Oct 26th, 8:30 AM - Oct 28th, 5:00 PM

International Conference on Computer Science and Communication Engineering

University for Business and Technology - UBT

Leadership and Innovation

Education | Research | Training | Consulting | Certification

PROCEEDINGS

8th UBT ANNUAL INTERNATIONAL CONFERENCE

26 - 28 OCTOBER

UBT Innovation Campus

INTERNATIONAL CONFERENCE ON **COMPUTER SCIENCE AND COMMUNICATION ENGINEERING**

Proceedings of the

8th Annual International Conference

International Conference Computer Science and Communication Engineering

Edited by

Edmond Hajrizi

October, 2019

Conference Book of Proceedings

International Conference

Pristina, 26-28 October 2019

ISBN 978-9951-437-85-1

International Conference on Business, Technology and Innovation
Pristina, Kosovo 26-28 October 2019

Editor: Edmond Hajrizi

Organizing Committee: [illegible]

Authors themselves are responsible for the integrity of what is being published.

Publisher,
UBT

Editor Speech of IC - BTI 2019

International Conference is the 8th international interdisciplinary peer reviewed conference which publishes works of the scientists as well as practitioners in the area where UBT is active in Education, Research and Development. The UBT aims to implement an integrated strategy to establish itself as an internationally competitive, research-intensive institution, committed to the transfer of knowledge and the provision of a world-class education to the most talented students from all backgrounds. It is delivering different courses in science, management and technology. This year we celebrate the 18th Years Anniversary. The main perspective of the conference is to connect scientists and practitioners from different disciplines in the same place and make them be aware of the recent advancements in different research fields, and provide them with a unique forum to share their experiences. It is also the place to support the new academic staff for doing research and publish their work in international standard level. This conference consists of sub conferences in different fields: - Management, Business and Economics - Humanities and Social Sciences (Law, Political Sciences, Media and Communications) - Computer Science and Information Systems - Mechatronics, Robotics, Energy and Systems Engineering - Architecture, Integrated Design, Spatial Planning, Civil Engineering and Infrastructure - Life Sciences and Technologies (Medicine, Nursing, Pharmaceutical Sciences, Physcology, Dentistry, and Food Science),- Art Disciplines (Integrated Design, Music, Fashion, and Art).

This conference is the major scientific event of the UBT. It is organizing annually and always in cooperation with the partner universities from the region and Europe. In this case as partner universities are: University of Tirana – Faculty of Economics, University of Korca. As professional partners in this conference are: Kosova Association for Control, Automation and Systems Engineering (KA – CASE), Kosova Association for Modeling and Simulation (KA – SIM), Quality Kosova, Kosova Association for Management. This conference is sponsored by EUROSIM - The European Association of Simulation. We have to thank all Authors, partners, sponsors and also the conference organizing team making this event a real international scientific event. This year we have more application, participants and publication than last year.

Congratulations!

Edmond Hajrizi,

Rector of UBT and Chair of IC - BTI 2019

CONTENTS

Password typo correction using discrete logarithms

Nikola K. Blanchard

Digitrust, Loria, Universit~ de Lorraine

Abstract. As passwords remain the main online authentication method, focus has shifted from naive entropy to how usability improvements can increase security. Chatterjee et al. recently introduced the first two typo- tolerant password checkers, which improve usability at no security cost but are technically complex. We look at the more general problem of computing an edit distance between two strings without having direct access to those strings — by storing the equivalent of a hash. We propose a simpler algorithm for this problem that is asymptotically quasi-optimal in both bits stored and exchanged, at the cost of more computation on the server.

Keywords: Usable security · Passwords · Discrete logarithm

Introduction

Despite recent advances in biometric authentication [12] and account linking [2], passwords are still the main method of authentication used online and will prob- ably remain so in the near future. Countless studies have been written on the pit- falls of password-based authentication [11], with users creating bad passwords [4] or repeatedly dodging security measures [15,10]. Failing to login is increasingly frustrating, and forgetting one's password is now about as frustrating as for- getting one's keys [5]. To improve usability, some services like Facebook have discreetly adopted typo correction for the 2-3 most frequent typos, such as for- getting the caps lock or capitalising the first character on mobile [9].
In an innovative paper in 2016 [6], Chatterjee et al. discovered that authenti- cation failures could turn 3% of the users away, but that a vast majority of errors comes from a few simple typos. They also developed a system called TypTop [7], which is efficient both computationally and memory-wise, and corrects up to 32% of typos. This system works by keeping a cache of allowed password hashes corresponding to the frequent typos made by the user, and updates this cache at each successful authentication. Those systems can actually have a positive impact on security as they make long passwords — which are more error-prone — much more usable, lowering the cost of using highly secure passwords.
We look at the general problem of storing information on the server that can allow typo correction while preventing an adversary in control of the server from computing the passwords from the stored information.

Main results. We introduce a metric called the keyboard distance, and a protocol to compute this distance (or the Hamming distance) between a queried string and a secret string, without it being possible to find the secret string in polynomial time (assuming the security of the discrete logarithm). This is non-trivial, as it was shown in [3] that any distance computation protocol can find the original password in a polynomial number of queries, which we prevent by having queries of non-uniform complexity.

Keyboard distance and algorithm

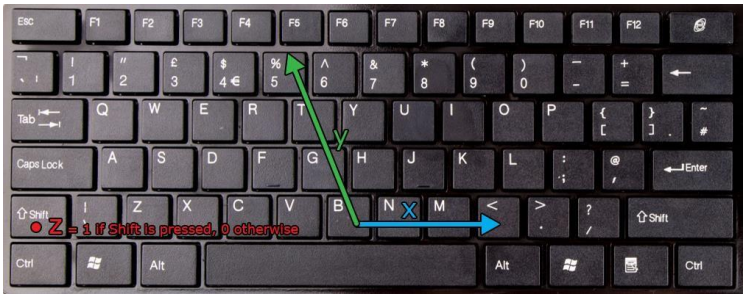


Fig. 1. Keyboard coordinate system, starting at the bottom left. The string "Arc" has coordinates ((1, 1, 1), (4, 2, 0), (3, 0, 0)).

Before the algorithm, we must first introduce a distance between strings which, although simple, is not generally used. Let's consider a keyboard, with a standard QWERTY layout, as in Figure 1. The 48 main keys of the keyboard and the different characters they can create can easily be modelled by a 3-dimensional coordinate system. The first dimension corresponds to the horizontal position of the key (or the row), the second dimension to the vertical (the diagonal column), and the third dimension to the modifiers, here only considering Shift although it could easily be extended. This forms a subset of a 14 x 4 x 2 latticel as shown in Figure 1.

Definition 1. *Let s be a string of length n. The string coordinates of s are defined as the sequences $(x_i)_{1\leq i\leq n}$,$(y_i)_{1\leq i\leq n}$ and $(z_i)_{1\leq i\leq n}$, where (x_i, y_i, z_i) are the coordinates of the i-th letter in the previous coordinate system.*

Definition 2. *Let s and s' be strings of identical length n. Let the keyboard distance between s and s' be defined as the L^1-distance between their string coordinates: $d(s, s') = \sum_{1\leq i\leq n} (|x_i - x'_i| + |z_i - z'_i| + |z_i - z'_i|)$.*

By this definition, the distance between homomorphic and homimorphic is 1, but the distance between homomorphic and Bomomorphic is 3, the same as the distance between homomorphic and homomor;jkc.

The expected distance between two random n-character strings is then $\frac{59707}{10296}$ n, or about 58 for 10-character keymashes.

Definition 3. *Let s be a string of length n, and let $(x_i)_{1\leq i\leq n}$,$(y_i)_{1\leq i\leq n}$ and $(z_i)_{1\leq i\leq n}$ be its string coordinates. Let p_i be the i-th prime number. We define the integral representation $X(s)$ of s as*

$$X(s) = \prod_{1\leq i\leq n} p_i^{x_i} \times p_{i+n}^{y_i} \times p_{i+2n}^{z_i}$$

To follow the example in the figure, the integral representation of "Arc" $2 \times 3^4 \times 5^3 \times 7 \times 11^2 \times 17 = 291579750$. The integral representation of "ArC" is $2 \times 3^4 \times 5^3 \times 7 \times 11^2 \times 17 \times 23 = 291579750 \times 23$.
We can now define the a cryptographic protocol to detect typos, inspired by the Diffie-Helman key exchange. Intuitively, we take a random element in a group and put it to the X-th power, where X is dependent on the password. Because of the function's structure, it is easy to compare the elements corresponding to two closely related strings. The security lies in the assumed hardness of computing the discrete logarithm.

Data: Username string U, Salt string S, Password string P
Group G, Pseudorandom number generator f
Result: An element $g_0 \in G$ as the "hashed" password sent to the server
begin
Compute the string coordinates $(x_i, y_i, z_i)_{1\leq i\leq |P|}$ of P
$X \longleftarrow \prod_{1\leq i\leq n} p_i^{x_i} \times p_{i+n}^{y_i} \times p_{i+2n}^{z_i}$; $Y \longleftarrow U + S$; $N \longleftarrow f(Y)$
Let g be a pseudorandom element g of G computed from N
Transfer $g_0 \longleftarrow g^X$ to the server

Algorithm 1: Key-setting/sending discrete logarithm algorithm

Remark 1. Alternatively, we could use a more intuitive definition, with $X(s) = \prod_{1\leq i\leq n} p_{3i-2}^{x_i} \times p_{3i-1}^{y_i} \times p_{3i}^{z_i}$. This way, strings that include others as prefixes have integral representations that are multiples of the prefixes' integral representations. As we only consider strings of constant length, this leads to higher values of $X(s)$ with no real advantage. On a standard keyboard, for a string s of length 10, $X(s) < 2^{966}$ with the second definition whereas $X(s) < 2^{768}$ with the first (and $X(s) < 2^{853}$ for length 12). In all cases, they are in expectation quite above 2^{250}, which is enough to prevent discrete logarithm attacks on small exponents [8].

Remark 2. The PRNG in the algorithm does not require a high level of security, and can simply be any algorithm to get an element from a set of pseudorandom bits — such as a PCG algorithm [13].

Data: Group G, constant D, maximum length n, nt $g_0, g_1 \in G$
Result: Keyboard distance between the passwords if it's less than D.
begin
  for i from 1 to D **do**
    for j from 0 to i **do**
      $L_0 \longleftarrow [\,]$
      $L_1 \longleftarrow [\,]$
      foreach $1 \leq a_1 \leq a_2 \leq ... \leq a_j \leq 3n$ **do**
        $X_0 \longleftarrow \prod_{a_k} p_{a_k}$
        $g' \longleftarrow g_0^{X_0}$
        $L_0 \longleftarrow \text{Concatenate}(L_0, g')$
      foreach $1 \leq b_1 \leq b_2 \leq ... \leq b_{i-j} \leq 3n$ **do**
        $X_1 \longleftarrow \prod_{b_k} p_{b_k}$
        $g' \longleftarrow g_1^{X_1}$
        $L_1 \longleftarrow \text{Concatenate}(L_1, g')$
      foreach $g' \in L_0$ **do**
        if $g' \in L_1$ **then**
          return i
  return REJECT

Algorithm 2: Distance-checking discrete logarithm algorithm

Remark 3. The reason why we compute two lists of elements is that computing errors where a_i is greater than expected is easy, as $g^{Xp_i} = (g^X)^{p_i}$. Computing errors the other way around is actually akin to computing a discrete logarithm in the group. As such, the distance computation in this algorithm always goes from the "smaller" to the "bigger" password, which can thankfully be mixed when the keyboard distance is greater than 1.

Security and performance

The security of this algorithm directly comes from the discrete logarithm assump- tion: computing P from gO corresponds exactly to solving the discrete logarithm with the promise that the solution is a 3n-smooth number - for potentially high n in case of added padding. To implement it in practice, one would have to be careful to choose an appropriate group [1]. A cyclic group of order P with P a 2048-bit prime should be enough for now, and a similar algorithm could be adapted for elliptic curves.

With this framework, the login queries are all of the same format - a single element of the group. This could lead to a proof of optimality in terms of space and communication bits required, depending on the group used in practice. It also means that faking an id is not easier than the hardest typo-tolerant frame- work that accepts the same typos. As the size of the group is much greater than the general password space, the discrete logarithm assumption also implies that bruteforcing the password is the best avenue of attack.

Besides the fact that it only allows the correction of substitution errors, the main downside of this algorithm is the time needed to compute the distance.

This is still acceptable on the client side, where the main hurdle is squaring an element at most 1600 times in a large group. Using efficient libraries, this can be done in less than 10ms. However, the server-side computation is where the cost becomes prohibitive. For strings of length 12, checking whether they are at distance 1 takes at most 72 exponentiation operations, or less than 500 squaring operations, doable in a few ms. At distance 2, computation already

takes 35 times more operations, which is on the edge of noticeable from the client- side. Checking whether they are at distance 3 (probably the highest reasonable distance for typos) is, alas, prohibitive, taking at least a few seconds. Using the trinomial revision, the number of expected exponentiations at distance D n is on average

$$\frac{1}{2}\sum_{i=0}^{D}\left(\binom{3n}{i}\binom{3n-i}{D-i}\right) = 2^{D-1} \times \sum_{i=0}^{D}\binom{3n}{D} \geq \frac{1}{2}\left(\frac{6n}{D}\right)^{D}.$$

The algorithm can also be adapted to compute Hamming distances, by checking all possible values for variants on a single letter instead of going by increasing keyboard distance.

Discussion

It was proved in [3] that black boxes that compute arbitrary distances between strings such as the one studied here are vulnerable to attacks with at most poly(n) queries, and with 0(n) queries against the Hamming distance. The for- mula above illustrates why our method is not concerned by those lower bounds: although a linear number of queries would be enough to find the original string from the computed distances, most of those couldn't be computed because of their potentially exponential cost.

A second lower bound shown in [3] concerns the minimal number of commu- nication and storage bits to obtain n bits of entropy, showing that in both cases, n - o(n) bits are necessary. In our case, we store and send a single element of the group, and the security is that of a discrete logarithm attack against the group. We then have a quasi-linear time complexity for current commonly used groups, with real values currently corresponding to an overhead of a factor between 10 and 20.

Some questions remain, such as whether it is possible to obtain a linear storage or communication complexity (or whether stronger lower bounds are provable otherwise). Moreover, the typos corrected here only concern the Ham- ming or keyboard distances, and don't allow complex typos such as exchange of adjacent letters. It would be interesting to check whether the method could be expanded to more complex distance functions. Finally is also one potential risk that requires investigating with this method. The discrete logarithm assumption concerns normal elements of the group. However, the elements considered here are not random elements but X-th powers, with B-smooth X, for $101 < B < 181$. Although B-smooth numbers are essential in discrete logarithm problems [14], there seems to be no attack so far where X being B-smooth is an issue.

References

1. Adrian, D., Bhargavan, K., Durumeric, Z., Gaudry, P., Green, M., Halderman, J.A., Heninger, N., Springall, D., Thome, E., Valenta, L., VanderSloot, B., Wustrow, E., Zanella-Beguelin, S., Zimmermann, P.: Imperfect forward secrecy: How diffie- hellman fails in practice. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. pp. 5-17. CCS '15, ACM, New York, NY, USA (2015). https://doi.org/10.1145/2810103.2813707
2. Batista, G.C., Miers, C.C., Koslovski, G.P., Pillon, M.A., Gonzalez, N.M., Simplicio, M.A.: Using Externals IdPs on OpenStack: A Security Analysis of OpenID Connect,

Facebook Connect, and OpenStack Authentication. In: IEEE 32nd In- ternational Conference on Advanced Information Networking and Applications - AINA. vol. 00, pp. 920-927 (5 2018). https://doi.org/10.1109/AINA.2018.00135
3. Blanchard, N.K.: Usability: low tech, high security. Ph.D. thesis, Institut de Recherche en Informatique Fondamentale (2019)
4. Bonneau, J.: The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In: IEEE Symposium on Security and Privacy. pp. 538-552 (5 2012)
5. Centrify: Centrify password survey: Summary. Tech. rep., Centrify (2014), https://www.centrify.com/resources/5778-centrify-password-survey-summary/
6. Chatterjee, R., Athayle, A., Akhawe, D., Juels, A., Ristenpart, T.: pASSWORD tYPOS and how to correct them securely. In: IEEE Symposium on Security and Privacy. pp. 799-818. IEEE (2016)
7. Chatterjee, R., Woodage, J., Pnueli, Y., Chowdhury, A., Ristenpart, T.: The typtop system: Personalized typo-tolerant password checking. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 329-346. CCS '17, ACM, New York, NY, USA (2017)
8. Guillevic, A., Morain, F.: Discrete Logarithms. In: Mrabet, N.E., Joye, M. (eds.) Guide to pairing-based cryptography, p. 42. CRC Press - Taylor and Francis Group (Dec 2016), https://hal.inria.fr/hal-01420485
9. Lambert, P.: The case of case-insensitive passwords (6 2012), https://web.archive.org/web/20190310221858/https://www.zdnet.com/article/the-case-of-case-insensitive-passwords/
10. Lipa, P.: The security risks of using "forgot my pass-word" to manage passwords (2016), https://web.archive. org/web/20170802185615/https://www.stickypassword.com/blog/the-security-risks-of-using-forgot-my-password-to-manage-passwords/
11. Ma, W., Campbell, J., Tran, D., Kleeman, D.: Password entropy and password quality. In: 4th International Conference on Network and System Security. pp. 583-587 (9 2010). https://doi.org/10.1109/NSS.2010.18
12. Memon, N.: How biometric authentication poses new challenges to our security and privacy [in the spotlight]. IEEE Signal Processing Magazine 34(4), 196-194 (2017)
13. O'Neill, M.E.: PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. ACM Transactions on Mathematical Soft- ware (2014)
14. Pomerance, C.: The role of smooth numbers in number theoretic algorithms. In: International Congress of Mathematicians. Citeseer (1994)
15. Ur, B., Noma, F., Bees, J., Segreti, S.M., Shay, R., Bauer, L., Christin, N., Cranor, L.F.: I added '!'at the end to make it secure": Observing password creation in the lab. In: Proceedings of the 11th symposium on usable privacy and security (2015)

An Overview of Big Data Analytics in Banking: Approaches, Challenges and Issues

Fisnik Doko[1], Igor Miskovski[1], Miroslav Mirchev[1]

[1] Univ. Ss. Cyril and Methodius, Faculty of Computer Science and Engineering, Skopje

Abstract. Banks are harnessing the power of Big Data. They use Big Data and Data Science to drive change towards data and analytics to gain an overall competitive advantage. The Big Data has potential to transform enterprise operations and processes especially in the banking sector, because they have huge amount of transaction data. The goal of this paper is to give an overview of different approaches and challenges that exists in Big Data in banking sector. The work presented here will fulfill the gap of research papers in the last five years, with focus on Big Data in central banks and credit scoring in central banks. For this paper, we have reviewed existing research literature, official reports, surveys and seminars of central banks, all these related directly or indirectly to Big Data in banks.

Keywords: Big Data, Banking, Credit scoring

Introduction

There is no single official definition for Big Data, it is referred as collection and analyses of large volumes of structured and unstructured data, potentially in real time to create value for companies. Concept of Big Data in financial context is different from other industries [l]. Popular approach for the pillars of Big Data are the Gartner 3V model, and the IBM infographicl that provides an overview of the components of "Big Data" by adding a fourth "V". Big Data is mostly defined with the following properties known as "five Vs" [2], [3]:

1. Volume: in today's connected world, huge amounts of data is being created every second, from tweets to videos and photos to bank transactions. Financial services [4] during the years handle high volume of data and always have been with the biggest datasets.
2. Variety: the data can range from structured to unstructured. Big Data technology allows users to analyze not only the structured data we find in the financial sector, but also the more complex unstructured data that is becoming more relevant in order to discover new conclusions and findings.
3. Veracity: refers to the trustworthiness of the data in Big Data, especially if it is obtained from third-party public sources. Veracity burden can rise exponentially with data volumes.
4. Velocity: is a concept which deals with the speed of the incoming data from different sources, it is the data is frequently updated and can be quickly analyzed with possibility to real time analyzes.
5. Value: by predicting new trends based on the analysis of data, banks and financial services can create value for customers by offering them new services.

Big Data is the modern emerging field where technology and Data Science provide new ways of extracting value from the ocean of new information. The ability to effectively manage

insights and extract knowledge is a key competitive advantage [5]. Finance industry uses Data Science for supporting and predicting credit scoring and trading and in operations via fraud detection [6].
Banks need to leverage the benefits of the volume of data they collect and the digital revolution trends to provide better adapted services to their customers in an increasingly competitive digital world [7].
In Section 2 we overview some of the usages and benefits of Big Data in banking industry. In Section 3 we overview the major challenges in the banking industry for adopting Big Data. Then, in Section 4, we describe the usage of Big Data in central banks. After that, in, in Section 5, we provide related work for credit scoring in the last years. In Section 6 we summarize the conclusions.

