ZoOm® - Customer Internal Testing Guidance

Updated April 16th, 2019

We have world-leading 3rd-party security test results and plenty of social proof that ZoOm is intuitive and convenient for users around the world. However, prospective customers often tell us that they need to conduct their own internal tests as well. We understand, and are eager to offer advice on how to test scientifically, learn the proper controls, and evaluate the test results. To help, we've written the following guidance for teams who need to conduct internal ZoOm tests.

Testing ZoOm With Designated Testers

These can be employees or individuals from a focus group, it doesn't matter as long as they agree to follow the testing framework and there a few grounds rules put in place first.

1. Genuine user testing
   a. Testing Organization will instruct each of its Individual Testers to perform the enrollment flow in ZoOm Demo App.
      i. Testing Organization should instruct Individual Testers:
         - To treat this as a usability test and try their best to succeed
         - To read and follow the on-screen enrollment tips
         - To follow instructions and retry if asked to
         - Not to attempt spoofs
         - Not to create more than one account
         - Not to use more than one device or let others use their device
         - Not enroll more than one identical twin
   b. Testing Organization should instruct the Individual Testers to perform **N** number of **authentication attempts** in normal office lighting (or any other lighting the Testing Organization decides they would like to test)
      i. An authentication attempt is not a single session, but a *good faith attempt to perform an authentication*. The ZoOm Retry Screen (or a customized screen created by Testing Organization) can be configured to help the user overcome a prohibitive environmental condition. ZoOm sessions are quick, and the system keeps users in the camera UI on fails to enable instant retries. A user attempting up to three ZoOm sessions in good faith should be able to successfully gain access, as environmental issues can typically be fixed by simply turning 90 degrees.
      ii. In other words, if User A initiates an authentication "attempt", success occurs if authentication was successful within three retries. The "attempt" should not be considered a failure if a single session is performed and the user quits.
      iii. Testing Organization should record the number of sessions performed per authentication attempt, and we should review the average number of sessions

needed to achieve success for users. This will be useful to compare to password/captcha and/or other security-related usability metrics, as well as weighed against overall time and cost of these related metrics versus the quickness/ease of invoking an instant retry to the user.

   c. Testing Organization should use ZoOm's continuous learning ("on" by default in the REST API)
   d. Perfect labeling (i.e., "ground truthing") of all sessions is the most accurate, but is also the most time consuming and requires expertise and special tools to perform correctly. The best way to get true results for this portion of the test **without** ground truthing is to ensure that during this period of time no spoofs are attempted. If spoofs are being attempted, the system will see many rejects and there will be no way to know why without manually verifying every session, and extracting spoofs and corresponding data; otherwise the results will be skewed.

2. Ask Individual Testers for feedback after each ZoOm enrollment and authentication attempt
   a. Usability/Ease of Use/Worked As Expected rating: 1-5 stars.
   b. Aggregate these ratings to get user experience feedback.
   c. Ask for Additional notes if desired.

   Note: In a recent UX test with the largest bank in Europe, results for the ZoOm user experience far exceeded customer expectations going into the project.

3. Spoof testing
   a. Again, perfect labeling (i.e., "ground truthing") of all sessions is the most accurate, but the most costly and requires expertise and special tools to perform correctly. The best way to get true results for this portion of the test **without** ground truthing is to ensure that during this spoof capture time period no genuine captures are being performed. If genuine captures are being performed and are succeeding, there will be no way to see why without manually tracking each and every session, resulting in skewed results.
   b. Please note, with [number of Individual Testers] people in the test, and even with only a few specifically devoted to spoofing, it may be hard to not accidentally perform a ZoOm session on a real person that succeeds. Per the above, if you are not perfectly labeling every session, the above success on a real person could be seen as a "false accept" or a "successful spoof" in Testing Organization's metrics if the utmost care is not taken.
   c. To minimize chance of error, FaceTec suggests spoof testing is only done by designated Testing Individuals before or after the Genuine test. The main reason is to ensure that all the sessions will be spoofs so you can safely aggregate results. If you are not labeling/ground truthing the results of each session, as ISO/NIST labs do when they perform liveness testing, you will not know the difference between true accept, false accept, true reject, and false reject.
   d. iBeta/NIST Test Overview:
      https://www.ibeta.com/perfect-score-for-facetec-facial-authentication-software-in-ibeta-presentation-attack-certification/

