Title: Applying a Modified Discrimination Model to Enhance Defense and Sensor Systems Security

Version: *"edit22" March 18,2013*

Primary Topic-10: Cyberspace Management

Alternate Topic-8: Networks and Networking

Alternate Topic:-7: Architecture, Technology, and Tools

Submitted to: DoD's 18th International Common, Control Research Technology Symposiums, June 19-21, 2013, Alexandria Virginia

By: Dr. Buddy H. Jeun, Ph.D (Engineering)
Sensor Fusion Technology, LLC
4522 Village Springs Run
Dunwoody, GA 30338
mail: jeunb@bellsouth.net
Telephone 678-662-9556

And John Younker, M.S (Engineering)
Sensor Fusion Technology, LLC
4522 Village Springs Run,
Dunwoody, GA 30338
email: john.ynkr@gmail.com
telephone: 678-223-3205

Primary Point of Contact: Dr. Buddy H. Jeun,

Abstract:

National defense security systems constantly face cyber attack, espionage, and hacking. The primary objective of this paper is to explore the application of a modified discrimination model to replace current encrypted user id and password authentication.

This discrimination model will be based on minimum distance as compared to the traditional discrimination model which is base on maximum probability.

The modified discrimination model treats user id and password as a multi-sensor information fusion technology problem. The model converts the input user id and password into a digital pattern feature vector. The model then processes the newly converted vector for distance between all known feature vectors store in the secure knowledge database. The new pattern vector with the minimum distance generate by the modified discrimination model will be the authorized person.

The new application will be demonstrated using mathematical simulated data. The new application will be verified by comparing its results with the widely known multi-sensor correlation model from multi-sensor information fusion technology

Contents

Introduction

The primary objective of this paper to explore the application of a modified discrimination model to enhance defense and sensor systems security. In an unstable global environment with regional political and ideological conflicts, our national security and sensor systems are under constant cyber attack, espionage, and hacking.

The current method of protecting sensitive information is using user id and password as used in personal computers, institutional, governmental, and national defense systems. Correct user id and password is required to access secured information. However, user id and password can be falsified and hacked. The proposed application provides a technical solution to protect the user id and password method by employing a modified discrimination model to provide a positive verification of user id and password.

An architectural block diagram consisting of seven modules explains the proposed application:

1) user id and password input module
2) sensor conversion module
3) feature vector extraction module
4) knowledge database management module
5) modified discrimination model
6) decision module
7) positive Identification module

The functions of the modules are, at any given time, input user id and password into the conversion module. User id and password are treated as a multi-sensor information fusion technological problem. Convert user id and password to a digital feature vector. Transfer the extracted feature vector to the modified discrimination model. Compare the distance between this new converted feature vector and all feature vectors in the knowledge database. Finally, select the input feature vector with the shortest distance as the true feature vector.

For purposes of explaining the new application, this presentation will apply the traditional multi-sensor correlation model to the same problem. When applying that model, the feature vector having correlation coefficient of “one” will be verified as the true feature vector. This verification process further demonstrates the proper function of the modified discrimination model as it provides positive identification of the true feature vector for the user id and password.