Advantages of Big Data in banking

Big Data and Data Science in banking refers to the field of applying software and technology in combination with advanced algorithms and methods to gain more insights, to make informed decisions, or to predict risk and revenue. Initial efforts are focused on gaining new insights from existing data and then by newly available sources of information with top priority customer centric objectives [4].
The following shows a list of advantages of Big Data in banking process and Data Science use cases that have the highest impact on the banking sector harnessing benefits of these technologies in 2018-2019 [8] [9]:
• Risk management. Big Data can be targeted to an organization's needs and applied to enhance different risk domains: credit risk, liquidity risk, operational risk and market risk [10].
• Customer insights, experience and analytics. Banks are looking to leverage large amounts of customer data across multiple service delivery channels to discover customer behavior patterns and increase knowledge for consumers [11].
• Enhanced fraud detection. Big Data technologies enable real-time analytics on larger datasets and correlation of data from multiple sources to determine fraud more efficiently [12].
• Algorithmic trading and forecasting stock market. Combining of various data sets from multiple markets and geographies provides enhanced view of market which can generate trade signals, trade execution, profit and risk exposure.
• Predictive analytics. Reveals patterns in the data that foresee the future event that can happen, through understanding social media, news trends, and other data sources, predicting prices and customers lifetime value and the market moves.
Following are Big Data use cases for marketing in the banking industry [13]:
• Sentiment analytics. Monitor social media to increase marketing success and to adjust marketing tactics correctly, identify high influence customers in social media because they are critical to fulfil goals [14].
• Customer 360. Identify the customer profile using more attributes to investigate the habits and to build complete holistic customer profile. Understanding the product engagement of the customer to send the correct marketing message.
• Customer segmentation. Big Data enables faster and sharper classification of customers into various segments that share similar characteristics or behaviors
based on consumer behavior and different attributes [15].
• Next best offer. Allows an organization to increase its opportunities by predicting what the customer wants next using recommendation system to predict customer preferences (like Netflix) based on linking of historical transactions to products.

Challenges and issues

Adopting transformational potential of Big Data is very complex process which implies many changes in IT ecosystems of banks. There are numerous challenges can face during this integration such as infrastructure, information privacy and storage cost. Some generalized challenges that should be overcome to have successful Big Data project are given in the following text.

Privacy and security. Handling Big Data is more scalable and flexible in cloud, but the privacy and security regulations often restrict this movement decision, also there are reputation consequences. Big Data has faced criticism for overstepping privacy boundaries. Ensuring Big Data projects retain their integrity and trust is critical to avoiding public embarrassment, mistrust, and liability. Big Data analytics is limited with numerous regulatory compliance considerations for data protection and privacy issues which impact the analyses because of the individuals being able to decline banks to use their personal data from processing in certain circumstances. Ethical issues should be reviewed in such scenarios, because predictive analyses will identify people with low social conditions which can be treated worse [16].

Storage and processing issues. While banking structured data are continually growing, the unstructured data is growing faster and is becoming more important source for customer insights [4]. This increases the need for having unstructured terabyte databases. The conventional data management techniques are no longer enough to handle the massively large, high-velocity, heterogeneous financial dataset [17]. Uploading large data in cloud or using cloud data for real time analyses with data that reside on-premise, it has processing performance impact that may not be the right choice. The biggest issues when using cloud are privacy and regulatory implications and regulations.

Technical and architectural challenges. Data is rapidly increasing hence it is very important to use appropriate techniques and technology that can handle such vast amount of large and variety complex datasets, which requires new infrastructure components like Hadoop, NoSQL, Map Reduce, where banking industry is lagging behind their peers in other industries [18]. Reading/writing data in enterprise SAN storage and replicating to redundant copies is wasteful for Big Data. These challenges
and the chances for failure are easily handled by the Hadoop Distributed File System
and by the Map Reduce programing paradigm. High-level development environments that abstract from processing architecture, are programming languages such as R and Python which are used to discover new insights from data [19].

Analytical challenges. Big Data itself does not create value until analytical capabilities are used that require skills to use them. Banks need skilled staff data scientists to benefit from Big Data opportunities and overcome governance issues. Because of strict governance rules in banking, banks are missing special working positions for data scientists. Also because of their standards and regulations banks are behind the other technologies in analytical capabilities of text and sentiment analyses,
voice and video. Also, the lack of multiple data types is factor where banks lag behind other industries.

Big Data in central banks

Big Data in central banks is of high volume because data are reported in granular basis on transaction level for example credit data for person or security by security, velocity is argued with the frequent gathering of these data. Drawback is that central banks have more general aggregated data and the data sets are often just numerical.

Central banks already have large datasets of statistics, structured data and information, which are regularly used into their decision-making process. Credit registry is often the largest data sets maintained by some central banks.
Experience of Central banks with Big Data is surveyed from BIS-IFC Big Data Survey [20]. As reported in 2015 there wasn't clear understanding of the "Big Data" definition. The primary focus to central banks has been has been accessing and processing the data [20].
One promising survey conducted by Central Banking in association with BearingPoint during mid-2017. It proves that Big Data is an active area for new projects in central banks and key Big Data projects in central banks is reported development of credit registers, followed by administrative sources and consolidation of internal systems [21]. In the IFC annual report 2018 [22] the feedback from central banks is concerned with the complex privacy implications of dealing with Big Data. The last
survey conducted in 2018 by Central Banking in association with BearingPoint [23]
reports on the approach central banks take towards Big Data. Central banks started looking for external sources to obtain Big Data. Complex privacy implications of dealing with Big Data and other sources as privacy laws evolve with the time and Big Data in central banks is predominantly regarded as useful for research and there is no evidence involvement in policymaking.

Credit scoring

Credit risk is the probability of loss due to a borrower's failure to make payments on any type of debt. The survey of Economist Intelligence Unit shows that most of the banks surveyed in six continents, especially Europe, are primarily concerned with the credit risk [10]. Models can be improved when there are other data sources with the ability to combine and join data from multiple sources. A well-trained system can then perform the credit-scoring tasks and help employees work much faster and more accurately.
Paper [24] shows that the neural network-based credit scoring model is more effective in screening default applications. Paper [25] introduces a two-step loan credibility prediction system which uses Decision Tree Induction algorithm for prediction. Paper [26] presents a new combination approach based on consensus of six classifier that creates a group ranking. Paper [27] shows that the usage of client's payment data is very important factor to enhance credit score prediction. Paper [28] demonstrates that most significant attributes in determining the outcome of a credit application are income, years of employment, credit score and whether or not the applicant had defaulted on a prior credit account. Paper [29] shows that data from social media is big factor in determination of credit score. The newest papers [30] and [31] presents greater insight on the usage of Big Data for credit scoring models, and how this technology is overcoming traditional credit scoring models. This provides proven base that banks should develop new credit models and introduce new data sources that will significantly enhance the model. They propose Big Data models which will include various data for the clients including where they shop, their purchases, their online social network profiles and other factors that aren't directly related with credit default.

Conclusion

Banks will realize value by effectively managing and analyzing the rapidly increasing data and putting the right skills and tools to better understand their operations and customers. By harnessing Data Science tools and technologies, banks can more effectively inform strategic

decision-making, reducing uncertainty and eliminating analysis-paralysis. Despite all the advantages of Big Data, it has a set of limitations when it comes to implementation in banking. Retail and central banks are using Big Data in a variety of ways that they treat as a Big Data, but they have to overcome all the mentioned challenges to be in hop with technology and to gain maximum benefit from Big Data. However, banks are still lagging in implementing Big-data credit- scoring tools will emerge as a way to ensure greater efficiency in underwriting while expanding access to the underbanked and to historically selected groups. Big Data is the reality and is going to stay there for a long time. Banks need to continuously revise policy and regulatory standards to can adopt new technologies. The above analysis shows that there are still research challenges to develop at all levels of the Big Data chain and involve a wide set of different technologies. Despite the technological aspects, there are organizational, cultural, and legal factors that dictate how banks will continue the road of implementing Big Data operations and business development

References

1. B. Fag and P. Zhang, "Big data in finance," in In Big data concepts, theories, and applications, Springer, Cham, 2016, pp. 391-412.
2. I. Kalbandi and J. Anuradha, "A brief introduction on Big Data 5Vs characteristics and Hadoop technology," Procedia Computer Science 48, pp. 319-324, 2018.
3. S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends [point of view].," Proceedings of the IEEE, vol. 103.2, pp. 143-146, 2015.
4. I. I. f. B. Value, "Analytics: The real-world use of big data in financial services," IBM, 2016.
5. S. Dubey and S. Nainwani, "https://www.cognizant.com/InsightsWhitepapers/ Banking-on-Data-Science.pdf," Cognizant, 2019. [Online].
6. G. E. F. D.-M. a. J. D. A.-L. Melo-Acosta, "Fraud detection in big data using supervised and semi-supervised learning techniques.," IEEE Colombian Conference on Communications and Computing (COLCOM), no. IEEE, 2017.
7. S. Totman, "Fintech Business," July 2017. [Online]. Available: https://www.fintechbusiness.com/blogs/759-big-data-and-customer-engagement. [Accessed 2019].
8. Srikanth, "Techiexpert," June 2018. [Online]. Available: https://www.techiexpert.com/top-data-science-use-cases-in-finance/. [Accessed May 2019].
9. E. Moldavskaya, "Intetics," Intetics Inc, November 2018. [Online]. Available: https://intetics.com/blog/top-5-machine-learning-use-cases-for-the-financial-industry. [Accessed May 2019].
10. "Retail banks and big data: Big data as the key to better risk management," The Economist Intelligence Unit, 2014.
11. K. Egetoft, "Digitalist Magazine," January 2019. [Online]. Available: https://www.digitalistmag.com/customer-experience/2019/01/29/data-driven- analytics-practical-use-cases-for-financial-services-06195123. [Accessed May 2019].
12. L. S V S S and S. D. Kavila, "Machine Learning For Credit Card Fraud Detection System.," International Journal of Applied Engineering Research, vol. 13, pp. 16819-16824, 2018.

13. Y.-P. SHEE, D. CROMPTON, R. HILLE and S. PETTER MÆHLE, "EVRY," [Online]. Available: https://www.evry.com/globalassets/insight/bank2020/bank-2020- --big-data--- whitepaper.pdf. [Accessed May 2019].
14. M. Pejić Bach, Ž. Krstić, S. Seljan and L. Turulja, "Text Mining for Big Data Analysis in Financial Sector," Sustainability, vol. 11, 2019.
15. "Medium," ActiveWizards, May 2018. [Online]. Available: https://medium.com/activewizards-machine-learning-company/top-9-data-science- use-cases-in-banking-6bb071f9470c. [Accessed May 2019].
16. A. Chandani and M. Mehta, "BANKING ON BIG DATA: A CASE STUDY," ARPN Journal of Engineering and Applied Sciences, vol. 10, 2015.
17. D. Bholat, "Big Data and central banks," Big Data & Society, 2015.
18. A. Jaiswal and P. Bagale, "A Survey on Big Data in Financial Sector," International Conference on Networking and Network Applications (NaNA), no. IEEE, 2017.
19. R. Elshawi , S. Sakr, D. Talia and P. Trunfio, "Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service.," Big Data Research., 2018.
20. "Central Banks use of and interest in big data," Irving Fisher Committee on Central Bank Statistics, October 2015. [Online]. Available: http://www.bis.org/ifc/publ/ifc- report-bigdata.pdf. [Accessed May 2019].
21. "Big data in central banks: 2017 survey," Central Banking, October 2017. [Online]. Available: prod.centralbanking.bb8.incinsight.net/central-banks/economics/data/3315546/big-data-in-central-banks-2017-survey. [Accessed May 2019].
22. "2018 IFC Annual Report," Irving Fisher Committee on Central Bank Statistics, 2018. [Online]. Available: https://www.bis.org/ifc/publ/ifc_ar2018.pdf. [Accessed May 2019].
23. E. Glass, "Central Banking," Big data in central banks: 2018 survey results, August 2018. [Online]. Available: https://www.centralbanking.com/central-banks/economics/data/3661931/big-data-in-central-banks-2018-survey-results. [Accessed may 2019].
24. A. Byanjankar, M. Heikkilä and J. Mezei, "Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach," IEEE Symposium Series on Computational Intelligence, pp. 719-725, 2015.
25. S. M, "Two step credit risk assessment model for retail bank loan appli-cations using decision tree data mining techniques," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5, no. 3, pp. 705-718, 2016.
26. M. Ala'raj and M. F. Abbod, "Classifiers consensus system approach for credit scoring," Knowledge-Based Systems, vol. 104, pp. 89-105, 2016.
27. E. Tobback and D. Martens, "Retail credit scoring using fine-grained payment data.," Journal of the Royal Statistical Society Series A (Statistics in Society), 2019.
28. A. Smith, D. Khaneja and B. Maher, "Credit Approval Analysis using R," 2017.
29. J. Lohokare, R. Dani and S. Sontakke, "Automated Data Collection for Credit Score Calculation based on Financial Transactions and Social Media," International Conference on Emerging Trends & Innovation in ICT (ICEI), pp. 134-138, 2017.
30. "A review of credit scoring research in the age of Big Data," JOURNAL OF FINANCIAL REGULATION AND COMPLIANCE, vol. 26, no. 3, 2018.
31. M. Hurley and J. Adebayo, "Credit scoring in the era of Big Data," Yale Journal of Law and Technology, vol. 18, pp. 148-201, 2017.

32. J. M. Cavanillas, E. Curry and W. Wahlster, New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe., Springer, 2016.

How Do Kosovo Firms Utilize Business Intelligence? An Exploratory study

Ardian Hyseni

Faculty of Economics, University of Ljubljana,
Kardeljeva pl. 17, 1000 Ljubljana, Slovenia

Abstract. The purpose of this paper is to examine how Kosovo firms utilize business intelligence to analyze their data and better understand the past, present and the future of their firm. Business intelligence offers firms the ability to analyze large amount of data quicker and more effective. Decision making has become easier with the use of BI tools, but making the right decision at right time has become vital for the firms performance. This research will contribute on business industry by providing evidence how Kosovo firms utilize the BI system, what BI tools and what business values and processes BI offers to the firms.

Keywords: Business Intelligence, Data Analysis, BI Tools, Decision Making

Introduction

In today's global economy, the amount of data has increased immensely, however the problem for business analysts is not the amount of data, but selecting the appropriate data. Researches on the outcomes of BI are sparse and so far have addressed only costs and benefits associated with BI and the effects of BI on the performance competitive advantage [1]. Despite the implementation growth of BI globally, new trends of BI are emerging and new opportunities and benefits are created. Unfortunately in the area of utilization of BI systems, there is very little research to consider the factors affecting the utilization of BI in Kosovo firms. The researches have generally argued that BI Investments induce better business performance [2]. To obtain maximum benefit from BI, systems need to be used effectively [3]. Even though many firms have an interest in implementing BI solutions, a lack of researches has limited the understanding of the importance of using or implementing BI solutions. According to Forbes executive management, operations and sales the main drivers for BI adoption, while dashboards, reporting, end-user self-service, advanced visualization, and data warehouse are the top five BI processes for 2018 [4]. This study will show how Kosovo firms utilize BI comparing to rest of the world.

Methodology

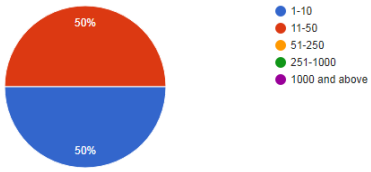
This paper is a qualitative research study; an exploratory study with unstructured questions. Questions prepared for this exploratory study will give an answer on the utilization of BI within Kosovo firms. This study prepared will help to understand how Kosovo firms utilize BI and create business value. And finally, to learn about the future plan on investments on BI system, and in what trends or challenges will firms be mainly focused on.

Data Gathering

Data analysts from four different firms were interviewed in depth, firms participating in this exploratory study were from different industries. After the interviews were taken, interviews were transcript for further analysis. Overall there were four data analysts participating on the research. The number of the respondents for qualitative study is enough to give valuable insight on the utilization of BI within Kosovo firms and their plans for the future investments on BI system.

Results

After interviewing four participants in depth, data was ready to be analyzed and the results gathered are promising and valuable enough to get a clear insight of the utilization of BI within Kosovo firms.



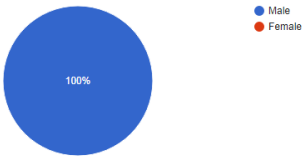
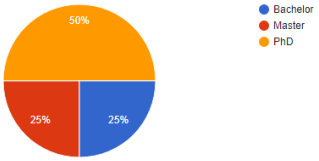


Fig 1.Firm size Fig 2.Gender

As shown on the Figure 1, the firms that participated in the study are with small and medium size comprising of 1-10 employees and 10-50 employees, and the participant were all of male sex as shown in Figure 2.



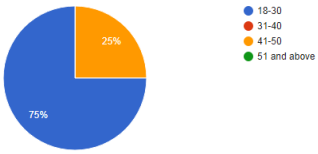


Fig 3.Education Fig 4. Age

In Figure 3, is shown the education level of the participants, two of the respondents were PhD and two others with master and bachelor background, and the age of the three participants was between 18-30 and only one of them was beyond 40s, this gives a clear indication that younger generations are much more used and adopted with new BI trends and technologies.

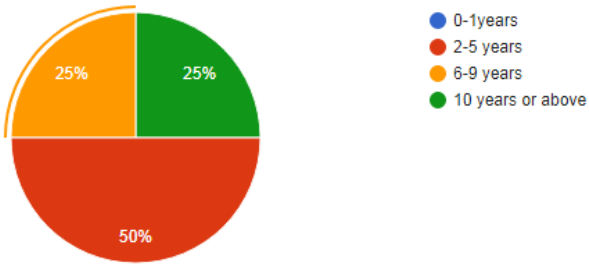


Fig 5. Work Experience with BI

In Figure 5, is shown the participants experience with data analysis and BI solutions, what plays crucial impact on the quality of this study.

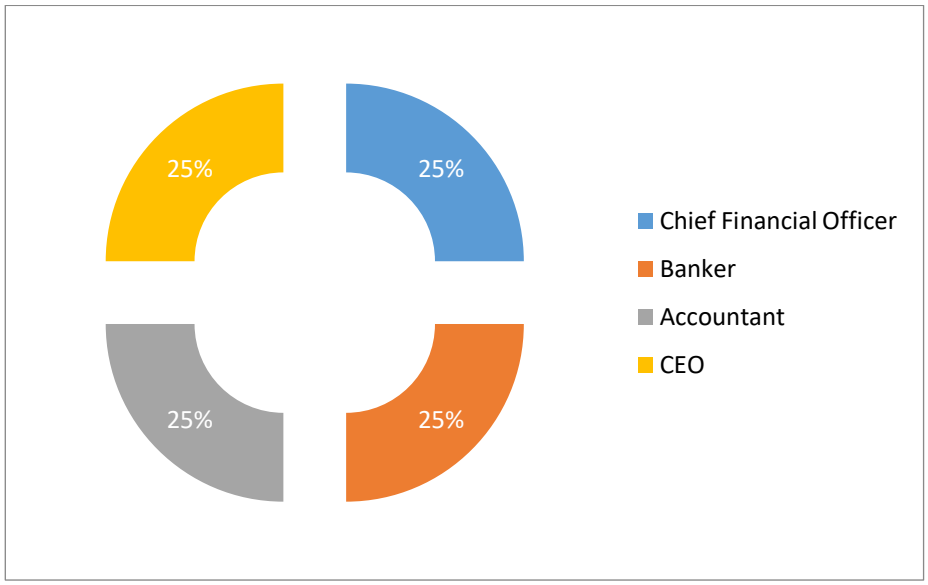


Fig 6. Users

Four respondents job position Chief Financial Officer, Banker, Accountant and CEO as shown on Figure 6, all working in different business sectors.

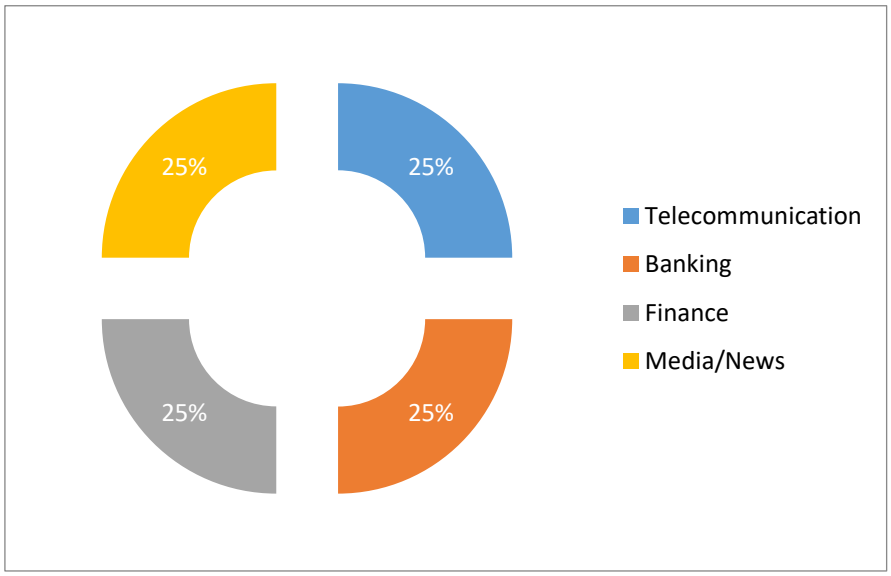


Fig 7. Industry

Respondents in this participating in this study were from Telecommunication, Banking, Media/News and Finance industry as shown on Figure 7.

