GA-Based User Identity Management System

Prof. Dipankar Dasgupta, IEEE Fellow

Director: Center for Information Assurance

Center website: cfia.memphis.edu
Dipankar Dasgupta’s Ph.D Thesis
Structured Genetic Algorithms in Search & Optimization, 1993

BOOKS PUBLISHED

1997
Evolutionary Algorithms in Engineering Applications

1998
Artificial Immune Systems and Their Applications

2008
Immunological Computation: Theory and Applications

2017
Advances in User Authentication
My Research Publications

Dasgupta’s research citation statistics as shown in Google Scholar (accessed on July 7, 2018).

Conducting multidisciplinary/collaborative research
Agenda

- User Identity Verification: Authentication
- Multi-Factor Authentication (MFA)
- Active/Continuous Authentication
- Adaptive Multi-Factor (A-MFA)
  - Overview: Goal & Objectives
  - Design of A-MFA Framework
  - A-MFA Prototype System
  - Use Cases for A-MFA
- Cyber Identity Ecosystem
- Summary
Authentication

- Authentication is the critical safe guards against illegal access to computing systems.
  - the process of giving individuals access to system objects based on their identity.

- Ensures that the individual is who he or she claims to be.
  - But says nothing about the access rights of the individual.

- Challenges
  - Correctly identify authorized users in particular Operational Settings.
  - Take appropriate action on demand basis to prevent unauthorized access.
Password-Based Authentication

- Single-factor
  - Username-password. (most widely used as of now!)

- Issues
  - Mostly targeted by the attackers
  - If this single channel is compromised, the users are denied of the service until it is restored
  - Recent advancement of computer processing power, makes to check all possible cases in a short amount of time
  - Difficult to remember for a wide variety of websites
Need for Multi-Factor: Sample Scenario

Currently use the same authentication method to authenticate the user.
What the User knows
- Password, PIN, pass phrases

What the User has
- Smart card, digital certificate, driver’s license

Who the User is
- Fingerprint, iris scan, voice recognition

Where the User is
- GPS, IP address of user

Two Factor
- Generally Password along with SMS for verification code
Authentication Types

- IP-Geolocation
- Password
- SMS
- Transaction Signing
- Smartcard
- Biometrics
Product - Microsoft Azure

1. Users sign in from any device using their existing username/password

2. Users must also authenticate using their phone or mobile device before access is granted.
## Different MFA products in Market Today

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Vendor</th>
<th>Factors</th>
<th>Features</th>
<th>Source (Website)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SecureAuth IdP</td>
<td>SecureAuth</td>
<td>Two factors and SSO (out of 20)</td>
<td>Mobile, cloud, web or VPN</td>
<td><a href="http://www.secureauth.com">www.secureauth.com</a></td>
</tr>
<tr>
<td>RSA SecureID</td>
<td>RSA</td>
<td>Two factors</td>
<td>Software (smartphones, tablets and PC) and hardware authenticators</td>
<td><a href="http://www.emc.com/security/rsa-securid.htm">http://www.emc.com/security/rsa-securid.htm</a></td>
</tr>
<tr>
<td>Safenet</td>
<td>SafeNet</td>
<td>Two factors</td>
<td>Cloud, Password + SMS/Hardware Token</td>
<td><a href="http://www.safenet-inc.com/multi-factor-authentication/?tabnum=2">www.safenet-inc.com/multi-factor-authentication/?tabnum=2</a></td>
</tr>
<tr>
<td>SecurEnvoy</td>
<td>SecurEnvoy</td>
<td>Two Factor</td>
<td>Tokenless (One-swipe, SMS Preload, Soft Token, Voice Call, Email Preload)</td>
<td><a href="http://www.secureenvoy.com/">www.secureenvoy.com/</a></td>
</tr>
<tr>
<td>Symantec O3</td>
<td>Symantec</td>
<td>Cloud identity and access control (Two Factor authentication)</td>
<td>Cloud applications (set policies for groups, persons, devices) [security control point]</td>
<td><a href="http://www.symantec.com/page.jsp?id=O3">www.symantec.com/page.jsp?id=O3</a></td>
</tr>
<tr>
<td>Microsoft Azure</td>
<td>Microsoft</td>
<td>Multi factor (Phone call, SMS and Password)</td>
<td>On premises and cloud authenticationsMobile Device + user-id and password</td>
<td>azure.microsoft.com/en-us/services/multi-factor-authentication/</td>
</tr>
<tr>
<td>Deepnet DualShield</td>
<td>Deepnet Security</td>
<td>Two factors out of 10 different methods</td>
<td>SMS, Voice, Mobile App, Face, Keystroke, Smart Cards</td>
<td><a href="http://www.deepnetsecurity.com/products/dualshield/">www.deepnetsecurity.com/products/dualshield/</a></td>
</tr>
<tr>
<td>Swivel Secure</td>
<td>Swivel Secure</td>
<td>SSO + two factor</td>
<td>Mobile App, SMS, tokens, Telephony, Browser</td>
<td><a href="http://www.swivelsecure.com/">www.swivelsecure.com/</a></td>
</tr>
<tr>
<td>miniOrange Strong Authenticator</td>
<td>miniOrange</td>
<td>SSO + two factor</td>
<td>14 different authentication types</td>
<td>miniorange.com/strong_auth</td>
</tr>
</tbody>
</table>
Current MFA trends

