Continuous Authentication by Typing Characteristics

Luke Knepper
Computer Systems Lab 2009-2010

Abstract

This project will examine various applications of authenticating users by use of their typing characteristics. The purpose of this project is to determine the effectiveness of analyzing the user's typing patterns in order to ensure user security. Log-in-time authentication and continuous authentication based on the user's typing characteristics will both be explored, analyzed, and tested. One application will analyze the user's typing characteristics when the user attempts to gain access to, or log-in to, a system. Another application will measure the user's keystroke data while the user uses the program, and then feed the typing data through a neural network to determine authentication status. This process will be beneficial to user security because it will ensure that an intruder does not gain wrongful access to a user's account while that user is logged in to the system.

Background and Introduction

Typing patterns differ by person. People naturally hold down specific keys for specific times and take longer between different keystrokes. These typing characteristics can be, and have been, used for authentication purposes. In this project, I propose and test the accuracy of using typing authentication methods to boost security. Rather than Measuring the typing characteristics while a user types a simple passphrase and/or username, which are both stati, as commercial systems do, I will test the Accuracy of a method which generates dynamic (i.e. random) text which the User is prompted to type. The system then measures the keystroke data while They type this dynamic text. This offers a considerable advantage because A keylogger can no longer record and playback a simple typing sequence to Gain access to the system, and instead will have to mimic the user's typing Exactly. Additionally, I will create and test an algoirthm which will continuously Monitor the user's typing characteristics during program use and look for a Change in the characteristics, which would signal that an intruder has Wrongrfullly gained access to the user's system, and then will lock the user out. Both of these applications will be beneficial to user security.
Previous research that has been found on the accuracy of authenticating by Typing patterns while using static phrases mostly concurred on two results: Neural-network algorithms are the optimal approach for this goal, and such Methods are on average 80-90% accurate. These findings suggest that Typing characteristic authentication can offer a powerful boost to security.
Neural networks are modeled after the human brain. They are composed of Nodes (i.e. neurons) which take inputs from previous nodes, compute a simple function from these Inputs, multiply the result by a given weight and then pass the result on to the next nodes (see Fig. 4 for a simple diagram).

Procedure

In the first quarter, I completed a proof of concept program to test the theory that Users can be differentiated by their typing patterns. This program prompts two users to type a sentence and records their typing data while they do that. It trains A neural network with their typing data to distinguish the two users. It then prompts the users to choose a user from between themselves, identity unknown To the computer, to type a third sentence. It then feeds the mystery users' typing data through the neural network and determines whether the first or second user was the mystery user based on the result fromthe network. The program has an accuracy of 18 / 20 = 90%, which shows that the concept is accurate enough to be used on a larger scale, but still leaves much to be desired. (See Fig. 1)
In the second quarter, I completed a data collection program (See Fig. 2) which has collected more than 1,500 data samples on which to run tests. The program Prompts the user with a short amount of text and asks them to type it. While They are typing, it records their keystrokes in the following format:
"Key-# / Key-direction / Time-in-millis." This format is flexible and allows for Different types of programs to measure different characteristics from the data. Once the user is done, the program sends the typing data off to the server, Where it is stored in the form of a text file via a PHP script.
A program to simulate continuous authentication application is near completion (See Fig. 3). This program emulates an instant messaging setting where the user Interacts with a pre-programmed bot which asks the user questions. The Program measures the user's typing characteristics when the user responds, Runs these data through a neural network made from the original user's Typing characteristics, and then raises the warning level each time the network Outputs a result which does not match the original user's characteristics. If the warning level reaches a critical point, signalling intrusion, the system locks down.