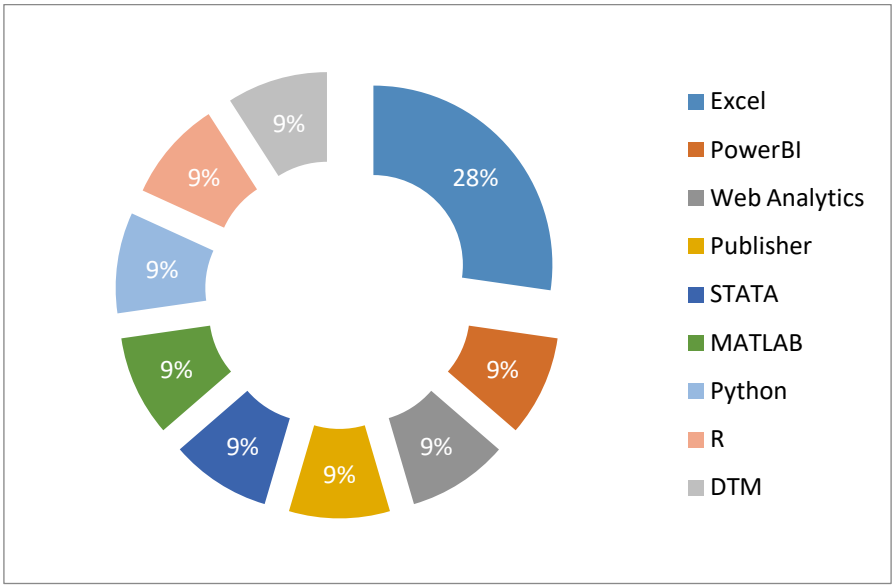


Fig 8. BI tools used by firms

As shown on the figure 8, the most used tool for data analysis remains Excel and proceeded by other tools which are used in different industries for different purposes.

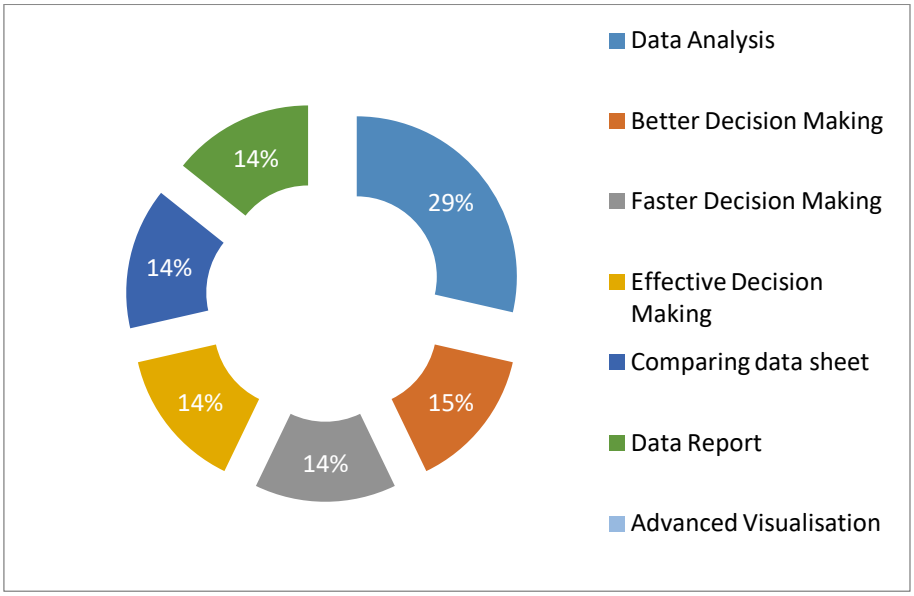


Fig 9. Main purpose of BI usage

As shown on the fig 9, results show that main purpose of BI is Data Analysis, and followed better decision making, faster decision making, effective decision making, comparing data sheet, reporting and advanced visualization. Based on the answers from the participants these are main processes used by Kosovo firms, but in order to get in-depth insight there is needed a larger number of respondents from different industry sector.

Conclusions

This paper provides Kosovo firms with valuable information about utilization of BI, and the results from this research have shown that Kosovo firms are well aware of the benefits from BI. This paper has important contribution on empowering the decision making departments, with information about BI, The decision makers can use these results provided by the paper, for their future business intelligence investment plan. As a conclusion, the use of BI will ensure firms better data analysis, decision making and as well competitive advantage.

References

1. Bernhard Wieder, Maria-Luise Ossimitz: The Impact of Business Intelligence on the Quality of Decision Making – A Mediation Model. Conference on ENTERprise Information Systems / CENTERIS 2015, October 7-9, 2015
2. M. Petrini, M. Pozzebon. Managing sustainability with the support of business intelligence: The Journal of Strategic, 18 , 178-191. (2009).

3. 3. A. Burton-Jones, C. Grange; From Use to Effective Use: A Representation Theory Perspective. Information Systems Research, 24, 632-658. (2013).
4. Forbes. The State of Business Intelligence, https://www.forbes.com/sites/louiscolumbus/2018/06/08/the-state-of-business-intelligence-2018/, 2018

Semantic Web Technologies

Dukagjin Hyseni

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. This study attempts to highlight the great importance of developing Semantic Web as one of the best discovery of better data management and presentation within the WWW. Since the W3C's was discovered, initially providing classic web content as web 1.0 that had link / hyperlink of document`s location, then web 2.0 as web-applications have more advanced technologies to connect data, and finally semantic web as extension of web 3.0 also known as Linked Data.
The results show that in addition to the rapid development of the Semantic Web, the demand to use its features by data publishers and data readers is rapidly expanding due to the time saving to publish multiple times the same data on other web pages.
Moreover, we will present the features of the Semantic Web, its technologies, development history, advantages and weaknesses, the potential benefits, and so on, including standards, frameworks, and programming languages that are being used in its development like: RDF (Resource Description Framework), XML etc.

Key words: Semantic web, Linked Data, W3C, RDF, XML.

Introduction

Since Tim Berners-Lee's original idea for a global system of interlinked hypertext documents from 1989, the World Wide Web has grown into the world's biggest pool of human knowledge. Over the past few years, the Web has changed the way people communicate and exchange information. In 1998, the size of the Web was estimated to exceed 300 million pages with a growth rate of about
20 million per month. The real size of the Web today is difficult to measure, although Web search indices cite a lower band number of unique and meaningful Web pages. The explosion of Web documents and services would not be so critical if users could easily retrieve and combine the information needed. Since Web documents are at best semi-structured in simple natural language text, they are vulnerable to obstacles that prevent efficient content retrieval and aggregation. An increasing problem is the number of languages used on the Web.
The Semantic Web term was popularized by Tim Berners-Lee and later elaborated in 2001. The first part of his vision for the Semantic Web was to turn the Web into a truly collaborative medium—to help people share information and services and make it easier to aggregate data from different sources and different formats. The second part of his vision was to create a Web that would be understandable and processable by machines. The key to machine-processable data is to make the data smarter. Smart data continuum of Semantic Web is consist from: Text and Databases, XML documents for single domains, Taxonomies, Ontologies and automated reasoning.
Semantic Web refers to the W3C`s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules

for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL and SKOS.

Semantic web technologies

Currently, the World Wide Web is primarily composed of documents written in HTML (Hyper Text Markup Language), a language that is useful for publishing information. HTML is a set of "markup" symbols contained in a Web page intended for display on a Web browser. During the first decade of its existence, most of the information on the Web is designed only for humanconsumption. Humans can read Web pages and understand them, but their inherent meaning is not shown in a way that allows their interpretation by computers.
The information on the Web can be defined in a way that it can be used by computers not only for display purposes, but also for interoperability and integration between systems and applications. One way to enable machine-to-machine exchange and automated processing is to provide the information in such a way that computers can understand it. This is precisely the objective of the semantic Web – to make possible the processing of Web information by computers.
"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." (Berners-Lee, Hendler et al. 2001). "The next generation of the Web will combine existing Web technologies with knowledge representation formalisms". (Grau, 2004)
The third common use of the term Semantic Web is to identify a set of technologies, tools and standards which form the basic building blocks of a system that could support the vision of a Web imbued with meaning. The Semantic Web has been developing a layered architecture, which is often represented using a diagram first proposed by Tim Berners-Lee, with many variations since.

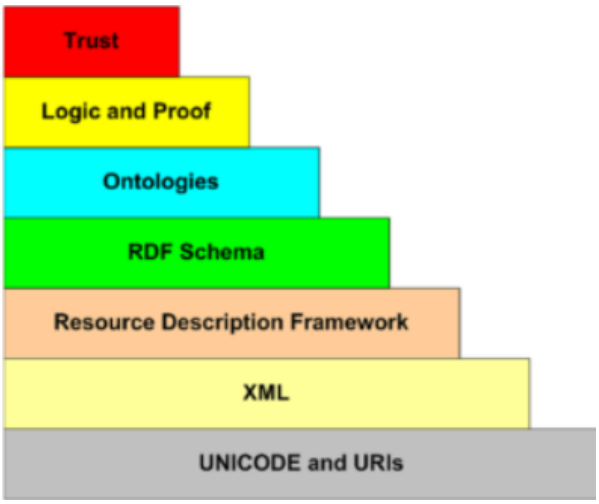


Figure 1: Semantic Web layered architecture

Resource Description Framework (RDF)

The Semantic Web aims to build a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries. It proposes to use RDF as a flexible data model and use ontology to represent data semantics. Currently, relational models

and XML tree models are widely used to represent structured and semi-structured data. But they offer limited means to capture the semantics of data. RDFS and OWL ontologies can effectively capture data semantics and enable semantic query and matching, as well as efficient data integration.

The Resource Description Framework (RDF) – now around for over a decade already as well – is the basic data model for the Semantic Web. It is built upon one of the simplest structures for representing data: a directed labeled graph. An RDF graph is described by a set of triples of the form (Subject Predicate Object), also called statements, which represent the edges of this graph. RDF's flat graph-like representation has the advantage of abstracting away from the data schema, and thus promises to allow for easier integration than customized XML data in different XML dialects: whereas the integration of different XML languages requires the transformation between different tree structures using transformation languages such as XSLT or XQuery. While the normative syntax to exchange RDF, RDF/XML, is an XML dialect itself, there are various other serialization formats for RDF, such as RDFa, a format that allows to embed RDF within (X) HTML, or non-XML representations such as the more readable Turtle syntax; likewise RDF stores (e.g. YARS2) normally use their own, proprietary internal representations of triples, that do not relate to XML at all.4

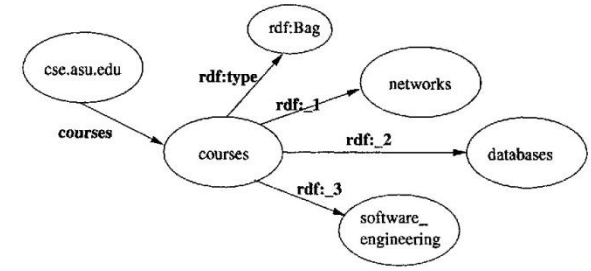


Figure 2. Graph model of RDF/XML implementation

SPARQL

SPARQL a query language for RDF is used to return and manipulate data from databases that are stored in Resource Description Format. It is the recommendation of (DAWG) Data Access Working Group on RDF under World Wide Web Consortium, and is recognized as one of the key technologies of the semantic web.

The SPARQL Protocol and RDF Query Language (SPARQL) is the de-facto standard used to query RDF data. While RDF and the RDF Schema provide a model for representing Semantic Web data and for structuring semantic data using simple hierarchies of classes and properties, respectively, the SPARQL language and protocol provide the means to express queries and retrieve information from across diverse Semantic Web data sources. The SPARQL query language was developed for the RDF layer of the Semantic Web architecture (see Fig.6.1). The query language has been developed without considering the other core languages of the Semantic Web, namely RDFS, OWL and RIF. The SPARQL is a matching graph pattern language. It defines a set of graph patterns, the simplest being the triple pattern. A triple pattern is like a normal RDF triple but with the possibility of a variable instead of an RDF term in the subject, predicate, or object positions. SPARQL introduces the notion of variable binding, that is, a pair (variable, RDF term), where variable is a query variable of interest indicated by ? or $ and the RDF term is the value assigned to the variable after the query has been executed. Similar to the namespace mechanism used for writing RDF/XML, SPARQL allows the definition of prefixes for namespaces. Prefixes are used inside a query to increase its readability. SPARQL introduced a set of constructs and clauses that could be part of a query.

The SELECT clause identifies the variables to appear in the query results and the WHERE clause provides the basic graph pattern to match against the data graph.

```
PREFIX ex: <http://example.org/> .
SELECT ?name
WHERE
{
ex:john, ex:hasName, ?name .
}
```

Figure 3 – Simple SPARQL query

Web Ontology Language (OWL)

OWL stands for Web Ontology Language and was founded on 2004. OWL is such a language for modeling such complex knowledge, expressive representation languages based on formal logic are commonly used. This also allows us to do logical reasoning on the knowledge, and thereby enables the access to knowledge which is only implicitly modeled. Since 2004 OWL is a W3C recommended standard for the modeling of ontologies, and since then has seen a steeply rising increase in popularity in many application do-mains. Central for the design of OWL was to find a reasonable balance be-tween expressivity of the language on the one hand, and efficient reasoning, i.e. scalability, on the other hand. This was in order to deal with the general observation that complex language constructs for representing implicit knowledge usually yield high computational complexities or even undecidability of reasoning, and therefore unfavorable scalability properties. OWL has three sublanguages: OWL Full, OWL DL and OWL Lite. OWL documents are used for modeling OWL ontologies. Two different syntaxes have been standardized in order to express these. One of them is based on RDF and is usually used for data exchange. It is also called OWL RDF syntax since OWL documents in RDF syntax are also valid RDF documents. The basic building blocks of OWL are classes and properties, which we already know from RDF(S), and individuals, which are declared as RDF in-stances of classes. OWL properties are also called roles, and we will use both notions interchangeably.

```
<rdf:RDF
   xmlns:owl ="http://www.w3.org/2002/07/owl#"
   xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:xsd ="http://www.w3.org/2001/XMLSchema#">
   <owl:Ontology rdf:about="">
        <rdfs:comment>An example OWL ontology</rdfs:comment>
        <owl:imports rdf:resource="http://www.mydomain.org/persons"/>
        <rdfs:label>University Ontology</rdfs:label>
   </owl:Ontology>
   <owl:Class rdf:ID="academicStaffMember"></owl:Class>
   <owl:Class rdf:ID="associateProfessor">
        <rdfs:subClassOf rdf:resource="#academicStaffMember"/>
    </owl:Class>
    …
</rdf:RDF>
```

Figure 4 – OWL Example

The OWL language provides mechanisms for creating all the components of an ontology: concepts, instances, properties (or relations) and axioms. Two sorts of properties can be

defined: object properties and data type properties. Object properties relate instances to instances. Data type properties relate instances to data type values, for example text strings or numbers. Concepts can have super and sub concepts, thus providing a mechanism for subsumption reasoning and inheritance of properties.

Development and usability of Web Semantic

Since time that Lee has developed semantic web, there are several version of development of semantic web. Thirty years after the birth of the Web, to have a Web linking applications, things, people, data, etc. Tim Berners-Lee insisted very early on the need to provide on the Web "more machine oriented semantic information, allowing more sophisticated processing" (Berners-Lee et al., 1994). To bootstrap that evolution Tim Berners-Lee then proposed in September 1998 a "Semantic Web Road map" (Berners-Lee, 1998) giving 20 years ago the blue prints of the architecture of the Semantic Web. In 1999 the first versions of RDF and RDFS were published by the W3C and the vision of a Semantic Web was then made visible to a broad audience in 2001 with an article in the Scientific American (Berners-Leet al., 2001). This well-known article presents the Semantic Web as an extension of the existing document-based Web with a Web of structured data and formal semantics better enabling computers and people to work in cooperation. A few years later, Tim Berners-Lee will be again instrumental in pushing what can be seen as a first wave of deployment of the Semantic Web with the Linked Data principles and the Linked Open Data 5-star rules (Berners-Lee, 2006) leading to the publication and growth of linked open datasets weaving a Web of Linked Data. It all started with the first international Semantic Web Working Symposium (SWWS), a workshop held in Stanford, Palo Alto, the 30th of July and 1st of August 2001. The following year the symposium became the International Semantic Web Conference (ISWC) series. Nowadays, Semantic Web not only has its conferences (e.g. ISWC, ESWC, SemTech, SemWeb.Pro) and journals (e.g. Semantic Web Journal, Journal of Web Semantics) but is also an established topic of older conferences and journals from other domains (e.g. The Web Conference WWW, VLDB, EKAW, IJCAI/ECAI, WI, etc.). 7

Semantic Web from year to year is developing rapidly thanks to the development of various platforms for designing and building web semantic technologies. Based on this technological development, usability of web semantic is increasing rapidly. This usability we can verify also by the huge number of Facebook users with 2.45 billion, Google averages over 2 trillion Google searches a day per year 2019 or over 63,000 search queries done per second and some others web- sites that are example and type of web semantics or linked data on web.8 Regarding to the data of usage of web semantics, it seems that usage of web-semantics is too high so we are going to present some graphs of using web semantics.

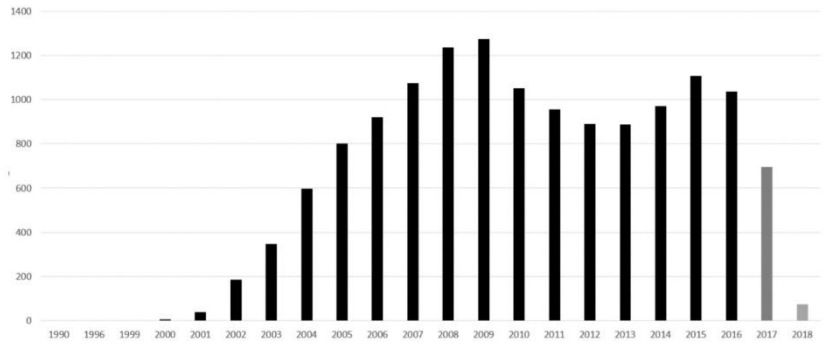


Figure 5 - Yearly distribution of 14157 documents found on the Web of Science on the topic "Semantic Web", note

that at the moment of the survey the years 2017 and 2018 were largely incomplete

Figure 6 - Distribution per country of 14157 documents found on the Web of Science on the topic "Semantic Web"

Conclusion

This paper has primarily described how RDF data store need to be represented as a schema in which the prepositions will be used as properties. The properties will be then interpreted by the SPARQL engines. The property based schema can extend to OWL ontologies which are RDF serialized.

The SPARQL engine will however only return search based on preposition when the underlying data set is got the appropriate schema. Thus all data sets of the future will need to be RDF repositories. It is therefore a subject of research to have software agents which can convert Non- RDF datasets to RDF data sets. The agents can a part of the SPARQL engine itself which can interpret Non-RDF datasets to be RDF data sets only for the purpose of query.

REFERENCES

1. Berners-Lee, T., J. Hendler, et al. (2001). The Semantic Web. Scientific American. May 2001
2. Berners-Lee (1998). Semantic Web Road Map. W3C.Available at: http://www.w3.org/DesignIssues/Semantic.html[last accessed 25/12/19]
3. Vijayan Sugumaran, Jon Atle Gulla, Applied Semantic Web Technologies, CRC Press (New York), 4-5
4. Jorge Cardoso & Amit Sheth, The Semantic Web and its Applications, Springer (Berlin, 2006)
5. Dr Brian Matthews & Leigh Dodds, Semantic Web Technologies, CCRLC Rutherford Appleton Laboratory (England, 2005), 4
6. Axel Polleres, Semantic Web Technologies: From Theory to Practice, Technischen Universitat Wien (Galway, 2001)
7. Sneha Kumari & Dr.Rajiv Pandey & Amit Singh & Himanshu Pathak, Sparql: Semantic Information Retrieval by Embedding Prepositions, IJNSA, (Lucknow, India, 2014), 49
8. Dieter Fensel & Federico Michele Facca & Ioan Toma & Elena Simperl, Springer Heidelberg Dordrecht London New York, (2011), 94 – 95
9. Fabien Gandon, A Survey of the First 20 Years of Research on Semantic Web and Linked Data, 10.3166/ISI.23.3-4.11-56ff.ffhal-01935898, (2018) 12
10. Ben Adida, Mark Birbeck, Shane McCarron, and Steven Pemberton. RDFa in XHTML: Syntax and Processing. W3C recommendation, W3C, October 2008. Available at: http: //www.w3.org/TR/rdfa-syntax/, last accessed 26/12/109
11. Steinmetz, N., Lausen, H., Brunner, M.: Web service search on a large scale. In: Proceedings of the 7th International Joint Conference on Service Oriented Computing (ICSOC 2009)

An Application for Transforming Google Classroom into Learning Management System

Edmond Jajaga

[1]UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. Google Classroom has been widely used by education institutions worldwide. It offers the possibility to manage course works between teachers and students in a paperless way. However, it does lack the support of the features usually offered by a Learning Management System e.g. course administration, documentation, tracking, reporting, and delivery. In this paper we present an application which uses Google Classroom API to enable course administration and analytics. The solution is validated in a couple of higher education institutions.