Effectiveness of MFA as a potent tool to tackle BYOD security complexity benefits the market.

Rise in smartphone thefts spurs use of MFA on mobile devices.

Cloud services need MFA to establish customer trust and increase cloud adoptability.

- Amazon, Google, Yahoo, Dropbox, Facebook, LinkedIn, Twitter, Microsoft uses two factors to access their online services for authentication.


Ballooning Demand for Public Cloud Services Expands the Addressable Market for MFA

Global Market for Public Cloud Computing Services (In US$ Million) by Geographic Region

<table>
<thead>
<tr>
<th>Region</th>
<th>2015</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>120293.9</td>
<td>191534.6</td>
</tr>
<tr>
<td>Europe</td>
<td>75980.74</td>
<td>29673.46</td>
</tr>
<tr>
<td>Asia</td>
<td>47203.54</td>
<td>55110.19</td>
</tr>
<tr>
<td>Rest of World</td>
<td>13743.47</td>
<td>7461.726</td>
</tr>
</tbody>
</table>

37% of organisations now use multi-factor authentication for a majority of employees – up from 30% last year.

By 2016, 56% of organisations expect the majority of users to rely on multi-factor authentication.
Why we should care?

Aside from the fact that all companies should take their customer data security seriously, not having adequate authentication mechanisms in place increases the potential of corporate PII breach risks including:

- **Legal Liability**
  - Government Enforcement Action
  - Class Actions
  - Individual Actions

- **Reputational Exposure**

- **Business Consequences**

- **Typical Breach Costs**
  - Outside Counsel
  - Credit Monitoring
  - Security & Technology upgrades
  - Defence costs
  - Fines
  - Settlements

- **Sec/Shareholder Issues**
- **Employee/Customer Issues**
Use of Multi-factor Authentication (MFA)

- Provide different choices to the user during authentication to verify their identity.
  - However, all the factors may not be available in all operating conditions.
- Come with a fail-safe feature in case of any authentication factor gets compromised
  - Users should be authenticated utilizing the other non-compromised modalities.

• Concerns:
  - How to choose a better set of authentication factors out of all possible choices in any given operating environment.
  - The choice of an appropriate set of authentication factor determines the performance of the MFA
The selection procedure should not follow (having bias towards) any pattern that can be used by the attackers.

The process should make the consideration of previous selection of the authentication factors to avoid repetitive use of the same factors.

### Modes of Auth. Factor Selection

<table>
<thead>
<tr>
<th>Static</th>
<th>A predefined set of modalities for any given environment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic</td>
<td>A set of modalities chosen dynamically at different time triggering event for authentication.</td>
</tr>
<tr>
<td>Dynamic</td>
<td><strong>Random</strong> Modalities are chosen in any random order at the time of authentication.</td>
</tr>
<tr>
<td></td>
<td><strong>Adaptive</strong> Modalities are chosen based on current system settings and previously selected modalities.</td>
</tr>
</tbody>
</table>
Adaptive Multi-Factor Authentication (A-MFA)

- This greatly enhances security without changing the user experience.

- However, when an unauthorized user attempts to gain access with stolen credentials and the additional factors and behaviours normally seen don't line up, the login is prevented and challenged.