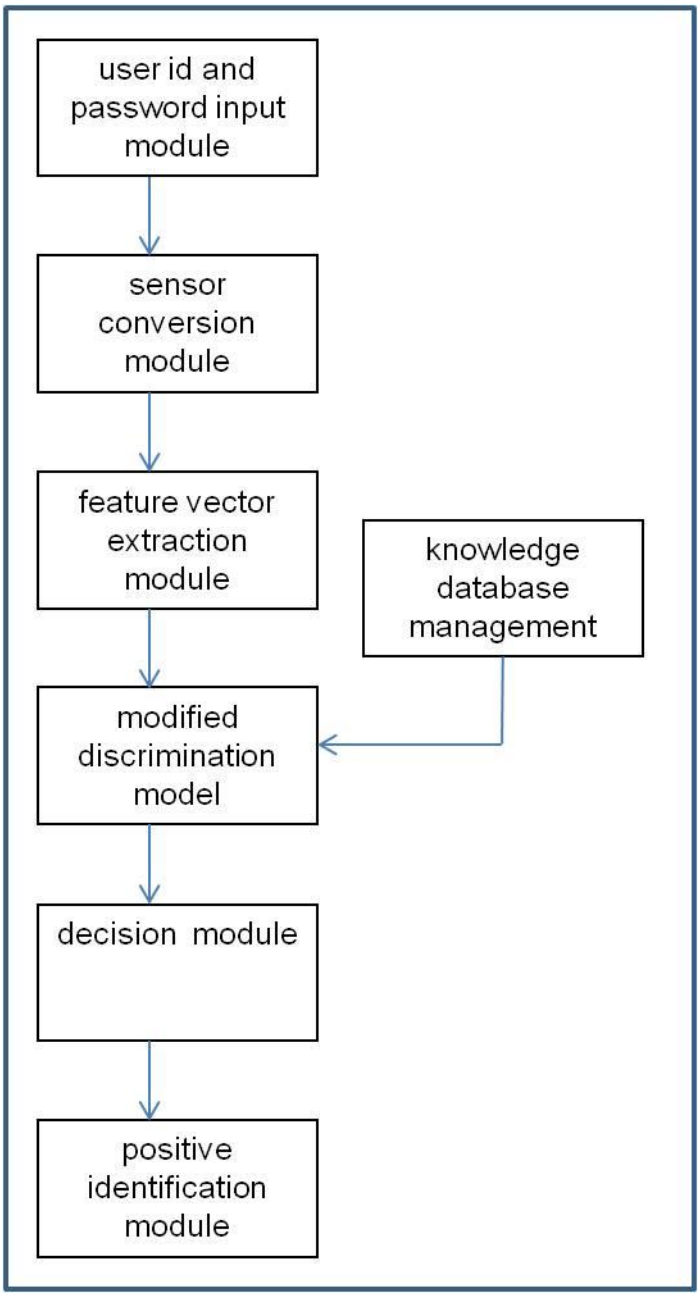

Figure 1 Modules

In the following sections of this paper, the modified discrimination model will be expressed as the function D. The multi-sensor correlation model will be expressed as the function R.

Traditional Discrimination Model

The traditional discrimination model has been known for a long time in the fields of science, engineering, and sociology. IBM's SPSS and UCLA's BMPD are commercial computer software packages that use the discrimination model for analysis.

The Bayesian conditional probability theory is one of the classical discrimination models. Mathematically, the Bayesian probability model for the discrimination and classification application can be expressed as following:

$$\boldsymbol{P}\left(\frac{\boldsymbol{T_k}}{\boldsymbol{S}}\right) = \frac{\left\{\boldsymbol{P}\left(\frac{\boldsymbol{S}}{\boldsymbol{T_K}}\right) \times \boldsymbol{P}(\boldsymbol{T_K})\right\}}{\left\{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left[\boldsymbol{P}\left(\frac{\boldsymbol{S}}{\boldsymbol{T_i}}\right) \times \boldsymbol{P}(\boldsymbol{T_i})\right]\right\}}$$

where:

$\boldsymbol{P}\left(\frac{\boldsymbol{T_k}}{\boldsymbol{S}}\right)$ is the probability of T_k given that T_k is in S

$$\boldsymbol{P}\left(\frac{\boldsymbol{S}}{\boldsymbol{T_k}}\right) = \frac{1}{(n \times \sqrt{2\pi})} \times e^{-\{(T-Y)\times(T-Y)^T\}}$$

$\boldsymbol{T} = (t1, t2, t3, \ldots tn)$ as feature vector in S

$\boldsymbol{Y} = (y1, y2, y3, \ldots yn)$ as feature vector in S

$\boldsymbol{P}(T_k)$ is the probability density function of T_k

In general, the Bayesian model produces a very accurate result for the unit variate normal assumption. However, when the feature vector is not of multi-variate distribution, the probability function becomes unknown. Therefore, the probability estimation will be complicated, and the result will not be useful. For that reason, the traditional discrimination model needs to be modified for this application.

