IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

A Proposal on Initial Remote User Enrollment for IVR-based Voice Authentication Systems

Umut Can Çabuk[1], Tuba Şenocak[1], Eyyüp Demir[1], Abdullah Çavdar[1]

Dept. of Electrical-Electronics Engineering, Erzincan University, Erzincan, Turkey[1]

Abstract: In the last decade, Interactive Voice Response (IVR) systems gained popularity among customer care, phone banking and many other commercial services. It helps to reduce conversation and dialing overheads, guides customers and automates call forwardings. On the other hand, many sensitive phone and online services, including banking operations and e-government applications, require strict and accurate user/customer authentication schemes. Although not being preferred so far, with regards to the recent developments on the voice recognition techniques, IVR emerges as a brilliant solution for authentication, too. When using IVR for voice authentication, all potential users must be enrolled and voiceprints (records) must be collected prior to start of the service. However, for scenarios where users are located in geographically distant places, it is very hard, if not impossible, to gather all the users in enrollment offices. So, this initial enrollment process should better be done remotely. In this study, we come up with a systematic method that allows institutions to enroll their users and record their voiceprints remotely, without compromising reliability of further authentication.

Keywords: Voice authentication, voice recognition, IVR, remote user enrollment, e-voting.

I. INTRODUCTION

The big boom of the broadband internet in the early 2000s and the mobile (cellular) internet nearly after a decade, not only provided enhanced multimedia, but also enabled online delivery of many commercial and governmental services with help of advanced cybersecurity techniques.
Many of these services, provide and/or require access to the sensitive information (for both people and institutions), which can be used for abuses and frauds, if not handled securely. Hence, it is a must to build systems in a way, so that only authentic users should be allowed to use and privacy measures should be considered.
In this respect, because of the security (and/or costs of security investment) concerns some services, such as e-voting, did not gain widespread popularity.
A reliable e-voting scheme, in addition to security and privacy, should be feasible to deploy and, should be very easy to use. Following subsections explain such a use case and one of the main problems that came into the open and addressed in this study, namely, remote enrollment of users (voters). While the solution proposed can be adopted to many use cases, we consider an e-voting (incl. online elections) scenario for the rest of the study.

A. Scenario and Assumptions

The e-voting scheme we have mentioned above, can either utilize a web site, a phone call or SMS services to collect the votes. However, in either case, authentication of the voter (the person who visits the web site or who answers the call) should be done to ensure that the person who should be voting, is the one who actually is voting. A widespread method, namely, use of passwords or PIN codes, does not ensure a personal authenticity and, in many cases, provides very weak protection. A good and easy way to achieve this is utilizing interactive voice response (IVR) systems, which are configured to recognize and differentiate people's voice (which is actually a biometric data), by making voice authentication calls to voters, prior to the final voting process.
Accuracy or success rates of such systems are out of the scope of this work. On the other hand, the initial setup of such systems require intensive attention, since pre-recording the voice data of voters is challenging without taking the records in-person, which would extremely be infeasible.

B. Problem

In order to use a person's voice as a biometric authentication entry on a non-secure channel (i.e. a standard phone call), the voice should have been recorded via a secure channel beforehand. In fact, this problem can be generalized to many other authentication types. For example, if a smart card (i.e. national ID) will be used as the authentication source, than it should surely be delivered to the right owner, by making an ID check and/or getting a wet signature, which is equivalent to a secure channel. A genuinely good method to do the voice pre-recording is making it in-person, with the presence of legal witnesses. However, this is obviously not possible in many cases, where voters are located in geographically diverse places. Likewise, usage of passwords, which can be broken, leaked or guessed, is not a good

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

solution, either. Thus, there is a solid need for a scheme to ensure the authentication of the voters in a reliable, scalable and secure fashion.

II. BACKGROUND & RELATED WORKS

In this section, we provide relevant background information and provide some previous related works.

A. Authentication

In earlier times, authentication was a simpler process. With the development of computer and internet technologies, authentication process has become more complicated. Because of the complex and anonymous form of the World Wide Web, some sensitive private information can easily be stolen by attackers remotely, without interaction. The term authentication mainly includes user authentication that is between human and machines, and device authentication that is between machines [1]. Authentication process can be defined as a precondition to grant access to resources in the system and it is used to identity verification of the user device or other presence in the computer systems [2].

