sciendo

BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **23**, No 2
Sofia • 2023 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2023-0016

Defending Against Identity Threats Using Risk-Based Authentication

Lalitha Sravanti Dasu, Mannav Dhamija, Gurram Dishitha, Ajith Vivekanandan, V. Sarasvathi

Department of Computer Science and Engineering, PES University, Bangalore, India
E-mails: dasusravanti@gmail.com manav16dec@gmail.com dishi.lucky@gmail.com ajithsuba2001@gmail.com sarsvathiv@pes.edu

***Abstract*:** *Defending against identity-based threats, which have predominantly increased in the era of remote access and working, requires non-conventional, dynamic, intelligent, and strategic means of authenticating and authorizing. This paper aims at devising detailed risk-scoring algorithms for five real-time use cases to make identity security adaptive and risk-based. Zero-trust principles are incorporated by collecting sign-in logs and analyzing them continually to check for any anomalies, making it a dynamic approach. Users are categorized as risky and non-risky based on the calculated risk scores. While many adaptive security mechanisms have been proposed, they confine identities only to users. This paper also considers devices as having an identity and categorizes them as safe or unsafe devices. Further, results are displayed on a dashboard, making it easy for security administrators to analyze and make wise decisions like multifactor authentication, mitigation, or any other access control decisions as such.*

***Keywords*:** *Identity, zero trust principles, Risk-based authentication, Health-posture, Adaptive and dynamic access control model.*

1. Introduction

In today's cloud and remote working era, users (staff & administrators) access a significant number of corporate IT assets, such as business and enterprise applications, servers, etc., through laptops or mobile devices from outside the corporate-governed network boundary. In cloud computing, the gap between insiders and outsiders is very ambiguous and in certain cases, the outsiders become insiders [17]. This paradigm shift has brought several cyber security challenges to firms, and one of the biggest challenges has been Identity-based threats and exploitation, leading to several security breaches. The attack surface (externally exposed apps, user identities, user devices, emails, etc.) has increased significantly, and trusting and granting access based on static controls (which always provide the same results) is still a big risk. Thus, dynamic solutions that also consider the context and changing

environment and make access control policies accordingly are required. Traditional security measures would just include single-factor authentication, where just a username and password would suffice to authenticate a user and assume that the user is legitimate. Multi-factor authentication exists but is still not widely used and adapted. After authentication, what is the surety that the user will refrain from malicious and risky activities, and what is the surety that the device used for logging in is a compliant one? This shows that conventional security measures are static, and there is a need to adopt an approach that is dynamic and adaptive to the context.

The traditional approach makes a lot of assumptions about the users, and there is a need to incorporate the newly emerged concept of "ZERO TRUST PRINCIPLE". An identity is a set of things that define or characterize someone or something. For example, a person's identity includes the information they use to authenticate themselves, such as their username and password, and their level of authentication. An identity may be associated with a user, a device, an application, or something else. Just imagine if device identity is neglected, and in the era of remote working where employees are outside the company's bounded network and security norms, it would not even take minutes for any sort of malware to spread to other devices connected over the network, allowing attackers to escalate their privileges. Therefore, there is a need to enforce more dynamic approaches to securing identities, which could be made possible by techniques like Risk-Based Authentication (RBA) and zero-trust principles. Zero trust is a security model that enables organizations to provide secure access to their resources by emphasizing "never trust, always verify". It is based on three principles: verify explicitly (fully authenticating and authorizing every request before access is granted), least privileged access (authorizing a user only with the minimum rights that they require) and assume a breach (organizations can plan additional layers of security with this mindset).

As an integral and reliable solution, identity security safeguards the identities of an organization. This indicates that under certain circumstances, any identity, be it an IT admin, third-party vendor, remote worker, device, or application, can become privileged, creating an attack vector for an organization's most valuable assets. This makes it necessary for an identity security strategy based on privileged access management to protect all human and machine identities across the critical asset access cycle. Each identity must be accurately authenticated, authorized with the proper permissions, and given structured access to privileged assets as part of an identity security strategy. All these actions must be done in a way that can be audited to ensure the integrity of the entire process. Identity security should enable organizations to safeguard access across any device, anywhere, and at the appropriate time, preventing them from having to pick between security and productivity.

Identity-based attacks have long been regarded as an important channel for organizations to safeguard. However, in recent years, there has been a significant increase in the number and types of identities in use. Companies, for example, have quickly adopted cloud-based technology and services to provide appealing digital experiences for their consumers to gain a competitive advantage. This leads to an increase in attacks, such as DDoS attacks, due to the increase in devices joining the network because of the increasing popularity of cloud computing and the Internet of

Things [16]. There has also been a rise in support for a distributed and remote workforce [14]. All these developments escalated dramatically in 2020 when only companies with strong digital businesses thrived. At the same time, attackers are refining their techniques and developing new approaches, all of which result in new and extended danger dimensions. Thus, despite the distinctive properties, which the cloud has, its security aspects must get more attention in order to protect the cloud and maintain its sustainability [18].

Traditional password-based authentication has proved to be weak time and again [2-4]. Passwords can easily be guessed by exploiting a victim's Personally Identifiable Information (PII) thus putting the victim at risk.

One hacked account is generally the beginning of a security crisis. Once an attacker has gained access, they can raise privileges or obtain intelligence that will aid them in achieving their objectives. Therefore, identification has become the new security perimeter. To lessen the danger of a data breach, make it more difficult for attackers to steal identities while arming yourself with technologies that make it easier to detect compromised accounts. Hence, there is a need for stronger defence mechanisms in protecting identities.

Authentication refers to the process of proving that a person is who he/she says he/she is. Risk-based authentication is a type of authentication that tries to match the needed authentication credentials to the perceived risk of the connection or authorizations requested. On one hand, the goal is to lessen the authentication burden on users and give a better experience, while on the other, robust authentication is enforced when it is most needed [1]. The dynamic aspect of RBA is that it does not produce the same outcome for every authentication. Unlike static username and password approaches, it takes into consideration the context of the authentication and, most importantly, the risk associated with every circumstance, making it a very adaptive method to defend identities.

