоду часу, ймовірність автентифікації користувача може істотно підвищуватися і, як показують проведені експерименти, досягає 97%. З іншого боку, не рекомендується використовувати занадто довгі вирази в якості ключової фрази, так як це призводить до того, що користувач починає «осмислено» виконувати набір тексту, що призводить до зниження ймовірності.
Автентифікація користувача здійснюється шляхом вирішення задачі виключення грубих помилок у спостереженнях і рішення задачі перевірки гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей.

Алгоритм рішення задачі виключення грубих помилок у спостереженнях:

Нехай **y = {y_1, y_2, ..., y_n}** - множина тимчасових інтервалів між натисканням клавіш, п - число тимчасових інтервалів.
Необхідно вирішити задачу перевірки значимості елементів **y_i** множини **y (i = 1,.. n)**. Для цього сформуємо множину **y '= y ^ {y_i}**, тобто всі математичні параметри розраховуються без урахування елемента **y_i** в вихідної множині **y`**$_n$ **n`**- число елементів у множині **y`**, **n`= n-1.**
Розраховуємо математичне очікування **M_i**

$$M_i = \frac{\sum_{k=1}^{n'} y'_k}{n'};$$

розраховуємо дисперсію

$$S_i^2 = \frac{\sum_{k=1}^{n'} (y_k - M_i)^2}{n' - 1};$$

Розраховуємо середньоквадратичне відхилення

$$S_i = \sqrt{S_i^2};$$

Розраховуємо коефіцієнт Стьюдента $\mathbf{t_p}$

$$t_p = \left| \frac{y_i - M_i}{S_i} \right|;$$

Для числа ступенів свободи **n`-1** і рівня значимості **α = 0.05** визначаємо табличний коефіцієнт Стьюдента $\mathbf{t_T}$. Якщо $\mathbf{t_p} > \mathbf{t_T}$ то елемент $\mathbf{y_i}$ - відкидається і розрахунок починається знову до тих пір, поки **i ≠ n**. В іншому випадку елемент $\mathbf{y_t}$ оголошується значимим.

Алгоритм рішення задачі перевірки гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей.

Вирішуємо задачу перевірки однорідності двох вибіркових дисперсій. Нехай **n** - число елементів у множині тимчасових інтервалів.

- S_1^2 - Вибіркова дисперсія для еталонних параметрів.
- S_2 - Вибіркова дисперсія для безлічі тимчасових інтервалів, отримана в режимі автентифікації, тоді

$$S_{\max}^2 = \max\left(S_1^2, S_2^2\right); \quad S_{\min}^2 = \min\left(S_1^2, S_2^2\right);$$

Розраховуємо коефіцієнт Фішера - розрахунковий $\mathbf{F_p}$

$$F_p = \frac{S_{\max}^2}{S_{\min}^2};$$

Для числа ступенів свободи **n-1** і рівня значимості **α = 0.05** визначаємо табличний коефіцієнт Фішера $\mathbf{F_T}$. Якщо $\mathbf{F_p} > \mathbf{F_T}$, то дисперсії неоднорідні, в іншому випадку дисперсії рівні.

Вирішуємо задачу перевірки гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей.

Нехай $\mathbf{y = \{y_1, y_2, ..., y_n\}}$ - множина тимчасових інтервалів між натисканням клавіш отриманих в режимі автентифікації, **n** – число тимчасових інтервалів. Нехай $\mathbf{\kappa_e}$ - число еталонних спроб набору ключової фрази, $\mathbf{x_\alpha = \{x_1, x_2, ..., x_n\}}$ - множина еталонних часових інтервалів ($\mathbf{\lambda = 1, k}$), тоді

$$M_{x_\lambda} = \frac{\sum_{i=1}^{n} x_{\lambda_i}}{n}, \qquad M_y = \frac{\sum_{i=1}^{n} y_i}{n};$$

Тепер необхідно перевірити, чи можна з певною ймовірністю p вважати, що розбіжність між $\mathbf{M_{xy}}$ і $\mathbf{M_y}$ викликані випадковими причинами. Для цього по таблиці розподілу Стьюдента (рівень значимості $\boldsymbol{\alpha = 1 - p}$, число ступенів свободи **n-1**) визначимо табличне значення $\mathbf{t_T}$, після чого розрахуємо величину $\mathbf{t_p}$. Якщо $\mathbf{t_p > t_T}$, то розбіжність не випадкова.

$$S^2_{x_\lambda} = \frac{\sum_{i=1}^{n} (x_{\lambda_i} - M_{x_\lambda})^2}{n-1}; \qquad S^2_y = \frac{\sum_{i=1}^{n} (y_i - M_y)^2}{n-1};$$

$$S = \sqrt{\frac{((S^2_{x_\lambda})^2 + (S^2_y)^2) \cdot (n-1)}{2n-1}};$$

$$t_p = \frac{\left| M_{x_\lambda} - M_y \right|}{S \cdot \sqrt{\frac{2}{n}}};$$

Нехай **r** - число позитивних рішень задачі, тобто що розбіжність випадкова, тоді $\mathbf{P = r / \kappa_e}$ - оцінка вірогідності, що користувач є автором еталонних характеристик.
Необхідно пам'ятати, що з плином часу клавіатурний почерк користувача буде зазнавати зміни. У цьому випадку рекомендується після кожної успішної спроби автентифікації зберігати розраховану величину S^2_y в списку еталонних характеристик користувача.

Розрахунок помилок 1-го та 2-го роду.

На підставі отриманих результатів з перевірки гіпотез, необхідно розрахувати помилки 1-го та 2-го роду.
Помилки 1-го роду визначаються як вірогідність того, що легітимний користувач не буде вірно ідентифікований:

$\mathbf{P_1 = N_1 / N_0}$

Де $\mathbf{N_0}$ – загальна кількість спроб пройти процедуру автентифікації легітимним користувачем, $\mathbf{N_1}$ – кількість невдалих спроб.
Помилки 2-го роду визначаються як вірогідність того, що нелегітимний користувач буде ідентифікований як легітимний:

$\mathbf{P_2 = N_2 / N_0}$

Де $\mathbf{N_0}$ – загальна кількість спроб пройти процедуру автентифікації нелегітимним користувачем, $\mathbf{N_2}$ – кількість вдалих спроб.

5. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Бочаров П. П., Печинкин А. В. Теория вероятностей. Математическая статистика. - 2-е изд. - М.: ФИЗМАТЛИТ, 2005. - 296 с. .
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. Учеб. пособие для втузов.— 2-е изд., стер.— М.: Высш. шк., 2000.— ISBN 5-06-003830-0. 480 с: ил.
3. Мыльников С.В. Азы биометрии.. – М: Н-Л, ISBN 978-5-94869-040-7; 60 с: 2007.
4. Лакин Г. Ф. Биометрия: Учебное пособие для биол. спец. вузов- 4-е изд., перераб. и доп. М.: Высш. шк., 1990. — 352 с.: ил.