1) Public Key Authentication:

Public key authentication, unlike its counterpart: private key authentication, suggests use of two different (but complementary) keys for cryptographic operations, namely; the private key and the public key. These two keys are related to each other, but it is not feasible to derive one from another using known mathematical methods, because of the extremely high complexity. When the public key used for encryption, then the private key can be used for decryption. This provides confidentiality. Yet, when the private key is used for encryption, then the public key is used for decryption. This scheme provides authentication and allows usage of digital signatures [3].

B. Multi-factor Authentication (MFA)

Multi- factor authentication is a concept that combines following three authentication approaches:

- Knowledge Factor Authentication
- Possession Factor Authentication
- Inherence Factor Authentication

To ensure a high level security, any user has to be succeed in these three stages. Previously, two-factor authentication was seen as sufficient. Nowadays, services and the user requests are more complicated so MFA systems are realized as three or more factor authentication [4]. Knowledge-based methods are predicated on data which is assumed to be known only by real users themselves, such as PIN (personal identification number) codes, passwords or secret questions, which are preset by the users, or given to the users by the service provider via e-mail, post, SMS or in person [1]. Possession factor approach is based on things that are belongings or physical possessions of the users. Generally, smart cards and USB dongles are used to hold and prove authenticity of the user information. These cards mostly include some personal data [5]. Short messaging service (SMS) based authentication as a recent application of possession factor approach, was developed to combine knowledge factors and possession factors to enhance the authentication security. Here, the user must have access to a specific phone (number) that he/she formerly declared. It also enables use of one-time passwords (OTP), which are used to mitigate the excessive reuse problem of the password and PIN based authentication method [3]. Inherence factors include features those are specific for each person in the world, so that (ideally) no two people may have the same property for such a feature. These factors are mostly based on biometric data, such as fingerprints, voice, retina, iris etc. For users, to verify their biometric features will always require special (or integrated) hardware. Biometric features provide high level of security (for authentication), yet it is much better to combine this with a password in order to build multi-factor authentication [3].

C. Interactive Voice Response System

Interactive Voice Response (IVR) technology is a customizable automatic digital telephone system that provides interaction between users and other computer systems [6]. It basically implies a voice conversation between a user and a computer interface, without any intervention of a phone operator. The basic working principle of IVR relies on collecting relevant information (commands) from the user and to direct them to appropriate targets. In this system, dual tone multi frequency (DTMF) keyboard strokes, human voice or both of them can be used as input [6]. The response of this system can be in the form of voice answer, some kind of intended operation (like top-up) fax, SMS or call back. [7]. When IVR is integrated with voice/speech recognition systems, then it can also be used to provide user authentication.

D. Voice Recognition

Speech (or voice) recognition systems essentially computer programs (with auxiliary peripherals) those hear, record and identify voice commands and relevant speeches of users. Working principles and some more details regarding such

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

systems can be found elsewhere [8, 9]. They mainly have two main types, as text-dependent and text-independent systems. The main principle of the text-dependent system is expecting the same text (it can be a keyword or a sentence) from the user for training and recognition process. While a text-dependent system has such a constraint, a text-independent system does not require the training test in the recognition process. Feature extraction, similarity analysis and selection are the common functions of these two systems. Feature extraction task is related to adjusting a set of coefficients in a predictive system by using the spectral envelope of the recorded voice. Similarity analysis is realized by computing the regression between some coefficients regarding the voice records. These coefficients are obtained by comparing two voice samples [10].

E. Related Works

Though IVR systems gain vast popularity, voice authentication is still not widespread. Nevertheless, security for the remote enrollment processes is rarely considered in such applications of IVR.

Wayman et al. in their book, though state voice authentication is a smart solution for many applications, point out that the remote enrollment process can be problematic, since there will be some doubts regarding the identity being claimed is correct and the person who is behind the computer is who he/she is claimed to be [11]. However, they do not propose any solution for this problem. A USA based credit bureau, Equifax Inc. also uses a voice authentication system for call center operations. Though they explain how this system works and provides security in their white paper, they do not mention any countermeasure regarding the user authentication and verification in the enrollment phase [12].

