Two-factor authentication for voice assistance in digital banking using public cloud services

Vassil Vassilev
London Metropolitan University
London, UK
v.vassilev@londonmet.ac.uk

Anthony Phipps
London Metropolitan University
London, UK
tsaphip1@londonmet.ac.uk

Matthew Lane
Cloud Hint Engineering Ltd
Cambridge, UK
mattjmlane@gmail.com

Khalid Mohamed
London Metropolitan University
London, UK
khalid145@hotmail.co.uk

Artur Naciscionis
London Metropolitan University
London, UK
a.naciscionis@londonmet.ac.uk

***Abstract*—This paper presents the results of a project on using voice assistants to enable financial and commercial operations which require authentication with increased level of security. Several prototypes of authentication service with two-factor authentication for the purpose of voice-controlled digital banking and online payments have been developed at the Cyber Security Research Centre of London Metropolitan University. They utilize the power of the cloud technology and the attractiveness of the devices equipped with voice assistant software, such as Amazon Alexa, to provide secure method for two-factor authentication, which meets the recent standards for security of the operations in digital banking.**

Keywords—Voice Assistants, Two-factor Authentication, Cloud Services, Infrastructure as Code, Continuous Integration/Continuous Deployment

I. INTRODUCTION

According to the national statistics nearly 10% of the workforce in UK lacks digital skills and at least 5% of the population have disabilities which limit their ability to use digital technologies. At the same time, Amazon sold hundreds of millions of devices equipped with their voice assistant software **Alexa**, and **Google Assistant** is now available on more than 500 million devices. The technological advances in voice recognition give the opportunity to close the gap between the current generations and the millennials, who according to the predictions will soon abandon completely the classical text-based authentication in favour of other media.

A number of banks in UK have considered adopting the voice assistants for banking to a different degree of success. Some of them, like Barclays, have chosen to use biometric methods of identification [2], while other, like NatWest, use separate beacon for independent authentication [3]. The biometric methods rely on the built-in features of the device, used for banking - typically mobile phone equipped with dedicated software such as Apple **Siri** or Microsoft **Cortana**, while the beacon-based methods employ the full power of the voice assistant serviced, typically provided by the vendor's own cloud. Although the first method can be seen as an ultimate goal, some well-known examples of technical difficulties in facial and voice biometrics do not allow these methods in their current state of development to be endorsed for control of banking operations. The work reported here presents a beacon-based authentication solution developed at the Cyber Security Research Centre of **London Metropolitan University**, which works with devices equipped with Amazon's voice assistant **Alexa** [4] and backed by the vendor services on Amazon's own cloud, **AWS** [5]. It requires an additional component for the authentication, a *beacon*, which is used for confirmation and implements an independent secondary authentication for increased security. The current implementation allows two different devices to be used as a beacon – a standard smartphone, which will be preferable beacon device for most of the users, and a special dash-button, which enables users with tactile impediments, elderly people who struggle to operate mobile devices and people with specific disabilities to do voice banking.

II. HOW WOULD WE LIKE IT

From the very beginning of our work we had two strong user requirements: the authentication should be done in a way which complies with the security requirements of PSD2 directive of EU [6], and the procedure should be as frictionless and simple as possible. The working scenario which illustrates such a procedure is depicted on Fig. 1.