- The selection of multiple authentication factors are conducted adaptively considering

[Operating devices] [Connected Media] [Surrounding Conditions/Environment]
Biometric Presentation Attacks

Biometrics
Public Place with shared wifi
Overall Concept of A-MFA
A-MFA: Overall Goal and Objectives


Objective 1: Pairwise Preference Information on Trustworthiness
- Develop Trustworthy Model with Probabilistic Constraints
- Error rates of individual factors

Objective 2: Cardinality of the Auth. factors
- Previous history of selected factors
- Constraints for the given environment settings
- Selected Authentication factors in a given environment settings
- Pairwise Preference Information on Trustworthiness

Objective 3: Build users’ profile with different captured auth. factors
- Implement the Trustworthy Model
- Implement Adaptive Selection Approach
- Evaluation of the system with different settings

Objective 4: Build a Pilot web-service to authenticate
- Implement A-MFA
- Run user-study on Pilot Framework
**Popular A-MFA Modalities & their features**

**M1: Face Recognition**
It is computed through face geometry features. Features include different points in Lips, eyes, brow and cheek, Crows-feet wrinkles nasal root wrinkles.

**M2: Finger Print**
Three level of features are used for this modality. Level 1 features show macro details of the ridge flow shape, Level 2 features (minutiae point) are discriminative enough for recognition, and Level 3 features (pores) complement the uniqueness of Level 2 features.

**M3: Password**
Password is the most common modality. It can be stored in hashed form and matched with the input by hashing the given password as string matching. Password can be made with alpha-numeric characters along with some special characters.

**M4: CAPTCHA**
It is used to prevent automated software to perform actions and can discriminate between human and bots. a CAPTCHA features an image file of slightly distorted alphanumerical characters. It also has read out feature for users with visually impaired.
**M5: SMS**
SMS feature is used to send the pass-code to any phone number and that code is valid for a short period of time. The phone number should be registered to the system a-priori basis.

**M6: Voice recognition**
It uses pitch and different formant features (F1, F2 and F3). The pitch of the speech signal contains crucial information about the intonation pattern. The formants represent the articulators of the speech signal where the resonant frequencies are generated.

**M7: Keystroke pattern**
This modality detects the pattern of the keystrokes. The features used for this techniques are: mean latency and standard deviation of digraphs [A combination of two letters representing one sound], mean duration and standard deviation of keystrokes.
In this chart the further away the characteristic is from the center, the better is the biometric technique.

So for instance keystroke scan and signature scan are low cost, require very little effort, and are not intrusive at all, however they are not distinctive.

On the other end of the spectrum, retina scan and iris scan, provide very high distinctiveness, however they are both expensive, and intrusive.
In this work, an authentication factor is defined as

(i) Single feature of an authentication modality;
(ii) Any combination of features of an authentication modality;
(iii) Combination of multiple features of different authentication modalities.

Key Term

- $M_k (k \in \mathbb{Z}^+)$ be the $k^{th}$ authentication modality and $\{M_k : f_{k,i}\}$ be its $i^{th}$ feature.
- $\{\{M_k\} : \{f_{k,i}\}_{i \in \mathbb{Z}^+}\}_{k \in \mathbb{Z}^+}$:
- $i^{th}$ features of different combinations of $\{M_k\}_{k \in \mathbb{Z}^+}$. 
The first features of $M_1$ and $M_2$: \{M_1 : f_{1,1}\} and \{M_2 : f_{2,1}\}.

They are considered as two authentication factors (according to (i)).

\{M_1 : f_{1,1}, f_{1,2}\} is one authentication factor (according to (ii))

- combinations of \{M_1 : f_{1,1}\} and \{M_1 : f_{1,2}\}

\{M_1, M_2 : f_{1,1}, f_{2,1}\} is considered as one authentication factor (according to (iii))

- combination of \{M_1 : f_{1,1}\} and \{M_2 : f_{2,1}\}.
Calculate the trustworthy value of combined factors from individual trustworthy values illustrated.