Modified Discrimination Model

The mathematical expression for the modified discrimination model we are proposing is as follows:

$D(X,Y) = \{ (X - Y) * (X - Y)^T \}$ [Buddy H. Jeun, 1980]

Where

X = (x1, x2, x3, ……xn) is a feature vector

Y = (y1, y2, y3, ……yn) is a feature vector

D(X, Y) is the distance between feature vector X and the feature vector Y

The properties of the model are:

D (X, Y) can be equal to or greater than zero

D (X, Y) can be equal to or less than a positive number

Mathematically, the above properties are expressed as:

$0 \leq D (X, Y) \leq \text{Delta}$

where Delta is a positive number.

The possible decisions of the model are:

1) If D(X,Y) = 0, then feature vector of X is equal to the feature vector of Y

   For example suppose:

   feature vector of X = (1, 1, 1, 1, 1, 1, 1, 1)

   feature vector of Y = (1, 1, 1, 1, 1, 1, 1, 1)

   Then, since $D(X, Y) = (X – Y)*(X – Y)^T = 0$, the feature vector of X is equal to the feature of Y

2) Otherwise, if D(X,Y) is greater than zero, then the feature vector of X is not equal to the feature vector of Y.

Knowledge DataBase of User IDs and Passwords

The knowledge database contains the true reference information for positive identification and classification of all system users. For this particular application, the user id of each allowed person can be the true full name, or social security number, or personal e-mail address as chosen by the user. The password associated with the given user id is typically a series of decimal digits with special characters as chosen by the user.

For concept demonstration and verification purposes in this presentation, simple mathematical equations will be used for the information stored in the knowledge database. We will use simply straight lines of different slope and intercept.

For example, suppose the knowledge database contains user id and password for three authorized persons X, Y and Z.

Person X will be represented by X = (2T + 3)

Given T = (1, 2, 3, 4) we have X = (5, 7, 9, 11)

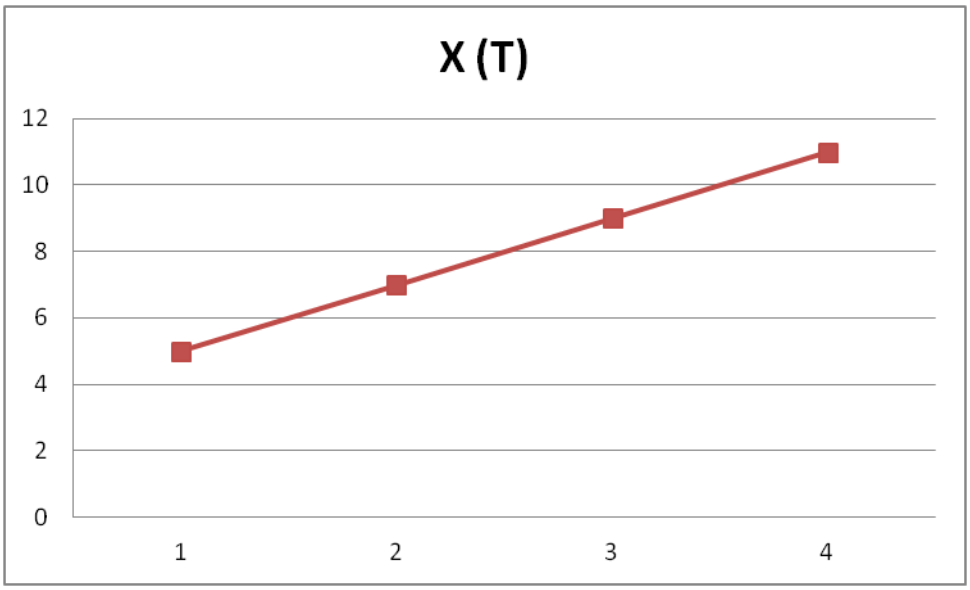

Figure 2 Person X's Account

Now convert the feature elements into binary digits as follows:

decimal 5 is (0, 1, 0, 1)

decimal 7 is (0, 1, 1, 1)

decimal 9 is (1, 0, 0, 1)

decimal 11 is (1, 0, 1, 1)

The required feature vector for person X in terms of binary digits is:

X = {0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1}

Similarly, the feature vector for person Y will be as follows:

Y = (3T – 1) as sample data.