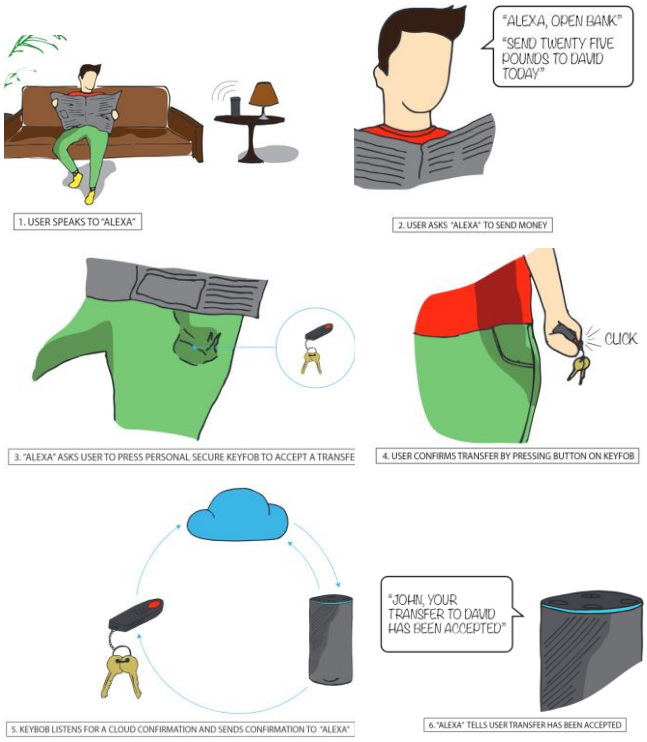


Fig. 1 Beacon-based authentication for voice banking

In this scenario the user initiates the banking transaction in a straightforward manner, just by talking to the voice assistant. We assume that the voice assistant device is registered on the vendor's cloud, which supports the first level of authentication against the registration. As part of the transaction the cloud service sends a request to the beacon device which confirms the transaction, thus providing

secondary authentication and the transaction proceeds with the bank. When the transaction commits, the cloud sends a confirmation notification to the beacon device. A brief comparison of this method to some alternative methods for authentication is shown in Table I bellow.

TABLE I. COMPARISON OF AUTHENTICATION METHODS

	BEACON	PWD	TOKEN	BIO	BEHAV
easy to use	✓		✖		✓
add device		✓	✖	✖	✓
not sensitive	✓		✓	✖	✖
high security	✓	✓	✓		
reliability		✓		✖	✖
speed			✓		✖
cost		✓	✖	✖	
trending	✓	✖		✓	

It might be expected that the password authentication is the easiest to use, but due to the need to remember it is not the case for people with memory deficiencies. This problem is magnified for all users when the security requirements mandate a dynamic password. In a voice context, the users may be asked to remember and input only different characters of their password so that each challenge yields a different but verifiable result. In addition, there can be some reluctance on behalf of the user to call out their password (or parts thereof) to a voice assistant. The biometric and behavioral methods, on the other hand, despite their attractiveness are too sensitive and because of this not reliable enough. The beacon method is similar to the token method, preferred by many banks for increasing the security of operation, but is much easier to use while providing the same level of security. Although it requires some additional costs its maintenance is less expensive that the other methods which require additional devices due to their relative simplicity and the possibility to use general-purpose devices such as smartphones instead.

III. TECHNOLOGY ENABLERS

In order to comply with the requirements of the PSD2 directive we had to construct a solution which meets two conditions: *first*, that there are two separate and independent communication channels for identification and authentication, which protects from potential security treats, and *second*, that both the identification and the authentication are completed within one single communication session to prevent social engineering fraud which might exploit existing logical vulnerabilities. In order to meet these requirements we investigated many currently available devices, communication protocols, signal processing features and information processing support.

A. The Voice Assistant

Although currently there are several general-purpose voice assistants and a large number of specialized chatbots in more or less satisfactory state of development, the market of consumer electronics is dominated by two vendors – Amazon and Google. Both vendors offer a whole range of devices supporting a variety of protocols (WiFi, USB, ZigBee, etc.) with a price tag from under 30£ to several hundreds. In both cases the devices are equipped with the same voice assisting software, which runs on the device but for more sophisticated applications it is backed by specialized services running on the vendor's cloud [4]. They have their own advantages and disadvantages and although some may see the Google platform as better choice the Amazon platform has its own advantages.

Our solution currently uses the Amazon platform due to historic reasons, but the recent announcements for cross-platform compatibility of some of the supporting services on the cloud makes this choice less critical.

B. The Cloud

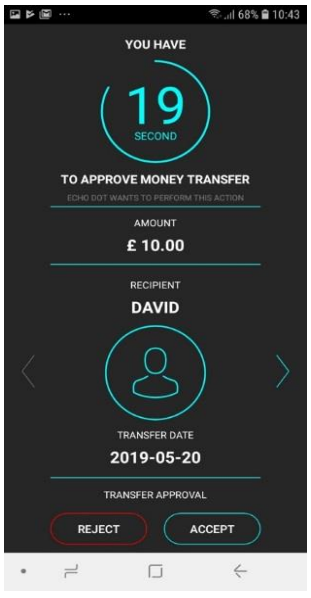
Contemporary cloud technology supports diverse applications but particular interest for us are the cloud service providers which market both the voice assisting devices and cloud services to support them. Our current preference is the platform of Amazon through its AWS cloud, which at the time we started this research provided better support for use of voice assistance with authentication.

