June 08, 2018

Defending against attacks on biometrics-based authentication

Tanmay Wadhwa

Neil Dhillon

Defending against attacks on biometrics-based authentication

ABSTRACT

Many devices include biometrics-based user authentication in addition to secret-based authentication. While secret-based authentication involves a precise match with the known secret, biometrics-based authentication involves fuzzy matching that verifies that the input is similar to known biometrics within an acceptable threshold level of difference. As a result, biometrics-based authentication techniques are susceptible to attacks in which a malicious actor attempts to authenticate as the user via biometrics data that is crafted carefully to be similar to the stored biometrics within the threshold.

The techniques of this disclosure guard against such attacks by use of a generative adversarial network (GAN) where random perturbation is added to the received biometrics input for a dynamically determined number of test iterations. The matching threshold value and the number of test iterations can be dynamically determined. If the authentication test during each of the iterations is passed by the perturbed biometrics input, the user providing the biometrics input is authenticated. Otherwise, the device falls back to secret-based authentication.

KEYWORDS

- Biometrics
- Authentication
- Fingerprint identification
- Face identification
- Adversarial attack
- Generative adversarial network

BACKGROUND

Users are often required to authenticate prior to using a device, such as a smartphone or a tablet. Many devices provide secret-based authentication that relies on the user knowing some pre-specified secret information, such as a password, PIN, or pattern. User input is compared with the pre-specified secret information known to the device. The user is authenticated if the input is an exact match with the known information.

In addition, many devices include biometrics-based authentication that relies on scans of one or more physical attributes, such as fingerprints, face, iris, etc. The biometrics are provided via one or more of the device sensors, such as fingerprint reader, camera, touchscreen, iris scanner, etc. Similar to secret-based authentication, the device compares the biometrics input with the corresponding biometrics information provided earlier by the user, e.g., at a time of registering the biometric for use during authentication.

Unlike the precise matching involved in comparing the input with the known secret, the matching between input and known biometrics information is a fuzzy process. Instead of an exact match, the process verifies that the input is similar to the known biometrics within an acceptable threshold level of difference. Successful authentication requires that the difference between the input and the known biometrics be less than or equal to a reasonable threshold value. Owing to the fuzzy matching process, biometrics-based authentication methods can be susceptible to attacks in which a malicious actor attempts to authenticate as the user via carefully crafted biometrics input. While such data may not be a precise match with the biometrics of the user, it can be crafted to be similar to the user's biometrics within the threshold limits, thus allowing the attacker to authenticate as the user.

DESCRIPTION

The techniques of this disclosure guard against attacks on biometrics-based authentication by implementing a generative adversarial network (GAN) along with dynamically determined and adjusted values for the matching threshold level and the number of test iterations. When a user request for biometrics-based authentication is received by the authentication processor, e.g., within the device, along with an encoded biometrics scan of the body part to be used for authentication, the matching threshold value and the number of test iterations are dynamically determined.

The dynamic calculation of the matching threshold and the number of test iterations can be based on consideration of a number of factors (obtained with user consent) such as a current location of the user device on which authentication is being performed, location history of past successful authentication events, average value of historical matching threshold levels, time elapsed since the last successful strong authentication, e.g., non-biometrics based authentication, etc. Additionally, the dynamic calculation of the number of test iterations can take into account the hardware and software capabilities of the user device such that the authentication operation can be carried within latency limits to ensure a positive user experience.

The received encoded biometrics input is first provided to a machine learning classifier, such as a neural network, to compare the input to the known biometrics of the user and calculate a classification score. If the classification score is below the matching threshold, the authentication is deemed to have failed and an authentication error is displayed on the user device. If the classification score meets the matching threshold, the encoded input biometrics data is passed on to the GAN.

The GAN is configured to add random perturbation or noise to the received encoded biometrics input. The perturbed biometrics data is relayed back to the machine learning classifier to recalculate the classification score using the perturbed biometrics data. The classification score for the perturbed biometrics data can vary from the score for the original encoded biometrics input, e.g., the score may remain the same or may reduce.

If the perturbation results in a reduction in the classification score such that the score falls below the matching threshold, a potential adversarial attack is inferred. In such cases, the biometrics authentication is deemed to have failed and the user device falls back to requiring strong non-biometrics authentication. The process of adding perturbation, recalculating the classification score, and checking against the matching threshold is carried out iteratively until a potential adversarial attack is inferred or classification score stays above the matching threshold for the set number of test iterations.