The motivation of the article lies in the fact that "identity is the first line of defence", that is, if detection is done at the basic front-line levels like authentication and before the attack could propagate to the network layers, it could potentially save a lot of time and effort [15]. The project is relevant from an industrial standpoint. The goal is to come up with and consider the aptest and appropriate risk factors, and risk estimation and evaluation techniques to defend against risky threats to the identities and mitigate them.

The rest of the paper is organized as follows: existing work on RBA is discussed in Section 2. Section 3 covers the proposed methodology. The result of the methodology implemented is presented in Section 4 and discussed in Section 5.

2. Related work

Authors in [5] propose a risk-based authentication solution for the problem of "lack of strength in non-continuous authentication in web applications". It involves the use of risk-based authenticators and dynamic risk engines. It mainly focuses on maintaining a high authentication frequency in which the user's authenticity is verified continuously but is transparent to the user. In situations of high risk, stronger

authentication is performed. This paper incorporates appropriate authentication based on the risk score derived from analysis of the identity's behavior, making it context-aware. Although it uses dynamic authentication, it only focuses on consumers or external entities, ignoring the possibility of an insider threat, which is likely to occur. This paper only considers the user identity context. However, the managed device health context is also important, especially for those internal users who may perform privileged tasks.

In [6], the authors provide a systematic review and examination of the state-of-the-art risk-based access control model. This model provides an in-depth understanding of the dynamic access control models that are needed. Other techniques that have been used include risk factors and risk estimation techniques. The paper provides detailed risk factors that are to be considered for risk-based control. It proposes nearly five risk estimation techniques and areas where they might be applicable and where they are not. The methods discussed are both risk- and context-aware, making them more specific to the threat under examination. Although this paper provides insight into the various approaches that could be used for risk-based access control, it does not mention any practical, implementable approaches to back up their study. Besides, the paper focuses mainly on attribute-based access controls, which are largely subject- and environment-based. This is good in terms of authorization and access controls, but it is very complex to gather, maintain, and leverage many attributes.

In [7] light is shed upon the ways in which cybercriminals are continually shifting the modes of attack and the best ways to combat those attacks are determined. It draws on insights, data, and signals from Microsoft, including the cloud, endpoints, and the intelligent Edge. Great insights are revealed about the consequences of the increase in the threat landscape and the change in the attack surface in the remote and hybrid working models. It emphasizes using zero trust principles, which provide the least privileged access and always assume a breach. This report provides zero trust controls for the six pillars: identities, applications, endpoints, network, infrastructure, and data. It also provides us with some riveting statistics on compromised users by various attack categories like password spray, phishing, and breach replay. It also has brought to light the fact that there has been a rise in the number of phishing emails. It claims that the security defence mechanism is far superior on the cloud than on-premises. However, large enterprises still deal with significant on-premises legacy systems, and the zero trust controls approach for this on-premises landscape has not been adequately addressed.

Authors in [8] analyze how Risk-Based Authentication is performed on eight online platforms, including such like Amazon, Facebook, Google, LinkedIn, and Twitch, based on IP address, user agent string, language, time parameter, and display resolution. The paper classifies these platforms into Single-Feature Models, Multi-Feature Models and VIP Models. The basic methodology it uses is to create accounts on these inspected online services and observe the behavior when accessing the services using these accounts in a variety of scenarios. The main benefit of this approach is the extent to which the analysis of these sites has been performed. The authors have created over 200 accounts for the inspection of the targeted online

platforms. Moreover, these online services have been trained over a period of 3.5 months. This does not consider all RBA approaches that could have been implemented by social media sites, like canvas fingerprinting. Their focus has been only on changing IP addresses, user agent strings, languages, time parameters, and display resolution. Furthermore, the risk scores have been estimated beforehand for the scenarios provided, making the approach static. In an ideal scenario, the risk scores would be calculated dynamically to ensure better detection of risks.

In [9], the writers present an intelligent risk-based authentication method based on temporal access behavior for general applications on mobile devices. Logs have been collected and features extracted based on user access to applications, and the data has been trained using random forest classifiers. This paper implements a dynamic approach by running the authentication implicitly in the background on smartphones. The classification of applications is purely based on the frequency of usage of the application, and this is not a good indicator since more important applications like banking applications could be used less frequently. Though the proposed methodology seems convincing, it has not been tested on real smartphone devices.

In [10], the author Kim Phan provides an analysis of the trustworthiness of user roles and system assets to improve the resilience of A-MFA systems. Improving the accuracy and complexity of adaptive MFA systems is critical for the system administrator. Techniques like risk-based authentication and the Dynamic Risk Engine have been used. Through this paper, a methodology is proposed that increases the robustness of the present A-MFA. This paper provides a methodology to calculate the trustworthiness of a user, which aids in identifying circumstances where consumers' trustworthiness rises. False rejections and inaccuracies in biometric technology have been observed. In addition, the load on users while utilizing this approach for adaptive MFA has not been measured.

Authors in [11] perform multiple methods of re-authentication and compare them to each other. The basic idea is to take the State Of The Art (SOTA), SUBJect (SUBJ), and LINK (LINK) methods and perform them independently on a set of volunteers to understand which is the most effective and reliable method out of the three. The following concept is to compare the many types of attacks that could be used against these security solutions. The full evaluation has been carried out in the presence of many users, which could aid in the detection and observation of non-functional security issues. The statistical method of evaluation has also helped in checking the effectiveness of a security method. Since the test has been carried out on a sample, it cannot always be said to represent the full population. Moreover, since consumers disable re-authentication when prompted too frequently in the real world, statistical data may have overlooked emotional tendencies.

