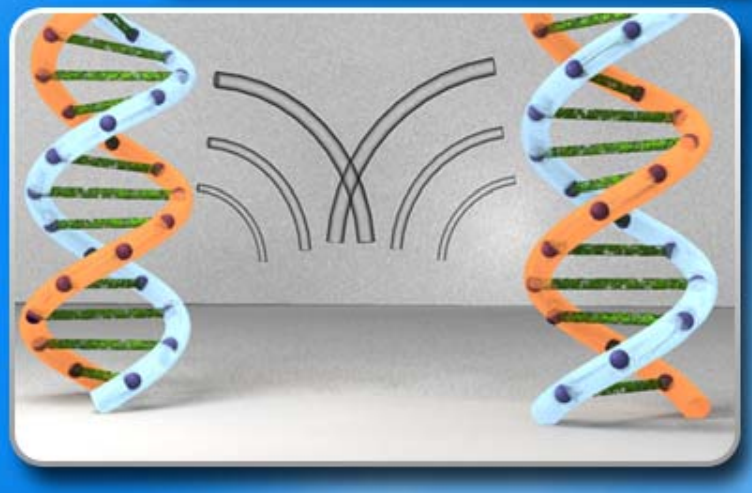
GECCO HUMIES - 2009

Behavior based User Authentication on Smart Phones



Muhammad Shahzad[1], Saira Zahid[1], Syed Ali Khayam[1,2], Muddassar Farooq[1]

[1] Next Generation Intelligent Networks Research Center
National University of Computer & Emerging Sciences
Islamabad, Pakistan
http://www.nexginrc.org

[2] School of Electrical Engineering & Computer Sciences
National University of Sciences & Technology
Islamabad, Pakistan
http://wisnet.seecs.edu.pk

Citations

Muhammad Shahzad, Saira Zahid, Muddassar Farooq, "A Hybrid GA-PSO Fuzzy System for User Identification on Smart Phones", Genetic and Evolutionary Computation Conference (GECCO), July, 2009, Montréal, Canada. (In Press)

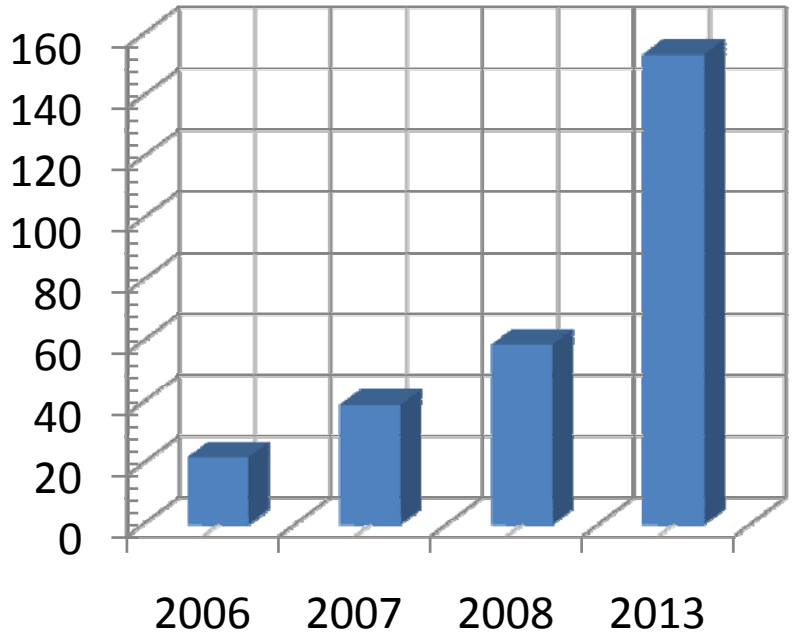
Saira Zahid, Muhammad Shahzad, Syed Ali Khayam, Muddassar Farooq, "Keystroke-based User Identification on Smart Phones", 12th International Symposium on Recent Advances in Intrusion Detection (RAID), Sept, 2009, Brittany, France. (In Press)

Smart Phones Market

BCC Research group published a 149 page report : “Global Market for Smart Phones and PDAs” in May 2009. Its digital copy costs $4850

- 2008: Global market of Smart Phones generated $58.7 billion
- 2013: expected to increase to $153.3 billion

SUMMARY FIGURE
PROJECTED GLOBAL SALES FOR SMARPHONES, 2006-2013
($ MILLIONS)



User Identification on Smart Phones: A Hot Issue

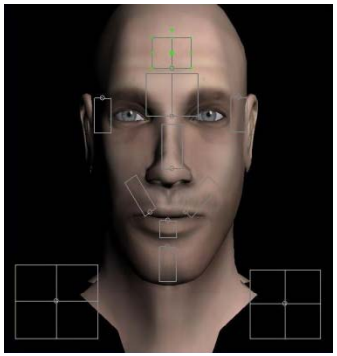
Mobile computing devices combine three extremely potent concepts

Information Security on mobile phones is a serious concern of both users and manufacturers

- Red herring mobiles scream for help: UK-based mobile security company adds security to mobile phones, October 2006.
- Ernst and Young Global Information Security Survey 2005 : Report on the Widening Gap