Nachappa et al.in their paper, propose a framework for a speaker recognition system, which utilizes phone calls [13]. They introduce a straightforward enrollment procedure, which do not include additional steps to verify the identity of the person calling (user). But, they focus more on further (regular) voice authentication phase, rather than the enrollment. Hence, they stated that such a voice authentication system may not be suitable for services that require high level of security. They advised use of passwords in combination with speaker recognition. Another related publication by Miguel-Hurtado et al. introduces a comprehensive study regarding a novel voice recognition system and shows results of some user-centric analyses [14]. Yet, they do not enforce any specific authentication process for the initial enrollment workflow. IVR systems may provide authentication for GSM provider processes. Turkcell, which is the oldest GSM operator of Turkey enables users to make transactions faster via IVR. For voice authentication, Turkcell requires their users to repeat the sentence “Turkcell beni sesimden tanır.” which translates as “Turkcell recognizes me from my voice”. Turkcell stores these records in its database. Thanks to the records, the users do not have to remember any passwords for making transactions including tariff changes and top-ups. Therefore, the users can make all transaction by using his/her own voice [15]. As Nuance Inc. (a sub-contractor company) reported, in order to provide authentication for the enrollment process, Turkcell checks if the caller owns the phone number, which he/she wants to make use of some services. That is, while still better than nothing, obviously not a strong protection.

III. PROPOSED SOLUTION

Within the scope of this study; we have developed a comprehensive algorithm that prevents a hypothetical fraudster, who is assumed to have stolen some personal information as well as the registered phone of an eligible voter, from imitating the voter. The directive is designed to eliminate the counterfeit attempts to make initial registration of unauthorized/imitated people.

A. Involved Parties

While a fully operational e-voting mechanism might require inclusion of much more parties, like legal/governmental bodies; the prelude phase introduced in this paper requires cooperation of 4 parties, who are the (customer) organization, the network operator, the voter, and the voting agency.

The customer organization can be any kind of real or legal person (such as an association, a corporation or a public institution), who plans and requests support to hold elections in the near future. The voter, is obviously the member of the customer organization who is eligible to vote via the pre-built web system. The network operator, though having no direct role on the elections, does a very important task, namely it hosts and runs the IVR system in order to call the pre-known voters, intercedes the initial authentication process by making an identity verification, and finally collects the voters’ voice data. Lastly, the voting agency is the public or private company, who gives support to the customer organization by arranging all the formalities and technical infrastructures. It is also responsible for taking any security and privacy measures, such that the elections will be completed without any fraud nor any question. The computer systems regarding recording, storage and counting of the votes, are located at the agency.

B. Process Flow

The flow diagram given in the Figure 1 below, presents the main idea and the operation of our proposed enrollment solution. The whole enrollment process consists of 6 main steps including preparations.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