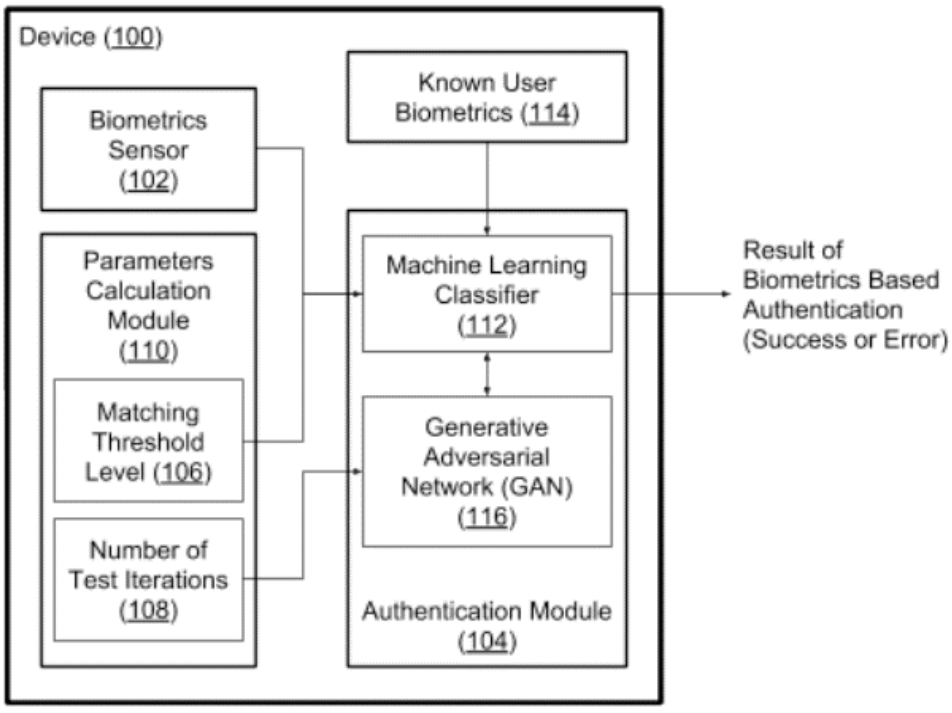

Fig. 1: Utilizing a GAN to defend against attacks on biometrics-based authentication

Fig. 1 shows the implementation of the techniques of the disclosure. A user utilizes a biometric sensor (102) for authenticating to a device (100). The corresponding biometrics of the user are captured by the sensor, encoded, and sent to an authentication module (104) on the device. The authentication modules is also provided values of matching threshold (106) and number of test iterations (108) as determined by a parameter calculation module (110).

Within the authentication module, the encoded biometrics and the matching threshold level are first passed to a machine learning classifier (112). The encoded biometrics are compared to the previously stored known biometrics of the user (114) by the machine learning classifier. The classifier may be implemented as a neural network. Based on the comparison, a classification score is assigned to the encoded biometrics input. The score indicates the extent to which the encoded biometrics input matches the previously stored known biometrics. If the classification score is lower than the matching threshold level, an authentication error is generated and displayed to the user. If the classification score is higher than the matching threshold level, the encoded biometrics signal is relayed to the GAN module (116) along with the number of test iterations.

A small amount of random perturbation or noise is added by the GAN module to the received encoded biometrics input. The resulting perturbed biometrics signal is passed back to the machine learning classifier for recalculating the classification score. Simultaneously, a counter that keeps track of the number of test iterations is decremented. Classification score of the perturbed biometrics input falling below the matching threshold level is considered an indication of a potential attack. In such a case, biometrics-based authentication is deemed to have failed. The user is then requested to authenticate via a stronger non-biometrics based authentication method, such as a shared secret.

If the classification score of the perturbed biometrics input stays above the matching threshold level, the process of adding perturbations via GAN is repeated until the score falls below the matching threshold level or the counter for the number of test iterations reaches zero, whichever is earlier. If the classification score stays above the matching threshold level after carrying out the perturbation process for the number of test iterations, the original encoded biometrics input is considered legitimate and the user is authenticated to the device.

The matching threshold level and the number of test iterations appropriate for any biometrics-based authentication request can be specified according to the requirements of the developer or the user. With user consent, these two parameters can be calculated dynamically for each biometrics-based authentication event as described above. For example, an attempt to unlock the device in a previously unknown location can require meeting a higher matching threshold value, which makes the authentication requirements stricter, e.g., since unfamiliarity may signal a potential attack.

Dynamic calculation of the number of test iterations additionally takes into account the latency introduced by the GAN operation. Given that a user is likely to attempt to unlock a device multiple times a day, it is important to maintain latency within acceptable limits that ensure a smooth and positive user experience. For example, the number of test iterations may be set such that the authentication operation can complete in less than 100 milliseconds. Since devices vary greatly in computational performance owing to differences in hardware and software, calculation of the number of test iterations may be based on measured performance benchmarks for the device. For example, devices with dedicated hardware for machine learning tasks may be set to perform more test iterations than those that lack such hardware. In some

implementations, any of the functions and modules utilized for implementing the techniques of this disclosure may be located external to the device.

The techniques of this disclosure utilize the same processes used by the attackers to detect and defend against the attacks. When the input biometrics data is genuine (generated via scanning the user's body part via the corresponding biometrics scanner), the random small perturbation added by the GAN module leads to a slight reduction in the classification score. However, the reduction is small enough such that the classification score of the perturbed biometrics signal stays above the matching threshold level, thus having no impact on the outcome of the authentication decision.

In contrast, random small perturbations added to biometrics data generated for an attack will likely result in the same score as the original input if the amount of perturbations added is not sufficiently large or, after sufficient perturbations are added, will yield a score substantially lower than the original score, thus falling below the matching threshold level and generating an authentication error. This disparity in results obtained from genuine and adversarial inputs arises due to the sensitivity of the machine learning classifier such that, in comparison to genuine biometrics data, adversarial inputs are less robust and more prone to yielding low classification scores upon the addition of noise. The amount of noise as well as the source of noise utilized for adding random perturbations may be varied in different implementations of the techniques of this disclosure. Moreover, various parameters involved in implementing the techniques of this disclosure, such as the matching threshold level and the number of test iterations, employ appropriate relevant heuristics, such as historic values of match threshold levels, to achieve an optimal balance between security and user experience.

The techniques of this disclosure can be implemented entirely in software, or in a combination of software and hardware. Therefore, the techniques can be deployed on older devices as well as extended to improve authentication experience in future devices by taking advantage of future improvements in hardware and software performance.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

The techniques of this disclosure guard against attacks on biometrics-based authentication by use of a generative adversarial network (GAN) where random perturbation is added to the received biometrics input for a dynamically determined number of test iterations. The matching threshold value and the number of test iterations can be dynamically determined. If the authentication test during each of the iterations is passed by the perturbed biometrics input, the user providing the biometrics input is authenticated. Otherwise, the device falls back to secret-based authentication.