e. NIST/ISO Lab Spoof/Presentation Attack Detection Full Test Results: https://www.zoomlogin.com/wp-content/uploads/2018/08/FaceTec_ZoOm_iBeta_ISO_PAD_Level_1_Certification-Letter_and_Report_08202018_Distributable.pdf

f. Please see the spoof section below for more details on spoof testing and methodologies suggested for this portion of the test.

4. Imposter testing

a. This is the scenario from meeting “try individual Tester 1’s account with individual Tester 2’s live 3D face”

b. One common way to test for imposters would be to find similar looking pairs of people and put glasses, hats, makeup, etc. on them.

c. Another perfectly valid method for testing imposter attacks is to randomly pair people together and have them attempt to access each other's accounts.

d. Session and attempt details must be recorded to get accurate feedback metrics. Similar to the above tests, the best way to get accurate results without labeling all sessions is to do the testing during a period where you know, and can guarantee, that all attempts being performed will be imposters.

e. Testing Organization should be aware that performing tests where FaceMaps from sessions are cross-compared is extremely difficult outside of our bespoke internal tools and proprietary compute cloud. This type of testing requires rigorous ground-truthing and thorough verification of all data and statistical systems. If Testing Organization chooses to test match performance via programmatic cross-comparison of FaceMap data, we suggest requesting an in-depth testing methodology review by FaceTec before collecting and recording the data.

5. Imposter + spoof testing

a. This testing combines spoof testing and imposter testing.

b. An example of this type of testing would be using a photo of Tester 1’s face to try to access Tester 2’s account (An increased security threat vector in more homogeneous countries).

6. Measuring results

a. Usability ratings: Testing Organization should decide what an acceptable 1-5 rating average is based on relative to the commonality between the Individual Testers and Testing Organization’s users. Rather than False Reject Rate, legitimate Testing Organization users being able to use the biometric platform and rating the platform as usable should be the top priority in real-world deployments.

i. In our compute cloud and testing harnesses, we target under 2% FRR when we run all genuine user sessions against all corresponding genuine user FaceMaps. This is without continuous learning, which further contributes to a lower FRR. The true user could experience a FRR as high as 5% on the first couple sessions and then drop below 1% after 10+ sessions, which will have allowed continuous learning to fully complete the FaceMap.

        ii. In general, most false rejects and enrollment/authentication issues are an indication the user is not following instructions and/or is in an inherently difficult lighting situation they cannot fix, rather than a failure of the ZoOm algorithms to classify the human as real.

    b. For spoof testing: Zero spoof attempts should be falsely accepted. This includes photos, videos (including of the user ZoOming), masks, dolls, 3D printed heads, wax figures, etc.
    c. For imposter testing: Zero Imposters should be falsely accepted as the real user, no matter how similar they look. The only caveat to this is that identical twins have about a 1/500 FAR with ZoOm on average, increasing with the age of the subjects.
    d. For imposter + spoof testing: Zero spoof attempts using artefacts of similarly looking imposters should falsely accept.