Given T = (1, 2, 3, 4) we have Y = (2, 5, 8, 11)

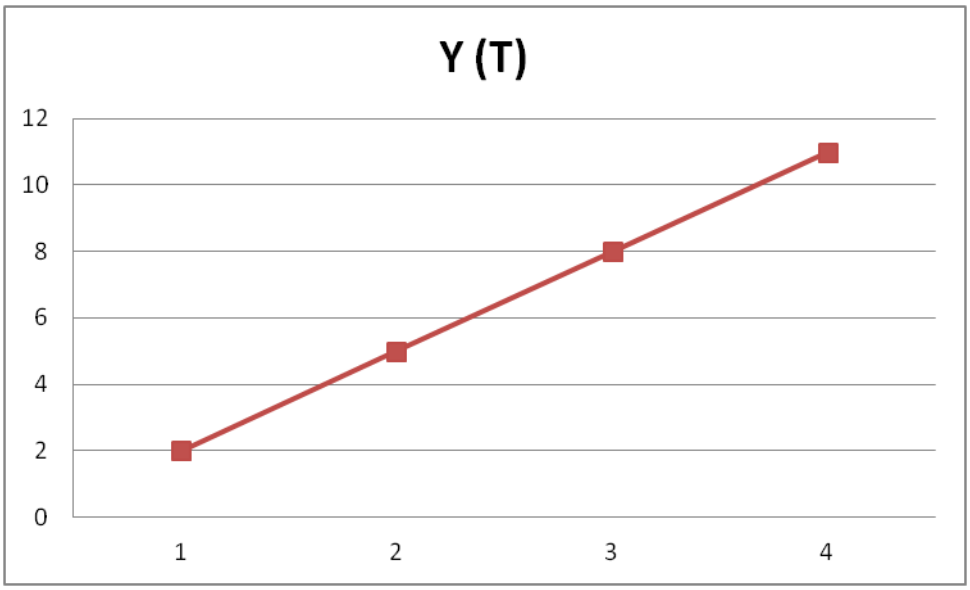

Figure 3 Person Y's Account

Now convert the feature elements into binary digits as follows:

Decimal 2 is (0, 0, 1, 0)

Decimal 5 is (0, 1, 0, 1)

Decimal 8 is (1, 0, 0, 0)

Decimal 11 is (1, 0, 1, 1)

The required feature vector for person Y in terms of binary digits will be as follows:

Y = { 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1 }

And similarly the feature vector for person Z will be generated as follows:

Z = (3T + 1) as sample data.

Given T = (1, 2, 3, 4) we have Z = (4, 7, 10, 13)