\[ T_{\text{dim}}(M_1, f_{1,1}) + T_{\text{dim}}(M_2, f_{2,1}) + T_{\text{dim}}(M_3, f_{3,1}) + \ldots + T_{\text{dim}}(M_n, f_{n,1}) \]

\[ T_{\text{dim}}(M_1, f_{1,1}) \sim N(0,1) \quad T_{\text{dim}}(M_2, f_{2,1}) \sim N(0,1) \quad T_{\text{dim}}(M_3, f_{3,1}) \sim N(0,1) \quad \ldots \quad T_{\text{dim}}(M_n, f_{n,1}) \sim N(0,1) \]

\[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}T_{\text{mi}}(M_1, f_{1,1})^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}T_{\text{mi}}(M_2, f_{2,1})^2} + \cdots + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}T_{\text{mi}}(M_n, f_{n,1})^2} = T_{\text{mi}}(M_1, M_2, \ldots, M_n, f_{1,1}, f_{2,1}, f_{3,1}, \ldots, f_{n,1}) \]

Trustworthy value of combined factors in a specific medium.
Machine Learning Algorithm

Evolution Environment

Genetic Algorithm Evolution Flow

Population

GA Operators

Evaluation

Fitness value

Mutation

Crossover

Reproduction
Genetic Pareto-Optimization
A Framework for A-MFA System

Adaptive Selection of multiple authentication factors

Effect of different medium

Effect of different devices

Vehicle Settings

Application Environment

Validated Users

Not Validated

VMs stores different authentication factors of users

Authentication factors (Modalities with features)

Login:

Driving Environment

UI
Some Details of A-MFA
Auth Modality Activation Pattern

Activated Modalities

Authentication Triggering Time

Time Elapsed  Device Changed  Media Changed  Time Elapsed  Device Changed  Time Elapsed  Media Changed

T1  T2  T3  T4  T5  T6  T7  T8

PWD  PWD  PWD  PWD
Illustration of Adaptive Selection Algorithm
Initial Experiments

- Dataset is created for 50 users as a test-bed for Adaptive-MFA System
  - **Face Dataset:**
    - 10 images for registration and 3~5 images for authentication purpose.
    - Faces94, faces95 dataset [1] are used
  - **Fingerprint Dataset:**
    - 3 images for registration and 2 images for authentication purpose.
    - CASIA Fingerprint Image Database Version 5.0 [2]
  - **Voice Dataset:**
    - 3 voice samples for registration and 1 voice sample for authentication.
  - **Keystroke Dataset:**
    - 5 keystroke samples for registration and 3 or more keystroke samples for authentication.
    - CMU dataset [3] is used.
  - Non-biometric data are generated programmatically.
  - The communication between client and server are done through https protocol which is basically an end-to-end encrypted communication while data-in-motion.

3. CMU dataset, Url: [http://www.cs.cmu.edu/~keystroke/](http://www.cs.cmu.edu/~keystroke/)
Multiple Factors verified
"Access granted"
## Authentication modalities incorporated in A-MFA System

<table>
<thead>
<tr>
<th>Knowledge-Based Modalities</th>
<th>Possession-Based Modalities</th>
<th>Biometric Modalities</th>
<th>Location-Based Modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Password</td>
<td>SMS Code</td>
<td>Face Recognition</td>
<td>GPS</td>
</tr>
<tr>
<td>Pass-phrase</td>
<td>TOTP Code</td>
<td>Fingerprint Recognition</td>
<td>IP address</td>
</tr>
<tr>
<td>Security Challenge Questions</td>
<td></td>
<td>Voice Recognition</td>
<td>MAC Address</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keystroke Recognition</td>
<td>Wi-Fi</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Triangulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cellular Triangulation</td>
</tr>
</tbody>
</table>
• The selection procedure should not follow (having bias towards) any pattern that can be used by the attackers.

• The process should make the consideration of previous selection of the authentication factors to avoid repetitive use of the same factors.
**Stress Test:** System accuracy given valid and imposter data and varying light and noise conditions

<table>
<thead>
<tr>
<th>Surrounding Conditions</th>
<th>Two-factor based Authentication</th>
<th>Three-factor based Authentication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valid Data</td>
<td>Imposter Data</td>
</tr>
<tr>
<td>Light</td>
<td>92%</td>
<td>0%</td>
</tr>
<tr>
<td>Noise</td>
<td>98%</td>
<td>0%</td>
</tr>
<tr>
<td>Light + Noise</td>
<td>94%</td>
<td>0%</td>
</tr>
<tr>
<td>Light - Noise</td>
<td>79%</td>
<td>8%</td>
</tr>
</tbody>
</table>
Auth-Spectra: Important Features