It is observed that most of the works have employed "risk history" (the user's previous risk values for a certain resource, whether good or bad) as a primary risk factor (refer to Section 3.1 for more information on risk factors) and have not incorporated "zero trust principles". Mostly a high-level emphasis has been given on the risk factors that contribute to the risk scores but the detailed risk estimation technique (how to calculate risk scores considering the risk factors) has not been

discussed. The device identity has not been taken under consideration while evaluating the risk. Moreover, the risk factors are application-specific and vary from paper to paper.

3. Proposed methodology

The proposed methodology involves a dynamic risk analysis that detects anomalies, i.e., any form of suspicious behavior. Fig. 1 demonstrates the proposed methodology and model architecture.

Fig. 1 follows a three-tier architecture, which includes storing log information in the backend database, processing the input, analyzing the data using the risk scoring algorithms to calculate risk scores in the middle layer, and presenting results on a dashboard in the front end. The top layer is completely related to the user information gathered. When end users interact with Salesforce or Azure portal, their activity logs are collected with the help of MS Azure Active Directory. Microsoft Intune helps capture the ID of the device from which the user logs in (MAC address). There exists a table in the database consisting of all users whose credentials have been compromised and sold on the dark web. These would be compared every time with the credentials encountered during login. If the login credentials exist in the compromised credentials table, it means that there is a high risk associated with these credentials. The middle layer deals with processing the information gathered. All the information collected would be in JSON format. The "Data collection and transformation parser" parses this JSON data into strings and feeds this into the MySQL database using the "DB adaptor". The stored information is then analyzed using risk score algorithms for various use cases. If further advanced analysis is required, REST API calls could be made to the Active Directory for further help. After the analysis is done, the SOC analysts and application administrators are notified via dashboards or emails.

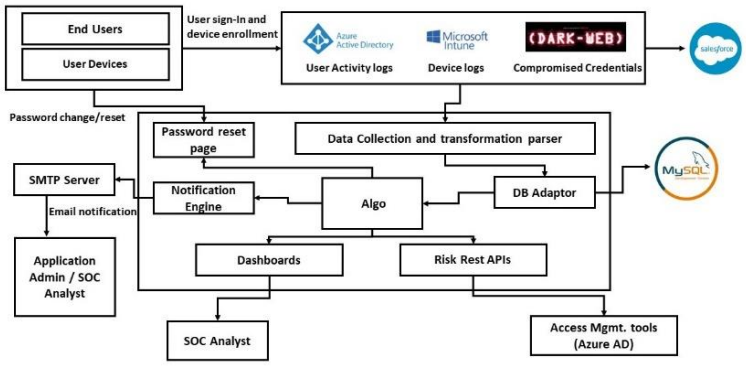


Fig. 1. Proposed architecture

This paper takes into consideration a few use cases that analyze the login logs of the users and produce a risk score. These risk scores are finally assigned weights based on their significance and a total risk score is derived. The SOC analyst is presented with a dashboard displaying the risk scores and details of all users and a decision on mitigating the user is taken.

3.1. Risk factors

Choosing risk variables that effectively influence access decisions is one of the key components of a risk-based access control approach. To calculate the risk value attached to the access request, a variety of risk variables can be used. These can include risk history, context, the benefits of the user (privileges a user acquires on being granted access to certain resources), etc. The proposed methodology has been implemented on a subset of real-world use cases, considering risk factors in this paper. Sections 3.1.1, 3.1.2, 3.1.3, 3.1.4, and 3.1.5 explain these use cases in detail.

3.1.1. Impossible travel

This risk factor makes use of user travel velocity based on the login time and location and assigns a travel risk score. This detects an impossible travel time between the locations and increases the risk score accordingly. The risk score can take any value in the interval 0-5. This algorithm calculates the risk score by comparing the velocity of the identified travel with the maximum possible travel velocity, which has been considered to be 240 m/s, the average speed of a flight [13]. Algorithm 1 shows the calculation of the risk score.

(1) $$\text{Risk Score} = 5 \times \frac{\text{Speed}}{\max \text{Speed}}.$$

Algorithm 1. Impossible Travel
Step 1. Begin
Step 2. *Input:* Geo points of consecutive login locations and time difference between login derived from sign-in logs using Azure
Step 3. *Output:* Travel Risk Score
Step 4. Calculate the speed of the possible travel
Step 5. If No change in location **Then**
Step 6. risk ← 0
Step 7. **return** risk
Step 8. End If
Step 9. If Change in location and time difference = 0 **Then**
Step 10. risk ← 5
Step 11. **return** risk
Step 12. End If
Step 13. maxSpeed ← 0.24
Step 14. risk ← 5 × (Speed/maxSpeed)
Step 15. return risk
Step 16. End

3.1.2. Anonymity Sign In

The Tor browser is a web browser that helps anonymize web traffic using the Tor network. In the case of an anonymous login, the entire login information is not available through the collected login logs, as the user might be using a Tor browser to conceal their identity. This behavior could be risky and must be mitigated. This information can be analyzed using the “riskEventType” attribute of the log.

Algorithm 2. Anonymity Sign In
Step 1. Begin
Step 2. *Input*: "riskEventType" field from sign in logs using Azure
Step 3. *Output*: If sign in was anonymous or not
Step 4. If riskEventType = "Anonymous" **Then**
Step 5. **return** true
Step 6. End If
Step 7. return false
Step 8. End

3.1.3. Atypical user Sign In

This use case focuses on the atypical behavior of user logins. Here, the user's logs are analyzed for abnormal and suspicious login attempts. The user's initial sign-in (during account creation) is compared with the last 10 logs. While verifying the last 10 sign-in logs, if consecutive logs match, we check if they match the initial sign-in. If the logs match the initial sign-in, we increase the confidence level by 0.5; if not, we increase it by 0.45. With every break in the consistency of the last 10 logs, the confidence is decreased by half. This algorithm ensures that if the latest 10 logs are all the same and match the initial sign-in, the user receives a confidence score of 5 (0.5×10), and if they are consistent but do not match the initial sign-in, a confidence score of 4.5 (0.45×10) is achieved. To check for consistency, we check for consistency in the location, browser, and device among the logs. Thus, this use case is divided into sub-use cases. Sections 3.1.3.1, 3.1.3.2, and 3.1.3.3 depict atypical location sign-in, atypical browser sign-in, and atypical device sign-in, respectively.