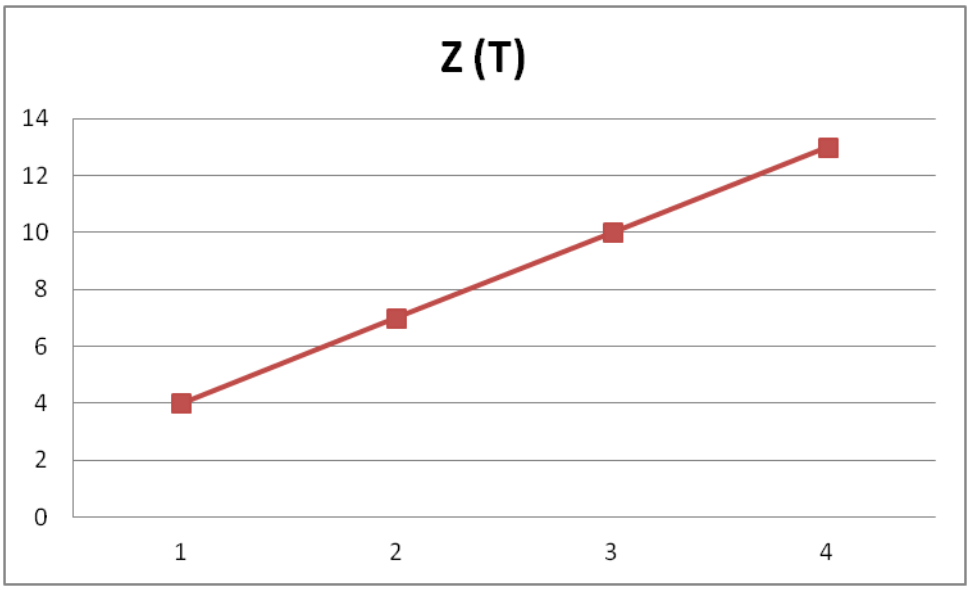

Figure 4 Person Z's Account

Converting the feature elements into binary digits as follows:

decimal 4 is (0, 1, 0, 0)

decimal 7 is (0, 1, 1, 1)

decimal 10 is (1, 0, 1, 0)

decimal 13 is (1, 1, 0, 1)

The required feature vector for person Z in terms of binary digits is:

Z = { 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1)

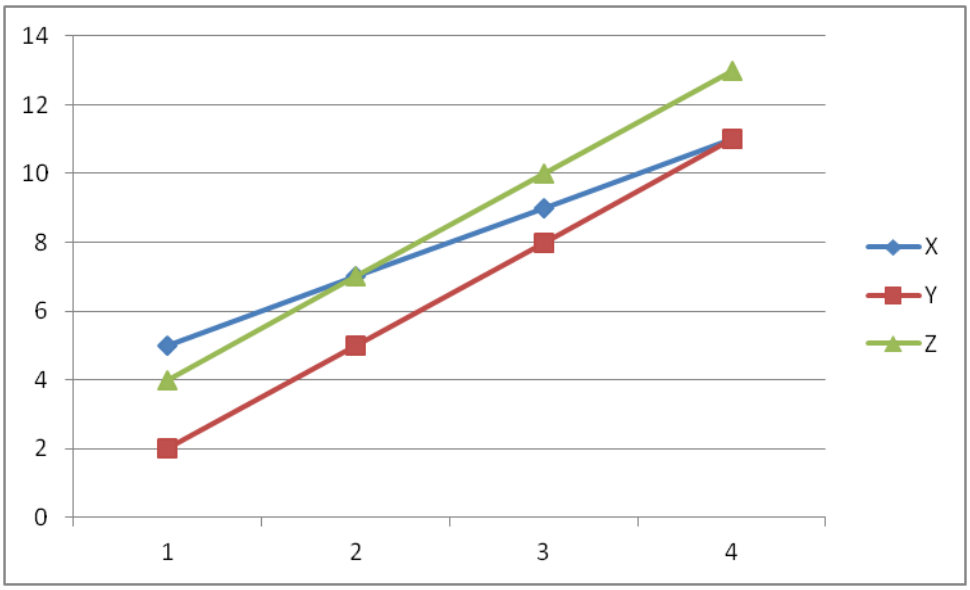

Figure 5 Three Accounts

Comparison to Multi-Sensor Correlation Model

Now look at the multi-sensor correlation model as a way of verifying the modified discrimination model.

Given feature vectors X and Y, the correlation coefficient of X and Y, can be expressed as follows: [Buddy H. Jeun, Alan Whittaker, 2002]

$$R(X, Y) = \{ (X \bullet Y) / ((X \bullet X) - (X \bullet Y) + (Y \bullet Y)) \}$$

Where :

R (X, Y) is the correlation coefficient of X and Y

X = (x1, x2, x3,…….xn) as the feature vector X

Y = (y1, y2, y3,…….yn) as the feature vector of Y

X●X is the dot product of the feature vector of X and itself

X●Y is the dot product of the feature vector of X and the feature vector of Y

Y●Y is the dot product of the feature vector of Y and itself

The properties of the multi-sensor correlation model are:

R (X, Y) is **greater** than or equal to -1 and

R (X, Y) is **less** than or equal to 1,

That is mathematically:

$-1 \leq R(X,Y) \leq 1$

The decision rules of the multi-sensor correlation model are mathematically shown by two feature vectors X and Y as follows:

X = (x1, x2, x3,…….xn)

Y = (y1, y2, y3,…….yn)

Now, if:

R (X, Y) is **equal** to **1** then X is most likely equal to Y

R (X, Y) is **less than 1** then X is most likely not equal to Y

Simulation and Verification of the Modified Discrimination Model

To demonstrate the application of the modified discrimination model, some simple mathematical simulated data is used in two cases considered as follows:

Case #1

Consider person A whose feature vector of user id and password is the same as person X's feature vector stored in the Knowledge Database.

Mathematically:

A = (0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1)

X = (0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1)

Y = (0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1)

Z = (0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1)

Substituting feature vector of A and feature vector of X, Y and Z into the modified discrimination model, we estimate the distances D(A, X), D(A, Y) and D(A, Z) as follows:

$D(A, X) = \{ (A – X) * (A – X)^T \}$

$D(A, X) = \sum\{ (A – X) \}$

$$D(A, X) = (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (1 - 1)^2$$
$$+ (0 - 0)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2$$
$$+ (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 1)^2$$

$D(A, X) = 0$

Therefore $D(A,X) = 0$

That is, the distance between feature vector A and feature vector X is zero.

Similarly:

$D(A, Y) = \{ (A - Y) * (A - Y)^T \}$

$D(A, Y) = 5$

Therefore $D(A, Y) = 5$, implying the distance between feature vector of A and the feature vector of Y is 5.

Similarly:

$D(A, Z) = \{ (A - Z) * (A - Z)^T \}$

$D(A, Z) = 5$

Therefore $D(A, Z) = 5$ implies the distance between feature vector of A and the feature vector of Z is 5.

According to the decision rule of the modified discrimination model, $D(A, X)$ is the smallest distance. Therefore, one can conclude that person A is positively identified as person X stored in the knowledge database. Person A should be allowed to access the secure information. Simulated data case #1 has demonstrated the power of positive identification.

Verification for case #1

Now the multi-sensor correlation model from the multi-sensor information fusion technology will be used to verify the accurate decision of modified discrimination model. This is done to prove that the feature vector of A is identically the same as the feature vector of X found in the secure knowledge database.

Mathematically, the multi-sensor correlation model can be expressed as follows:

$R(A, X) = \{ (A \bullet X) / ((A \bullet A) - (A \bullet X) + (X \bullet X)) \}$

Where

R (A, X) is defined as the correlation coefficient for feature vector of A and the feature vector of X.

A●X is the dot product of feature vector of A and the feature vector of X

A●A is the dot product of feature vector of A and itself

X●X is the dot product of feature vector of X and itself

Now substituting the feature vector of A and feature vector of X into the multi-sensor correlation model we get:

A = (0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1)

X = (0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1)

And A●X = (0*0) + (1*1) +(0*0) + (1*1)

+ (0*0) +(1*1) +(1*1) +(1*1)

+ (1*1) + (0*0) +(0*0) +(1*1)

+ (1*1) +(0*0)+(1*1) + (1*1)

A●X = 2 + 3 + 2 + 3

A●X = 10

Therefore A●X = 10

Similarly A●A = 10

And X●X = 10

Now substitute all of the dot products into the multi-sensor correlation model using the values given above we get:

R (A, X) = { (A●X) / ((A●A) – (A●X) + (X●X)) }

R (A, X) = (10) / (10 – 10 + 10)

R (A, X) = 10 / 10

R (A, X) = 1

Therefore, R (A, X) = 1 shows that the feature vector of A is positively identified as the feature vector of X. This decision by the multi-sensor correlation model is the same as the decision from the modified discrimination model. Both models have reached the same conclusion. That is, case #1, has been verified and the following statements are true and accurate:

(modified discrimination model) D (A, X) = 0

(multi-sensor correlation model) R (A, X) = 1

implying that the feature vector of A is positively identified as the feature vector of X.