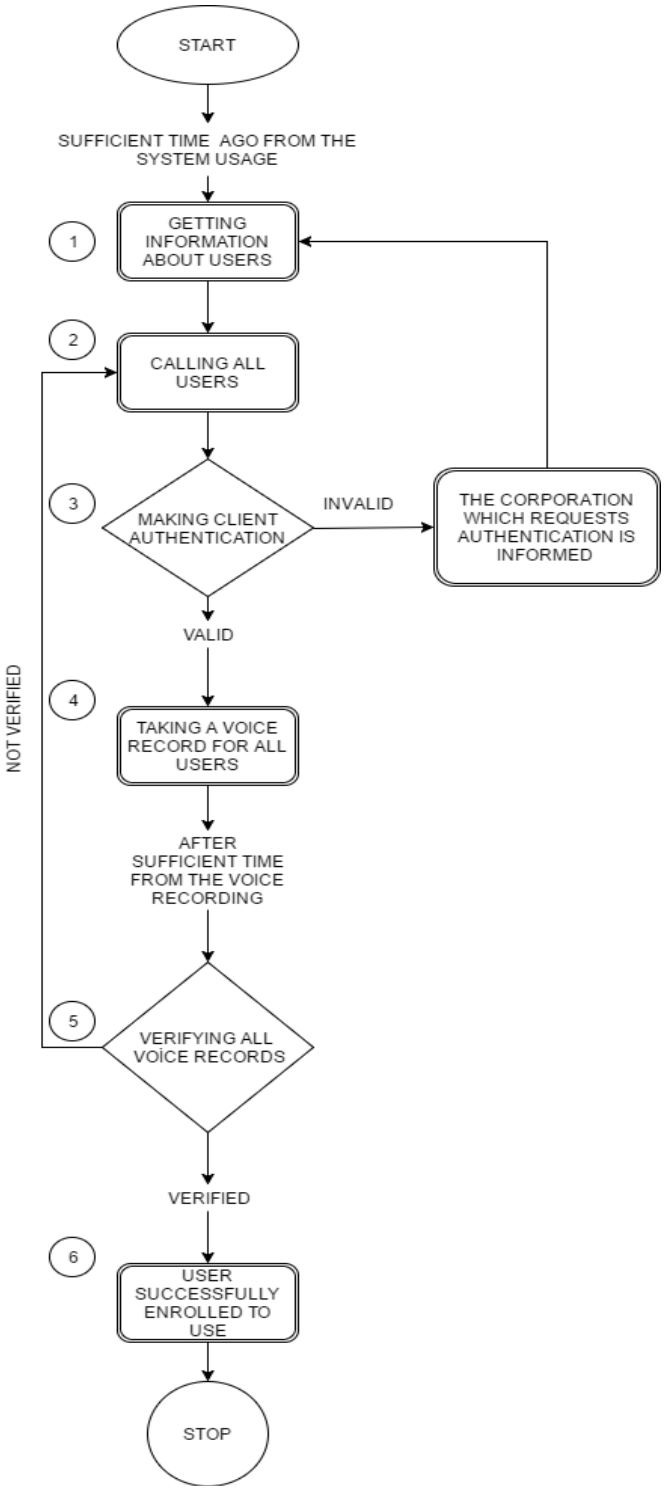

Fig. 1. Main process flow of the proposed enrollment procedure for one user

First, the organization who requires the authentication, which can be different than the company who has and provides the IVR system, shall prepare their member (user) records and install these to the IVR system as an external database. After the user database is integrated to the IVR system, some sufficiently long time (i.e. three months) before the intended service starts (i.e. online elections held, or customer survey started), the IVR provider company initiates an automated outbound call to all of these users to realize the authenticative enrollment of the users. In the third step, as a key operation, users who answered the call are asked several (at least two) specific questions, like mother's maiden name and other membership information including date of subscription, billing records etc. Use of a password is not intended since it could be hard to remember if the system will be used rarely. Plus, passwords generate another

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

enrollment problem for remote users. If a user cannot answer any of the questions, then the user shall not be enrolled for further authentication and organization authorities will be informed. If the user passes the question-based authentication, then his/her voice is recorded and a unique voiceprint is generated. This generated voiceprint shall be stored in databases of the IVR provider.

The fifth step is the second key point regarding the enrollment. Some random and sufficiently long time (i.e. one week) the after successful generation of the voiceprint, each user shall be called again. This time, they shall be asked to verify their voice by talking (for text independent IVR systems) or repeating a phrase (for text dependent IVR systems). If this authentication step is also successful, hence the user's enrollment process is completed, and the user is listed as ready. So, the user then can authenticate him/herself by his/her voice whenever required. If the user cannot pass this authenticative step, say another person is imitating the first one, then the voiceprint obtained in the third and fourth steps are removed and the user is transferred back to the second step to be sure about both the voiceprint is recorded from the correct person, and the person who verifies the voiceprint is the original owner of the recorded voiceprint. Optionally, the organization authorities can be informed, too. Thus, a two-stage authentication with multiple factors is achieved.

Whenever all (or most) of the users are enrolled, then the intended service (online elections, customer surveys, banking operations etc.) can be started by the organization authorities. During the service, for each use, each user should authenticate him/herself again with help of the voice authentication infrastructure of the IVR provider, which stores the generated voiceprint. The voiceprint should be stored at least as long as the service is active.