CONTINUOUS IDENTITY PROTECTION
Auth-Spectra is always on, ensuring the users' identity as long as they are connected to the system or service

JUST-IN-TIME DECISION
Authentication modalities are selected in real-time providing the optimal decision for any authentication attempt

CLOUD IDENTITY MANAGEMENT
Auth-Spectra provides a cloud based solution to administrate your existing services efficiently

MACHINE LEARNING APPROACH
Auth-Spectra utilizes machine learning techniques to analyze user behavior and select modalities based on intelligent algorithms

BROAD RANGE OF APPLICATIONS
Auth-Spectra is suitable for many applications including healthcare networks and online education

BUILT WITH THE FUTURE IN MIND
The Auth-Spectra framework is robust and flexible enough to meet your current and future authentication needs, such as adding new sensors and modalities

Patent # 9,912,657
Issue Date: March 6, 2018
Video of A-MFA Prototype Demo

A company using a similar Technology:
https://www.okta.com/learn/Adaptive-MFA
A-MFA invisibly can integrate hundreds of auth factors.

- Including behaviours, as an extra set of "factors".
- Evaluates if there is enough of a match with a user's known profile to allow the user to access a site or service without requiring the user to enter any additional factors.
Salehie and Tahvildari (February, 2018) introduce the questions for eliciting adaptation requirements: When to adapt? Why do we have to adapt? Where do we have to implement change? What kind of change is needed? Who has to perform the adaptation? How is the adaptation performed?
A person be authenticated on a regular interval through

- Physical aspects (example: fingerprint, face geometry, etc.)
- Interaction with the system (example: keystroke pattern, mouse movement, etc.)
- Existing context of the user (example: structural semantic analysis, forensic authorship, etc.)
- Experienced data usage (example: computational linguistics)
User Identity Profiling for Continuous Authentication

The system then re-checks using sensors that passed until the locally defined threshold is met.

Keystroke sensor reports "Not you?"

Modalities relating to how you behave
- Mouse Movement
- Keystroke Pattern

Modalities relating to the context you exist in
- Forensic Authorship
- Structural Semantic Analysis

Background Authentications Over Time

- 0 sec: User logs in
- At 3 sec: Using the mouse
- At 7 sec: Typing in Outlook
- 10 sec
- 12 sec: Updating a Document
- 13 sec
- 16 sec
- 19 sec
- 22 sec
- 25 sec
- 28 sec
- 31 sec
Identity Ecosystem Steering Group (IDESG)

IDESG Members

- More than 350 members and 65 universities over 12 countries

- Private sector group that works under the National Strategy for Trusted Identities in Cyberspace (NSTIC) initiative toward the goal of creating a trust framework that can replace passwords, allow individuals to prove online that they are who they claim to be, and enhance privacy [3].

- Identity Ecosystem Framework
  - A set of three core documents that describe the Identity Ecosystem and requirements, best practices, and approved standards needed to be considered in compliance with it.

A-MFA Applications:

- Continuous, high-confidence, identity authentication for:
- Banking, including online funds transfer
- Online testing in education and training settings
- Secure access to Electronic Medical Records
- Access to Sensitive sites by government employees and others.
- Internet of Things (IoT) sensory data access.
- Use in Blockchain Technology for access verification to Hyper ledger.
- Specific web services such as PayPal, Netflix and other paid services.

Deployable at different levels of Internet Computing:

- Application level (financial applications, email/business/personal applications, social applications)
- User level (root user, administrators, guest user)
- Document level (pdf containing application form, document containing proprietary information, image/video containing confidential and sensitive footage)
2018 IEEE Symposium on Computational Intelligence in Cyber Security (CICS 2018) at
2018 IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTELLIGENCE (IEEE SSCI 2018)
November 18- November 21, 2018, Bengaluru, India.
URL: http://ieee-ssci2018.org/cics.html/

DEADLINES:
• Special Track/Session Proposal: April 5, 2018
• Paper Submission: July 23, 2018

Symposium Chair: Dipankar Dasgupta, IEEE Fellow, The University of Memphis, USA
Co-Chair: Marco Carvalho, Florida Institute of Technology, USA
Co-Chair: Shamik Sural, Indian Institute of Technology, Kharagpur, India
THANK YOU!