3.1.3.1. Atypical Location Sign In

This is a category of the use case "Atypical User Sign In" described in Section 3.1.3. Here, abnormalities in the user's location are accounted for, and the user is assigned a location risk score.

Algorithm 3.1. Atypical location login in
Step 1. Begin
Step 2. *Input*: Sign in logs of the user
Step 3. *Output*: Location risk score
Step 4. initialCountry ← user's country in earliest sign in log (during account creation)
Step 5. locationConfidence ← 0
Step 6. If totalLogs < 10 **Then**
Step 7. $n \leftarrow$ totalLogs
Step 8. Else
Step 9. $n \leftarrow 10$
Step 10. End If
Step 11. For $i \leftarrow$ totalLogs $-n$ **To** $n - 1$ **Do**
Step 12. **If** country(i) = country ($i + 1$) **Then**
Step 13. **If** country(i) = initialCountry **Then**
Step 14. locationConfidence += 0.5
Step 15. **Else**

Step 16. locationConfidence += 0.45
Step 17. **End If**
Step 18. **Else**
Step 19. locationConfidence /= 2
Step 20. **End If**
Step 21. End For
Step 22. Risk ← 5 – locationConfidence
Step 23. return risk
Step 24. End

3.1.3.2. Atypical Browser Sign in

This is a category of the risk factor "Atypical User Sign In" described in Section 3.1.3. Here, abnormalities in the user's browsers are accounted for, and the user **is** assigned a browser risk score.

Algorithm 3.2. Atypical browser login in
Step 1. Begin
Step 2. *Input*: Sign in logs of the user
Step 3. *Output*: Browser risk score
Step 4. initialBrowser ← user's browser in earliest sign in log (during account creation)
Step 5. browserConfidence ← 0
Step 6. If totalLogs < 10 **Then**
Step 7. n ← totalLogs
Step 8. Else
Step 9. n ←10
Step 10. End If
Step 11. For i ← totalLogs −n **To** n − 1 **Do**
Step 12. **If** browser (i) = browser (i + 1) **Then**
Step 13. **If** browser (i) = initialBrowser **Then**
Step 14. browserConfidence += 0.5
Step 15. **Else**
Step 16. browserConfidence += 0.45
Step 17. **End If**
Step 18. **Else**
Step 19. browserConfidence /= 2
Step 20. **End If**
Step 21. End For
Step 22. Risk ← 5 – browserConfidence
Step 23. return risk
Step 24. End

3.1.3.3. Atypical Device Sign In

This is a category of the risk factor "Atypical User Sign In" described in Section 3.1.3. Here, abnormalities in the user's device are accounted for, and the user is assigned a device risk score.

Algorithm 3.3. Atypical device login in
Step 1. Begin
Step 2. *Input*: Sign in logs of the user
Step 3. *Output*: Device risk score
Step 4. initialDevice ← user's device in earliest sign in log (during account creation)
Step 5. deviceConfidence ← 0
Step 6. If totalLogs < 10 **Then**
Step 7. $n \leftarrow$ totalLogs
Step 8. Else
Step 9. $n \leftarrow 10$
Step 10. End If
Step 11. For $i \leftarrow$ totalLogs − n **To** $n - 1$ **Do**
Step 12. **If** device (i) = device (i + 1) **Then**
Step 13. **If** device (i) = initialDevice **Then**
Step 14. deviceConfidence += 0.5
Step 15. **Else**
Step 16. deviceConfidence += 0.45
Step 17. **End If**
Step 18. **Else**
Step 19. deviceConfidence /= 2
Step 20. **End If**
Step 21. End For
Step 22. Risk ← 5 – deviceConfidence
Step 23. return risk
Step 24. End

3.1.4. Compromised credentials

The dark web is a black market where transactions involving drugs, stolen credentials, credit card details, etc. are sold or brokered. This is a part of the world wide web that isn't visible to search engines and requires an anonymous browser like Tor to be accessed. Credentials hacked by hackers can be bought from the dark web by paying hundreds to thousands of dollars or even more. [12] shows the dark web scan of a sample company. Ideally, compromised credentials must be purchased from the dark web and stored in a database, but for implementation purposes, a few of the user's credentials are manually added into a table called "user compromised" as credentials that have been hacked and sold on the dark web. Using this risk factor, it is verified during user login if the credentials used by the user are compromised. If the account is found to be compromised, the user is redirected to the reset password page.

Algorithm 4. Compromised Credentials
Step 1. Begin
Step 2. *Input*: Username and password from user login page
Step 3. *Output*: Redirect to the necessary page and update password risk
Step 4. If user credentials are found in "user_compromised" table **Then**

Step 6. update password risk as 5 in database
Step 7. redirect to “reset password” page
Step 8. Else
Step 9. update password risk as 0 in database
Step 10. authenticate user and successfully login
Step 11. End If
Step 12. End

3.1.5. Login from Unsafe Device

A device is safe if it is registered on Azure AD and unsafe if it is not. Microsoft Intune helps collect device ID information. During login, if the sign-in logs contain a device ID associated with the respective user, the device is safe. If the "deviceId" field in the sign-in logs happens to be empty, this is an indication that the user is logging in from a device that is not registered, making it an unsafe device.