Keywords: Learning Management System, Google Classroom, Analytics

Introduction

Higher educational institutions are extensively looking for better approaches and methodologies for offering educational services. Learning Management Systems (LMSs) has been widely accepted as a de facto standard for providing educational digital services.
A Learning Management System (LMS) is a software application for the:
administration, documentation, tracking, reporting, and delivery of educational courses, training programs, or learning and development programs [1]. Commercial LMSs are too expensive, while in-house developed systems struggle to cope with technology changes. In our previous works [3, 4] we have described our previous versions of the in-house developed LMSs. The latest one leveraged the advantages of using online storages e.g. Google Drive [4]. As part of their G Suite for Education, Google has recently offered their free-of-charge platform Classroom (GC)
1. The provision of GC eclipsed our efforts of building the same features which were already offered in GC. GC helps students and teachers organize assignments, boost collaboration, and foster better communication. Except the cost, the other greatest advantage of this approach is that Google takes care of technology updates. However, GC is not a true LMS. It supports tracking and delivery of educational courses while partially supports course administration. But, it does not offer the required LMS features of documentation and reporting. In this paper, we describe an application which fulfills the missing features for transforming GC into a fully-featured LMS.

Conceptual design and implementation

Enriching GC with LMS features could be resolved in one of the two approaches: using apps that work with GC or develop a new application using Classroom API2. Regarding the first

approach, hundreds of education applications are available online3 that work with GC. They are mainly developed for primary and secondary school. Since we tried to give a solution for higher education institution, we followed the second approach. As depicted in Fig. 1, our application consists of the Reporting and Administration module, which will use Classroom API as per data retrieval service with GC.

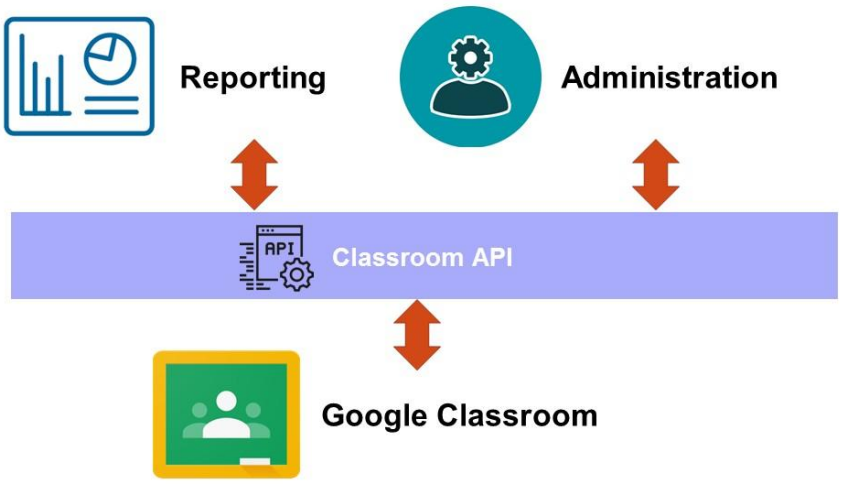


Fig. 1. The GC-based LMS application architecture.

At the time of writing this paper, Classroom API ver. 1offers the management of courses and aliases, course roster and coursework and assignment submissions. It offers a RESTful interface with ready examples provided on the official website for consuming the service within different platforms (.NET, Java, Python, Ruby, PHP etc.). Our solution was implemented as an ASP.NET MVC application, with SQL database backend. A dedicated controller, CourseManagerController.cs, deals with the responsibilities of the Administration module, while DashboardController.csdeals with the Reporting module.

The Administration Module

As described in Table 1, this module serves for a number of responsibilities. Course Administration sub module offers the methods for CRUD over GC courses from within our app. Classroom API object BatchRequest methods is also used for multiple courses creation. Typically, this method is called every start of a term. The administrator prepares a CSV file with a standardized format and uploads it to be read by the application. Then, a preview of the courses to be created is displayed and if the data are read correctly the admin can send the request to GC.
We have used a service account4 for authentication and authorization to access data in Classroom API. The owner of the course is set up to be the course instructor. Other participants in the course can be assigned/resigned by using the Roster Manager sub module.
The responsibility of delivering an announcement to all term courses is handled through the Announcements sub module. Eventually, a list of Material5 is also associated with the announcement. A Material object, is any of: a Google Drive file, YouTube video, a link or a Google Form.

Table 1. The Administration sub modules.

Module	Description
Course Administration	Offers CRUD operations on GC courses. Moreover, offers creation of multiple courses in batch from a CSV file.
Roster Manager	Manage course delegations including instructor and student's assignments to a course.
Announcements	Create and send announcement to one course or all current term courses.
Synchronizations	Synchronize local database with Google apps data including: user data and GC data.
Exports	Export Reporting data to Excel/PDF.

The application works offline i.e. it uses local database to store information about GC courses. The application database is updated every 24 hours through the Synchronizations sub module. Namely, the database is up to date with GC including: course owner's changes, course instructors and students list, course data and course materials published as per processing within the Reporting Module. Within the requests of the Reporting Module is also the feature of exporting data for university's higher authority business decisions. Thus, a module for data export in Excel was also developed. A representing UI from the Administration Module is presented in Fig. 2.

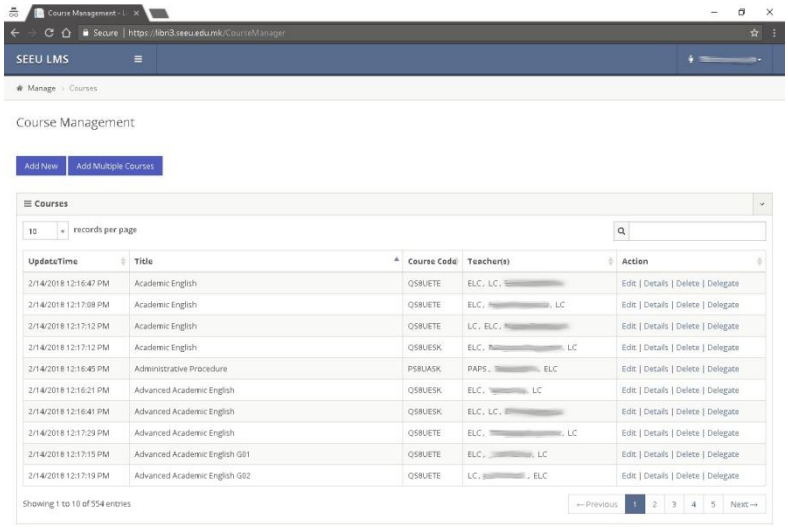


Fig. 2. The Administration Module page UI.

The Reporting Module

One of the most important features of an LMS as per the higher authorities of the university is the Reporting Module. This module should give an insight about the instructor and student's performance within a course. As shown in our previous work [2], our LMS usage is affected by level and resources of instructors' LMS usage.

As depicted in Table 2, this module mainly consists of two sub modules. The Dashboard module gives a personalized overview to the logged in user, based on his role. The system is

used by three kinds of users: administrators, university academic leaders and faculty deans or directors of specific departments [2].

Table 2. The Reporting sub modules.

Module	Description
Dashboard	The homepage of an authorized user. It displays a summary of the latest information on course delegations, published content within GC and course stats.
Analytics	Provides analytical information of GC usage.

As described in Fig. 3, this module gives information to a particular user about the current semester course's load of instructors, near real-time stats and feeds about the course. Thus, one can observe the materials published within a course in GC and some statistical metrics about the course e.g. has or has not a published syllabus, the number of assignments published, number of teaching materials, course level etc. The Output part in Fig. 4 loads published course materials based on different perspectives: term, department, user or course perspective. For example, if one clicks the Feeds button next to a particular course the Output part will be populated with the list of published works within that course.

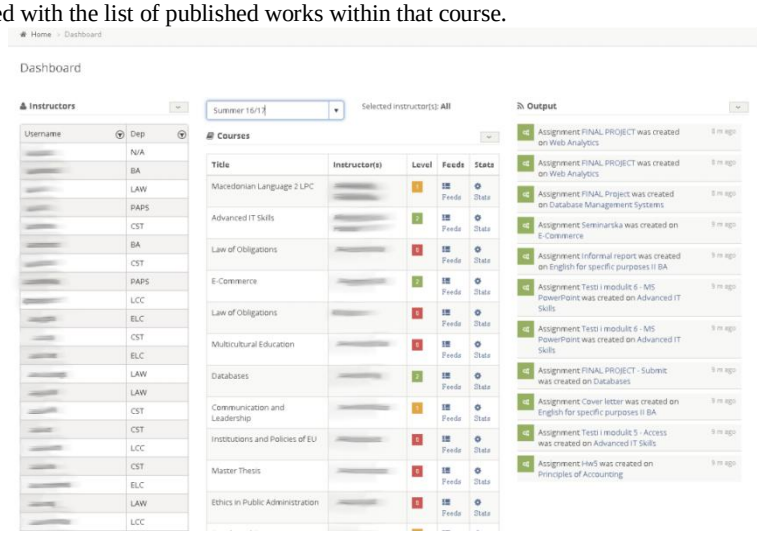


Fig. 3. The Dashboard page UI.

The Analytics Module is responsible for best describing and rendering of GC data usage. As depicted in Fig. 4, at the top right corner of the page an authorized user based on his role can see a chart of showing GC usage. GC usage is calculated from the division of the number of courses with at least one material published against all courses. The stats are calculated during the synchronization process and are available at any time to be served within the UI. The chart appearing at the top main part of the UI shows in which period the courses started to be activated i.e. when a first material got published within a course. In the main part of the page are placed the course level statistics separated within every faculty/department.

Validation

The proposed approach was validated in a couple of universities in the region. Namely, it was actively used about three years in the South East European University and for two years in Mother Teresa University. Our app made GC management very easy and flexible. The analytical data offered to the university management the much needed statistics of having an active insight on what was going on within every course. However, the main issues appeared on deciding if the course holds a syllabus or not and the algorithm about course level specification. Different alternative solutions were considered to find the best descriptive evaluation for these issues, which are out of the scope of this paper.

Fig. 4. The Analytics page UI.

Related Works

In the past we have been working with in-house developed LMSs [3, 4]. Following the provision of cloud-based and free-of-charge solutions we proceeded with GC. The GC-based LMS is cheaper than commercial LMSs, but with less tools. Moreover, it requires less development efforts for maintenance and has the advantages of using other compatible apps with GC.

The approaches of using Google Apps are mainly focused on providing a student information system (SIS) rather than LMS. Namely, Aladdin [5] is focused for primary school administration, while we are focused on LMSs for higher education. Alma and Additio App [6] have a powerful gradebook, including attendance tracking, but does not support course level specification.

As compared to open source LMSs our approach inherits the advantages of using GC: free cloud storage, powerful assignments module and user-friendly mobile version.

Conclusion

In this paper we described our approach on building an LMS based on GC. It fulfills the missing features for transforming GC into a fully-featured LMS. The solution is cost-effective and with short development lifecycle. As per future works we envision the development of the Gradebook Module integrated with GC's assessment's grades.

References

1. Ellis, Ryann K.: Field Guide to Learning Management, ASTD Learning Circuits (2009)
2. Abazi-Bexheti, L., Kadriu, A., Apostolova-Trpkovska, M., Jajaga, E. and Abazi-Hilli, H.: LMS Solution: Evidence of Google Classroom Usage in Higher Education. Business Systems Research 9(1), 31-43, (2018), http://dx.doi.org/10.2478/bsrj-2018-0003
3. Abazi-Bexheti, L., Jajaga, E. and Apostolova, M.: Online testing module in LMS", 35th International Conference on Information Technology Interfaces (ITI), pp. 181-186, (2013)
4. Abazi-Bexheti, L., Jajaga, E., Apostolova, M. and Ismaili, B.: An Experience in Integrating Learning Management System with user's Google Drive, MIPRO 2015 - 38th International Convention CE - Computers in Education, (2015)
5. Aladdin (homepage), https://www.aladdin.ie/, last accessed 07.01.2020
6. Alma and Additio App (homepage), https://www.getalma.com/, last accessed 07.01.2020

Hadamard's coding matrix and some decoding methods

Hizer Leka, Azir Jusufi, Faton Kabashi

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstact. In this paper, we will show a way to form Hadamard's code order $n\ 2^p$ (where p is a positive integer) with the help of Rademacher functions, through which matrix elements are generated whose binary numbers 0,1, while its columns are Hadamard's encodings and are called Hadamard's coding matrix. Two illustrative examples will be taken to illustrate this way of forming the coding matrix. Then, in a graphical manner and by means of Hadamard's form codes, the message sequence encoding as the order coding matrix will be shown. It will also give Hadamard two methods of decoding messages, which are based on the so-called Haming distance. Haming's distance between two vectors u and v was denoted by $d_{u,v}$ and represents the number of places in which they differ. In the end, four conclusions will be given, where a comparison will be made of encoding and decoding messages through Haming's coding matrices and distances.

Keywords: Hadamard's code, encoding, decoding, Rademache function, Hamming distance

Introduction

Definition 1.1. A Hadamard matrix of order n , ${}_T H_n$, is an $n\ n$ square matrix with elements 1 'shat n and -1's such $H_n H_n nI_n$, where I_n is the identity matrix of order n . [3]

Examples of Hadamard matrix order 1, 2 and 4 [3]:

$$H_1 = [1]\ ,\ H_1^{'} = [-1]\ ,\ H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\ ,\ H_2^{'} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}\ ,\ H_2^{''} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$$

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\ ,\ H_4^{'} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}\ ,\ H_4^{''} = \begin{bmatrix} -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}.$$

Hadamard's matrix of order n is generated by the following formula:

$$H_n = H_2 \otimes H_{n/2}$$

where $\otimes$ is the product of Kroneker.

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & . & . & a_{1n}B \\ a_{21}B & a_{22}B & . & . & a_{2n}B \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{m1}B & a_{m2}B & . & . & a_{mn}B \end{pmatrix} = (a_{ij})_{mn}$$

Example:

2

$$H_4 = H_2 \otimes H_2 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix}$$

$$H_8 = H_2 \otimes H_4 = \begin{bmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{bmatrix} \quad \text{and} \quad H_{16} = H_2 \otimes H_8 = \begin{bmatrix} H_8 & H_8 \\ H_8 & -H_8 \end{bmatrix}$$

Let u,v be two vectors in F^n. The Hamming distance between two vectors u and v, denoted by du,v is the number of the places in which they differ. For example, if u and v are defined as u 0,1,0,0 and v 1,0,0,1, then the Hamming distance between u and v is 3, i.e. du,v d0,1,0,0,1,0,0,1 3. [1]

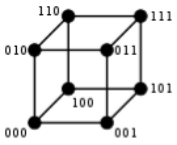


Fig.1.1

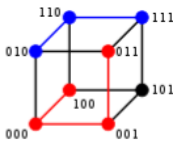


Fig.1.2

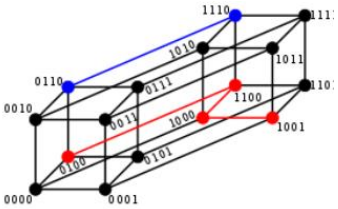


Fig. 1.3

Each non-zero message has a certain Hamming distance, which means that even the distance of the codes is also set. Hadamard's generated code forbids generating a Hadamard code from a Hadamard matrix, the rows of which constitute an orthogonal code set.

Definition 2. For k N, the k^{th} Rademacher function r_k : 0,1 1, 1is defined by $r_k t$ 1 $2_k t$, where t 0,1. [7]

Hadamard code and Encoding Matrices

Hadamard's code is an example of a linear code with binary digits that determines the length of code length messages. Hadamard's codes are orthogonal and belong to a linear class of codes. They are used as error correction codes which are very useful in delivering information over long distances or through channels where errors can occur in messages.

Definition 2.1 [6] (*Hadamard code*) Let $r \; N$. The generation matrix of Hadamard code is a 2^r *r* matrix where the rows are all possible binary strings in F^r .

Example.[6] For $r \; 2$, we have

$$G = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

which maps the messages to

$$Gx = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

In general, the Hadamard code based on the Hadamard matrix H_n , where $n = 2^k$, has a generator matrix that is $(k+1)=2^k$. The rate is $(k+1)/2^k$ - terrible, especially as k increases. The code can correct 2^{k-2} - 1 errors in a 2^k -bit encoded block, and in addition detect one more error- excellent. [4]

If $u = (u_1, u_2, , u_n)$ and $v=(v_1, v_2, , v_n)$ are vectors over Z_2 , define:

$$u \oplus v = (u_1 \oplus v_1, u_2 \oplus v_2, \Lambda, u_n \oplus v_n)$$
$$uv = (u_1 v_1, u_2 v_2, \Lambda, u_n v_n) \qquad [4]$$

In the following, we will use Radamecher functions to generate Hadamard's coding matrices of the order $n \; 2^p$ (where, p is a positive integer) as follows:

$$G_{p \times n} = \begin{bmatrix} R_p \\ R_{p-1} \\ . \\ . \\ . \\ R_2 \\ R_1 \end{bmatrix} = \begin{bmatrix} r_{p,1} & r_{p,2} & \cdots & r_{p,n} \\ r_{p-1,1} & r_{p-1,2} & \cdots & r_{p-1,n} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ r_{1,1} & r_{1,2} & \cdots & r_{1,n} \end{bmatrix}$$

where G_{pn} is $p \; x \; n$ the matrix generated, whose rows are p successive functions of Rademacher (sequences), which form a basis for Hadamard's matrices where $r_{ij} \; F_2 \; 0,1, i, \; j \; : i$ 1,2,...,*m* and *j* 1,2,..., *n* . Rademacher's functions were determined by German mathematician

Rademacher in 1922, [Rademacher, "Einige Sätze von allgenein orthogonal function," p. 112-138, (1922)). [1]

Rademacher functions with $n\,2^4$ =16 pulses are shown in figure(2.1), along with the sequence representation of the functions in the logical elements {0,1}, which are called Rademacher sequences.

Example 2.1[1]. The generator matrix for Hadamard matrix (code) of order two

i.e $n\,2$, $p\,1$ is :

$$G_{1\times 2} = [R_1] = [r_{1,2} \quad r_{1,2}] = [0 \quad 1]$$

Example 2.2 [1]and[5]. The generator matrix for Hadamard matrix (code) of order four

i.e $n\,2^2\,4$, (p =2) is :

$$G_{2\times 4} = \begin{bmatrix} R_2 \\ R_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

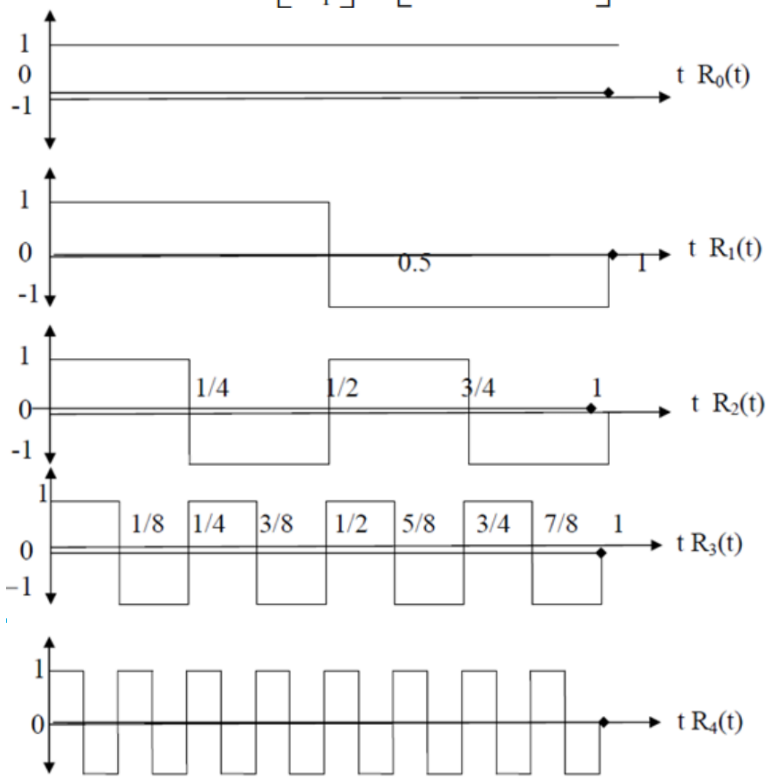


Fig. 2.1

$$R_0 = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$
$$R_1 = (0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1)$$
$$R_2 = (0,0,0,0,1,1,1,1,0,0,0,0,1,1,1,1)$$
$$R_3 = (0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1)$$
$$R_4 = (0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1)$$

Fig.(2.1): The graphs of R_0 , R_1 , R_2 , R_3 , R_4 Rademacher functions (Rademacher sequences). The encoding of p -tuple message sequences in to Hadamard sequences (Hadamard codewords) of length $n = 2^p$ is shown as follows.