Case #2

The feature vector of B is simulated by the simple mathematical equation as follows:

B = (T + 3)

Given T = (1, 2, 3, 4) we have B = (4, 5, 6, 7)

That is:

decimal 4 is (0, 1, 0, 0)

decimal 5 is (0, 1, 0, 1)

decimal 6 is (0, 1, 1, 0)

decimal 7 is (0, 1, 1, 1)

The feature vector of B in terms of binary digits can be represented as follows:

B = (0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1)

Now apply the modified discrimination model to see whether or not person B should be permitted to access the secure information.

Using the feature vectors of B, X, Y, and Z and substituting into the modified discrimination model produces the following values for D(B, X), D(B, Y) and D(B, Z):

$$D (B, X) = \{ (B – X) * (B – Y)^{T} \} = 9$$

$$D (B, Y) = \{ (B – Y) * (B – Y)^{T} \} = 8$$

$$D(B, Z) = \{ (B - Z) * (B - Z)^{T} \} = 12$$

Since none of D(B, X), D(B, Y) and D(B, Z) is zero, the feature vector of B does not pass the requirement set by the modified discrimination model. Therefore, person B should not be allowed access to the secure information.

Verification for case #2

As in case #1, using the multi-sensor correlation model from multi-sensor information fusion technology we verify the accuracy of case #2 as follows:

$$R(B, X) = \{ (B \bullet X) / (B \bullet B) - (B \bullet X) + (X \bullet X) \} = 4$$

$$R(B, Y) = \{ (B \bullet Y) / (B \bullet B) - (B \bullet Y) + (Y \bullet Y) \} = 3$$

$$R(B, Z) = \{ (B \bullet Z) / (B \bullet B) - (B \bullet Z) + (Z \bullet Z) \} = 2$$

Since none of the coefficients of the feature vector of B and X, Y and Z is one, person B should be denied access to the secure information. This verifies the accuracy of the decision by the modified discrimination model for case #2.

Conclusions

1. Simulated case #1, proves that the modified discrimination model can positively identify the secured person whose true user id and password is stored in the knowledge database. Therefore, those secured persons can be granted permission to access the secure information.
2. Simulated case #2, proves that for those persons who are not registered in the knowledge database, the modified discrimination model can positively identify them and deny them access to the secure information.
3. Since the modified discrimination model provides a powerful automatic system of positive identification, personal, institutional, governmental, and nationally secured information can be protected.
4. Since the modified discrimination model converts the user id and password into a feature vector based on the multi-sensor information fusion technology, and the knowledge database stored and secured the feature vector of user id and password, therefore the modified discrimination model can minimize the risks from hacking.
5. The modified discrimination model provides a computer algorithm that is easily implemented and embedded into the personal, institutional, governmental, and national security systems.

References

1. [Geoffrey J. McLachlan, 1992], Discriminant Analysis and Statistical Pattern Recognition (Wiley series in probability and statistics)

2. [Andrew R. Webb, Keith D. Copsey, 2011], Statistical Pattern Recognition, John Wiley & Son, New York

3. [Buddy H. Jeun, 1980], The Design and Implementation of An Improved Multivariate Classification Scheme. Ph.D. Dissertation, Department of Electrical Engineering, University Of Missouri, Columbia, MO.

4. [Buddy H. Jeun, Alan Whittaker, 2002] Multi-Sensor Information Technology Applied To The Development Of Smart Aircraft. 7th ICCRTS, The joint conference of the U.S. Department of Defense, and the Canadian Department of National Defense. Que'bec City, Canada.

5. [Jeun, Younker, Hung, 2003] A Nuclear Plume Detection and Tracking Model For The Advanced Airborne Early Warning Surveillance Aircraft. 8th ICCRTS, Defense University, Washington, DC. June 17-19,2003.

6. [K. Fukunaga, 1972] Introduction To Statistical Pattern Recognition, New York, Academic, 1972

7. [P. A. Lacenbruch, 1975] Discrimination Analysis, Hafner Press.