Algorithm 5. Login from Unsafe Device
Step 1. Begin
Step 2. *Input*: Login logs of user
Step 3. *Output*: Whether user logged in using registered device or not
Step 4. deviceID ← device ID obtained from logs
Step 5. If deviceID not empty **Then**
Step 6. **return** true
Step 7. End If
Step 8. return false
Step 9. End

3.2. Calculation of final risk score

The final risk score of the user is calculated using the risk scores derived from the use cases. These scores are given weights based on the priority and importance they hold in detecting an identity threat.

The use cases “unsafe device” and “anonymous sign-in” do not have a risk score associated with them and have not been added to the final risk score. However, these risk factors have been indicated on the dashboard when a user is further examined. This is because these risk factors carry a Boolean value of yes or no. These risk factors, though important, would not justify their impact on the risk score. For example, if we consider the 'unsafe device' use case, the risk score can be 0/5 or 5/5. However, if a legitimate user logs in from an unregistered device, carrying a risk score of 5/5 would not be justified. Hence, these risk factors are accounted for in the “total number of violations” and are mentioned on every user’s examination page. On the other hand, the “compromised credentials” use case is accounted for in the total risk score since a user whose credentials are compromised and available on the dark web is at great threat, and a score of 5/5 is justified. It is possible to consider unsafe devices and anonymous sign-in for the final risk score by giving them a very small weight, but this is avoided here in this paper. Below is the derivation of the total risk score.

Let:

w1 = Weight of travel risk in final risk score
w2 = Weight of location risk in final risk score
w3 = Weight of browser risk in final risk score
w4 = Weight of device risk in final risk score
w5 = Weight of password risk in final risk score
r1 = Travel risk obtained using (1) and Algorithm 1
r2 = Location risk obtained using Algorithm 3.1
r3 = Browser risk obtained using Algorithm 3.2
r4 = Device risk obtained using Algorithm 3.3
r5 = Password risk obtained using Algorithm 4

(2) $\text{TotalRisk} = (w1 \times r1 + w2 \times r2 + w3 \times r3 + w4 \times r4 + w5 \times r5).$

The total risk can have a maximum value of five and each of the individual risk scores can have a maximum value of five. Hence from (2), the sum of all weights=5.

Based on the importance of the use case in determining if an identity threat has occurred, this paper has considered the weights to have a relation to the form:

(3) $$w1 > w2 > w3 = w4 > w5.$$

(3) describes the relative weights of the risk scores. w1 (travel) must be the highest followed by w2 (location). The weight assigned for browser and device risks must be much lesser than location. Password risk is weighted the least. With the help of (3), the weights have been calculated using harmonic series. w1 is assigned an initial value of 1 and the rest follow a harmonic series as shown in (4).

(4) $$x\left(1 + \left(\frac{1}{2}\right) + \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right) + \left(\frac{1}{16}\right)\right) = 5.$$

Solving for *x* in (4),

(5) $$x = \frac{16 \times 5}{33} = \frac{80}{33}.$$

Substituting *x* in (4),

(6) $$w1 = \frac{80}{33},$$

(7) $$w2 = \frac{40}{33},$$

(8) $$w3 = \frac{20}{33},$$

(9) $$w4 = \frac{20}{33},$$

(10) $$w5 = \frac{5}{33}.$$

Substituting (6), (7), (8), (9), and (10) in (2),

(11) $$\text{TotalRisk} = \frac{(80 \times r1 + 40 \times r2 + 20 \times r3 + 20 \times r4 + 5 \times r5)}{33 \times 5}.$$

3.3. User mitigation

A “Risk Investigation” group is created in Azure Active Directory. The analyst can add any user found to be suspicious or risky to the risk investigation group by clicking the mitigate button on the dashboard. Further investigation and action can be taken by the admin on these users. Fig. 2 shows the users added to the risky group on Azure AD.

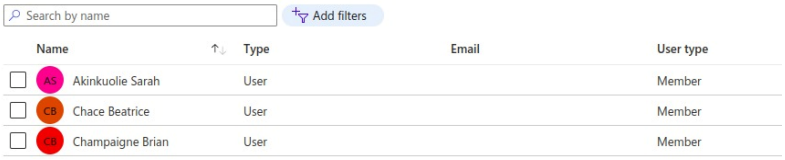


Fig. 2. Risk investigation group on Azure AD

4. Results and discussions

4.1. Implementation

50 users were onboarded into an Azure AD tenant. User details and sign-in logs on Azure AD are collected using Java 11 with the help of graph APIs when users interact with Salesforce. These logs are added to the database using "MySQL connector-java-8.0.11" connector for Java and risk scores are calculated upon analyzing the logs collected. These risk scores are then updated in the database. A final risk score is calculated for each user based on the scores obtained in the use cases.

APIs accessed via Tomcat 10 server are created for getting and putting data into the database. Finally, a dashboard has been created with the help of Jakarta servlets to access the findings and bootstrap to present them. The project has been built using Maven with the help of Eclipse IDE.

Fig. 3. Admin dashboard

- The admin dashboard is shown in Fig. 3 where all the users' information is displayed.
- The rows of this table can be sorted based on various attributes like username, initial login device, initial login location, the total number of login attempts, confidence level, risk level, and number of violations (number of risk factors having

score > 3 and any other Boolean violation) by the user. This enables the admin to sort based on risk scores or the number of violations and take required actions on risky users.

- The “examine” button provided along with each user provides more information about the user.
- If suspected risky, the admin can mitigate using the “mitigate” button against each user. This adds the user to the “Risk Investigation” group created on Azure AD and the admin can take further actions on the user.
- In Fig. 3, the users highlighted in green are safe and those in red require further verification. These four users will be examined in detail in the following sections.

4.2. Risky user examination

- It can be seen from Fig. 3 that Sarah has a final risk score of 3.26/5 which makes her a risky user as the risk score>3.
- When the “Examine” button is clicked more details of the respective user are displayed as shown in Fig. 4.