AWS is a cloud service provider that provides software as a service [5]. Everything that is traditionally available in a data centre only is currently available with AWS. It provides computational, storage and communication services which enable users to quickly provision the resources that are needed for server-side applications without significant capital investment. Specifically, for its voice assisting devices AWS provides a dedicated service, which allows converting the spoken messages into text for further analysis [4]. In addition, AWS supports general-purpose lambda-functions, which can be used for implementing the identification, authentication and authorization services crucial for digital banking [6].

C. The Beacon

The role of a beacon can be played by a variety of devices with diverse capabilities and utilitarian purposes, but in order to maximize the value proposition it is important to give preference to both general-purpose devices, used by the majority of the potential users, and simple and cheap specialized devices to cover the users with special needs. The external outlook of the two beacons is shown on Fig. 2.

For the first beacon we used standard Android smartphone, supported with the proprietary messaging API of Google for secure communication with the cloud [8]. This way we managed to maintain an interrupted session throughout the entire authentication process.



a) Smartphone as a beacon *b) Dash-buttion as a beacon*

Fig. 2 Smartphone and dash-button for secondary authentication

Our solution for the second beacon is more complex due to the impossibility to use all necessary APIs within a single session. We constructed our own button device, similar to the original Amazon dash-buttons used for online ordering of pre-selected products, but unlike the button of Amazon, we had to develop also our own protocol for transmitting the identification information to the cloud via the voice assistant.

IV. How does it work

Our solution has a distributed architecture, spanning across several devices (voice assistant, beacon), locations (at the user, on the cloud, inside the bank) and services (streaming, messaging, gateways, storage, etc.).

A. Architecture

The generic architecture of the cloud-side application is shown on Fig. 3 in Appendix A. The identification information is provided by the user devices – voice assistant (for primary authentication) and beacon (for secondary authentication). The authentication is performed on the cloud by comparing the registered information, stored in a cloud-side database, with the actual information, supplied by the end devices in real-time. The final authorization of the bank operation is performed only after the bank approves the request.

B. Cloud Services

Amazon Alexa Skill Kit: This service implements bi-directional speech-to-text translation with automatic maintenance of sessions between the voice assistant and the human user. At the same time, it constrains the communication channel to voice communications only. Due to this constraint, we were forced to implement our own encoding of the identification information needed for authentication in the case of button-based beacon.

Amazon Lambda: Amazon Lambda is a computational service that runs your code in response to events, such as signing in, retrieving data, sending data, and so on. The service handles all of the scaling, patching, analytics and administration of the cloud infrastructure, necessary to run the code. We used the lambda services for several different purposes – to handle the interaction with the bank, to store and retrieve the registration information, to push notifications onto the beacon and to handle the requests and the responses.

Amazon Gateway: Amazon API Gateway is a service that allows developers to create a RESTful API that is able to handle HTTP communications by triggering the execution of the corresponding Lambda service.

Amazon DynamoDB: Amazon DynamoDB is a NoSQL database service on AWS which is used for storing and retrieving registration information for the purpose of authentication. It can be substituted by any other cloud-supported database, but the use of DynamoDB is preferable because it is fully managed service on AWS, which greatly simplifies the process of development and maintenance.

Detailed component diagram of the cloud-side software is shown on Fig. 4 in Appendix A.

C. Beacons

Due to the different degree of support for programming required functionality on different devices the two beacon devices are programmed using completely different kits.

Google Firebase Cloud Messaging (FCM): Google Firebase Cloud Messaging is a push notification messaging system that allows developers to send messages to a compatible user's device on a variety of platforms in point-to-point, multicasting and subscription-based broadcasting mode. Used for programming the push communications between the smartphone and the cloud.

ESP32 Microcontroller: Simple, cheap but powerful microcontroller for programming low cost IoT solutions, especially appropriate for encoding, processing and transmission of sound signals by low-consumption devices [9]. We employed it for controlling the communication between the dash button, the cloud and the voice assistant and programmed it in Python using an Arduino Kit.