For $m\ n\ 1$, we write the binary of m as :

$m_{b\,i}, _{i1}, , _{1}, _{0}$, then $H_m\ m_b * G_{pn}$, ku $_i F_2$, i, i 0,1,2, , p . m_b

is p - tuple message sequences and $H_{\ m}$ is $m\ th$ Hadamard sequence (codeword). Hadamad matrices (codes) of order n 2,4,8,16are shown in tables 1,2,3 and 4 respectively.[1]

Table 1 : Hadamard matrix(code) of order n = 2,($H_{(2,1)}$ code)

Integer (m)	1-tuple message sequence ($(m)_b$)	Hadamard codeword $H_m = (m)_b\ G_{1\times 2}$
0	(0)	H_0 = (0,0)
1	(1)	H_1 = (0,1)

Table 2 : Hadamard matrix(code) of order n = 4, ($H_{(4,2)}$ code)

Integer (m)	2-tuple message sequence ($(m)_b$)	Hadamard codeword $H_m = (m)_b\ G_{2\times 4}$
0	(0,0)	H_0 = (0,0,0,0)
1	(0,1)	H_1 = (0,0,1,1)
2	(1,0)	H_2 = (0,1,0,1)
3	(1,1)	H_3 = (0,1,1,0)

Table 3 : Hadamard matrix(code) of order n = 8, ($H_{(8,3)}$ code)

Integer (m)	3-tuple message sequence ($(m)_b$)	Hadamard codeword $H_m = (m)_b\ G_{3\times 8}$
0	(0,0,0)	H_0 = (0,0,0,0,0,0,0,0)
1	(0,0,1)	H_1 = (0,0,0,0,1,1,1,1)
2	(0,1,0)	H_2 = (0,0,1,1,0,0,1,1)
3	(0,1,1)	H_3 = (0,0,1,1,1,1,0,0)
4	(1,0,0)	H_4 = (0,1,0,1,0,1,0,1)
5	(1,0,1)	H_5= (0,1,0,1,1,0,1,0)
6	(1,1,0)	H_6= (0,1,1,0,0,1,1,0)
7	(1,1,1)	H_7= (0,1,1,0,1,0,0,1)

Table 4 : Hadamard matrix(code) of order n = 16 , ($H_{(16,4)}$ code)

Integer (m)	4-tuple message sequence $((m)_b)$	Hadamard codeword $H_m = (m)_b G_{4\times16}$
0	(0,0,0,0)	H_0 = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
1	(0,0,0,1)	H_1 = (0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1)
2	(0,0,1,0)	H_2 = (0,0,0,0,1,1,1,1,0,0,0,0,1,1,1,1)
3	(0,0,1,1)	H_3 = (0,0,0,0,1,1,1,1,1,1,1,1,0,0,0,0)
4	(0,1,0,0)	H_4 = (0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1)
5	(0,1,0,1)	H_5= (0,0,1,1,0,0,1,1,1,1,0,0,1,1,0,0)
6	(0,1,1,0)	H_6= (0,0,1,1,1,1,0,0,0,0,1,1,1,1,0,0)
7	(0,1,1,1)	H_7= (0,0,1,1,1,1,0,0,1,1,0,0,0,0,1,1)
8	(1,0,0,0)	H_8= (0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1)
9	(1,0,0,1)	H_9= (0,1,0,1,0,1,0,1,1,0,1,0,1,0,1,0)
10	(1,0,1,0)	H_{10}= (0,1,0,1,1,0,1,0,0,1,0,1,1,0,1,0)
11	(1,0,1,1)	H_{11}= (0,1,0,1,1,0,1,0,1,0,1,0,0,1,0,1)
12	(1,1,0,0)	H_{12}= (0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0)
13	(1,1,0,1)	H_{13}= (0,1,1,0,0,1,1,0,1,0,0,1,0,1,0,1)
14	(1,1,1,0)	H_{14}= (0,1,1,0,1,0,0,1,0,1,1,0,1,0,0,1)
15	(1,1,1,1)	H_{15}= (0,1,1,0,1,0,0,1,1,0,0,1,0,1,1,0)

Hadamard Decoding methods :

In this section,we will introduce two methods for decoding Hadamard codwords: Let w be received word.

Method (1) :

Find the closest codeword $u \in H_{(n,p)}$ such that:

$$d(w,u) \le d(w,v), \forall v \in H_{(n,p)}.$$

Method (2) :

This method composed of two steps:

Step 1 :

Compute

$$S = H_{(n,p)} * w^T.$$

Step 2:

If $S = \theta$ (where , θ is a zero vector),then the received word is a codeword in Hadamard code $H_{n,p}$,but,if S ,the receivedword w is received in error.Ih order to find the location of error in w , we compaired S with the each column of Hadamard code which gives the location of error in w.

For example ,if the original message is (1,1,0), by using Hadamard code of order n 8 , then the encoded message is H_6 0,1,1,0,0,1,1,0. Let the encoded message H after the error be w 0,1,0,0,0,1,1,0. We decode it as follows :

By 1st method :

$$d(w,H_0)=3, d(w,H_3)=5, d(w,H_6)=1$$
$$d(w,H_1)=3, d(w,H_4)=3, d(w,H_7)=5$$
$$d(w,H_2)=5, d(w,H_5)=3$$

We see that $d(w,H_6) \le d(w,H_i)$, $\forall i,\ i = 0,1,...,7$, and thus H_6 to have been transmitted. is the codeword that is most likely to have been transmitted.

By 2nd method :

$$S = H_{(8,3)} * w^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

S is similar to third column of Hadamard code of order n = 8 , therefore we can see that the error was in the third place of w , and we write w =(0,1,1,0,0,1,1,0). Since, $\omega \in$ H(8,3) code, therefore we can see that the original message was (1,1,0).

Conclusions

1. Generating or rpresenting of Hadamard matrices (codes) from using Rademacher functions (sequences) is easy to find.
2. Using the Kronecker product method, coding Hadamard matrices is very quick and easy.
3. A new algorithm is given in section four which as we think is very efficient than Hamming method. It can be straightforward to implement.
4. Both the Hamming codes and the Hadamard codes are actually special cases of a more general class of codes: Reed-Muller codes.

References

1. Hameed k. Dawiod Khalid H. Hameed “On representation of Hadamard Codes” AL- Fatih Journal . No . 32 .2008,
2. Falkowski,B.J. and Sasao T., “Unified algorithm to generate Walsh functions in four ifferent orderings and its programmable hardware implementations”, IEE proc. Vis.Image process., .152,No.6,December 2005.
3. Hong-Yeop Song “Examples and Constructions of Hadamart matrices” Yonsei University, Seoul 120-749, Korea , June 2002.
4. Hadamard code, Massoud Malek, California State University, East Bay
5. Rademacher,H.,”Einige Sätze von allgemeinen orthogonal funktionen”,Math.Annal.,112-138, 1922.

6. Yuan Zhou and Kaiyuan Zhu, “Hamming and Hadamard Codes” CSCI-B609: A Theorist’s Toolkit, Fall 2016 Oct 6.
7. Jordan Bell,” Rademacher functions”, Department of Mathematics, University of Toronto, July 16, 2014.
8. Yaroslavsky,L.P.”Digital holography and digital image processing: principles,methods, algorithms”, KluwerAcademic,Boston,2003.
9. Walsh,J.L.“Aclosed set of normal orthogonal functions” ,Amer.J.Math.45,pp.5-24,1923.

AI leverage in easing the 5G complexity and enhancing 5G intelligent connectivity

Xhafer Krasniqi

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. As 5G era is approaching fast and pre-commercial 5G tests and trials are happening everywhere around the world, one of the key challenges for carriers and 5G providers is to maintain and operate the network complexity required to meet diverse services and personalized user experience requirements. This maintenance and operation have to be smarter and more agile in 5G than it was in previous generations. AI and ML can be leveraged in this case to ease 5G complexity and at the same time enhance the intelligent connectivity between diverse devices and tiny end points, e.g. IoT sensors.
Machine learning and AI algorithms can be used to digest and analyse cross-domain data that would be required in 5G in a much more efficient way enabling quick decision and as such easing the network complexity and reducing the maintenance cost. The cross-domain data includes geographic information, engineering parameters and other data to be used by AI and ML to better forecast the peak traffic, optimize the network for capacity expansion and enable more intelligent coverage through dynamic interference measurements.
This paper provides an overview of 5G complexity due to its heterogenous nature and the key role of AI and ML to ease this complexity and enhance the intelligent connectivity between diverse devices with different requirements. The focus of this paper will be on the key aspects of AI and ML application in 5G and the key benefits from this application. Finally, this paper will analyse the overall performance of 5G in terms of coverage and latency compared with traditionally operated networks.

Keywords: 5G, AI, ML, IoT, Sensors, Network Slicing, Virtualisation, MTC.

Introduction

There is no doubt that 5G is expected to be a complex network since it will be the main mechanism for transporting the avalanche of data generated by billions of devices and IoT sensors connected through 5G. Having to handle all of this data and huge number of connected devices will add further to the complexity of 5G and the chaos brought by 5G and as such would require a new technology that would help handling this data and would make sense of this complexity and chaos. This new technology needed and required by 5G is AI. On the other side, AI will also need 5G and its massive data in order for the AI to be successful and enhance the AI algorithm. The more data sets that are used as inputs to machines learning the more accurate the outcome patterns will be.

5G Network

5G is the latest cellular network generation that is currently under development and still at an infancy stage. It is going through a number of pre-commercial tests and trials and it is at an early adoption stage with a very limited coverage. It has been deployed by a number of operators, but this is only the early version of it, the LTE-based
5G or also known as NSA (Non-StandAlone) implementation.
5G network is very heterogeneous in nature and it is a platform that interacts with different networks and different radio technologies that serve various devices with different resource requirements, as shown in the figure below.

Figure 1, 5G heterogeneity

5G Complexity

Main things that would make 5G complex and would differentiate it from other previous networks in terms of complexity are the heterogeneity of 5G, its network topology, design and propagation models together with user's mobility and usage patterns. Some or all of these parameters were also part of the previous networks and generations, but not in the same scale as 5G.
5G aims to address three major requirements as shown in the diagram below:

- Massive broadband (eMBB)
  - In the range of 10 Gbps
- Very high number of connections (mMTC)
  - In the range of a million per km2
- Extremely reliable and low latency (uRLLC)
  - In the range of 1 ms

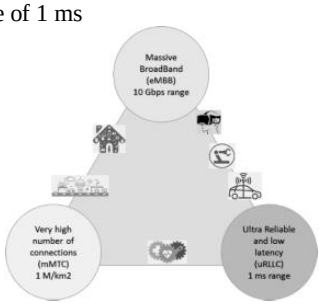


Figure 2, 5G requirements

AI

AI is the science of training systems to emulate human tasks through learning and automation. AI is not a new technology since it has been around since the 1950s, but it is only recently finding its place in mainstream applications as a result of the rapid increase of IoT data volume, high-speed connectivity and high-performance computing [8].
It is the AI that can augment greatly the value of IoT by making use of all the data from huge number of devices to drive to improve the ML algorithms and make machines learn.

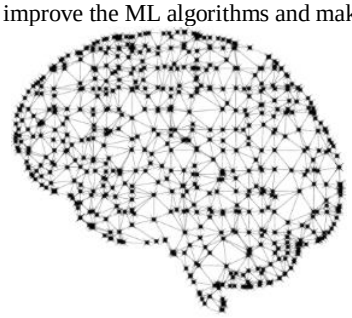


Figure 3, AI connectivity

AI in the context of 5G

In the context of 5G and the enhancement of mobile networks, AI and ML are interchangeably used, but they differ from each other. AI is a broad concept that does certain tasks in a smart way and closer to humans. It
relies on ML to collect the data and analyse the pattern from which the software system learns and improves and this makes machines smarter. On the other side, ML is a subset of AI and is seen as an application of AI to allow machines access to data and learn for themselves. ML is also known as a current-state-of-the-art of AI.

When it comes to the use of AI/ML in 5G and whether this can help 5G to handle its complexity, there are two approaches:

- Basic approach where AI/ML is used to perform some basic tasks based on some pre-set algorithms
- More advanced approach where AI/ML is used to be more context aware and learn from the surrounding situations and acts accordingly
    - This approach has emerged with the popularity of Internet and the huge amount of the generated information
    - Instead of teaching the computer to do everything, it is better to code them to think like humans and give them access to the huge information enabled by Internet and 5G

Internet of Things (IoT)

Is a system of devices connected together. By devices in the context of IoT is meant any device from tiny sensors to wearable and smartphones.

These devices can talk to each other and gather a lot of information that when analysed can create a lot of useful information that can be used by different entities to perform a specific task or learn from those information.
In the context of this paper, the information gathered from these IoT devices can be used by AI in order for the AI algorithm to learn.

Figure 4, IoT devices

What functions of 5G can AI help with?

Number of functionalities and services provided by 5G is so high that managing and configuring them manually may become a bottleneck and would require some form of automation. This form of automation would be difficult without AI. An example of such a functionality would be the network slicing, see figure x, as one of the main 4G and 5G functions, that currently are manually configured. Once the 5G deployment starts rolling out, the number of network slices will be much higher and manual configuration would affect operator's abilities to provide this service. Use of AI in this case would simplify the configuration drastically by filtering and routing backend traffic based on device needs and as such enhance operator's abilities to provide this service.
Another case where AI could be leveraged is to enhance device abilities through a better understanding of their surroundings, i.e. to improve contextual awareness. The improvement of contextual awareness can only be achieved with 5G as a technology with a very low latency.
Other case where AI and ML can help the 5G is with the beam selection and steering and by identifying and computing the strongest beam or strongest set of beams, i.e. beam reference signal (BRSRP). Based on these computed beams, the serving cell site decides if the UE needs to switch and when to switch to a neighbouring cell and what specific beam to connect to.

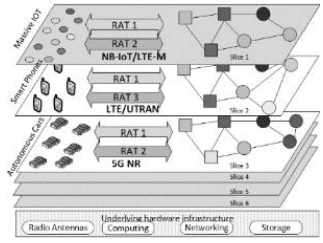


Figure 5, Network slices

The key advantages that operators gain from AI and ML for 5G enables are as following:

- High level of automation leading to minimum or no involvement of human interaction and as such more efficient service to offer
- Traffic aggregation and traffic steering from different access networks through different applications
- Capability to provision resources to different use cases, e.g. network slice, with different QoS
- requirements in a very dynamic way
- Collection of real time information and construction of a complete user profile including user subscription, QoS requirement, network performance and other events and logs

Intelligent Connectivity

Is a new concept built on a fusion of three major technologies, 5G, IoT, which is meant to serve as a means to accelerate the development of disruptive digital services.
This new concept facilitates connection of devices through a fast and low latency mobile network, that is 5G, collects digital information through the machines and sensors, which is the function of IoT then analyses and contextualizes by AI/ML and finally generates meaningful outcome useful for the users. This would enable new transformational capabilities in most of the industry sectors, e.g. transport, manufacturing, healthcare, public safety, security etc.

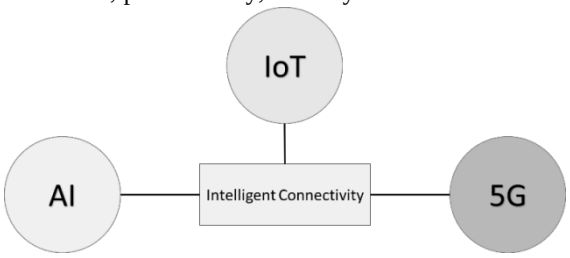


Figure 6, Intelligent connectivity

Areas of applications of AI in 5G

The number of areas in mobile networks and in 5G specifically where AI could apply is huge, but some of the main areas where AI could be used are [5]:
Network planning-where operators will use AI to improve the network capacity planning which will lead to cost reduction and better performance of the network.
AI and machine learning methods can be applied to predict and forecast the traffic by detecting traffic patterns and as such learning online and helping with the automation of decisions. This avoids the need for over- provision as it is the case with the traditional network capacity planning.
Management of network performance- allows a network controller to learn from experience while it enhances the network.

Management of customer experience- AI will help to improve and manage the customer experience by using the IoT data that reveal important consumer insights in the context of real-time situations. This helps the consumers and provides them with an experience tailor-made to their life.
Management of product life cycle- Artificial intelligence helps to manage and improve the product lifecycle by leveraging the data that describe the current and historical product insights, root causes and correlations, future outcomes and recommended improvements. This helps to transform the product lifecycle from a data management tool to an intelligent decision-making system.
Management of network itself – Artificial intelligence and machine learning can help the software-defined networks to learn how to manage themselves by using operational data. This requires some training of the network to manage itself for some initial and simple operations. Initial application of AI to network operations includes some security functions, such as security attack mitigation, automated path selection, and some basic operations running on 'auto-pilot'.
Management of revenue – Adoption of AI in 5G is already happening and will reduce the capital expenditure and as such improve the revenue stream.

Conclusion

This paper highlighted the key features and benefits of using AI/ML in 5G and how can 5G also impact the AI development by facilitating the access to relevant data. The paper also highlighted that areas with the greatest benefit from AI in mobile communications are the management of cost and network efficiencies and the improvements to customer experience.
However, to make a full use of AI/ML in 5G, the development of standardised interfaces to enable access to relevant data is needed, and this is anticipated to be the main challenge. Another challenge is to explore and examine the use of AI for customer experience optimization and to automate network operations including network planning and network management.
Some of the key points of this paper are that AI is being deployed in networks, will be critical for customer service and will help the network operates to regain the investment.
As for the future work in this field, all the interested stakeholders, such as industry, academia and research, should increase the efforts to transfer the intelligence on the end devices and end things that would make them smart things as opposed to only access the relevant data and make the network smart.

References

1. 1. Eugonio Pasqua: "How 5G, AI and IoT enable "Intelligent Connectivity", https://iot-analytics.com/how-5g-ai-and-iot- enable-intelligent-connectivity/, IoT Analytics, February 2019
2. 2. ElectronicDesign: "Everything Gets Smarter When 5G and AI Combine", https://www.electronicdesign.com/industrial- automation/brave-new-world-everything-gets-smarter-when-5g-and-ai-combine
3. 3. Omkar Dharmadhikari: "Leveraging Machine Learning and Artificial Intelligence for 5G", https://www.cablelabs.com/comp-over-docsis-femtocells-in-the-age-of-vran, CableLabs, June 2019

4. 4. “Employing AI techniques to enhance returns on 5G network investments”, https://www.ericsson.com/49b63f/assets/local/networks/offerings/machine-learning-and-ai-aw-screen.pdf, Ericsson, 2019
5. 5. “How will AI enable the switch to 5G”, https://www.ericsson.com/en/networks/offerings/network-services-and- automation/ai-report, Ericsson, 2019
6. 6. “When AI and 5G Combine, Watch For a New Generation of Applications”, https://www.aitrends.com/ai-and-5g/when- ai-and-5g-combine-watch-for-a-new-generation-of-applications/, AItrends.com, August 2019
7. 7. “The Artificial Intelligence of Things”, https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers- ebooks/sas-whitepapers/en/artificial-intelligence-of-things-110060.pdf, SAS, 2018
8. 8. Vince Jeffs: “Artificial Intelligence and Improving the Customer Experience”, https://www.pega.com/system/files/resources/2019-09/ai-and-improving-cx-en.pdf, PEGA, 2017

Vulnerability of passwords consisting of Numerical Repetitive Sequences in the WPA2 protocol

Genc Gregor Kelmendi, Edmond Hajrizi

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. Protocols that govern wireless security WPA2/WPA have been proven much more secure in comparison to their predecessor WEP. However, the human factor jeopardizes the solidity of cryptography by implementing passwords consisting of programmatically predictable numerical structures such as
1234512345, 11114444, 999888777 and so on. The methods presented in this paper are effective in decrypting such passwords within seconds using ordinary
processor power. The prevalence of this vulnerable practice in the Prishtina (Kosovo) region is estimated to be 15.7% in 89 tested wireless routers. Under this study such types of passwords are termed and referred to as Numerical Repetitive Sequences or NRS. The paper defines NRS mathematically, identifies
NRS types, composes formulas for calculating variations, presents algorithms to generate NRS, and proposes tools to implement attacks. The methods documented in this study should only be used for educational purposes.

Keywords: wireless password, network security, mathematical sequences, arithmetic progression, handshake attack

Introduction

Among the additional security that WPA/WPA2 protocols provided compared to the WEP predecessor, was the minimum password length requirement of 8 characters. This was intended to condition users to implement passwords more resistant to brute-force attacks. But by not validating the type of inserted characters, a security hole is left where the password can consist of repeatable characters and sequences, and therefore rendering them predictable. The length criterion alone cannot effectively avoid unsafe practices. A password consisting of 8 characters of the digit 51 is as unsafe as a password consisting of only one character of number 5. Similarly, a password of type "11223344" is as unsafe as a 4 character password of "1234 ". Character replication or repetition can be easily simulated programmatically, and therefore does not increase the entropy and strength of the password.
Passwords consisting of numbers only are particularly vulnerable to brute-force attacks. This is because the decimal base of numbers consist of only 10 elements (0-9), and the variations are significantly smaller when compared to alphabets which have a richer amount of elements. When these numeric passwords are also limited to certain templates, they even further endanger the security of the system. For example, passwords 1234512345, 11114444, or 999888777 are not only numerical by nature, but also suffer from replications and fixed progressions (incremental or decremental). Unfortunately, such passwords are often applied in the wireless domain. After taking a closer look, we realize that these types of passwords are essentially arithmetic sequences that are replicated for a given number of times, therefore for an easier

reference they will be referred to as Numerical Repetitive Sequences (NRS). In this paper NRSs will be mathematically defined, algorithmically generated, and implemented as a dictionary attack against the captured WPA handshake.