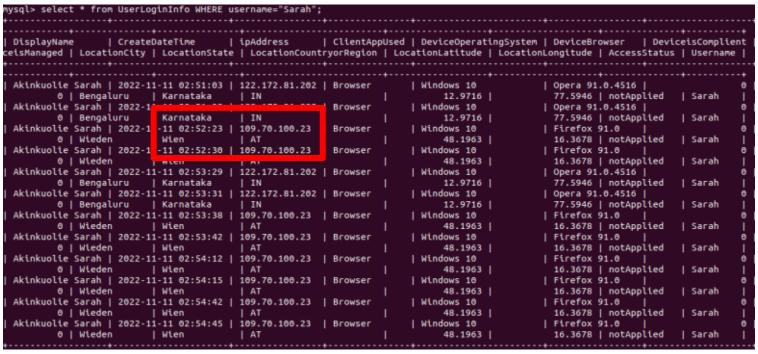
Fig. 4. More Information about user Sarah

- Anonymous Login shows “YES” implying that Sarah would have used a TOR browser to login in the last few attempts. Therefore, there is a need to enforce MFA or verify further if it is genuinely Sarah who is using the TOR browser or a hacker who is using her account anonymously.


```
"id": "4deddba4-453f-412f-b058-dd4f5a790200",
"createdDateTime": "2022-11-11T02:54:42Z",
"userDisplayName": "Akinkuolie Sarah",
"userPrincipalName": "sarah@dishipesloutlook.onmicrosoft.com",
"userId": "43e63117-f746-4b8e-a6df-6515d49727b8",
"appId": "c44b4083-3bb0-49c1-b47d-974e53cbdf3c",
"appDisplayName": "Azure Portal",
"ipAddress": "109.70.100.23",
"clientAppUsed": "Browser",
"correlationId": "0344e6ed-047b-450a-8c49-8d04ab85d1df",
"conditionalAccessStatus": "notApplied",
"isInteractive": true,
"riskDetail": "none",
"riskLevelAggregated": "high",
"riskLevelDuringSignIn": "high",
"riskState": "atRisk",
"riskEventTypes": [
    "anonymizedIPAddress"
],
"riskEventTypes_v2": [
    "anonymizedIPAddress"
],
```


Fig. 5. Sarah’s login record indicating anonymous login

- In Fig. 5, the “riskEventTypes” field shows “anonymizedIPAdress” which is an indication of an anonymous sign-in.
- The graph in Fig. 4 also shows Sarah has got an extremely high-risk score of 5 in the use case “impossible travel” alone and moderate risk scores in the other use cases making the travel risk extremely high affecting the final risk score.



```
mysql> select * from UserLoginInfo WHERE username="Sarah";
| DisplayName       | CreateDateTime      | ipAddress       | ClientAppUsed | DeviceOperatingSystem | DeviceBrowser    | DeviceisComplient
ceisManaged | LocationCity | LocationState | LocationCountryorRegion | LocationLatitude | LocationLongitude | AccessStatus | Username |
| Akinkuolie Sarah | 2022-11-11 02:51:03 | 122.172.81.202 | Browser       | Windows 10            | Opera 91.0.4516 |                 0
        0 | Bengaluru    | Karnataka     | IN                      |          12.9716 |           77.5946 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-1                                         Browser       | Windows 10            | Opera 91.0.4516 |                 0
        0 | Bengaluru      Karnataka     | IN                      |          12.9716 |           77.5946 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-1 -11 02:52:23 | 109.70.100.23    Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden         Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-1 -11 02:52:30 | 109.70.100.23    Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden                                                 |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:53:29 | 122.172.81.202 | Browser       | Windows 10            | Opera 91.0.4516 |                 0
        0 | Bengaluru    | Karnataka     | IN                      |          12.9716 |           77.5946 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:53:31 | 122.172.81.202 | Browser       | Windows 10            | Opera 91.0.4516 |                 0
        0 | Bengaluru    | Karnataka     | IN                      |          12.9716 |           77.5946 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:53:38 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:53:42 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:54:12 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:54:15 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:54:42 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
| Akinkuolie Sarah | 2022-11-11 02:54:45 | 109.70.100.23  | Browser       | Windows 10            | Firefox 91.0    |                 0
        0 | Wieden       | Wien          | AT                      |          48.1963 |           16.3678 | notApplied   | Sarah    |
```

Fig. 6. Sarah’s sign-in logs

- Fig. 6 depicts sign-in logs of Sarah. It is observed that she logged in from Karnataka, India, and within a very less time difference, she even logged in from Wilen, Austria which is practically not possible. As there is a maximum velocity it takes to travel from India to Austria, but the velocity calculated from the collected logs is excessively higher than the maximum one. Therefore, the travel risk of this user calculated using the risk scoring algorithms devised is high and this could also hint at a DDOS situation.

4.3. Safe user examination

- User Karthik as observed in Fig. 3 shows that he has a high confidence level which makes it obvious that the final risk score is very low implying that this user is not a risky user.
- Fig. 7 depicts the risks related to all the use cases, so the location, browser, and device risks have a score lesser than 2 making them non-risky.