IV. SWOT ANALYSIS

In this section, we present a detailed analysis of positive and negative aspects of using an implementation of the proposed scheme.

A. Strengths

To collect people in the registration centers for recording voiceprints and verifying their identity is more expensive rather than realizing these processes by using phones/mobiles (via IVR). Besides the cost savings, the proposed system also increases the accessibility for people such as patients, disabled or abroad. The usage of phone calls ensures the time efficiency, too. On the other hand, this scheme provides a high degree of security (in terms of authentication), since it relies on two cascaded multi-factor authentication steps. Because of the system's fast post-processing ability and security, the recorded voice data can be stored like digital signatures and also can be used repeatedly for different operations. Another strength of the scheme is its aptitude for further developments of e-voting, e-governance, banking, digital signatures etc.

B. Weaknesses

Per to our observations, it seems very hard to convince people (incl. users/customers and institutions) that such a remote enrollment mechanism is secure enough to provide authentication for further online and phone-based services. This problem, in fact is valid for many online services like e-voting. Use of open-source software is not a sufficient solution here, since many users won't have time nor capability of checking the reliability of the system.

Another major drawback is the absence of hard copy records, like receipts, regarding the result of such an enrollment procedure. People may want to secure themselves with wet-signed documents because of their habits. Nevertheless, for legal situations, it may be more useful to store signed enrollment receipts rather than having a digital data like an SMS confirmation or an e-mail.

C. Opportunities

It is foreseen that the proposed mechanism enables three major opportunities for further applications and developments, via customized integrations. First, it allows fully remote enrollment for geographically distant users, which will increase participation for any kind of service. Because, the enrollment procedure will not require any physical effort for office visits. Second, any service that uses this mechanism will be able automate this enrollment procedure and will need much less workload for office staff. Plus, much more user can be reached cheaper. Hence, the system will be more scalable. Last but not least, many e-government services like e-voting will be easier to initiate. This opportunity also fetches to private sector companies and nonprofit organizations, who want to hold any kind of elections, like board elections or user/customer/personnel surveys.

D. Threats

Apart from the weaknesses; after analyzing different use cases, as well as e-voting, we come up with two severe threats: operator frauds and voice imitation. A malicious person may want to authenticate himself as someone else, by imitating the target person. Likewise, a malicious person, group or entity in the operator company may change or

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

distort the records in order to prevent someone from using the system, which requires the authentication, or may create a fake user. The latter is a more significant threat since imitating or replaying someone else's voice is mostly distinguishable by modern IVR systems, plus our method consists of a second verification phase. But, detecting operator frauds are much harder. Here, different signature mechanisms can be used to protect the integrity of the recorded user data. Plus, regulatory bodies must control the operators and enforce licensing.

V. CONCLUSION

In this conceptual work; we have proposed a two-step multi-factor authenticative scheme formed as a remote user enrollment procedure for online services that use IVR-based voice authentication technology.

The method proposed makes it easy to enroll potential users to the intended online service. And while it allows fully remote enrollments (without requiring to visit an office), it also considers security at a very high level. Plus, there is also no need for any kind of special hardware nor smart card, too. Since there are two steps regarding the authentication of a new user to be enrolled, which are separated with large time slots, the security level becomes increases. Thus, systems involving any kind of e-voting, sensitive surveys and many commercial operations can be set up securely, easier and cheaper. Likewise, this scheme will also make it easier or more feasible to develop many e-government and e-democracy services, as we mentioned in one of our previous works [16]. However, we should note two drawbacks; First, overall security of the system is limited by the accuracy of the membership (subscription) information stored at the organization. Second, accuracy and precision of the IVR system also constraints the overall security, which seems as less concerning as these systems improve rapidly.

As future works, we will build solutions for logical and technical issues in other subparts of a full e-voting system and make trials in cooperation with network operators. We will also work on mitigating the weaknesses and threats stated in the previous section. Reliability and success rates of IVR-based voice recognition is another topic for further studies.

ACKNOWLEDGMENT

We kindly thank to staff at Turkcell Technology and Turkcell Global Bilgi companies for their valuable support.