D. Working scenario

With minor simplifications, the scenario for voice-controlled bank operation, such as checking balance or sending money using the smartphone as a beacon looks like the following:

- The user will initiate the voice transaction by issuing an oral statement to the voice enabled device. The Alexa software will convert the speech to text and will send it to the Alexa Skill Kit service on AWS for analysis. The Alexa Skill Kit service will analyze the text and will seek a suitable action that was predefined during development. Once found, the Alexa Skill Kit will pass the information to the Alexa Lambda service for authentication and further processing.
- The Alexa Lambda service will break down the information into components to retrieve the request. Once ready, it will create a pending action which will be saved temporarily in DynamoDB. The Alexa Lambda service will then send a request for confirmation to the user's registered smartphone via push notification using Google Firebase Cloud Messaging service and will start a timer, giving the user to confirm or reject the requested operation.
- The user feedback will be then sent back to AWS via the Gateway, which will dispatch it to Confirmation Lambda service. After checking the integrity of the message for preventing intervention from "man-in-the-middle", the service will update the status of the operation put on hold and after authorization from the bank will send back to the user a receipt.

The scenario for voice banking using the button is similar, except that the identification information needed for authentication is send to the voice assistant rather than directly to the cloud. In that case the information is encoded into a sound signal with multi-frequency modulation. Once received by the respective Lambda service, this information is then decoded and used for user authentication.

V. How did we develop it

Our research project was executed in two separate stages. During the first phase, we did an intensive feasibility study to select technologies and tools needed. Alongside, we identified some of the problems which we faced later - session maintenance, non-sound information support and necessity for prevention of potential interference from third parties inside the transactions, for example.

During the second phase we employed an experimental agile software development methodology, which was critical for the success of the project as a whole. Two particular methodological borrowings from the DevOps community allowed to finish the project in time – utilization of the concept of Infrastructure-as-code (IaC) [10] – see Fig. 5 in Appendix B for and the use of Continuous Integration/Continuous Deployment (CI/CD) [12]. The first helped enormously by automation of the process of designing of the software components and composition of the software services needed for implementing them through the use of *Terraform* [11], while the second was vital for supporting

incremental development in the presence of frequent changes as the code grew up on GitLab.

VI. What can be done next

The solution we developed demonstrated technological feasibility and viability during intensive user testing using both the general-purpose device (smartphone) and the dedicated device (dash button). We are now confident that using voice assistants for digital banking is not only an attractive idea, but also a viable alternative which will serve both the current trends and the specific needs of many people.

Of course, there are a number of directions for further development before the voice banking becomes practice. Some of them are related to the need for adopting standards and can be implemented immediately, like the migration of the authentication service to Open Banking platform [14]. Other require removing the dependence from the vendor-specific platforms and might be more difficult to achieve because the effect on business. The major vendors of voice assistance will be hardly volunteering for standardization and creating of an open platform due to the uncontrolled changes in their own products. The good news, however, is the information that some discussions between leading vendors and major high street banks are already starting taking place.

An additional problem is the reluctance of the banks to send data outside and host it on the public cloud. In our solution no transactional information is stored on the cloud, just minimal registration information. In principle, it is possible to remove completely the customer information from the cloud but this would delay enormously the time for completing the transactions and will additionally lower the reliability. This problem can be resolved if the banks implement the authentication service on their own managed clouds, but again – this will affect the business of the vendors of voice assisting hardware and software. Until when they come to terms with this our solution is a viable opportunity that could provide convenient and secure access to digital financial services to all.

Acknowledgment

The research reported here has been possible thanks to a grant from UK Department DCMS under Innovate UK program Cyber ASAP. The user testing of the prototype has been funded by Lloyds Banking Group, but all results and opinions are solely of the authors and in no way represent an official position of the bank.

References

[1] Office fro National Statistics (2019) [https://www.ons.gov.uk/employment-andlabourmarket/peopleinwork].

[2] IBS Intelligence (2018), Siri makes its way to Barclays as voice-controlled payments become mainstream [https://ibsintelligence.com/ibs-journal/siri-makes-way-barclays-voice-controlled-payments-become-mainstream/].

[3] Which? (2019), NatWest trials 'voice banking' with Google Assistant [https://www.which.co.uk/news/2019/08/natwest-trials-voice-banking-with-google-assistant/].