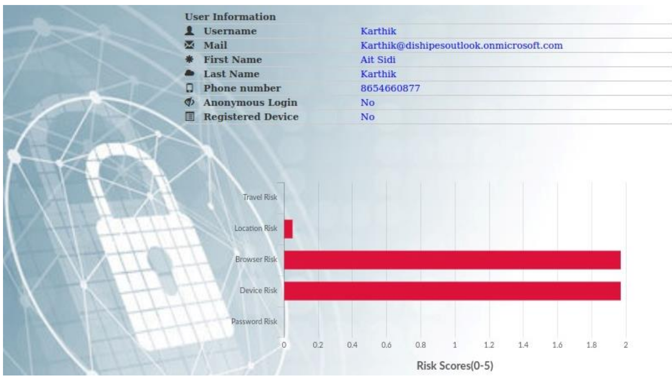
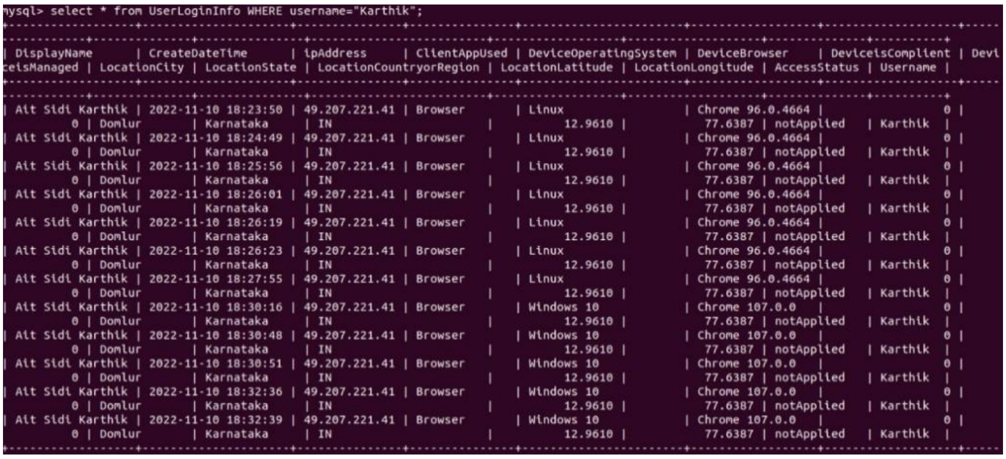


Fig. 7. More Information about user Karthik



mysql> select * from UserLoginInfo WHERE username="Karthik";
DisplayName	CreateDateTime	ipAddress	ClientAppUsed	DeviceOperatingSystem	DeviceBrowser	DeviceisComplient	DeviceisManaged	LocationCity	LocationState	LocationCountryorRegion	LocationLatitude	LocationLongitude	AccessStatus	Username
Ait Sidi Karthik	2022-11-10 18:23:50	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:24:49	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:25:56	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:26:01	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:26:19	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:26:23	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:27:55	49.207.221.41	Browser	Linux	Chrome 96.0.4664	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:30:16	49.207.221.41	Browser	Windows 10	Chrome 107.0.0	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:30:48	49.207.221.41	Browser	Windows 10	Chrome 107.0.0	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:30:51	49.207.221.41	Browser	Windows 10	Chrome 107.0.0	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:32:36	49.207.221.41	Browser	Windows 10	Chrome 107.0.0	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik
Ait Sidi Karthik	2022-11-10 18:32:39	49.207.221.41	Browser	Windows 10	Chrome 107.0.0	0	0	Domlur	Karnataka	IN	12.9610	77.6387	notApplied	Karthik

Fig. 8. Karthik log details

4.4. Safe device

- As shown in Fig. 9, the dashboard displays user Colby's device as a "registered device" indicating that the device, which he used to sign in, is safe and registered.
  - A device is considered registered if Microsoft Intune can collect the deviceId in the sign-in logs as shown in Fig. 10.

User Information
Username Colby
Mail Colby@dishipesoutlook.onmicrosoft.com
First Name Andreola
Last Name Colby
Phone number 2369900313
Anonymous Login No
Registered Device Yes
Travel Risk
Location Risk
Browser Risk
Device Risk
Password Risk
0
0.5
1
1.5
2
2.5
3
Risk Scores(0-5)

Fig. 9. Registered device

"createdDateTime": "2022-11-15T08:27:37Z",
"userDisplayName": "Andreola Colby",
"userPrincipalName": "colby@dishipes1outlook.onmicrosoft.com",
"userId": "d244b3eb-7b35-4f21-bf86-e4eb4fb64b36",
"appId": "c44b4083-3bb0-49c1-b47d-974e53cbdf3c",
"appDisplayName": "Azure Portal",
"ipAddress": "49.205.32.143",
"clientAppUsed": "Browser",
"correlationId": "1ee27f61-ba0b-447c-a5de-b076a54fd42d",
"conditionalAccessStatus": "notApplied",
"isInteractive": true,
"riskDetail": "none",
"riskLevelAggregated": "none",
"riskLevelDuringSignIn": "none",
"riskState": "none",
"riskEventTypes": [],
"riskEventTypes_v2": [],
"resourceDisplayName": "Windows Azure Service Management API",
"resourceId": "797f4846-ba00-4fd7-ba43-dac1f8f63013",
"status": {
"errorCode": 0,
"failureReason": "Other.",
"additionalDetails": null
},
"deviceDetail": {
"deviceId": "86cde253-da96-4455-87f6-5e82bc334f09",

Fig. 10. Colby deviceId

4.5. Unsafe device

In Fig. 11, “Registered Device” field displays “NO” as in the graph explorer as seen in Fig. 12, the “deviceId” field is empty indicating that the user Beatrice has logged in from an unregistered device making it an unsafe device.

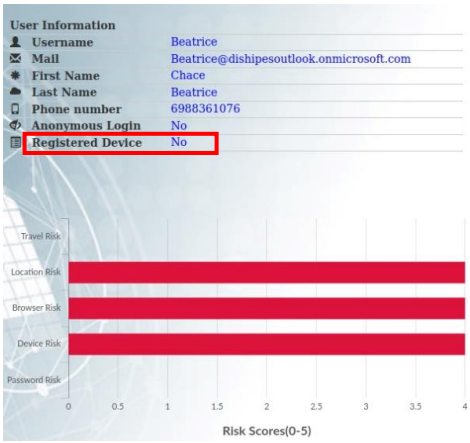


Fig. 11. Unregistered device

Fig. 12. Beatrice empty deviceId

4.6. Compromise credentials

Fig. 13 depicts Table “user_compromised” which contains all the compromised credentials and Fig. 14 contains the current credentials of the users. When user Michael tries to log in using his credentials, he would be asked to reset his password and would be redirected to the reset password page as depicted in Figs 15 and 16, respectively.