REFERENCES

[1] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," in Proceedings of the IEEE, vol. 91, no. 12, pp. 2021-2040, Dec 2003. DOI: 10.1109/JPROC.2003.819611

[2] G. Stocksdale. "NSA glossary of terms used in security and intrusion detection". SANS Institute Resources. 1998. [Online]. Available: http://www.sans.org/newlook/resources/glossary.htm

[3] Panse, Deepa, and P. Haritha. "Multi-factor Authentication in Cloud Computing for Data Storage Security". Int. Journal of Advanced Research in Computer Science and Software Engineering 4.8 (2014): 629-634.

[4] Hung, Tran Cong, Nguyen Thanh Tri, and Ho Nhut Minh. "An Enhanced security for government base on multifactor biometric authentication". Int. Journal of Computer Networks & Communications (IJCNC) Vol.8, No.6, Nov. 2016. DOI: 10.5121/ijcnc.2016.8605

[5] A. Kholmatov. "Privacy protecting biometric authentication systems". Diss. Sabanci University, Istanbul, 2008.

[6] S. S. M. Saqquaf, S. Araballi, P. Bhagyalakshmi, S. S. Mahadeek and B. M. Varsha, "Dynamically automated interactive voice response system for smart city surveillance". 2016 IEEE Int. Conf. on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, pp. 1176-1180. DOI: 10.1109/RTEICT.2016.7808017

[7] I. Patel, Pretesh B., and Tshilidzi Marwala. "Interactive voice response field classifiers". Systems, Man and Cybernetics, 2008. SMC 2008. IEEE Int. Conference on. IEEE, 2008.

[8] I. V. Vasyltsov, M. P. Karpinskyy and S. B. Kavka, "The structure of the voice authentication system". The Experience of Designing and Application of CAD Systems in Microelectronics, 2003. CADSM 2003. Proc. of the 7th Int. Conf., 2003, pp. 490-491. DOI: 10.1109/CADSM.2003.1255129

[9] G. T. Tsenov and V. M. Mladenov, "Speech recognition using neural networks". 10th Symposium on Neural Network Applications in Electrical Engineering, Belgrade, 2010, pp. 181-186. DOI: 10.1109/NEUREL.2010.5644073

[10] Venayagamoorthy, K. Ganesh, V. Moonasar, and K. Sandrasegaran. "Voice recognition using neural networks". Communications and Signal Processing, 1998. COMSIG'98. Proc. of the 1998 South African Symposium on. IEEE, 1998.

[11] A. Jain, D. Maltoni, D. Maio and J. Wayman. 2005. "Biometric Systems Technology, Design and Performance Evaluation". Springer-Verlag London limited, 2005.

[12] "Progressive Authentication - Voice Biometrics". Equifax Inc. 2012. [Online]. Available (as of 04 July 2017): http://www.equifax.com/technology_analytics/anakam/documentation/Anakam_Voice_Authentication.pdf

[13] M.N. Nachappa, A.M. Bojamma, C.N. Prasad, Nithya M. "Automatic Speaker Verification System". International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) Vol. 1, Issue 3, July 2014, PP 26-32

[14] O. Miguel-Hurtado, R. Blanco-Gonzalo, R. Guest and C. Lunerti, "Interaction evaluation of a mobile voice authentication system". 2016 IEEE Int. Carnahan Conference on Security Technology (ICCST), Orlando, FL, 2016, pp. 1-8. DOI: 10.1109/CCST.2016.7815697

[15] "Case study: Turkcell Global Bilgi - Nuance VocalPassword™ Deployment Achieves Industry-Leading Adoption Rates". Customer Care Solutions from Nuance. Nuance Communications Inc. 2011. [Online]. Available (as of 05 July 2017): http://www.speechouse.com/dokuman/nuance_turkcell_VocalPassword.pdf

[16] U. C. Çabuk, A. Çavdar and E. Demir, "E-Democracy: The Next Generation Direct Democracy and Applicability in Turkey", in Proc. INET'16, Ankara, 2016, paper 28. [Online]. Available (as of 05 July 2017): http://inet-tr.org.tr/inetconf21/bildiri/28.docx