Its in your Hands, like its in your Eyes and Face

It is believed that keystrokes of people are distinct from each other just like their faces, finger prints, and eyes



Doesn't require any extra hardware for identification

Keystrokes based Analysis: A long sought solution

A lot of research effort was focused towards keystroke-based user identification in 80s, 90s, and the current decade

Most of the work has been done in desktop domain

In 80s

- Umphress and Williams 1985 6% FAR
- Legget and Williams 1988 5% FAR

In 90s

• Bleha, Slivinsky, and Hussein	1990	8.1%	FRR,	2.8%	FAR
• Joyce and Gupta	1990	16.67%	FRR,	0.25%	FAR
• Legget and Williams	1991	11.1%	FRR,	12.8%	FAR
• Obaidat and Sadoun	1999				

Recently Clarke and Furnell in 2007, worked on Mobile Phones 18%FAR, 29% FRR

Keystrokes based Analysis: A long sought solution

The accuracy of existing solutions is quite far from acceptable

All the work done till now has been in static controlled environments and the best results had 5% error rate

We tested the schemes of earlier researchers on mobile dataset

- Provided abysmal accuracy: Even the best had 30% errors in our dynamic uncontrolled environment

Multiplexed Keypads

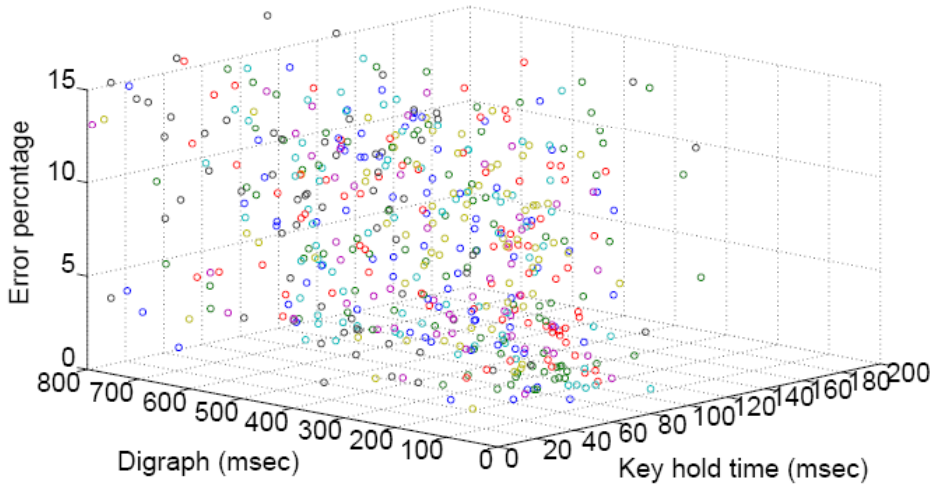
A Problem of Bio-inspired Classification

No one realized that this is a problem of bio-inspired classifier

- Every one was using machine learning classifiers
- Once we realized that conventional classifiers are not working, we decided to go unconventional

We designed a novel bio-inspired classifier and the improvement in results was not incremental: the accuracy improved **dramatically**

Diffusion in the dataset was handled using **Fuzzy Logic**



A Problem of Bio-inspired Classification

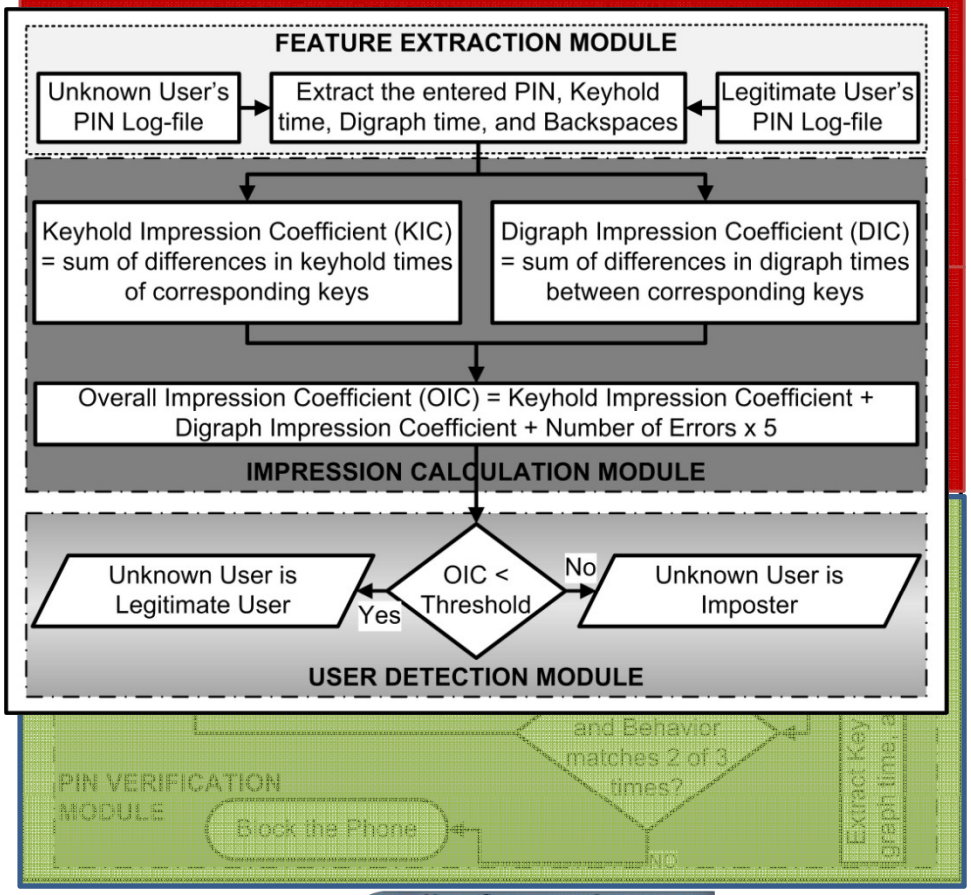
Used GA, PSO and Hybrid of PSO and GA to optimize fuzzy