NRS Definition

Numerical Repetitive Sequences (NRS) is the numerical string of the length *l* that replicates the finite arithmetic sequence of the form:

$$\{a_i\}_{i=1}^{n} \; with \; a_{i-1} + d \qquad (1)$$

Where the first term $a1$ and the upper limit *l* meet the conditions

$$1 \leq n \leq 10 \; ; Mod(l, n) = 0 \qquad (2)$$

$$a_1, l, n, d \in \mathbb{N}^0 \mid 0 \leq a_1 \leq 9 \qquad (3)$$

While distance *d* is defined as

$$0 \leq a_1 + [d(n-1)] \leq 9 \qquad (4)$$

Such as $d = 0$ if and only if $n = 1$, generating constant arithmetic sequence

$$d = 0 \Leftrightarrow n = 1 \qquad (5)$$

If $n < l$ arithmetic sequence or each term is replicated *l/n* times successively, thus producing the numerical repetitive sequence with length *l*.

Types of NRS

Definition 1. If each term is replicated separately *l/n* times, the sequence is labeled as internal sequence, and symbolically is denoted as S_I.

Definition 2. If entire sequence is replicated *l/n* , the sequence is labeled as external sequence, and symbolically is denoted as S_E.

Definition 3. The S sequence is a singular sequence if $n = l$ and is symbolically denoted as S_S. The singular sequence is neither internal nor external. Since the upper bound is equal to the length of the sequence $n = l$, neither sequence nor terms can be replicated.

Definition 4. The S sequence is a constant sequence if $d = 0$ and is symbolically denoted as $S0$. The constant sequence is both an internal and an external sequence at the same time. Since d is 0, it follows from equation (5) that the string has only one term. Thus since $\{a\} = a1$ the whole sequence and the corresponding terms are repeated l times, satisfying the definitions 1 and 2. Primes greater than 10 have only constant sequences.

Definition 5. The S sequence is an incremental sequence if $d > 0$ is symbolically denoted as $S+$.

Definition 6. The S sequence is a decremental sequence if $d < 0$ is symbolically denoted as $S-$.

Examples

Example 1. Let there be S a numerical repetitive sequence of the length $l = 9$ of the arithmetical sequence expressed in a recursive form as $\{a_i\}_{i=1}^{n}$ with $a_{i-1} + d$, where $n = 3, a_1 = 4$ and $d = -2$. The arithmetic sequence generated by these values is $\{4,2,0\}$ while the replication value is 3, because

$$\frac{l}{n} \Rightarrow \frac{9}{3} = 3 \tag{6}$$

If the entire string is replicated 3 times we get the sequence $S_E = 420420420$, thus producing an external sequence. If each term is replicated separately 3 times we obtain the sequence $S_I = 444222000$, thus producing a sequence with internal replication.

Example 2. Generate an incremental sequence S_E^+ with external replication of the length $l = 12$. The possibilities for the upper n limit according to equation (2) are $\{1,2,3,4,6\}$. Let us pick $n = 6$ and the first term as $a_1 = 4$. It follows from the equation (4)

$$0 \leq a_1 + [d(n-1)] \leq 9 \Rightarrow 0 \leq 4 + 5d \leq 9 \tag{7}$$

Options for the distance are $\{0,1\}$. We cannot pick 0 since the problem specifies to generate an external sequence. Thus we pick $d = 1$. The following terms are: $a_2 = 4 + 1 = 5$; $a_3 = 5 + 1 = 6$; $a_4 = 6 + 1 = 7$; $a_5 = 7 + 1 = 8$; $a_6 = 8 + 1 = 9$. From this we get the arithmetic sequence $\{4,5,6,7,8,9\}$. The replication value is

$$\frac{l}{n} \Rightarrow \frac{12}{6} = 2 \tag{8}$$

External replication produces the sequence S: 456789456789

Analysis and Variations Formulas for NRS

Various lengths of sequences determine different values of variables *n*, *d* and *a*1 thus producing different variations for every length and consequently disables the application of classic formulas for calculating variations. In this analysis we will compile formulas that calculate separately the variations of constant, singular, and internal/external sequences, and summing the total sum at the end.

Formulas for calculating variations of constant sequences

Each length *l* has 10 constant sequences because each number is integer with 1. It follows that for each length there are 10 constant variations, a variation for each value 0-9 replicated as long as the sequence is long, and this can be marked as:

$$|S_K| = 10 \tag{9}$$

Length consisting of prime numbers greater than 10 only have constant sequences,

which means their variation is always 10, as demonstrated by Example 5 and the following proof.

Proof. By definition prime numbers have no other factors except 1 and themselves. According to equation (2) in Chapter 2, the NRS definition limits the upper limit not greater than or equal to 10, consequently the only allowed upper limit for prime numbers remains to be number 1. When the upper limit is 1, the distance is 0 and it follows that the sequence is always constant with an amount of 10 variations.

Formulas for calculating variations of singular sequences

Given the equation (2) n cannot be greater than 10 and also the sequence is singular only if l is equal to n, then it logically follows that lengths greater than 10 have no singular sequences. Following formula calculates variations of singular sequences:

$$|S_s| = 2(11 - l)^+ \mid l \geq 6 \qquad (10)$$

From this formula we note that if l is greater than 10 the result will be negative (or zero), and the function denoted by the symbol + gets only the positive part2, as it is also illustrated in the equation (16), where it follows that the set of singular sequences for the respective length is an empty set.

$$|S_S| = \emptyset \Leftrightarrow l > 10 \qquad (11)$$

Formulas for calculating variations of internal and external sequences

Values of n produce fixed variations regardless of the length of the sequence, the length simply determines how many times those variations need to be replicated. Thus, e.g. the upper limit n = 4 produces 48 internal and external sequences, even when the length l is 8 or 12 or any other length that is divisible to the limit n = 4. This demonstrates that NRS passwords are unsafe regardless of their length, and increasing the length does not effectively enrich the variation pool. Table 1 illustrates the variations given for each value of variables n with the potential to produce internal or external sequences.

Table 1. Variations of ***SI***, for a given ***n***.

	S_I^+	S_E^+	S_I^-	S_E^-	Total
n= 2	45	45	45	45	180
n= 3	20	20	20	20	80
n= 4	12	12	12	12	48
n= 5	8	8	8	8	32
n= 6	5	5	5	5	20
n= 7	4	4	4	4	16
n= 8	3	3	3	3	12
n= 9	2	2	2	2	8
n= 10	1	1	1	1	4

The formula that calculates the internal and external variations for a given n is

$$|S_{I,E}| = 4\left[\sum_{d=1}^{k} 10 - d(n-1)\right]^+ \qquad (12)$$

Where k is the maximum distance supported by the given n, as defined by equation (4).

Total variations of NRS for a given length

The formula that calculates the total variations for a given length is:

$$\overline{S_l} = \sum_{n\leq 2}^{l-1}\left[4\left[\sum_{d=1}^{k} 10 - d(n-1)\right]^{+}\right] + 2(11-l)^{+} + 10 \quad \forall\, n|l \tag{13}$$

Example 4. Calculate the total variations for length 9.

$$\begin{aligned} |S| &= 4[10 - 1(3-1) + \ 10 - 2(3-1) + 10 - 3(3-1) + \\ &10 - 4(3-1)] + \ 2(11-9)^{+} + 10 \\ &= 4[10 - 2 + \ 10 - 4 + 10 - 6 + 10 - 8] + \ 2(2)^{+} + 10 \\ &= 4[8 + 6 + 4 + 2] + 4 + 10 \\ &= 94 \end{aligned} \tag{14}$$

Explanation 1. The formula conditions the upper limit to be divisive with l. Thus, the factors of 9 are {1,3,9}. The first sigma starts from number 2 to $l - 1$ which is 8. From 2 to 8 is only one number that is proportional to 9, and it is 3. For upper limit 3 the second sigma is applied starting at distance 1 and is increased up to 4. Any increment greater than 4 it returns an empty set. For example, if k is 5 then $10 - 5(3 - 1) = 0$. In this manner $k = 4$. After calculating the Sigma's, the singular variations 4 and the constant variations 10 are added.

Example 5. Calculate the total variations for length 29.

$$\begin{aligned} |S| &= \ 2(11-29)^{+} + 10 \\ &= 2(-18)^{+} + \ 10 \\ &= 2(0) + 10 \\ &= 10 \end{aligned} \tag{15}$$

Explanation 2. The formula conditions the upper limit to be a factor of l. But, number 29 being a prim number does not have any factor within the definition of sigma (withing the range $n \geq 2$ and $l - 1$). As such first sigma does not apply, and consequently neither the second. The last part of the formula is replaced with actual values and is calculated. Since the result is negative -18, only the positive part is obtained which is 0 by using the following function from the reference [8]:

$$f^{+} = \frac{|f| + f}{2} \tag{16}$$

Algorithm for the Generation of NRS List

The following is a function of the C# language that returns the entire NRS list for a given length:

```
public string GenerateSPN(int l) {
  int d, n, r, term = 0;
  string va = "", sequence = "", output = "";

  for (int ni = 1; ni <= l; ni++) {
    if ((l % ni) == 0) {
      n = ni;
      for (int ai = 0; ai <= 9; ai++) {
        for (int di = -9; di <= 9; di++) {
          if (di == 0)  { if (n != 1) continue; }
          else if (n == 1) { di = 0; }

          if (ai + (di*(n - 1)) <= 9 & ai + (di*(n - 1)) >= 0) {
            d = di;
            r = l / n;

            for (int i = 1; i <= n; i++) {
              term = ai + (i - 1) * d;
              va += term;
              for (int j = 1; j <= r; j++) {sequence += term;}
            }

            if (n == 1 || d == 0) {
              output += sequence + "\n";
            }
            else {
              string sequence_ext = "";
              for (int j = 1; j <= r; j++) {sequence_ext += va;}
              output += sequence + "\n";
              output += sequence_ext + "\n";
            }
            sequence = "";
            va = "";
          }
          if (n == 1) break;
        }
      }
    }
  }
  return output;
}
```

Methodology of Attack

The methodology for decrypting NRS passwords of WPA and WPA2 wireless router consists of the following steps:

a) Generate the list of NRSs for lengths of 8 to 20 and is stored in a text file.
b) The victim wireless handshake is captured using the wifite tool.
c) The handshake is attacked by importing the text file in the pyrit
d) If the handshake is decrypted using one of the passwords generated as NRS, the attack will be considered successful.

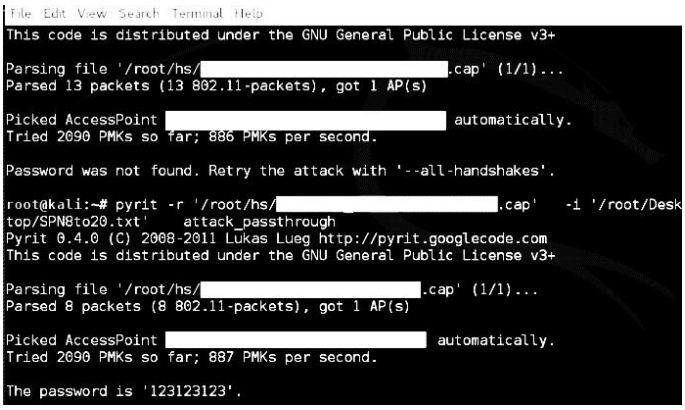


Fig. 1. Example of successful decryption.

The SPN password in Figure 1 successfully broken is 123123123. This is a password of length 9, with distance 1, external replication of value 3, initial term 1 and upper limit 3.

We also note that 887 passwords were tried per second. The pyrit tool has completed the entire NRS list with 2090 combinations in 2.4 seconds. The CPU processor used for these attacks is the Core2 Duo P8600 @ 2.4 Ghz. Although it is a relatively old
2009 processor with a sub-average processor power it has managed to try NRS
passwords up to 20 characters within 2-3 seconds.

The Prevalence of the Vulnerability

Statistics in the following table 2 are based on 89 handshakes captured in 5 different locations. Handshakes are tried to be decrypted with NRS list of the length from 8 to
20 with a total variation pool of 2090 passwords. Out of the 89 tested wirelesses, 14
NRS passwords were successfully extracted, accounting for approximately 16% of all wireless tested

The frequency of successfully decrypted NRS passwords is as follows: 12345678 (6 times); 123456789 (2 times); 0000011111 (1 time); 22223333 (1 time); 77778888 (1 time); 1231231 23 (1 time); 12341234 (1 time) and 777888999 (1 time).

Table 2. Statistics on the success of WPA handshake decryption grouped by location, and the respective total score.

	Loc. 1	Loc. 2	Loc. 3	Loc. 4	Loc. 5	Total
Amount of handshakes	25	9	14	6	35	89
Decrypted	3	0	2	3	6	14
Failed	22	9	12	3	29	75
Success rate	12%	0%	14.3%	50%	17.1%	15.73 %

Conclusion

The efficiency of NRS list is very high in the wireless domain. With a variation of only 2090 passwords it managed to decrypt the WPA / WPA2 handshake successfully in 15.7% of the cases. Additional studies and statistics are needed to verify this success rate in other international regions.

Acknowledgment. Acknowledgments to Prof. Dr. Edmond Hajrizi, Msc. Naim Preniqi and PhD.can. Besnik Qehaja.

References

9. ČVUT Fakulta Elektrotechnická, Math Tutor - Sequences - Theory - Introduction, [http://math.feld.cvut.cz/mt/txta/1/txe3aa1c.htm].

10. Nipissing University, Finite Series Tutorial, [http://calculus.nipissingu.ca/tutorials/finiteseries.html].
11. Khan Academy, Explicit formulas for arithmetic sequences | Algebra (article), [https://www.khanacademy.org/math/algebra/sequences/constructing-arithmetic-sequences/a/writing-explicit-formulas-for-arithmetic-sequences].
12. Wolfram|Alpha, Wolfram|Alpha Examples: Mathematics, [http://www.wolframalpha.com/examples/Math.html].
13. Wikipedia, Arithmetic Progression, [https://en.wikipedia.org/wiki/Arithmetic_progression].
14. Dirk Van de moortel, Permutations - Variations - Combinations, [http://users.telenet.be/vdmoortel/dirk/Maths/PermVarComb.html].
15. Wikipedia, Lagrange Interpolating Polynomial, [https://en.wikipedia.org/wiki/Lagrange_polynomial].
16. Wikipedia, Positive and negative parts, [https://en.wikipedia.org/wiki/Positive_and_negative_parts].
17. Armend Shabani, Strukturat Diskrete

Performance comparison of the TCP methods to control congestion

Salem Lepaja, Shpresë Sadiku Maxhuni

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. Congestion Control is one of the main functions of TCP to prevent and control traffic congestion in communication networks. Hence,in this paper we analyzed and compared performances of the TCP NewReno, TCP Westwood and TCP Vegas methods to control congestion in networks. Performance analysis in terms of: throughput, goodput, number of lost packets, and the cwnd dynamics are carried out by means of simulation using NS3 simulator. We have considered two scenarios. In the first scenario we have measured performances of each of the three methods operating individually, whereas in the second scenario performances are measured when they share the bottleneck link. Simulation results show that in both scenarios TCP NewReno outperforms two other methods, followed by TCP Vegas. Whereas, in the second scenario, sharing the link resource, TCP Westwood shows better performances than TCP Vegas.

Keywords: TCP, congestion control, performances.

Introduction

Transport layer is responsible for end-to-end communication between application pro- cesses running on different hosts. It is considered as the most important layer from the perspective of the applications due to the services that this layer offers to them. Transport layer distinguishes two modes of data delivery: the reliable and the unrelia- ble. The reliable delivery is based on feedbacks from the receiver to the sender through an Acknowledgment (thereinafter referred to as ack) for every received segment. When an ack is not received, the sender concludes that the segment is lost and retransmits it, after that segment's timer has expired. The unreliable delivery means that there is no guarantee that segments will arrive in time, error-free and in the correct order. Transport layer uses the TCP (Transmission Control Protocol) for the reliable deliv- ery and the UDP (User Datagram Protocol) for the unreliable data delivery.
Fig.1 illustrates the mechanisms of acks, timers and retransmissions that TCP uses in order to provide reliable data delivery.

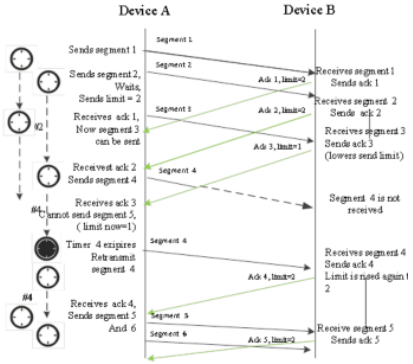


Fig. 1. TCP Mechanism with Acks, timers and retransmission

Congestion Control in TCP

Congestion control is one of the key functions of the transport layer that is implemented in TCP. Congestion refers to a network state where extended delays are introduced as a result of datagram overcrowding in one or more routers. Two main conditions that cause congestion in networks are:

1. When data from a communication channel with higher throughput have to go through a channel with lower throughput.
18. The router has a lower egress throughput than the sum of its ingress throughputs.

Congestion Control Algorithms

The original version of TCP did not include any mechanisms for congestion control. After a series of "congestion collapses" that occurred in 1980s, Van Jacobson in his paper [1], proposed four algorithms to address this problem:

1. Slow Start
2. Congestion Avoidance
3. Fast Retransmission
4. Fast Recovery

Although these algorithms are independent of each other, in practice are always implemented in combinations.

Slow Start. At the initial stage of a TCP connection, the data are sent with a lower rate that increases depending on the rate of the arrival of acks from the receiver [2].

Congestion Avoidance. Slow start ends when Congestion window (thereinafter referred to as cwnd) exceeds slow start threshold - ssthresh (usually set to 65535 bytes). This is when the congestion avoidance phase begins and the cwnd now increases line- arly, until congestion occurs.

Fast Retransmission. When three or more duplicate acks are received for the same segment, the sender will immediately retransmit the lost segment, without waiting for a timer timeout. This leads to a significant improvement of the transmission efficiency.

Fast Recovery. Duplicate acks signal that there is a mild congestion in the network, because out of order segments are still arriving at the receiver. This is why there is no need do go back to the Slow Start phase and reduce halfway the cwnd, but it can be continued with the same value that was used before duplicate acks arrived.

Simulation Environment

There are a number of TCP methods to control congestion in networks, like: TCP Reno, New Reno, Vegas, SACK, BIC, CUBIC, Westwood [2, 3, 4]. In this paper we have simulated and analyzed three of them: TCP New Reno, Vegas and Westwood. The per- formance analyses and comparison of these three methods, in terms of throughput, goodput, packet loss and congestion window, were carried out by means of simulations using NS-3 simulator [5, 6]. Our network topology, as shown in Fig.4, comprises of two routers and two groups of three hosts each, connected to one of the routers. We have experimented with two different scenarios:

1. Scenario I: When the three methods are operating individually.
2. Scenario II: The three methods sharing a single link

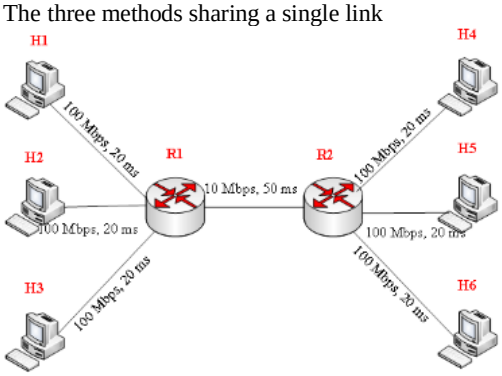


Fig. 2. Network Topology

In Tab.1 are shown the used simulation parameters.

Table 1. Simulation parameters.

Host-Router link Bandwdth	100 Mbps
Router-Router link Bandwdth	10 Mbps
Host-Router link delay	20 ms
Router-Router link delay	50 ms
Packet size	1.2 KB
Host-Router queue size	(100000*20)/Packet Size
Router-Router queue size	(10000*50)/Packet Size
No. of Packets	10000000
Error Rate	0.000001
Simulation Duration	100 s

Results and Discussion

Scenario I

In this scenario, we investigated the performance of the three above mentioned conges- tion control methods when they are performing separately i.e., when a single method uses the link at a time. The simulations were performed in terms of throughput, goodput and packet loss metrics.

Table 2. Simulation results: Scenario I.

	NewReno	*Westwood*	*Vegas*
Average Throughput	1.17 Mbps	0.85 Mbps	0.96 Mbps
Maximum Throughput	2.47 Mbps	1.51 Mbps	2.57 Mbps
Maximum Goodput	2.25Mbps	1.3Mbps	2.16 Mbps
Packet loss due to Congestion	10	14	14
Packet loss due to buffer overflow	0	0	0

As it can be seen from Table 2. TCP NewReno outperforms two other methods in all considered metrics, followed by TCP Vegas. Low congestion level and sufficient buffer capacity, makes it possible for TCP New Reno to utilize more efficiently the available bandwidth. In situations of high packet loss, TCP Westwood would have advantages due to its bandwidth estimation algorithm, where cwnd does not halve, instead it takes on values closer to real network capacity.