[4] Coates, D. (2019), Voice Applications for Alexa and Google Assistant, Manning Publications (2019); ISBN: 1617295310.

[5] Derry, R. (2019), AWS: AMAZON WEB SERVICES: The Complete Guide From Beginners For Amazon Web Services, Independently published (2019); ISBN: 1696347831.

[6] Wadia, Y., Gupta, U. (2017), Mastering AWS Lambda: Learn how to build and deploy serverless applications, Packt Publishing (2017); ISBN: 1786467690.

[7] European Commission (2015), Directive 2366: Payment services (PSD 2) [https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en].

[8] Moroney, L. (2017), The Definitive Guide to Firebase: Build Android Apps on Google's Mobile Platform, Apress(2017); ISBN: 1484229428.

[9] Kurniawan, A. (2019), Internet of Things Projects with ESP32, Packt Publishing (2019); ISBN: 1789956870.

[10] Morris, K. (2016), Infrastructure as Code: Managing Servers in the Cloud, O'Reilly Media (2016); ISBN: 1491924357.

[11] Brikman, Y. (2017), Terraform - Up and Running, O'Reilly (2017); ISBN: 1491977086.

[12] Fowler, M. (2018). Continuous Integration. [https://martinfowler.com/articles/continuousIntegration.html; accessed Jul. 2018].

[13] Michael, C. (2016), Open Data API [https://openbanking.atlassian.net/wiki/spaces/DZ/pages/16320694/Open+Data+API+Specifications].