(F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered.

Our proposed scheme reduced the errors to 2%

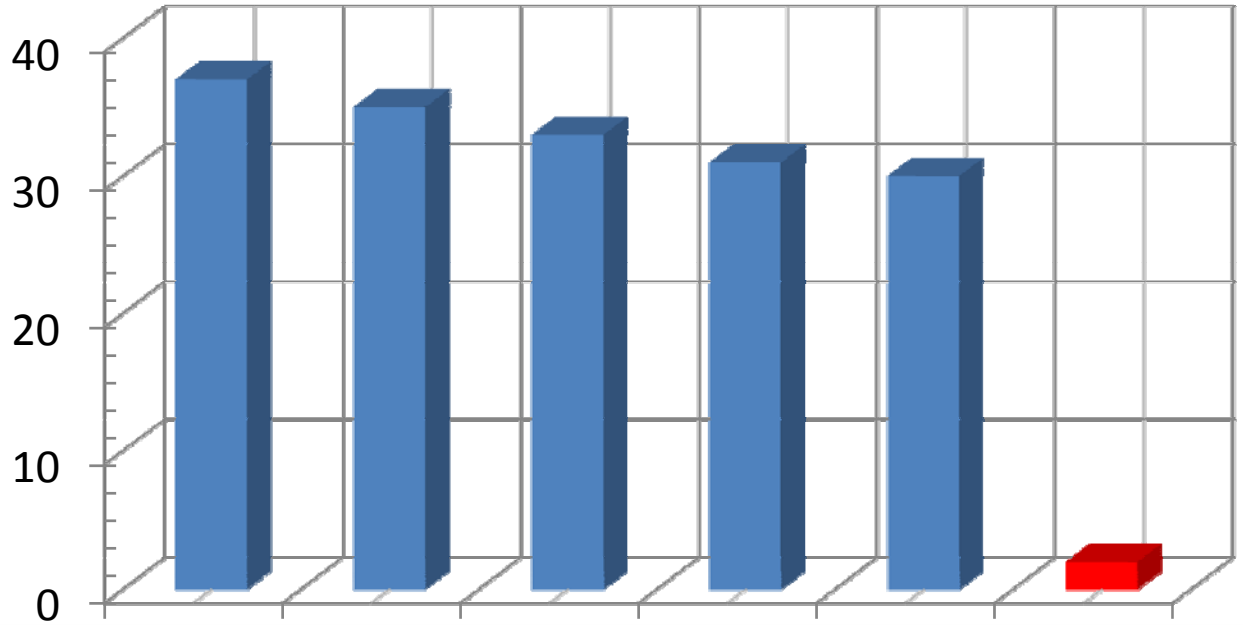
- In static Analysis we reduced errors to 0%

Algorithm for Verification Mode



10

Error %ages of classifiers on Mobile Datasets



(E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions.

The Best Entry

GECCO 2009 – Anonymous Reviewers' Comments

- I like this paper, it is clearly written with a sophisticated command of mechanics and style, it addresses a **real world problem** using **real world data**, uses **sound methodology**, and shows **promising results**.
- In any case, the classifier obtained by optimizing fuzzy rules using a hybrid GA/PSO method is quite interesting and might be worth a separate evaluation as a stand-alone classification tool, independent of the application.

RAID 2009 – Anonymous Reviewers' Comments

- The work is both **deep**, demonstrating an improvement over approaches that have been used on desktop machines, and **broad**, considering how to use their approach, and describing limitations of that approach.
- I thought that the topic was interesting and relevant, and that the constrained keypad of a mobile phone provides a natural opportunity! for good results.

The Best Entry

We collected keystrokes data of 25 users from diverse backgrounds

- It was very hard to convince people because of privacy concerns
- This is the first dataset of its kind available freely at http://www.nexginrc.org

Our problem addresses security so lets see how this system is taken in the community of its interest

- Paper accepted in RAID 2009: the conference that ranks among the top 5 security conferences in world

The problem is so hot that even RAID 2008 had a research paper on keystroke

(G) The result solves a problem of indisputable difficulty in its field.

Prototype Version

Prototype Version of the product is ready and you are welcomed to test it at the end of presentation session