3-27-2015

A Novel Behavior-Based User Authentication Scheme

Zeyu Wu
Indiana University - Purdue University Fort Wayne

A Password-less User Authentication Scheme

Zeyu Wu (Student) and Anyi Liu (Faculty Advisor)

Department of Computer Science - Indiana University Purdue University Fort Wayne

Abstract

In this project, we present a user-behavior-based authentication scheme, which completely removes the need for traditional alpha-numeric passwords. In order to be granted access to the system, the user must correctly pinpoint a secret location on a map, which is pre-defined by the user. In addition, the user's behavior as he/she navigates to the pre-defined secrete geographic location on the map are also used. To uniquely identify different users, a number of metrics have been extracted from each user's behaviors. Then, data mining algorithms are used to create a profile for each legitimate user based on their behaviors. Access will only be granted to the user who not only knows the secret location, but also behaves in his/her unique manner. Our evaluation results illustrate that our password-less authentication scheme can provide identify most users correctly without producing any error, while deny the intruders who behavior differently from the legitimate users.

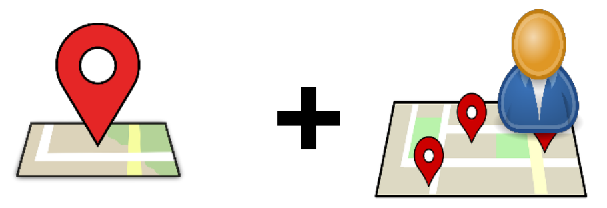
Motivation

Username/Password-Based Authentication

Traditional Authentication Scheme

Disadvantages

- Weak passwords are vulnerable to various attacks
  - Key-logger, brute-force, dictionary attacks, and bribing.
- Improperly stored passwords are prone to theft or tampering.
- The strong alphabetic passwords are difficult to remember.



Location-Based Authentication **Behavior-Based Authentication**

Our Authentication Scheme

Advantages

- Secret locations are easier to be remembered than passwords.
- Even if the intruder knows the secret location, it is still difficult for her to be granted with access.
- The security of authentication can be significantly improved, when combining location and behavior-based authentication.

Design Goals

Our scheme is designed to fulfill the following objectives:

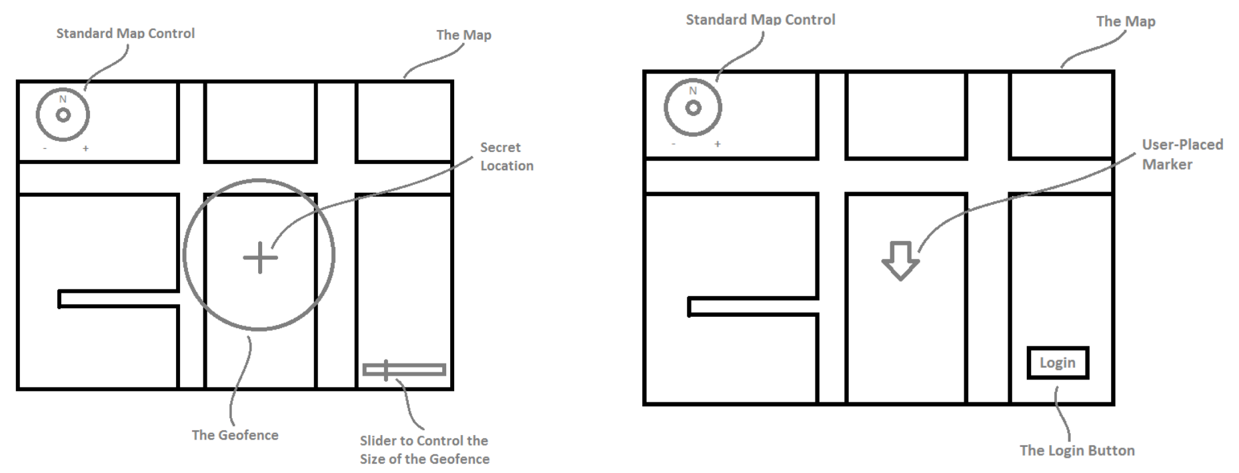
- Efficient formulation of behavior-based attributes.
- Low training and testing time.
- The chance that the intruder can be granted access should be negligible.
  - Low false positives (FP) rate and high true positives (TP) rate.

Methodology

Our methodology uses two-levels of authentication

Level 1: Location-Based

- The user logs in by navigating to the secret location he/she specified during the registration process.