Figures

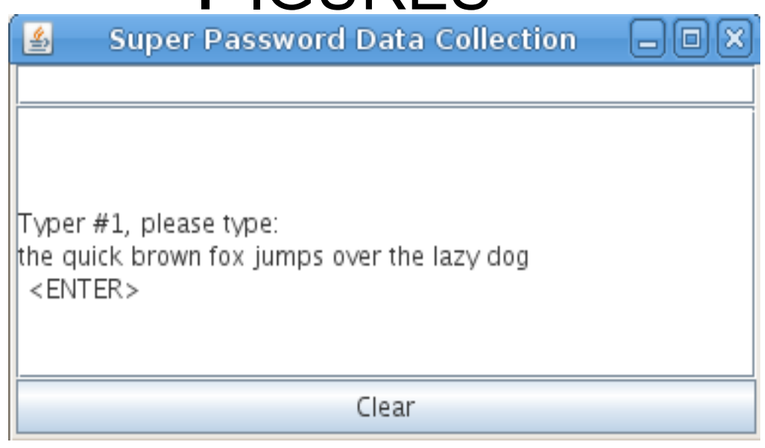

Fig 1: The proof-of-concept program

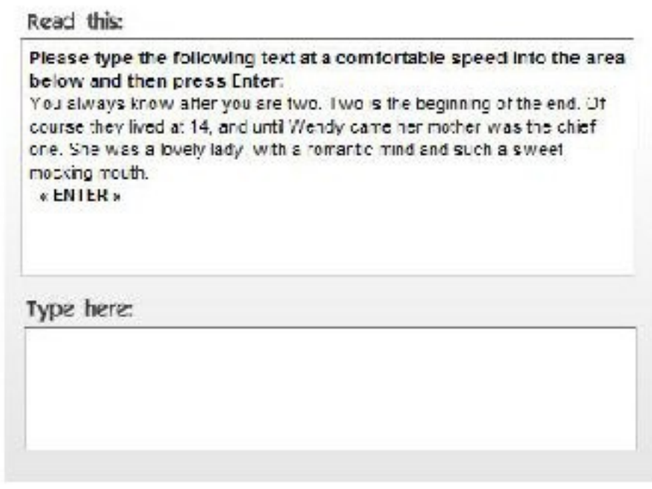

Fig 2: The data collection program

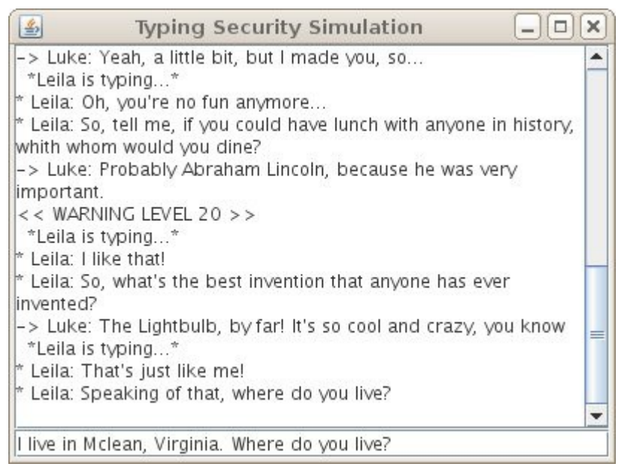

Fig 3: The continous authentication simulation program

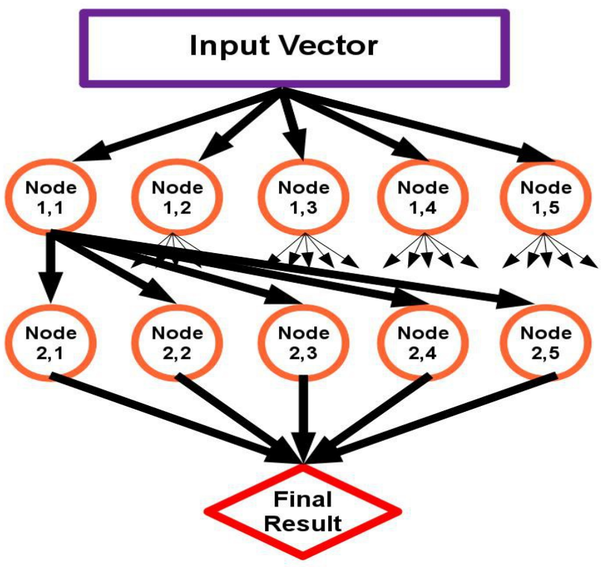

Fig 4: An outline of a simple neural network

Discussion

Three things will be determined from this experimentation:
1. Which network structure is most accurate for this purpose
2. Statistics on the accuracy of the neural network in this purpose
3. The optimal corpus size (i.e. length of text) for this purpose
In order to test these things, I will first need a lot of data to run Through the neural networks. I am collecting this data from my Data collection applet. An automated testing system will then be Developed to create (i.e. train) a neural network of each type for each set of data. Once the networks are created, they will be Tested by running different amounts of typing data from the Collected data through each neural network and recording the Ratio of success and failure (i.e. the accuracy). The ratios for Each network type and corpus size will then be compared to Determine the goals above.
Once this is done, a mock authentication system can be created and more experimentation can be done using this system.