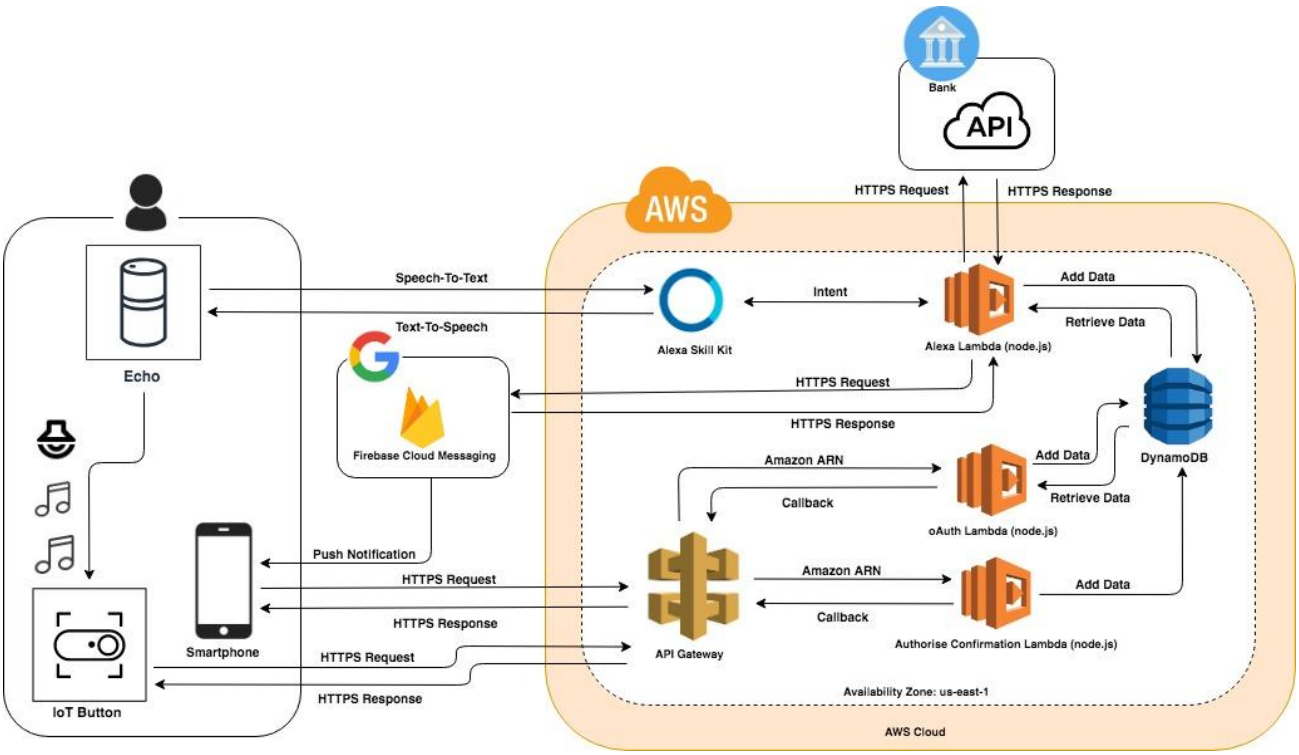


Fig. 3 Voice Authentication Architecture

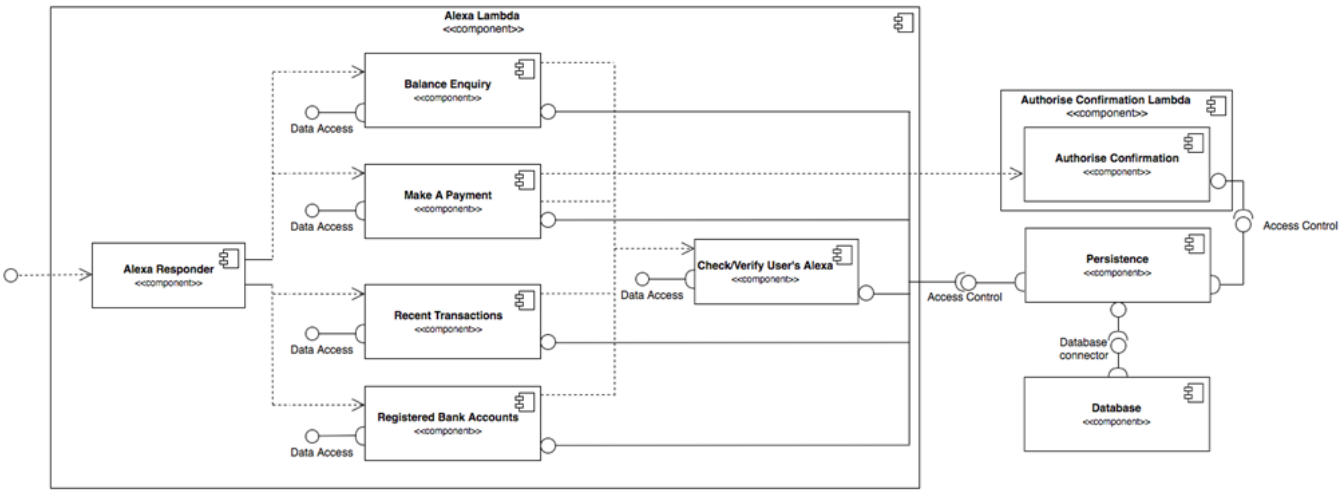


Fig. 4 Component Diagram of the AWS Services

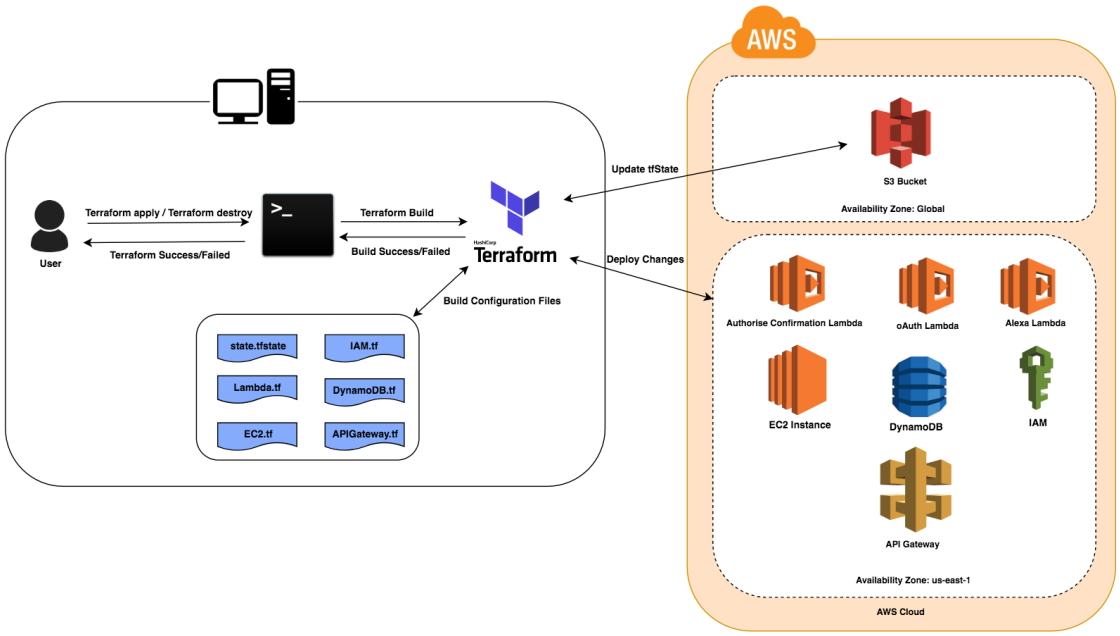