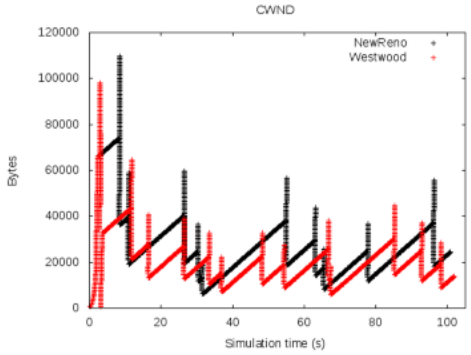


Fig. 3. Congestion Window

The better performance of New Reno can be seen from Figure 3, which illustrates the change of the cwnd as a function of simulation time for TCP New Reno and TCP West- wood.

Scenario II

Considering that in real life cases, like those on the Internet, multiple TCP methods to control congestion operate simultaneously in a combination fashion, it is important to measure the performance when they share the link.

The topology is the same as the one used in the first scenario, with the only change that now the connections of TCP Westwood and TCP Vegas start 20 seconds after the connection of TCP New Reno has started. The results of simulation for scenario II given in Table 3., show that NewReno offers significantly better throughput and goodput than other two methods, while the packet loss remains the same. In this scenario there is a disadvantage of TCP Vegas for the use of bandwidth when coexisting with New Reno and Westwood. This is as a result of the more conservative nature with which TCP Vegas increases the cwnd, when com- pared to the more aggressive algorithms of TCP Westwood and especially to the one of TCP New Reno. TCP Vegas has a proactive congestion control algorithm that reduces the cwnd at the initial stages of the congestion, which prevents TCP Vegas to utilize its share of bandwidth resources more efficiently.

Table 3. Simulation results: Scenario II.

	NewReno	*Westwood*	*Vegas*
Average Throughput	1.09 Mbps	0.97 Mbps	0.73 Mbps
Maximum Throughput	2.47 Mbps	0.86 Mbps	0.61 Mbps
Maximum Goodput	2.18 Mbps	0.78 Mbps	0.55 Mbps
Packet loss due to Congestion	11	11	11
Packet loss due to buffer overflow	0	0	0

Conclusions

This paper analyzed and compared the performances of three TCP methods to control congestion: TCP NewReno, TCP Westwood and TCP Vegas. Based on the simulation results, for the two scenarios considered, TCP NewReno performs the best among the three methods analyzed. For the first scenario, TCP NewReno is followed by TCP Ve- gas, whereas for the second one it is followed by TCP Westwood.
An explantion for such a performance degradation of TCP Vegas, is the too early de- crease of the CWND when the congestion is at its initial state. This prevents TCP Vegas from using its share of bottleneck resources. Our results are in the line with the fact that TCP NewReno is the most used method today to control congestion. Our further work in this area will be concentrated on investigating the optimal size of the cwnd and ssthresh value.

References

1. V. Jacobson and M. J. Karels: Congestion Avoidance and Control. University of California- Berkley, (1988).
2. M. Allman, V. Paxson, E. Blanton: TCP Congestion Control. In Request for Comments: 5681, Network Working Group, Purdue University (2009).
3. James F. Kurose, Keith W. Ross, Computer Networking, Pearson 2010, 6th Edition.
4. Andrew S. Tanenbaum, David J. Wetherall, Computer Networks, Prentice Hall, 2010, Fifth Edition.
5. https://www.nsnam.org/release/ns-allinone-3.27.tar.bz2
6. https://www.nsnam.org/docs/manual/ns-3-manual.pdf

Improvement of Gender Recognition using the Cosfire Filter Framework (Simulations Platform of Shape-Preserving Regression – PCHIP)

Virtyt Lesha[1] Arben Haveri[2] Jozef Bushati[3]

[1] „Luarasi“ University, Rruga e Elbasanit, Tirana, Albania
[2,3] University of Shkodra, Sheshi „2 Prilli“, Shkodra, Albania

Abstract. Biometrics is evolving every day more and more in technical sense and consequently faces with further challenges that become sharper. One of these challenges of is gender recognition that finds very important and key applications. In this paper, we consider the gender recognition process implemented through the Cosfire filter applied through Viola- Jones algorithm and simulated through the Matlab platform. Objective of this paper is improving the execution of gender recognition. The database contains 237 images of 128 to 128 pixels, where 128 are males and 109 are females. For each of them, gender recognition is performed by applying current and improved Viola-Jones algorithm and execution time for each of them is measured. Consequently, it is noticed that the execution time in the case of modified algorithm is lower than the first version. The change consists in intervening in recursive filtering by duplicating it. Furthermore, data obtained from both algorithms in question are processed through the Shape-Preserving Regression - PCHIP regression by giving respective equations and the coefficients of the determination and the respective residual plots performed by Matlab simulation test-bench. Recommendations can be issued in context of further execution time reduction of Viola-Jones algorithm applied on gender recognition.

Keywords: biometrics, execution, Viola-Jones, filtering

Introduction

Biometric applications are recognizing a very significant increase in almost every simulation domain at different levels or even application levels of different types, ranging from crucial to advertising (9). Consequently, research in biometrics and its various derivatives leads to the further development of applications that need performance improvements at the time of execution or adapting to different hardware parameters that may be necessary to execute certain software. which performs the said biometric function. Moreover, one of the current and most recent developments in the field of biometrics is gender recognition; this is a biometric derivative that still leaves a lot of room for research and development and furthermore it is still at the level of improving such parameters as the image quality against which the algorithm will be recognized, the execution time of the algorithm, etc (6). However, according to (4) one of the recently developed algorithms for gender recognition is the Viola-Jones algorithm applied in accordance with the COSFIRE Frame-Work filter. Although this object-detection algorithm is known to be a slow algorithm in the training process (2) it has a great advantage in speeding the object detection process. This algorithm uses the COSFIRE filter and therefore does not contain multiplications (8).

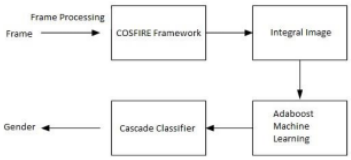


Fig. 1. The scheme of Gender Recognition process that uses Viola-Jones algorithm

In our paper we have addressed a performance analysis of the Viola-Jones algorithm in the context of using the COSFIRE filter platform for gender recognition taking into account runtime parameters. Specifically, a classic version of the Viola-Jones algorithm has been considered with the aim of improving it in sensing the execution time of the said algorithm. The modification in question was carried out through simulations in Matlab simulation platform tools and subsequent regressive comparisons were performed using Shape- Preserving Regression and reflecting the corresponding results of the determination coefficients.

Methodology

The algorithm in question was developed according to Viola-Jones and considers a database of 237 images containing faces of different genders ranging from 128 pixels to128 pixels; among these images are 128 male images and the rest are females and the objective is gender recognition and further is the improvement of the execution time performance of the Viola-Jones algorithm used to perform this gender recognition process. Specifically, the intervention was carried out in the frame processing phase and in the COSFIRE Framework as shown in the following code:

```
function tuple = computeTuples(inputImage,operator)
if ~iscell(operator)
    operatorlist{1} = operator;
else
    operatorlist = operator;

end
set = cell(0);
index = 1;
reflecting, rotating
for op = 1:length(operatorlist)
    params = operatorlist{op}.params;
    reflectOperator = operatorlist{op}
    for reflection = 1:2^params.invariance.reflection
        if reflection == 2
            if params.inputfilter.symmetric == 1
                reflectOperator.tuples(2,:) = mod(pi - reflectOperator.tuples(2,:),pi);
            else
                reflectOperator.tuples(2,:) = mod(pi - reflectOperator.tuples(2,:),2*pi);
            end
            reflectOperator.tuples(4,:) = mod(pi - reflectOperator.tuples(4,:),2*pi);
        end
        rotateOperator = reflectOperator;
        for psiindex = 1:length(params.invariance.rotation.psilist)
            if strcmp(params.inputfilter.name,'Gabor')
                if params.inputfilter.symmetric == 1
                    rotateOperator.tuples(2,:) = mod(reflectOperator.tuples(2,:) +
params.invariance.rotation.psilist(psiindex),pi);
                else
                    rotateOperator.tuples(2,:) = mod(reflectOperator.tuples(2,:) +
params.invariance.rotation.psilist(psiindex),2*pi);
                end
            end
```

The intervention is performed in the above cycle by "for" adding two more cycles, thereby increasing the complexity but on the other hand reducing the execution time in the COSFIRE Framework process.
Further, each of the 237 images measured the execution time of the classical as well as the improved algorithm and were placed in a database for comparison in a regressive analysis. The regressive analysis will consist of the Gauss8 adapted to Shape- Preserving Regression model and has this equation:

$$f(x) = a1 * \exp\left(-\left(\frac{x-b1}{c1}\right)^2\right) + a2 * \exp\left(-\left(\frac{x-b2}{c2}\right)^2\right) + a3 * \exp\left(-\left(\frac{x-b3}{c3}\right)^2\right) + a4 * \exp\left(-\left(\frac{x-b4}{c4}\right)^2\right) + a5 * \exp\left(-\left(\frac{x-b5}{c5}\right)^2\right) + a6 * \exp\left(-\left(\frac{x-b6}{c6}\right)^2\right) + a7 * \exp\left(-\left(\frac{x-b7}{c7}\right)^2\right) + a8 * \exp\left(-\left(\frac{x-b8}{c8}\right)^2\right) \quad (1)$$

Results

After applying the improved algorithm and regressive analysis, the corresponding results of the simulations carried out through the Matlab R2019b Toolbox platform are presented. Figure 2 below shows the Shape-Preserving Regression graph as well as the Residual Plots for the case when we applied the Viola-Jones default algorithm execution time for gender recognition for 237 images.

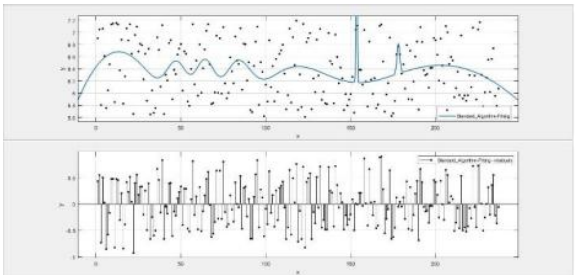


Fig. 2. The plot of regression of the standard Viola-Jones algorithm and the corresponding residual-plots

The figure 3 shows the Shape-Preserving Regression graph as well as the corresponding residual plots for the case where we applied the runtime measurement of the improved Viola-Jones algorithm for gender recognition to 237 images.

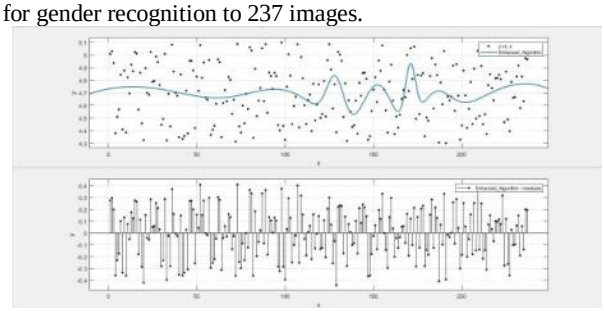


Fig. 3. The plot of regression of the improved Viola-Jones algorithm and the corresponding residual-plots

Below we have the regression coefficients that correspond to the equation (1):

Table 1. The table that shows the trendline coefficients according to the algorithm for the classic simulation model as well as the improved one

Coefficients (with 95% confidence bounds)	The Values of Standard Algorithm	The Values of Improved Algorithm
a1	1.804 (-1.682, 5.29)	0.4577
b1	107.6 (80.24, 135)	105.1
c1	37.8 (-5.901, 81.5)	35.84
a2	6.456 (6.3, 6.612)	4.737
b2	203.7 (183.8, 223.5)	10.83
c2	147.3 (9.712, 284.9)	211.6
a3	5.54 (1.684, 9.396)	0.5139
b3	6.243 (-12.58, 25.07)	152.5
c3	59.44 (7.723, 111.2)	12.82
a4	1.146 (-0.1672, 2.459)	3.445
b4	63.87 (56.3, 71.44)	261.3
c4	8.777 (-6.01, 23.56)	103.5
a5	5.261 (-6.373e+05, 6.374e+05)	0.2986
b5	153.5 (-285.7, 592.7)	184
c5	0.3766 (-1.295e+04, 1.295e+04)	13.22
a6	1.009 (-0.91, 2.929)	0.3774
b6	81.82 (68.66, 94.99)	128.4
c6	12.12 (-7.931, 32.17)	6.922
a7	0.8392 (-0.7852, 2.464)	0.4573
b7	48.27 (35.48, 61.06)	170.9
c7	8.877 (-7.213, 24.97)	4.109
a8	0.505 (-0.4078, 1.418)	0
b8	178.1 (176.3, 179.8)	-166.3
c8	1.143 (-1.183, 3.47)	29.18

Also, presented below are the supporting data of the trend-line performance and the execution time of which 237 images of the simulation model.

Table 2. The table that shows the performance of the data for the two models

	SSE	R-square	Adjusted R-square	RMSE
Standard Algorithm	45.51	0.0987	0.001375	0.4622
Improved Algorithm	10.73	0.07793	-0.02164	0.2244

Conclusions and Discussions

In this paper we discussed about the field of biometric gender recognition regarding the performance of the Viola-Jones algorithm. The importance of the Viola-Jones model, as an algorithm for biometric facial identification processes, plays an important role and provides space for continuous reconfigurations and modifications that improve the performance of the

models in question in different terms where one of them is the time of facial identification execution versus a certain database. In our model we have implemented a simulation model through Matlab software (R2019a), which realizes the gender recognition process including an open source database of 237 facial images (13). The purpose of this paper was to present a simulation model that deals inside it with a modified version of the Viola-Jones algorithm; the modeling was intended to give the results of this improvement of the Viola-Jones algorithm in terms of the execution time of identifying a facial image (5) (11).
The modification, carried out in the Viola-Jones algorithm, was the intervening in the respective cycles by introducing two "for" cycles into the algorithm. Also, the modification of this modeling was the intervention standard code that makes it possible to read the database that is further applied in the basic model algorithm (7). Also, to reflect the results in time, we simulated through Matlab a "Shape-Preserving Regression – PCHIP" algorithm model that generates a threefold reflection performance of the standard Viola-Jones algorithm versus the improved algorithm. Along with the graphical data of Shape-Preserving Regression – PCHIP approximation, specific and supporting parameters are also given, where they are considered as computable coefficients (3). Because of the Shape-Preserving Regression – PCHIP method simulations, it is concluded that the execution time of the facial image identification process for the standard Viola-Jones algorithm has extremes that show greater time values compared to the improved pattern (1). Disputes and discussions that arise from this study leave spaces for improvements in the terms of time needed to perform facial identification process performances as well as the size of the database needed to keep the face images; in relation to the database arises the challenge of designing models that go hand in hand with the improvement of the model's quality in relation to the execution (10)(12).

References

1. Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., Senior, A. W.: "Guide to biometrics", (Springer-Verlag, New York, 2004), doi: 10.1007/978-1-4757- 4036-3.
2. Bourlai, T.: "Face Recognition Across the Imaging Spectrum", (Springer International Publishing, Switzerland, 2016), doi: 10.1007/978-3-319-28501- 6.
3. Buhrow, W. C.: "Biometrics in support of military operations: Lessons from the battlefield", (CRC Press, Florida, USA, 2016).
4. Chityala, R., Pudipeddi, S.: "Image processing and acquisition using Python", (CRC Press, Florida, USA, 2014).
5. Das, R.: "The science of biometrics: Security technology for identity verification", (Routledge, New York, 2018), doi: 10.4324/9780429487583.
6. Datta, A. K., Datta, M., Banerjee, P. K.: "Face detection and recognition: Theory and practice", (Taylor & Francis, Florida, USA, 2015), doi: 10.1201/b19349.
7. Gonzalez, R. C., Woods, R. E.: "Digital image processing", 4th ed, (Pearson, New York, 2018).
8. Jain, A. K., Li, S. Z.: "Handbook of face recognition", (Springer-Verlag, London, 2011), doi: 10.1007/978-0-85729-932-1
9. Newman, R.: "Security and access control using biometric technologies", (Cengage Learning, US, 2009).

Applying SOA Approach to Financial Institution: Case Study

Agon Memeti, Florinda Imeri

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. A financial institution is an institution that conducts financial transactions, such as depositing money, taking out loans and/or exchanging currencies. Systems used by them like any other system tend to skew old, but, the replacement and integration of these systems is a difficult due to the heterogeneous nature. Due to this it is imperative to consider alternative infrastructure such as SOA (service-oriented architecture), which is seen as the best technology for internal and external interfaces, resulting cost reductions associated with its deployment; combined with standardized protocols, and increased interoperability among IT infrastructures. Studies shows that this flexible architecture will encourage innovation and increase the banks' ability to react to customer feedback a lot more swiftly. An approach to building information system for the financial institution, based on the technology of SOA is discussed in the paper. The financial institution for which we have proposed a SOA based architecture is Saving House Mozhnosti, a financial institution that exists in the financial market for 17 years. The Savings House serves micro, small and middle enterprises that belong in the sector of trade, services and production as well as physical persons and offers financial products and services such as crediting and savings. It is recognizable for the support to the people in need to get financial support for their good business ideas, belief in free initiative, individual creativity and personal responsibility. As one successful financial institution, there is no doubt it can highly benefit by SOA approach. Our goal is to act as a guide in helping this institution tackle the types of issues using a services-based approach, thus improving customer experience. The proposed approach will improve the manageability of the system, increase its speed and reliability and provide security.

Keywords: Infrastructures., SOA., transactions., technology.

Introduction

Financial institutions are faced with the continued need to increase the flexibility of the services they offer, whereas are being under intense pressure to reduce costs. Due to rapid changing of the marketplace, financial institutions should react effectively in timely manner[1]. To meet the demands of the customers, IT Infrastructure needs to be tailored creating thus a competitive advantage for the organization to complete business processes. SOA as an infrastructure, is a relatively new concept with an increase interest in it, both from industry and the academy[2]. There are a variety of statistics available from various magazines and technology analysts relating to the adoption of SOA in the industry, generally indicating the widespread acceptance of SOA. SOA delivers greater business values and competitive advantage in the marketplace by offering better alignment between IT and business processes with its attributes, such as reusability, agility, etc. It is aggregation of components that satisfy a design need. It comprises of components, services, and processes. Components are binaries that have a defined interface

whereas service is a grouping of components that finishes a job[3]. In this paper we will present an SOA based approach for a financial institution called Saving House Moznosti [4], a financial institution that operates for 17 years in the market of Republic of North Macedonia. It offers services to micro, small and middle enterprises, but it offers financial products and services to physical persons as well. In the market it is recognized for the financial support that it gives to the people to implement their business ideas, by offering initial capital. Financial institutions should focus on investing in such a way to provide tailored services and products to customers, but there are always some small gaps that lead to customer service issues. Some of the issues that our financial institution in study faces are:

- Automatically loan approval
- Renewal of FD (FixedDeposit)

But as any other solution, the proposed approach for the financial institution in this study has its own limitations. First, one very important issue is that what should be taken into the considerations is the fact that its clients are people who usually are working on agriculture sector, and literally not all of them own a device that enables them to fulfill a loan request. Another issue that must be considered is the financial situation of the company, whether it can implement it, as the implementation part is a bit difficult, as is it the nature of SOA.

Methodology

The online application and documentations for loan can be sent to a Client Relationship Officer through email, but, when customers want to get a loan, they must go by themselves to the offices of the Saving House so they will get the right information on the amount of money they can borrow. The approval process of approving the loan takes some time, which causes frustration to the customer which in some cases they may end up taking his/her business elsewhere. But, those are rules followed by client relationship officers who have a certain protocol for approving a loan and giving money to the customer. We think that this information can be controlled by customers themselves, if such a service is offered in a web-based way. First, the costumer will have to fill a form with his personal data, and from there he'll be segmented automatically to the system and check the possible loans he could borrow from this institution, and then apply by following the necessary procedures. This would be cost effective on both sides and would substitute the traditional way of getting information from the Client Relationship Officer in person. The second issue we listed has to do with the renewal of FD. So, after the fixed deposit matures, the system automatically renews the FD if not withdrawn, informing the customer manually, by cell phone, or not informing at all. We suggest that this can be done by reminders which the system will sent to the customers automatically few days before maturity. As we can see, we have the Customers and Client Relationship Officers as main actors for our approach. SOA would act as middleware between these two roles, thus creating flexible services.

Platform Design and Development

To give a better overview of this project initiation idea, we've used sequence diagram as type of UML diagram to describe the overall process of our proposed SOA Case in Study. As it can be seen from the Fig. 1, we've included both issues listed in problem statement.