username	password
Chris	Mw8DJDJsw2RzzUB
Max	Hi_ok_123
Lin	Lana8433
Mia	password111
Lowan	UsaRules@007
Rick	championsOfEngaland123
Michael	Gaga2390

Fig. 13. Compromised credentials table

username	password
Michael	Gaga2390
Judith	Qumu9471
Frank	!PyLk247S254
Bonalyn	g5aCj3f6%HeK
Charles	oRJc&4w44YUW
Donna	Tx609V489@JQ
Mia	r%9h9Q3EvHC4

Fig. 14. Users’ current credentials

Fig. 15. Reset password prompt

Fig. 16. Reset password page

5. Conclusion and future enhancements

This paper proposes a methodology to protect identities by defending them against any threats by considering identity (instead of network) as a security perimeter. A modern, strategic, and intelligent way of authenticating is introduced. The sign-in logs are collected and analyzed every "30 minutes", implying that the risk scores are also calculated and updated every 30 minutes. This approach justifies the dynamic aspect of the work done and incorporates the zero-trust principles of "keep verifying and always assuming a breach". Five real-time use cases have been chosen, and accordingly, risk-scoring algorithms have been devised. Thus, making this approach risk-based and adaptive to the changing environment, lets appropriate actions or decisions to be taken (mitigating risky users). Hence, even though a user has been authenticated with the right set of credentials, it has never been assumed that the user is legitimate. Utmost priority is given to the risk scores obtained by continually checking and verifying for any risky behavior or anomaly.

The scope of the paper is confined to a subset of the millions of real-time use cases, but there is scope to extend them to more real-time use cases. Users and devices (health postures) have been considered as identities, but service applications could also be taken into consideration. The future scope could also include a detailed exploration of access control decisions. At present, risky users are just mitigated, but more sophisticated actions can be taken against them.

References

1. Hassan, A., B. Nuseibeh, L. Pasquale. Engineering Adaptive Authentication. – In: Proc. of IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C'21), IEEE, 2021, pp. 275-280.
2. Lal, N. A., S. Prasad, M. Farik. A Review of Authentication Methods. – International Journal of Scientific & Technology Research, Vol. **5**, 2016, pp. 246-249.
3. Bonneau, J. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. – In: Proc. of IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 538-552.
4. Wang, D., Z. Zhang, P. Wang, J. Yan, X. Huang. Targeted Online Password Guessing: An Underestimated Threat. – In: Proc. of ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 1242-1254.
5. Steinegger, R. H., D. Deckers, P. Giessler, S. Abeck. Risk-Basedauthenticator for Web Applications. – In: Proc. of 21st European Conference on Pattern Languages of Programs, 2016, pp. 1-11.
6. Atlam, H. F., M. A. Azad, M. O. Alassafi, A. A. Alshdadi, A. Alenezi. Risk-Based Access Control Model: A Systematic Literature Review. – Future Internet, Vol. **12**, 2020, No 6, p.103.
7. Microsoft Digital Defence Report 2021. **https://www.microsoft.com/en-us/security/business/microsoft-digital-defense-report-2021**
8. Wiefling, S., L. L. Iacono, M. Dürmuth. Is This Really You? An Empirical Study on Risk-Based Authentication Applied in the Wild. – In: Proc. of IFIP International Conference on ICT Systems Security and Privacy Protection, Cham, Springer, 2019, pp. 134-148.
9. Ashibani, Y., Q. H. Mahmoud. An Intelligent Risk-Based Authentication Approach for Smartphone Applications. – In: Proc. of IEEE International Conference on Systems, Man, and Cybernetics (SMC'20), IEEE, 2020, pp. 3807-3812.

10. P h a n, K. Implementing Resiliency of Adaptive Multi-Factor Authentication Systems. – Culminting Projects in Information Assurance, Vol. **65**, 2018, p. 96.
11. W i e f l i n g, S., T. P a t i l, M. D ü r m u t h, L. L. I a c o n o. Evaluation of Risk-Based Re-Authentication Methods. – In: Proc. of IFIP International Conference on ICT Systems Security and Privacy Protection, Cham, Springer, 2020, pp. 280-294.
12. Compromised Credentials Report.
https://restech.solutions/wp-content/uploads/Dark-Web-Scan-Sample-Company.pdf
13. Vistas – How Fast Is a Private Jet?
https://www.vistajet.com/en/stories/fastest-private-jet/#:~:text=Most%20commercial%20aircraft%20typically%20fly,according%20to%20Flight%20Deck%20Friend.
14. Microsoft Report Shows Increasing Sophistication of Cyber Threats.
https://blogs.microsoft.com/on-the-issues/2020/09/29/microsoft-digital-defense-report-cyber-threats/
15. Gartner Identifies Top Security and Risk Management Trends for 2021.
https://www.gartner.com/en/newsroom/press-releases/2021-03-23-gartner-identifies-top-security-and-risk-management-t?s=09
16. V a t t i k u t i, S., M. R. H e g d e, M. M a n i s h, V. B o d d u v a r a m, V. S a r a s v a t h i. DDoS Attack Detection and Mitigation Using Anomaly Detection and Machine Learning Models. – In: Proc. of IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS'21), 2021, pp. 1-6.
17. K u m a r, P. R., P. H. R a j, P. J e l c i a n a. Exploring Security Issues and Solutions in Cloud Computing Services – A Survey. – Cybernetics and Information Technologies, Vol. **17**, 2017, No 4, pp. 3-31.
18. A l o s a i m i, W., M. Z a k, K. A l-B e g a i n, R. A l r o o b a e a, M. M a s u d. Mitigation of Distributed Denial of Service Attacks in the Cloud. – Cybernetics and Information Technologies, Vol. **17**, 2017, No 4, pp. 32- 51.

Received: 08.12.2022; Second Version: 26.04.2023; Third Verion: 04.05.2023;
Accepted: 12.05.2023