Fig. 5 Infrastructure-as-Code Service Architecture

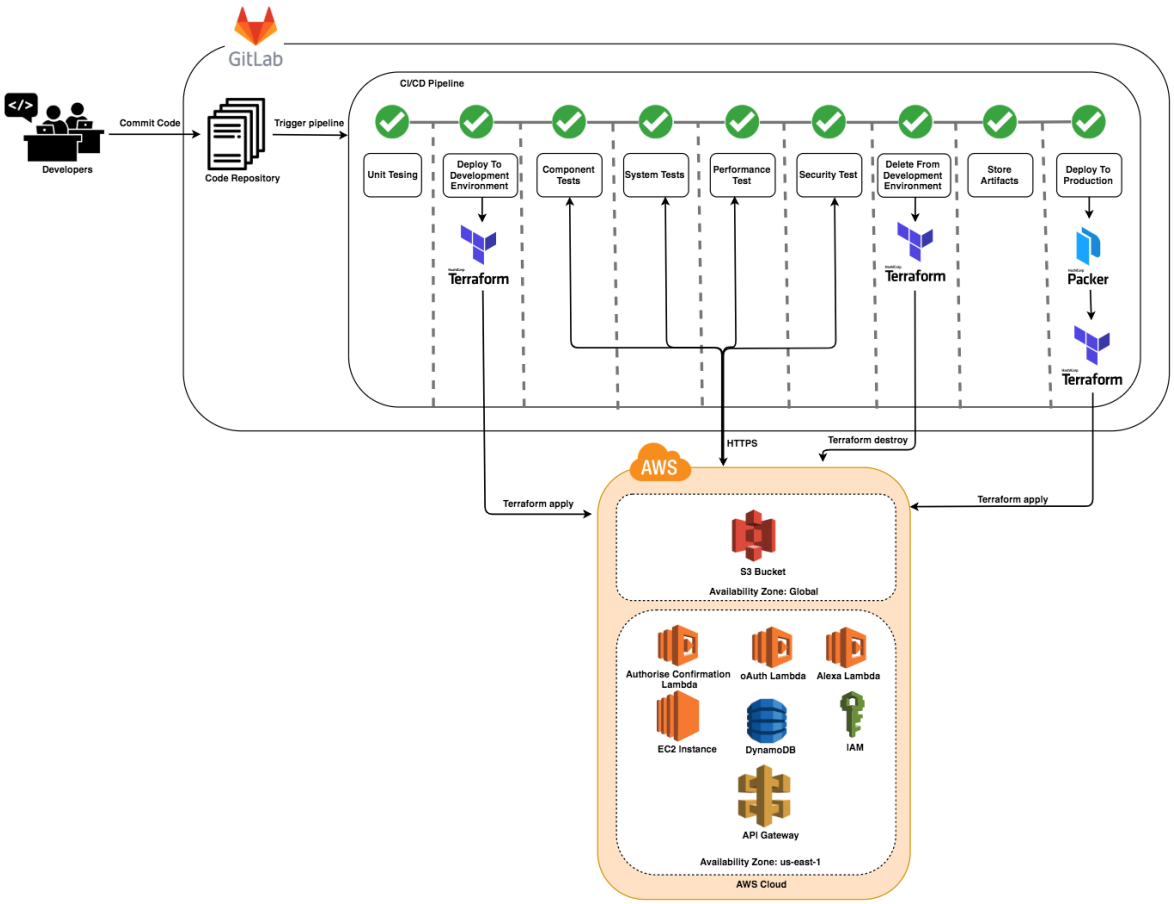


Fig. 6 Continuous Integration/Continuous Deployment using Public Code Repository