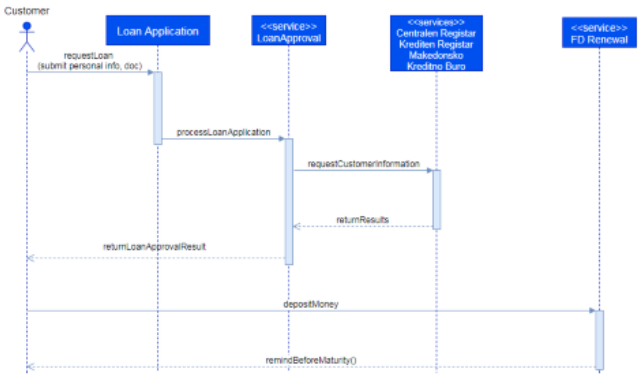


Fig.1. Sequence Diagram of the loan approval process and FD renewal

The process is as following. The customer as the main entity in our case decides to request a loan at our financial institution in study, Mozhnosti, which already offers the online loan application. The relevant parties from which Mozhnosti retrieve information are the following: Centralen Registar, Krediten Registar and Makedonsko Kreditno Buro that are included at the sequence diagram as specific services that already are used by client relationship officers during loan application process. In our case, the loan approval service changes the current rules. After the customer enter his personal information (including age), amount of money to loan, etc. and submits the necessary documents he makes the request. Age is very important, since the customer based on his age is automatically segmented by the system for the loan he can borrow, such as student loans, retiree loans, etc. The loan application will be processed from loan approval service. Loan approval service asks to retrieve information about the customer from three services as usually done by client relationship officers, in order to request credit rating for the customer. Each of institutions mentioned are represented as individual Web Services, such as Central Registar service, Krediten Registar service and Makedonsko Krediten Buro service, and they provide the loan approval service the necessary information, in order to be capable for deciding at the end of the process. Each financial institution has its own categories of criteria, such as occupation, finance, assurance, etc. And they are provided and gathered from these services, from where they are compared to the limit score of the institution to approve or reject the loan. Some of the categories may be more important than others, for instance in weighting finance versus assurance, finance would be more crucial in evaluation criteria. After retrieving the results from these services, the loan approval service decides for the loan application of the customer and returns the results, the loan is rejected or approved. Finally, customer receives result and as an opportunity given to him is to decide whether to accept or give up on the loan requested. The following process may continue in person, thus signing the generated loan and making a contract with the financial institution for the sequential flow of the payment way, if one decides to accept the loan.

Fixed deposit is another issue, that is included in the same diagram. So, after user has deposited an amount of money, the FD (Fixed Deposit) service will remind him before the deposit matures, in order to notify the customer for the status of the deposit, thus the customer is given opportunity to decide to withdraw or renew the fixed deposit by himself.

Programming Platforms and Tools

In order to enable the proposed idea, we'll propose also tools and programming platforms considered to be used. Since we have interaction between Web Services, BPEL1 process is

appropriate for our case. BPEL as a web service with an associated process definition defined in an XML-based language [5], will allow the web services to interconnect and share information, respectively loan approval service will interconnect and receive information from three credit rating services defined in the sequence diagram above, and therefore will make an appropriate decision. Undoubtedly, WSDL must be defined for the BPEL process. As a programming language, Java is suggested to be used as the best choice for development, since it is used more and more in several domains. One of the biggest areas where Java is used is web service development, due to its compatibility and portability. Worth noting is the fact that Java do not introduce breaking changes with new releases, it is a cross platform, and every solution written in any version of it will be running on all subsequent versions of the language [6]. SLA2 is one of the issues that must be taken in consideration. It is defined as "a commitment between a service provider and a client "[7]. And, it should ensure service availability, response time, quality, scalability, traffic levels, performance, etc. In order to prove functionality of services, the process must be tested, monitored and evaluated, if one wants to overcome challenges or problems that may face during development or after that. Also, security across users must be maintained.

Conclusion and Future Works

SOA is the future of banking and financial institutions. Although its steps have not begun yet, one thing is for sure, as soon as possible financial institutions in our country will begin to construct their architecture based on SOA opportunities, they will improve their performance in the marketplace, thus meeting the market demands. Innovation, effective cost and customer needs are three most important things that lead to a successful management of a financial institution. As soon as all the banks integrate SOA in their system, benefits will arise due to the reusability of services, interconnection, and so on. SOA is already used in banking industry across the world and has made big impacts. And, it will continue to revolutionize banking system. Our proposed approach is the very first step. This project idea represents a modest effort to improve the overall functioning of the financial institution in Republic of Macedonia. There is so much to be said and more to do, as improving the quality of banking should be one of the priorities in the development of countries like ours. Attempt to implement the proposed approach will remain as a future work.

References

1. “SOA , standards and IT systems : how will SOA impact the future of banking services ?S. Van Wyk, ,” BIAN.
2. “Impact of SOA Adoption with regard to Business Value : A study from South East Asia Bank, I. B. Sutawijaya, O. Steen, N. Holmberg, A. Olerup, and M. Wärja, ,” [2010].
3. “Using SOA for development of information system ‘ Smart city ,E. Duravkin,’” vol. 2, no. 11, p. 2010, [2010].
4. Profile of the Savings house Mozhnosti, “Можности - Дома.” 2019.
5. “Understanding Business Process using BPM and BPEL.S. C. Kaliyaperumal,” .
6. “Ranking the most in-demand programming languages in banking _ eFinancialCareers.” [2019].
7. “Service-level agreement.” [2018].

Analyzing the linearity of some operators

Faton Kabashi, Azir Jusufi, Hizer Leka, Flamure Sadiku

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. Linear operators occupy an important place in functional analysis and linear algebra, which are among the most important and substantive disciplines of mathematics, whose methods and results have created an indispensable apparatus for the development of numerical mathematics, theory of approximations, equations differential and especially mathematical physics and applied mathematics. Also, linear operators are a central object of study in vector space theory. A linear operator is a function which satisfies the conditions of additivity and homogeneity. Not every function is linear operators. We will try to explore some functions which are also linear operators.

Keywords: Function, linear operators, vectors, vector spaces

Vector spaces

Often in analytical geometry, mechanics, physics, etc., we come across object oriented objects called vectors. Those objects define linear actions, vector collection, and scalar vector multiplication. [1,4]

Thus, the set of free vectors in a straight line, plane, or ordinary three-dimensional space, with respect to the said actions, form special algebraic structures, which enjoy certain properties. These structures are called vector spaces. In the general case, by taking abstract objects, actions are defined and the conditions are formulated which must satisfy them [3].

A vector space X is an aggregate of elements, called vectors, u, v,... for which linear operations (addition u + v of two vectors u, v and multiplication α u of a vector u by a scalar α are defined and obey the usual rules of such operations. The scalars are assumed to be complex numbers unless otherwise stated (complex vector space). α u is also written as u α whenever convenient, and α^{-1}u is often written as u α . The zero vector is denoted by 0 and will not be distinguished in symbol from the scalar zero. Vectors u_1,..., u_n are said to be linearly independent if their linear combination $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_n u_n = 0$ only if $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$; otherwise they are linearly dependent [2].

The dimension of X , denoted by dim X , is the largest number of linearly independent vectors that exist in X . If there is no such finite number, we set dim X = ∞ . A subset M of X is a linear manifold or a subspace if M is itself a vector space under the same linear operations as in X .

The dimension of M does not exceed that of X . For any subset S of X , the set M of all possible linear combinations constructed from the vectors of S is a linear manifold; M is called the linear manifold determined or spanned by S or simply the (linear) span of S . According to a basic theorem on vector spaces, the span M of a set of n vectors u1,..., un is at most n - dimensional; it is exactly n -dimensional if and only if u1,..., un are linearly independent. There is only one 0-dimensional linear manifold of X , which consists of the vector 0 alone and which we shall denote simply by 0 [1,2].

Example 1.1. The set of all complex-valued continuous functions u : x → u(x) defined on an interval I of a real variable x is an infinite-dimensional vector space, with the obvious definitions of the basic operations α u + β v . The same is true when, for example, the u are restricted to be functions with continuous derivatives up to a fixed order n . Also the interval R may be replaced by a region' in the m dimensional real euclidean space R m [6].
Example 1.2. The set of all solutions of a linear homogeneous differential equation

$$u^{(n)} + a_1(x)u^{(n-1)} + \ldots + a_n(x)u = 0$$

with continuous coefficients a $_j$ (x) is an n -dimensional vector space, for any solution of this equation is expressed as a linear combination of n fundamental solutions, which are linearly independent [3].
Definition 1.1. Let X be an N -dimensional vector space and let x_1,..., x_N be a family of N linearly independent vectors. Then their span coincides with X , and each u ∈ X can be expanded in the form

$$u = \sum_{j=1}^{N} \xi_j x_j$$

in a unique way. In this sense the family {x j}is called a basis of X , and the scalars j are called the coefficients (or coordinates) of u with respect to this basis.
The correspondence u (ξ_j) is an isomorphism of X onto C N (the set of numerical vectors) in the sense that it is one to one and preserves the linear operations, that is, u → (ξ_j) and v → (η $_j$) imply $\alpha u + \beta v \to (\alpha\xi_j + \beta\eta_j)$ As is well known, any family x_1,..., x $_p$ of linearly independent vectors can be enlarged to a basis x_1,..., x $_p$, x $_{p+1}$,..., x_N by adding suitable vectors x $_{p+1}$,..., x_N .
Definition 1.2. For any subset S and S ' of X , the symbol S + S' is used to denote the (linear) sum of S and S ' , that is, the set of all vectors of the form u + u' with u ∈ S and u'∈ S' . If S consists of a single vector u , S + S' is simply written u + S' . If M is a linear manifold, u + M is called the inhomogeneous linear manifold (or linear variety) through u parallel to M .
The totality of the inhomogeneous linear manifolds u + M with a fixed M becomes a vector space under the linear operation $\alpha(u+M)+\beta(v+M)=(\alpha u+\beta v)+M$. This vector space is called the quotient space of X by M and is denoted by X / M . The elements of X / M are also called the cosets of M . The zero vector of X / M is the set M , and we have u + M = v + M if and only if u − v ∈ M . The dimension of X / M is called the codimension or deficiency of M (with respect to X) and is denoted by codim M . We have dim M + co dim M = dim X [4,6].

Linear operators. Matrix representations

Definition 2.1. Let X be a vector space. A complex-valued function f □u□ defined for u □ X is called a linear form or a linear functional if $f[\alpha u+\beta v]=\alpha f[u]+\beta f[v]$ for all u, v of X and all scalars α,β.
Example 2.1. If X = C N (the space of N -dimensional numerical vectors), a linear form on X can be expressed in the form: $f[u]=\sum_{j=1}^{N}\alpha_j\xi_j$ for $u=(\xi_j)$. It is usual to represent f as a row vector with the components α $_j$, when u is represented as a column vector with the components ξ $_j$. The $f[u]=\sum_{j=1}^{N}\alpha_j\xi_j$ is the matrix product of these two vectors.
Example 2.2. Consider whether linear operations from $\mathbb{R}^2$ $\mathbb{R}^2$ are linear operators:

a) $a)\ A((a, b)) = (\sin a, b)$
b) $b)\ A((a, b)) = (a - b, 0)$

Solution: We try additives and homogeneity:

Let it be $U_1 = (a_1, b_1)$ and $U_2 = (a_2, b_2)$ where $U_1, U_2 \in \mathbb{R}^2$

$$a)\ A(U_1 + U_2) = A\big((a_1, b_1) + (a_2, b_2)\big)$$
$$= A(a_1 + a_2, b_1 + b_2)$$
$$= \sin(a_1 + a_2), (b_1 + b_2)$$
$$\neq (\sin a, b)$$Type equation here.

Let $U_1 = \left(\frac{\pi}{2}, 0\right), U_2 = (\pi, 1)$

$$A(U_1 + U_2) = A\left(\left(\frac{\pi}{2}, 0\right) + (\pi, 1)\right)$$
$$= A\left(\frac{3\pi}{2}, 1\right)$$
$$= \left(\sin\frac{3\pi}{2}, 1\right) = (-1,1)$$
$$A(U_1) + A(U_2) = A\left(\frac{\pi}{2}, 0\right) + A(\pi, 1)$$
$$= \left(\sin\frac{\pi}{2}, 0\right) + (\sin\pi, 1)$$
$$= (1,0) + (0,1) = (1,1)$$

Although $(-1,1) \neq (1,1)$, thus $A(U_1 + U_2) \neq A(U_1) + A(U_2)$
then the given operation is not linear operators.

$$b)\ A(U_1 + U_2) = A\big((a_1, b_1) + (a_2, b_2)\big)$$
$$= A(a_1 + a_2, b_1 + b_2)$$
$$= (a_1 + a_2 - (b_1 + b_2), 0)$$
$$= \big((a_1 - b_1) + (a_2 - b_2), 0\big)$$
$$= (a_1 - b_1, 0) + (a_2 - b_2, 0)$$
$$= A(U_1) + A(U_2)$$

Thus, $A(U_1 + U_2) = A(U_1) + A(U_2)$

$$A(\lambda U) = A\big(\lambda(a_1, b_1)\big)$$
$$= A(\lambda a_1, \lambda b_1)$$
$$= (\lambda a_1 - \lambda b_1, 0)$$
$$= \lambda(a_1 - b_1, 0)$$
$$= \lambda(AU)$$

Thus, $(\lambda U) = \lambda(AU)$

Since additive and homogeneity are met then we say that it is linear operators [3].

Definition 2.2. Let X , Y be two vector spaces. A function T that sends every vector u of X into a vector v = Tu of Y is called a linear transformation or a linear operator on X to Y if T preserves linear relations, that is, if $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T u_1 + \alpha_2 T u_2$ for all u1, u2 of X and all scalars α_1, α_2 . X is the domain space and Y is the range space of T .

If Y = X we say simply that T is a linear operator in X . In this book an operator means a linear operator unless otherwise stated. For any subset S of X , the set of all vectors of the form Tu with u ∈ S is called the image under T of S and is denoted by TS ; it is a subset of Y . If M is a linear manifold of X , TM is a linear manifold of Y . In particular, the linear manifold TX of Y is called the range of T and is denoted by R(T) . The dimension of R(T) is called the rank of T ; we denote it by rank T . The deficiency (codimension) of R(T) with respect to Y is called the deficiency of T and is denoted by defT .
Thus rank T + def T = dim Y . For any subset S ' of Y , the set of all vectors u ∈ X such that Tu ∈ S' is called the inverse image of S ' and is denoted by T –1S' . The inverse image of 0 ⊂ Y is a linear manifold of X ; it is called the kernel or null space T of N (T). The dimension of N (T) is called the nullity of T , which we shall denote by nul T . We have rank T + nul T = dim X . To see this it suffices to note that T maps the quotient space X / N (T) (which has dimension dim X - nul T) onto R(T) in a one-to-one fashion. If both nul T and def T are zero, then T maps X onto Y one to one. In this case the inverse operator T –1 is defined; T –1 is the operator on Y to X that sends Tu into u . Obviously we have (T) = T. T is said to be nonsingular if T –1 exists and singular otherwise. For T to be nonsingular it is necessary that dim X = dim Y . If dim X = dim Y , each of nulT = 0 and def T = 0 implies the other and therefore the nonsingularity of T [2,6].

Linear operations on operators

Definition 3.1. If T and S are two linear operators on X to Y , their linear combination $\alpha S + \beta T$ is defined by $(\alpha S + \beta T)u = \alpha(Su) + \beta(Tu)$ for all u ∈ X , and is again a linear operator on X to Y .
Let us denote by ℜ(X ,Y) the set of all operators on X to Y Y; ℜ(X ,Y) is a vector space with the linear operations defined as above. The zero vector of this vector space is the zero operator 0 defined by 0u = 0 for all u ∈ X .
Problem 3.1. rank(S + T) = rankS + rankT .
The dimension of the vector space ℜ(X ,Y) is equal to NM , where N = dim X and M = dimY .
To see this, let $\{x_k\}$ and $\{y_j\}$ be bases of X and Y , respectively, and let Pjk be the operator on X to Y such that

$$P_{jk}x_h = \delta_{kh}y_j, \quad k,h = 1,...,N; j = 1,...,M .$$

These MN operators Pjk , are linearly independent elements of ℜ(X ,Y), and we have from $Tx_k = \sum_{j=1}^{M} \tau_{jk} y_j$, M = dim Y, yelds $T = \sum \tau_{jk} P_{jk}$. Thus $\{P_{jk}\}$ is a basis of ℜ(X ,Y), which proves the assertion. $\{P_{jk}\}$ will be called the basis of ℜ(X ,Y) associated with the bases $\{x_k\}$ and $\{y_j\}$ of X and Y , respectively. The last result shows that the matrix elements τ jk are the coefficients of the "vector" T with respect to the basis $\{P_{jk}\}$.
The product TS of two linear operators T , S is defined by (TS)u = T (Su) for all u ∈ X , where X is the domain space of S , provided the domain space of T is identical with the range space Y of S [4,5]. The following relations hold for these operations on linear operators :

(TS)R = T (SR), which is denoted by TSR

$$(\alpha T)S = T(\alpha S) = \alpha(TS), \text{ denoted by } \alpha TS$$

$$(T_1 + T_2)S = T_1S + T_2S$$

$$T(S_1 + S_2) = TS_1 + TS_2$$

$$\text{rank}(TS) \leq \max(\text{rank}T, \text{rank}S)$$

The algebra of linear operators

If S and T are operators on X to itself, their product TS is defined and is again an operator on X to itself. Thus the set $\Re(X) = \Re(X, X)$ of all linear operators in X is not only a vector space but an algebra. $\Re(X)$ is not commutative for dim $X \geq 2$ since TS = ST is in general not true. When TS = ST , T and S are said to commute (with each other). We have T 0 = 0T = 0 and T1 = 1T = T for every $T \in \Re(X)$, where 1 denotes the identity operator (defined by 1u = u for every $u \in X$). Thus 1 is the unit element of $\Re(X)$. The operators of the form $\alpha 1$ are called scalar operators and in symbol will not be distinguished from the scalars α . A scalar operator commutes with every operator of $\Re(X)$.
We write TT = T 2 , TTT = T 3 and so on, and set T 0 = 1 by definition. We have

$$T^m T^n = T^{m+n}, \left(T^m\right)^n = T^{mn}, \quad m, n = 0,1,2,...$$

For any polynomial $p(z) = \alpha_0 + \alpha_1 z + ... + \alpha_n z^n$ in the indeterminate z , we define the operator p T T nT= $\alpha 0 + \alpha 1 + ... + \alpha$. The mapping $p(z) \rightarrow p(T)$ is a homomorphism of the algebra of polynomials to $\Re(X)$; this means that p(z) + q(z) = r(z) or p(z)q(z) = r(z) inplies p(T) + q(T) = r(T) or p(T)q(T) = r(T) respectively. In particular, it follows that p(T) and q(T) commute.
If $T \in \Re(X)$ is nonsingular, the inverse T –1 exists and belongs to $\Re(X)$; we have T –1T = TT –1 = 1. If T has a left inverse T ' (that is, a $T' \in \Re(X)$ such that T 'T = 1), T has nullity zero, for Tu = 0 implies u = T 'Tu = 0 . If T has a right inverse T ' ' (that is, TT " = 1), T has deficiency zero because every $u \in X$ lies in R(T) by u = TT "u . If dim X is finite, either of these facts implies that T is nonsingular and that T ' = T –1 or T " = T –1 , respectively [4]. If S and T are nonsingular, so is TS and $(TS)^{-1} = S^{-1}T^{-1}$ For a nonsingular T , the negative powers T –n , n = 1,2,... can be defined by $T^{-n} = \left(T^{-1}\right)^n$. In this case

$$T^m T^n = T^{m+n}, \left(T^m\right)^n = T^{mn}, \quad m, n = 0,1,2,...$$

is true for any integers m, n .

References

1. Ramadan Zejnullahu, Analiza Funksionale, Prishtinë, 1998.
2. Emrush Gashi, Dukagjin Pupovci, Hapësirat Vektoriale, Prishtinë,
3. Emrush Gashi, Kursi i Algjebrës së Lartë, Prishtinë, 1976.
4. SvetozarKurepa, Funkcionalna Analiza. Elementi Teorije Operatora, Zagreb.
5. S. Alançiq, Hyrje në Analizën Reale dhe Funksionale, Prishtinë, 1986.

6. Tosio Kato, Perturbation Theory of Linear Operators, Springer-Verlag Berlin Heidelberg 1966, 1976

Katalogimi në botim – **(CIP)**
Biblioteka Kombëtare e Kosovës "Pjetër Bogdani"

004(496.51)"2019"(062)
62(496.51)"2019"(062)

International Conference Computer Science and Communication Engineering : proceedings of the 8th Annual International Conference Pristina, 26-28 october 2019 / editet by Edmond Hajrizi. - Prishtinë : UBT, 2020. – 79 f. : ilustr. ; 30 cm.

1.Hajrizi, Edmond

ISBN 978-9951-437-85-1

INTERNATIONAL CONFERENCE ON BUSINESS,TECHNOLOGY AND INNOVATION

CHAPTERS:

- Computer Science and Communication Engineering
- Management, Business and Economics
- Mechatronics, System Engineering and Robotics
- Energy Efficiency Engineering
- Information Systems and Security
- Architecture – Spatial Planning
- Civil Engineering, Infrastructure and Environment
- Law
- Political Science
- Journalism, Media and Communication
- Food Science and Technology
- Pharmaceutical and Natural Sciences
- Design
- Psychology
- Education and Development
- Fashion
- Music
- Art and Digital Media
- Dentistry
- Medicine & Nursing

Lagjja Kalabria p.n KS - 10000, Prishtinë
+383 38 541 400
+383 38 542 138

www.ubt-uni.net
conferences@ubt-uni.net

ISBN 978-9951-437-85-1