Level 2: Behavior-Based

- To construct the user behavior profile, we formulate seven attributes, which can efficiently quantify the characteristics of a users' behavior, when she interaction with the map.
- The attributes can be used by various data mining algorithms to construct behavior profile for a user.
  - We avoid using those attributes, which construct models that are either to general or too specific.

Name of Attributes	Description
Zoom Frequency Index	The number of zooming actions in a unit time
Zoom Level Index	The average zoom level of a session
Double Click Zoom Percentage	The percentage of time spent using double click zoom
Mouse Wheel Zoom Percentage	The percentage of time spent using mouse wheel zoom
Slider Zoom Percentage	The percentage of time spent using slider zoom
Lowest Zoom	The lowest zoom level of the session
Number of User Actions Until Lowest Zoom	The number of user actions taken before the lowest zoom level is reached

Implementation

We have Implementation our prototype with the following components:

- A web-based map interface for user data collection implemented using Google Map's JavaScript API.
  - 10 users are asked to navigate the map from the Great Pyramids at Giza to the Friends Circle at IPFW.
- A Java program is used to extract the user behavior-related attributes.
  - Average time taken to extract attributes from one user is 0.0266 seconds.
- Weka, a data mining software, is used to train, test, and visualize the data via various algorithms.

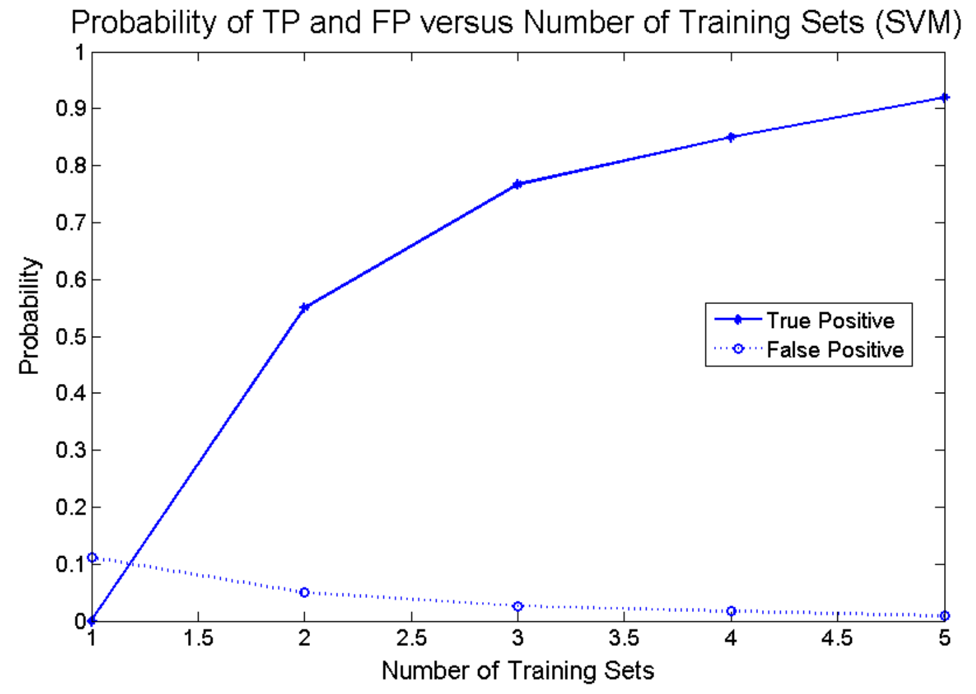
Results

Algorithm	Weighted Avg. FP	Weighted Avg. TP
SVM	0.009	0.92
Naïve Bayes	0.016	0.86
Random Forest	0.018	0.84

SVM Produces the best result with the lowest FP and highest TP.

Algorithm	Weighted Avg. ROC Area	Worst ROC Area
SVM	0.956	0.889
Naïve Bayes	0.987	0.964
Random Forest	0.982	0.882

Based on ROC area, Naïve Bayes is the algorithm with the best general performance.



TP increases and FP decreases as number of training sets increase.

- The secret location will have the strength equivalent to an 8-digit alpha-numeric password. (50 meters radius or smaller)
- With the addition of behavior-based authentication, the chance that a single random guess will success is one in 1600 trillion.

Discussion and Future Direction

Discussion:

- We might choose the SVM algorithm for our future experiment since it produces low FP rate and high TP rate.
- The majority of the user can be identified without producing any error.
  - False positives occurs in 40% of users.

Future Research:

- Application of our authentications scheme on wearable devices
- Exploring the possibility of uniquely identifying a user by his/her mouse movement alone

Acknowledgement

We thank the great comments provided by Dr. Jin Soung Yoo .