2D Profile Image Comparison to 3D ZoOm FaceMap

Comparison to pre-existing social media-style profile images does not add a layer of security as there have usually not been any authenticity checks on them. They were uploaded from the user's computer and not captured in real time from an engine that Testing Organization controls. Receiving a "non-match" between an insecure/unverified profile photo is not an indication of anything nefarious. Profile images will generally be unconstrained and exhibit high variations in:

- Pose (Women do tend to match lower when they pose with more angle)
- Illumination
- Expression
- Editing of the photo (i.e. filters & face tuning)
- Age differential; it could be the same person 10 years ago

These types of variations are solved for by Face Identification algorithms designed for surveillance (like AWS Rekognition), but these variables are not ideal for face authentication. We expect most of the profile images you will attempt to use will fail to be processed/acquired as they will not pass ZoOm's facial expression, quality, and head pose requirements.

Configuring the Test Correctly and ZoOm Behavior Topics

1. ZoOm versions
    a. Testing Organization should stay current with the web/browser/JS SDK version. The most recent sessions are from 6.7.1, which is 1-2 or more months ago. There have been many advances and improvements made that will be beneficial in these tests.
    b. An example of this is the creation of a tailored sample called the Onboarding and Authentication Sample, which could be a good baseline for the test's client side code.
    c. Video here -- https://drive.google.com/file/d/1sBSwWjgAXh4llZXks1MP_bUhS0FfW5lh/view?usp=sharing
2. Session ID

   a. Testing Organization is undoubtedly creating internal usage databases/logs to track performance and actions of users. It is *highly* recommended that Testing Organization stores the Session ID provided from the ZoOm web/browser/JS SDK alongside these transactions. In case of a discrepancy in the data, the Session ID can used to ensure the data is correct.
   b. Session ID is a required parameter of the REST API and all SDK samples are already configured to use it properly. The suggestion here is to also record this Session ID, as any internal bookkeeping Test Supervisor is doing to associate behavior with results.
   c. Session ID is documented in the ZoOm REST API Guide
      i. https://dev.zoomlogin.com/zoomsdk/#/webservice-guide

3. Retry Reason Legend --
   a. A Failure Reason Code value will be returned when a ZoOm sessions fails. This value can be recorded if you desire to a more specific reason for failures when they are experienced.

```
/**
 Represents the estimated reason that the subject failed the ZoOm session.
 */
typedef NS_ENUM(NSInteger, ZoomRetryReason) {
   ZoomRetryReasonGeneric = 0,
   ZoomRetryReasonBadLighting = 1,
   ZoomRetryReasonFaceAngle = 2,
   ZoomRetryReasonNotAvailable = 3,
   ZoomRetryReasonVeryBadLighting = 4
};
```

Spoofing & Liveness Detection

Indepth Liveness Detection Methodology:
https://www.zoomlogin.com/wp-content/uploads/2018/09/FaceTec_Liveness_Testing_Methodology.pdf

1. Guidance for due-diligence spoof testing against ZoOm:
   a. For enrollment/liveness check: Zero presentation attacks should generate a FaceMap that succeeds a call to /liveness (REST) or getLiveness() (Java).
   b. This includes:
      i. Government ID attacks
      ii. Printed photo attacks
         1. High res
         2. Low res
         3. Matte
         4. Glossy
         5. Head cutouts
         6. Faces with eyes or other portions cut out of face

        iii. Static image computer/phone screen spoofs
            1. From other devices / presented on various screens
            2. High res
            3. Low res
        iv. Video spoofs on computer/phone or using a projector
            1. From other devices / presented on various screens
            2. High res
            3. Low res
        v. Masks, dolls, wax figures, mannequin heads, and any other non-living representations of human faces

2. FaceTec spoof lab video tour
    a. Link: https://youtu.be/pG-aLdxsZI4
    b. During the video tour of our spoof lab we showcase:
        i. Tens of thousands of hand-made paper printouts, face cutouts, and various paper modalities.
        ii. Our “Auto Spoof” program runs against our ZoOm engine and automatically present spoofs to ZoOm to attempt bypasses -- hundreds of thousands of spoofs are tested per day via this automated system. And the system is continuously being updated to make more complex automatic spoofs.
        iii. Masks, dolls, disguises, 3D printed faces w/ makeup, and many more artifacts.